Ambardar frontmatter 12/5/98 10:21 AM Page 2

Analog and Digital Signal Processing

Second Edition

Ashok Ambardar

Michigan Technological University



Brooks/Cole Publishing Company
I(T)P® An International Thomson Publishing Company

Pacific Grove • Albany • Belmont • Bonn • Boston • Cincinnati • Detroit • Johannesburg • London
Madrid • Melbourne • Mexico City • New York • Paris • Singapore • Tokyo • Toronto • Wahington

PGR2026-00033
Krisp Technologies EX1024 Page 1

CONTENTS

LIST OF TABLES **xi**
PREFACE **xiii**
FROM THE PREFACE TO THE FIRST EDITION **xv**

1 OVERVIEW **1**
1.0 Introduction . . . 1
1.1 Signals . . . 1
1.2 Systems . . . 3
1.3 The Frequency Domain . . . 4
1.4 From Concept to Application . . . 7

2 ANALOG SIGNALS **8**
2.0 Scope and Objectives . . . 8
2.1 Signals . . . 8
2.2 Operations on Signals . . . 12
2.3 Signal Symmetry . . . 14
2.4 Harmonic Signals and Sinusoids . . . 16
2.5 Commonly Encountered Signals . . . 18
2.6 The Impulse Function . . . 21
2.7 The Doublet . . . 26
2.8 Moments . . . 28
Problems . . . 30

3 DISCRETE SIGNALS **39**
3.0 Scope and Objectives . . . 39
3.1 Discrete Signals . . . 39
3.2 Operations on Discrete Signals . . . 41
3.3 Decimation and Interpolation . . . 44
3.4 Common Discrete Signals . . . 47
3.5 Discrete-Time Harmonics and Sinusoids . . . 49
3.6 Aliasing and the Sampling Theorem . . . 52
3.7 Random Signals . . . 55
Problems . . . 59

PGR2026-00033
Krisp Technologies EX1024 Page 2

4 ANALOG SYSTEMS **68**
4.0 Scope and Objectives . . . 68
4.1 Introduction . . . 68
4.2 System Classification . . . 70
4.3 Analysis of LTI Systems . . . 76
4.4 LTI Systems Described by Differential Equations . . . 76
4.5 The Impulse Response of LTI Systems . . . 81
4.6 System Stability . . . 85
4.7 Application-Oriented Examples . . . 86
Problems . . . 89

5 DISCRETE-TIME SYSTEMS **96**
5.0 Scope and Objectives . . . 96
5.1 Discrete-Time Operators . . . 96
5.2 System Classification . . . 97
5.3 Digital Filters . . . 101
5.4 Digital Filters Described by Difference Equations . . . 103
5.5 Impulse Response of Digital Filters . . . 111
5.6 Stability of Discrete-Time LTI Systems . . . 115
5.7 Connections: System Representation in Various Forms . . . 116
5.8 Application-Oriented Examples . . . 117
Problems . . . 121

6 CONTINUOUS CONVOLUTION **130**
6.0 Scope and Objectives . . . 130
6.1 Introduction . . . 130
6.2 Convolution of Some Common Signals . . . 133
6.3 Some Properties of Convolution . . . 135
6.4 Convolution by Ranges (Graphical Convolution) . . . 138
6.5 Stability and Causality . . . 144
6.6 The Response to Periodic Inputs . . . 146
6.7 Periodic Convolution . . . 150
6.8 Connections: Convolution and Transform Methods . . . 151
6.9 Convolution Properties Based on Moments . . . 153
6.10 Correlation . . . 156
Problems . . . 160

7 DISCRETE CONVOLUTION **169**
7.0 Scope and Objectives . . . 169
7.1 Discrete Convolution . . . 169
7.2 Convolution Properties . . . 171
7.3 Convolution of Finite Sequences . . . 172
7.4 Stability and Causality of LTI Systems . . . 177
7.5 System Response to Periodic Inputs . . . 178

PGR2026-00033
Krisp Technologies EX1024 Page 3

7.6 Periodic Convolution . . . 180
7.7 Connections: Discrete Convolution and Transform Methods . . . 183
7.8 Deconvolution . . . 185
7.9 Discrete Correlation . . . 187
Problems . . . 189

8 FOURIER SERIES 197
8.0 Scope and Objectives . . . 197
8.1 Fourier Series: A First Look . . . 197
8.2 Simplifications Through Signal Symmetry . . . 201
8.3 Parseval's Relation and the Power in Periodic Signals . . . 205
8.4 The Spectrum of Periodic Signals . . . 207
8.5 Properties of Fourier Series . . . 210
8.6 Signal Reconstruction and the Gibbs Effect . . . 216
8.7 System Response to Periodic Inputs . . . 221
8.8 Application-Oriented Examples . . . 224
8.9 The Dirichlet Kernel and the Gibbs Effect . . . 226
8.10 The Fourier Series, Orthogonality, and Least Squares . . . 230
8.11 Existence, Convergence, and Uniqueness . . . 232
8.12 A Historical Perspective . . . 234
Problems . . . 238

9 THE FOURIER TRANSFORM 248
9.0 Scope and Objectives . . . 248
9.1 Introduction . . . 248
9.2 Fourier Transform Pairs and Properties . . . 252
9.3 System Analysis Using the Fourier Transform . . . 271
9.4 Frequency Response of Filters . . . 275
9.5 Energy and Power Spectral Density . . . 282
9.6 Time-Bandwidth Measures . . . 284
Problems . . . 289

10 MODULATION 300
10.0 Scope and Objectives . . . 300
10.1 Amplitude Modulation . . . 300
10.2 Single-Sideband AM . . . 310
10.3 Angle Modulation . . . 312
10.4 Wideband Angle Modulation . . . 316
10.5 Demodulation of FM Signals . . . 319
10.6 The Hilbert Transform . . . 321
Problems . . . 326

11 THE LAPLACE TRANSFORM 330
11.0 Scope and Objectives . . . 330
11.1 The Laplace Transform . . . 330

PGR2026-00033
Krisp Technologies EX1024 Page 4

11.2 Properties of the Laplace Transform . . . 332
11.3 Poles and Zeros of the Transfer Function . . . 339
11.4 The Inverse Laplace Transform . . . 340
11.5 The s-Plane and BIBO Stability . . . 344
11.6 The Laplace Transform and System Analysis . . . 347
11.7 Connections . . . 358
Problems . . . 360

12 APPLICATIONS OF THE LAPLACE TRANSFORM 367
12.0 Scope and Objectives . . . 367
12.1 Frequency Response . . . 367
12.2 Minimum-Phase Filters . . . 369
12.3 Bode Plots . . . 370
12.4 Performance Measures . . . 378
12.5 Feedback . . . 382
12.6 Application of Feedback: The Phase-Locked Loop . . . 387
Problems . . . 390

13 ANALOG FILTERS 398
13.0 Scope and Objectives . . . 398
13.1 Introduction . . . 398
13.2 The Design Process . . . 402
13.3 The Butterworth Filter . . . 403
13.4 The Chebyshev Approximation . . . 412
13.5 The Inverse Chebyshev Approximation . . . 421
13.6 The Elliptic Approximation . . . 427
13.7 The Bessel Approximation . . . 432
Problems . . . 440

14 SAMPLING AND QUANTIZATION 446
14.0 Scope and Objectives . . . 446
14.1 Ideal Sampling . . . 446
14.2 Sampling, Interpolation, and Signal Recovery . . . 456
14.3 Quantization . . . 460
14.4 Digital Processing of Analog Signals . . . 465
14.5 Compact Disc Digital Audio . . . 470
14.6 Dynamic Range Processors . . . 472
Problems . . . 475

15 THE DISCRETE-TIME FOURIER TRANSFORM 482
15.0 Scope and Objectives . . . 482
15.1 The Discrete-Time Fourier Transform . . . 482
15.2 Connections: The DTFT and the Fourier Transform . . . 483
15.3 Properties of the DTFT . . . 487

PGR2026-00033
Krisp Technologies EX1024 Page 5

15.4 The Transfer Function . . . 494
15.5 System Analysis Using the DTFT . . . 499
15.6 Connections . . . 501
15.7 Ideal Filters . . . 504
15.8 Some Traditional and Non-traditional Filters . . . 509
15.9 Frequency Response of Discrete Algorithms . . . 512
15.10 Oversampling and Sampling Rate Conversion . . . 515
Problems . . . 520

16 THE DFT AND FFT **535**
16.0 Scope and Objectives . . . 535
16.1 Introduction . . . 535
16.2 Properties of the DFT . . . 536
16.3 Connections . . . 542
16.4 Approximating the DTFT by the DFT . . . 544
16.5 The DFT of Periodic Signals . . . 546
16.6 The DFT of Nonperiodic Signals . . . 552
16.7 Spectral Smoothing by Time Windows . . . 555
16.8 Applications in Signal Processing . . . 563
16.9 Spectrum Estimation . . . 566
16.10 Matrix Formulation of the DFT and IDFT . . . 569
16.11 The FFT . . . 571
16.12 Why Equal Lengths for the DFT and IDFT? . . . 577
Problems . . . 581

17 THE z-TRANSFORM **592**
17.0 Scope and Objectives . . . 592
17.1 The Two-Sided z-Transform . . . 592
17.2 Properties of the z-Transform . . . 596
17.3 Poles, Zeros, and the z-Plane . . . 600
17.4 The Transfer Function . . . 602
17.5 The Inverse z-Transform . . . 605
17.6 The One-Sided z-Transform . . . 613
17.7 The z-Transform and System Analysis . . . 618
17.8 Frequency Response . . . 621
17.9 Connections . . . 624
Problems . . . 625

18 APPLICATIONS OF THE z-TRANSFORM **637**
18.0 Scope and Objectives . . . 637
18.1 Transfer Function Realization . . . 637
18.2 Interconnected Systems . . . 640
18.3 Minimum-Phase Systems . . . 642
18.4 The Frequency Response: A Graphical Interpretation . . . 645

PGR2026-00033
Krisp Technologies EX1024 Page 6

18.5 Application-Oriented Examples . . . 649
18.6 Allpass Filters . . . 658
18.7 Application-Oriented Examples: Digital Audio Effects . . . 660
Problems . . . 664

19 IIR DIGITAL FILTERS 673
19.0 Scope and Objectives . . . 673
19.1 Introduction . . . 673
19.2 IIR Filter Design . . . 674
19.3 Response Matching . . . 676
19.4 The Matched z-Transform for Factored Forms . . . 684
19.5 Mappings from Discrete Algorithms . . . 685
19.6 The Bilinear Transformation . . . 691
19.7 Spectral Transformations for IIR Filters . . . 694
19.8 Design Recipe for IIR Filters . . . 703
Problems . . . 707

20 FIR DIGITAL FILTERS 715
20.0 Scope and Objectives . . . 715
20.1 Symmetric Sequences and Linear Phase . . . 715
20.2 Window-Based Design . . . 720
20.3 Half-Band FIR Filters . . . 733
20.4 FIR Filter Design by Frequency Sampling . . . 736
20.5 Design of Optimal Linear-Phase FIR Filters . . . 740
20.6 Application: Multistage Interpolation and Decimation . . . 744
20.7 Maximally Flat FIR Filters . . . 748
20.8 FIR Differentiators and Hilbert Transformers . . . 749
20.9 Least Squares and Adaptive Signal Processing . . . 751
Problems . . . 754

21 MATLAB EXAMPLES 762
21.0 Introduction . . . 762
21.1 The ADSP Toolbox and Its Installation . . . 762
21.2 MATLAB Tips and Pointers . . . 763
21.3 Graphical User Interface Programs . . . 765
21.4 The ADSP Toolbox . . . 766
21.5 Examples of MATLAB Code . . . 769

REFERENCES 798
INDEX 801

PGR2026-00033
Krisp Technologies EX1024 Page 7

LIST OF TABLES

Table 1.1 Response of an RC Lowpass Filter . . . 3

Table 4.1 Form of the Natural Response for Analog LTI Systems . . . 77
Table 4.2 Form of the Forced Response for Analog LTI Systems . . . 77

Table 5.1 Form of the Natural Response for Discrete LTI Systems . . . 105
Table 5.2 Form of the Forced Response for Discrete LTI Systems . . . 105

Table 8.1 Some Common Spectral Windows . . . 220
Table 8.2 Smoothing by Operations on the Partial Sum . . . 228

Table 9.1 Some Useful Fourier Transform Pairs . . . 253
Table 9.2 Operational Properties of the Fourier Transform . . . 254

Table 10.1 Some Useful Hilbert Transform Pairs . . . 322

Table 11.1 A Short Table of Laplace Transforms . . . 331
Table 11.2 Operational Properties of the Laplace Transform . . . 333
Table 11.3 Inverse Laplace Transform of Partial Fraction Expansion Terms . . . 342

Table 12.1 Time Domain Performance Measures for Real Filters . . . 379

Table 13.1 3-dB Butterworth Lowpass Prototype Transfer Functions . . . 408
Table 13.2 Bessel Polynomials . . . 436

Table 14.1 Various Number Representations for $B = 3$ Bits . . . 461

Table 15.1 Some Useful DTFT Pairs . . . 486
Table 15.2 Properties of the DTFT . . . 487
Table 15.3 Relating the DTFT and Fourier Series . . . 501
Table 15.4 Connections Between Various Transforms . . . 503
Table 15.5 Discrete Algorithms and their Frequency Response . . . 513

Table 16.1 Properties of the N-Sample DFT . . . 536
Table 16.2 Relating Frequency Domain Transforms . . . 543
Table 16.3 Some Commonly Used N-Point DFT Windows . . . 558
Table 16.4 Symmetry and Periodicity of $W_N = \exp(-j2\pi/N)$ . . . 571
Table 16.5 FFT Algorithms for Computing the DFT . . . 573
Table 16.6 Computational Cost of the DFT and FFT . . . 577

PGR2026-00033
Krisp Technologies EX1024 Page 8

Table 17.1 A Short Table of z-Transform Pairs 594
Table 17.2 Properties of the Two-Sided z-Transform 597
Table 17.3 Inverse $\boldsymbol{z}$-Transform of Partial Fraction Expansion (PFE) Terms 609
Table 17.4 Properties Unique to the One-Sided z-Transform 613

Table 19.1 Impulse-Invariant Transformations 680
Table 19.2 Numerical Difference Algorithms 685
Table 19.3 Numerical Integration Algorithms 688
Table 19.4 Digital-to-Digital (D2D) Frequency Transformations 695
Table 19.5 Direct Analog-to-Digital (A2D) Transformations for Bilinear Design 697

Table 20.1 Applications of Symmetric Sequences 719
Table 20.2 Some Windows for FIR Filter Design 721
Table 20.3 Characteristics of Harris Windows 723
Table 20.4 Characteristics of the Windowed Spectrum 727

PGR2026-00033
Krisp Technologies EX1024 Page 9

PREFACE

In keeping with the goals of the first edition, this second edition of *Analog and Digital Signal Processing* is geared to junior and senior electrical engineering students and stresses the fundamental principles and applications of signals, systems, transforms, and filters. The premise is to help the student think clearly in both the time domain and the frequency domain and switch from one to the other with relative ease. The text assumes familiarity with elementary calculus, complex numbers, and basic circuit analysis.

This edition has undergone extensive revision and refinement, in response to reviewer comments and to suggestions from users of the first edition (including students). Major changes include the following:

1. At the suggestion of some reviewers, the chapters have been reorganized. Specifically, continuous and discrete aspects (that were previously covered together in the first few chapters) now appear in separate chapters. This should allow instructors easier access to either sequential or parallel coverage of analog and discrete signals and systems.

2. The material in each chapter has been pruned and streamlined to make the book more suited as a *textbook*. We highlight the most important concepts and problem-solving methods in each chapter by including boxed *review panels*. The review panels are reinforced by discussions and worked examples. Many new figures have been added to help the student grasp and visualize critical concepts.

3. New application-oriented material has been added to many chapters. The material focuses on how the theory developed in the text finds applications in diverse fields such as audio signal processing, digital audio special effects, echo cancellation, spectrum estimation, and the like.

4. Many worked examples in each chapter have been revised and new ones added to reinforce and extend key concepts. Problems at the end of each chapter are now organized into "Drill and Reinforcement", "Review and Exploration", and "Computation and Design" and include a substantial number of new problems. The computation and design problems, in particular, should help students appreciate the application of theoretical principles and guide instructors in developing projects suited to their own needs.

5. The MATLAB-based software supplied with the book has been revised and expanded. All the routines have been upgraded to run on the latest version (currently, v5) of both the professional edition and student edition of MATLAB, while maintaining downward compatibility with earlier versions.

6. The MATLAB appendices (previously at the end of each chapter) have been consolidated into a separate chapter and substantially revamped. This has allowed us to present integrated application-oriented examples spanning across chapters in order to help the student grasp important signal-processing concepts quickly and effectively. Clear examples of MATLAB code based on native MATLAB routines, as well as the supplied routines, are included to help accelerate the learning of MATLAB syntax.

PGR2026-00033
Krisp Technologies EX1024 Page 10

7. A set of new self-contained, menu-driven, graphical user interface (GUI) programs with point-and-click features is now supplied for ease of use in visualizing basic signal processing principles and concepts. These GUIs require no experience in MATLAB programming, and little experience with its syntax, and thus allow students to concentrate their efforts on understanding concepts and results. The programs cover signal generation and properties, time-domain system response, convolution, Fourier series, frequency response and Bode plots, analog filter design, and digital filter design. The GUIs are introduced at the end of each chapter, in the "Computation and Design" section of the problems. I am particularly grateful to Craig Borghesani, Terasoft, Inc. (`http://world.std.com/~borg/`) for his help and MATLAB expertise in bringing many of these GUIs to fruition.

This book has profited from the constructive comments and suggestions of the following reviewers:

- Professor Khaled Abdel-Ghaffar, University of California at Davis
- Professor Tangul Basar, University of Illinois
- Professor Martin E. Kaliski, California Polytechnic State University
- Professor Roger Goulet, Université de Sherbrooke
- Professor Ravi Kothari, University of Cincinnati
- Professor Nicholas Kyriakopoulos, George Washington University
- Professor Julio C. Mandojana, Mankato State University
- Professor Hadi Saadat, Milwaukee School of Engineering
- Professor Jitendra K. Tugnait, Auburn University
- Professor Peter Willett, University of Connecticut

Here, at Michigan Technological University, it is also our pleasure to acknowledge the following:

- Professor Clark R. Givens for lending mathematical credibility to portions of the manuscript
- Professor Warren F. Perger for his unfailing help in all kinds of TEX-related matters
- Professor Tim Schulz for suggesting some novel DSP projects, and for supplying several data files

Finally, at PWS Publishing, Ms Suzanne Jeans, Editorial Project Manager, and the editorial and production staff (Kirk Bomont, Liz Clayton, Betty Duncan, Susan Pendleton, Bill Stenquist, Jean Thompson, and Nathan Wilbur), were instrumental in helping meet (or beat) all the production deadlines.

We would appreciate hearing from you if you find any errors in the text or discover any bugs in the software. Any errata for the text and upgrades to the software will be posted on our Internet site.

Ashok Ambardar	Michigan Technological University
Internet:	`http://www.ee.mtu.edu/faculty/akambard.html`
e-mail:	`akambard@mtu.edu`

PGR2026-00033
Krisp Technologies EX1024 Page 11

FROM THE PREFACE TO THE FIRST EDITION

This book on analog and digital signal processing is intended to serve both as a text for students and as a source of basic reference for professionals across various disciplines. As a text, it is geared to junior/senior electrical engineering students and details the material covered in a typical undergraduate curriculum. As a reference, it attempts to provide a broader perspective by introducing additional special topics towards the later stages of each chapter. Complementing this text, but deliberately not integrated into it, is a set of powerful software routines (running under MATLAB) that can be used not only for reinforcing and visualizing concepts but also for problem solving and advanced design.

The text stresses the fundamental principles and applications of signals, systems, transforms and filters. It deals with concepts that are crucial to a full understanding of time-domain and frequency-domain relationships. Our ultimate objective is that the student be able to think clearly in both domains and switch from one to the other with relative ease. It is based on the premise that what might often appear obvious to the expert may not seem so obvious to the budding expert. Basic concepts are, therefore, explained and illustrated by worked examples to bring out their importance and relevance.

Scope

The text assumes familiarity with elementary calculus, complex numbers, basic circuit analysis and (in a few odd places) the elements of matrix algebra. It covers the core topics in analog and digital signal processing taught at the undergraduate level. The links between analog and digital aspects are explored and emphasized throughout. The topics covered in this text may be grouped into the following broad areas:

1. An introduction to **signals** and **systems**, their representation and their classification.
2. **Convolution**, a method of time-domain analysis, which also serves to link the time domain and the frequency domain.
3. **Fourier series** and **Fourier transforms**, which provide a spectral description of analog signals, and their applications.
4. The **Laplace transform**, which forms a useful tool for system analysis and its applications.
5. Applications of Fourier and Laplace techniques to **analog filter design**.
6. **Sampling** and the discrete-time Fourier transform (**DTFT**) of sampled signals, and the **DFT** and the **FFT**, all of which reinforce the central concept that sampling in one domain leads to a periodic extension in the other.
7. The ***z*-transform**, which extends the DTFT to the analysis of discrete-time systems.
8. Applications of digital signal processing to the design of **digital filters**.

PGR2026-00033
Krisp Technologies EX1024 Page 12

We have tried to preserve a rational approach and include all the necessary mathematical details, but we have also emphasized heuristic explanations whenever possible. Each chapter is more or less structured as follows:

1. A short opening section outlines the objectives and topical coverage and points to the required background.
2. Central concepts are introduced in early sections and illustrated by worked examples. Special topics are developed only in later sections.
3. Within each section, the material is broken up into bite-sized pieces. Results are tabulated and summarized for easy reference and access.
4. Whenever appropriate, concepts are followed by remarks, which highlight essential features or limitations.
5. The relevant software routines and their use are outlined in MATLAB appendices to each chapter. Sections that can be related to the software are specially marked in the table of contents.
6. End-of-chapter problems include a variety of drills and exercises. MATLAB code to generate answers to many of these appears on the supplied disk.

A solutions manual for instructors is available from the publisher.

Software

A unique feature of this text is the *analog and digital signal processing* (ADSP) software toolbox for signal processing and analytical and numerical computation designed to run under all versions of MATLAB. The routines are self-demonstrating and can be used to reinforce essential concepts, validate the results of analytical paper and pencil solutions, and solve complex problems that might, otherwise, be beyond the skills of analytical computation demanded of the student.

The toolbox includes programs for generating and plotting signals, regular and periodic convolution, symbolic and numerical solution of differential and difference equations, Fourier analysis, frequency response, asymptotic Bode plots, symbolic results for system response, inverse Laplace and inverse z-transforms, design of analog, IIR and FIR filters by various methods, and more.

Since our primary intent is to present the principles of signal processing, not software, we have made no attempt to integrate MATLAB into the text. Software related aspects appear only in the appendices to each chapter. This approach also maintains the continuity and logical flow of the textual material, especially for users with no inclination (or means) to use the software. In any case, the self-demonstrating nature of the routines should help you to get started even if you are new to MATLAB. As an aside, all the graphs for this text were generated using the supplied ADSP toolbox.

We hasten to provide two disclaimers. First, our use of MATLAB is not to be construed as an endorsement of this product. We just happen to like it. Second, our routines are supplied in good faith; we fully expect them to work on your machine, but provide no guarantees!

Acknowledgements

This book has gained immensely from the incisive, sometimes provoking, but always constructive, criticism of Dr. J.C.Mandojana. Many other individuals have also contributed in various ways to this effort. Special thanks are due, in particular, to

- Drs. R.W. Bickmore and R.T. Sokolov, who critiqued early drafts of several chapters and provided valuable suggestions for improvement.
- Dr. A.R. Hambley, who willingly taught from portions of the final draft in his classes.

PGR2026-00033
Krisp Technologies EX1024 Page 13

- Drs. D.B. Brumm, P.H. Lewis and J.C. Rogers, for helping set the tone and direction in which the book finally evolved.
- Mr. Scott Ackerman, for his invaluable computer expertise in (the many) times of need.
- At PWS Publishing, the editor Mr. Tom Robbins, for his constant encouragement, and Ms. Pam Rockwell for her meticulous attention to detail during all phases of editing and production, and Ken Morton, Lai Wong, and Lisa Flanagan for their behind-the-scenes help.
- The students, who tracked down inconsistencies and errors in the various drafts, and provided extremely useful feedback.
- The Mathworks, for permission to include modified versions of a few of their *m-files* with our software.

We would also like to thank Dr. Mark Thompson, Dr. Hadi Saadat and the following reviewers for their useful comments and suggestions:

- Professor Doran Baker, *Utah State University*
- Professor Ken Sauer, *University of Notre Dame*
- Professor George Sockman, *SUNY—Binghamton*
- Professor James Svoboda, *Clarkson University*
- Professor Kunio Takaya, *University of Saskatchewan*

And now for something completely different!

We have adopted the monarchic practice of using words like *we*, *us*, and *our* when addressing you, the reader. This seems to have become quite fashionable, since it is often said that no book can be the work of a single author. In the present instance, the reason is even more compelling because my family, which has played an important part in this venture, would never forgive me otherwise! In the interests of fairness, however, when the time comes to accept the blame for any errors, I have been graciously awarded sole responsibility!

We close with a limerick that, we believe, actually bemoans the fate of all textbook writers. If memory serves, it goes something like this

A limerick packs laughs anatomical,
In space that's quite economical,
But the good ones I've seen,
So seldom are clean,
And the clean ones so seldom are comical.

Campus lore has it that students complain about texts prescribed by their instructors as being too highbrow or tough and not adequately reflecting student concerns, while instructors complain about texts as being low-level and, somehow, less demanding. We have consciously tried to write a book that both the student and the instructor can tolerate. Whether we have succeeded remains to be seen and can best be measured by your response. And, if you have read this far, and are still reading, we would certainly like to hear from you.

PGR2026-00033
Krisp Technologies EX1024 Page 14

Chapter 1

OVERVIEW

1.0 Introduction

"I listen and I forget,
I see and I remember, I do and I learn."
A Chinese Proverb

This book is about signals and their processing by systems. This chapter provides an overview of the terminology of analog and digital processing and of the connections between the various topics and concepts covered in subsequent chapters. We hope you return to it periodically to fill in the missing details and get a feel for how all the pieces fit together.

1.1 Signals

Our world is full of signals, both natural and man-made. Examples are the variation in air pressure when we speak, the daily highs and lows in temperature, and the periodic electrical signals generated by the heart. Signals represent information. Often, signals may not convey the required information directly and may not be free from disturbances. It is in this context that signal processing forms the basis for enhancing, extracting, storing, or transmitting useful information. Electrical signals perhaps offer the widest scope for such manipulations. In fact, it is commonplace to convert signals to electrical form for processing.

The value of a signal, at any instant, corresponds to its (instantaneous) amplitude. Time may assume a continuum of values, t, or discrete values, nt_s, where t_s is a **sampling interval** and n is an integer. The amplitude may also assume a continuum of values or be *quantized* to a finite number of discrete levels between its extremes. This results in four possible kinds of signals, as shown in Figure 1.1.

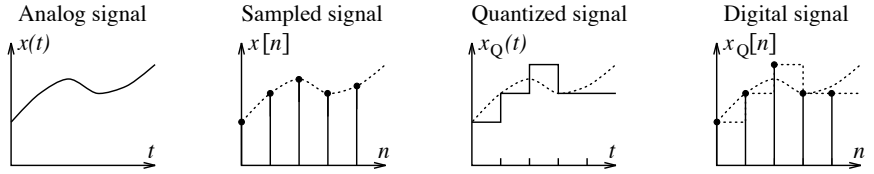


Figure 1.1 Analog, sampled, quantized, and digital signals

The music you hear from your compact disc (CD) player due to changes in the air pressure caused by the vibration of the speaker diaphragm is an **analog signal** because the pressure variation is a *continuous* function of time. However, the information stored on the compact disc is in digital form. It must be processed

PGR2026-00033
Krisp Technologies EX1024 Page 15

and converted to analog form before you can hear the music. A record of the yearly increase in the world population describes time measured in increments of one (year), and the population increase is measured in increments of one (person). It is a **digital signal** with discrete values for both time and population.

1.1.1 Signal Processing

Analog signals have been the subject of much study in the past. In recent decades, digital signals have received increasingly widespread attention. Being numbers, they can be processed by the same logic circuits used in digital computers.

Two conceptual schemes for the processing of signals are illustrated in Figure 1.2. The digital processing of analog signals requires that we use an analog-to-digital converter (ADC) for sampling the analog signal prior to processing and a digital-to-analog converter (DAC) to convert the processed digital signal back to analog form.

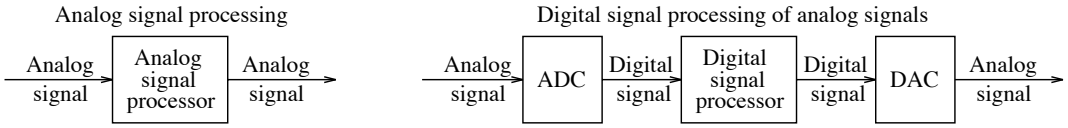


Figure 1.2 Analog and digital signal processing

Few other technologies have revolutionized the world as profoundly as those based on digital signal processing. For example, the technology of recorded music was, until recently, completely analog from end to end, and the most important commercial source of recorded music used to be the LP (long-playing) record. The advent of the digital compact disc has changed all that in the span of just a few short years and made the long-playing record practically obsolete. Signal processing, both analog and digital, forms the core of this application and many others.

1.1.2 Sampling and Quantization

The *sampling* of analog signals is often a matter of practical necessity. It is also the first step in **digital signal processing** (DSP). To process an analog signal by digital means, we must convert it to a digital signal in two steps. First, we must sample it, typically at uniform intervals t_s. The discrete quantity nt_s is related to the integer index n. Next, we must *quantize* the sample values (amplitudes). Both sampling and quantization lead to a potential loss of information. The good news is that a signal can be sampled without loss of information if it is *band-limited* to a highest-frequency f_B and sampled at intervals less than $\frac{1}{2f_B}$. This is the celebrated **sampling theorem**. The bad news is that most signals are not band-limited and even a small sampling interval may not be small enough. If the sampling interval exceeds the critical value $\frac{1}{2f_B}$, a phenomenon known as **aliasing** manifests itself. Components of the analog signal at high frequencies appear at (alias to) lower frequencies in the sampled signal. This results in a sampled signal with a smaller highest frequency. Aliasing effects are impossible to undo once the samples are acquired. It is thus commonplace to band-limit the signal before sampling (using lowpass filters).

Numerical processing using digital computers requires finite data with finite precision. We must limit signal amplitudes to a finite number of levels. This process, called **quantization**, produces nonlinear effects that can be described only in statistical terms. Quantization also leads to an irreversible loss of information and is typically considered only in the final stage in any design. The terms *discrete time* (DT), *sampled*, and *digital* are therefore often used synonymously.

PGR2026-00033
Krisp Technologies EX1024 Page 16

1.2 Systems

Systems may process analog or digital signals. All systems obey energy conservation. Loosely speaking, the **state** of a system refers to variables, such as capacitor voltages and inductor currents, which yield a measure of the system energy. The *initial state* is described by the initial value of these variables or initial conditions. A system is **relaxed** if initial conditions are zero. In this book, we study only **linear systems** (whose input-output relation is a straight line passing through the origin). If a complicated input can be split into simpler forms, linearity allows us to find the response as the sum of the response to each of the simpler forms. This is **superposition**. Many systems are actually nonlinear. The study of nonlinear systems often involves making simplifying assumptions, such as linearity.

1.2.1 System Analysis in the Time Domain

Consider the so-called RC circuit of Figure 1.3. The capacitor voltage is governed by the differential equation

$$\frac{dv_0(t)}{dt} + \frac{1}{\tau}v_0(t) = \frac{1}{\tau}v_i(t)$$

Assuming an uncharged capacitor (zero initial conditions), the response of this circuit to various inputs is summarized in Table 1.1 and sketched in Figure 1.3.

Table 1.1 Response of an RC Lowpass Filter

Input $v_i(t)$	**Response** $v_0(t)$
$A, \quad t \geq 0$	$A(1 - e^{-t/\tau}),\ t \geq 0$
$A\cos(\omega_0 t)$	$\dfrac{A\cos(\omega_0 t - \theta)}{(1+\omega_0^2\tau^2)^{1/2}}, \quad \theta = \tan^{-1}(\omega_0\tau)$
$A\cos(\omega_0 t),\ t \geq 0$	$\dfrac{A\cos(\omega_0 t - \theta)}{(1+\omega_0^2\tau^2)^{1/2}} + \dfrac{A\omega_0\tau}{1+\omega_0^2\tau^2}e^{-t/\tau},\ t \geq 0$

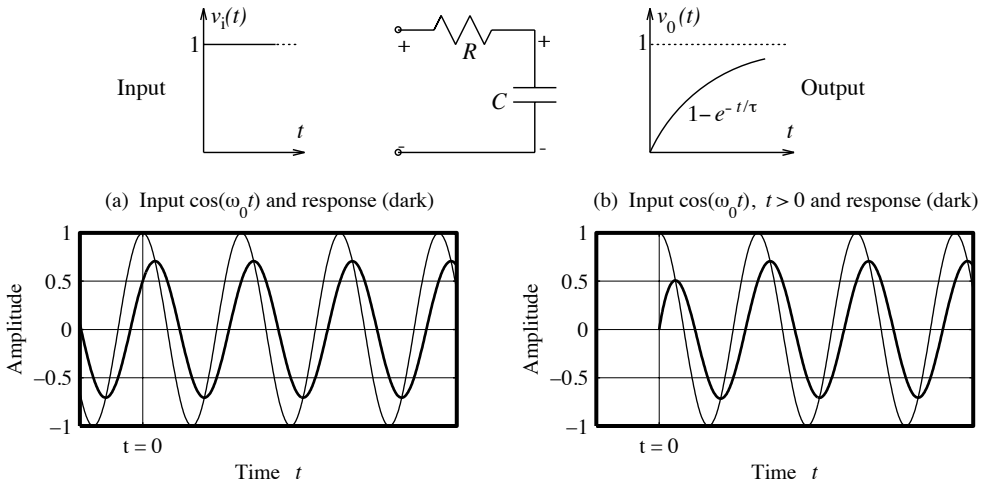


Figure 1.3 An RC circuit and its response to various inputs

PGR2026-00033
Krisp Technologies EX1024 Page 17

It is not our intent here to see how the solutions arise but how to interpret the results in terms of system performance. The cosine input yields only a sinusoidal component as the **steady-state response**. The response to the suddenly applied step and the switched cosine also includes a decaying exponential term representing the **transient** component.

1.2.2 Measures in the Time Domain

The **time constant** $\tau = RC$ is an important measure of system characteristics. Figure 1.4 reveals that for small τ, the output resembles the input and follows the rapid variations in the input more closely. For large τ, the system smooths the sharp details in the input. This smoothing effect is typical of a **lowpass filter**. The time constant thus provides one index of performance in the time domain. Another commonly used index is the **rise time**, which indicates the rapidity with which the response reaches its final value. A practical measure of rise time is the time for the step response to rise from 10% to 90% of the final value. As evident from Figure 1.4, a smaller time constant τ also implies a smaller rise time, even though it is usually difficult to quantify this relationship.

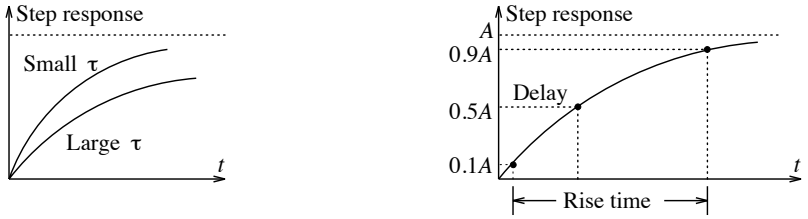


Figure 1.4 Step response of an RC circuit for various τ and the concept of rise time

1.3 The Frequency Domain

The concept of frequency is intimately tied to sinusoids. The familiar form of a sinusoid is the oscillating, periodic time waveform. An alternative is to visualize the same sinusoid in the **frequency domain** in terms of its **magnitude** and **phase** at the given frequency. The two forms are shown in Figure 1.5. Each sinusoid has a unique representation in the frequency domain.

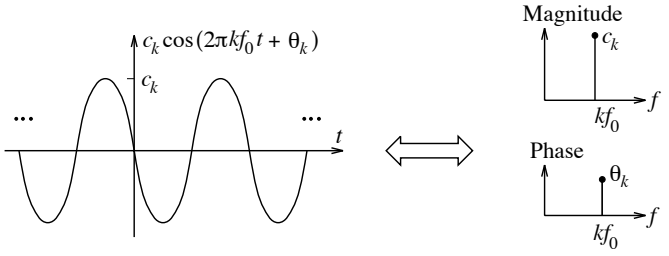


Figure 1.5 A sinusoid and its frequency-domain representation

PGR2026-00033
Krisp Technologies EX1024 Page 18

1.3.1 The Fourier Series and Fourier Transform

A periodic signal with period T may be described by a weighted combination of sinusoids at its fundamental frequency $f_0 = 1/T$ and integer multiples kf_0. This combination is called a **Fourier series**. The weights are the magnitude and phase of each sinusoid and describe a discrete spectrum in the frequency domain.

For large T, the frequencies kf_0 in the discrete spectrum of a periodic signal become more closely spaced. A nonperiodic signal may be conceived, in the limit, as a single period of a periodic signal as $T \to \infty$. Its spectrum then becomes continuous and leads to a frequency-domain representation of arbitrary signals (periodic or otherwise) in terms of the **Fourier transform**.

1.3.2 Measures in the Frequency Domain

The response of the RC circuit to a cosine at the frequency ω_0 depends on $\omega_0\tau$, as shown in Figure 1.6. For $\omega_0\tau \ll 1$, both the magnitude and phase of the output approach the applied input. For a given τ, the output magnitude decreases for higher frequencies. Such a system describes a **lowpass filter**. Rapid time variations or sharp features in a signal thus correspond to higher frequencies.

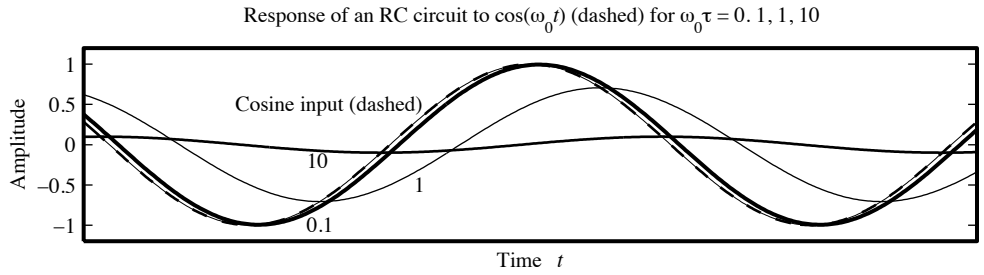


Figure 1.6 Response of an RC circuit to a cosine input for various values of $\omega_0\tau$

If the input consists of unit cosines at different frequencies, the magnitude and phase (versus frequency) of the ratio of the output describes the **frequency response**, as shown in Figure 1.7. The magnitude spectrum clearly shows the effects of attenuation at high frequencies.

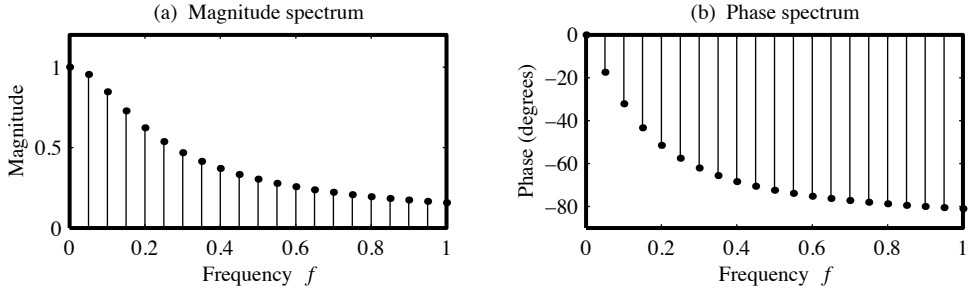


Figure 1.7 Frequency response of an RC circuit to unit cosines at various frequencies

At the frequency $\omega_0 = \omega_B = 1/\tau$, the magnitude of the frequency response is $1/\sqrt{2}$. This frequency serves as one definition of the **bandwidth**. More generally, the bandwidth is a measure of the frequency range that contains a significant fraction of the total signal energy.

PGR2026-00033
Krisp Technologies EX1024 Page 19

There are measures analogous to bandwidth that describe the time **duration** of a signal over which much of the signal is concentrated. The time constant τ provides one such measure.

The relation $\omega_B \tau = 1$ clearly brings out the reciprocity in time and frequency. The smaller the "duration" τ or the more localized the time signal, the larger is its bandwidth or frequency spread. The quantity $\omega_B \tau$ is a measure of the **time-bandwidth product**, a relation analogous to the uncertainty principle of quantum physics. We cannot simultaneously make both duration and bandwidth arbitrarily small.

1.3.3 The Transfer Function and Impulse Response

To find the frequency response of a continuous-time system such as the RC circuit, we must apply cosines with unit magnitude at all possible frequencies. Conceptually, we could apply all of them together and use superposition to find the response in one shot. What does such a combination represent? Since all cosines equal unity at $t = 0$, the combination approaches infinity at the origin. As Figure 1.8 suggests, the combination also approaches zero elsewhere due to the cancellation of the infinitely many equal positive and negative values at all other instants. We give this signal (subject to unit area) the name unit **impulse**. We can find the response of a relaxed system over the entire frequency range by applying a single impulse as the input. In practice, of course, we approximate an impulse by a tall narrow spike.

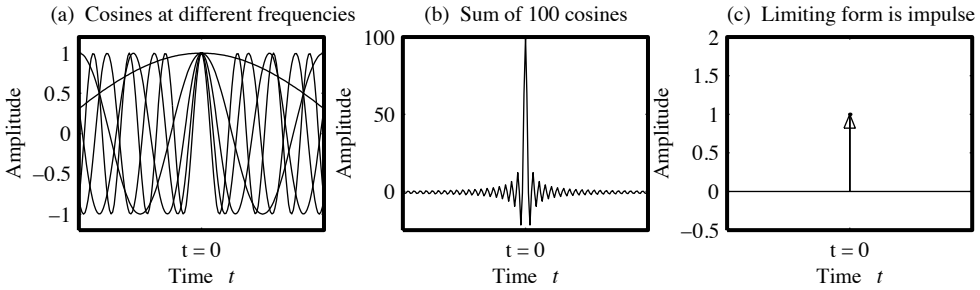


Figure 1.8 Genesis of the impulse function as a sum of sinusoids

The time-domain response to an impulse is called the **impulse response**. A system is completely characterized in the frequency domain by its frequency response or **transfer function**. A system is completely characterized in the time domain by its impulse response. Naturally, the transfer function and impulse response are two equivalent ways of looking at the same system.

1.3.4 Convolution

The idea of decomposing a complicated signal into simpler forms is very attractive for both signal and system analysis. One approach to the analysis of continuous-time systems describes the input as a sum of weighted impulses and finds the response as a sum of weighted impulse responses. This describes the process of **convolution**. Since the response is, in theory, a cumulative sum of infinitely many impulse responses, the convolution operation is actually an integral.

1.3.5 System Analysis in the Frequency Domain

A useful approach to system analysis relies on **transformations**, which map signals and systems into a transformed domain such as the frequency domain. This results in simpler mathematical operations to evaluate system behavior. One of the most useful results is that convolution is replaced by the much simpler

PGR2026-00033
Krisp Technologies EX1024 Page 20

operation of multiplication when we move to a transformed domain, but there is a price to pay. Since the response is evaluated in the transformed domain, we must have the means to remap this response to the time domain through an **inverse transformation**. Examples of this method include phasor analysis (for sinusoids and periodic signals), Fourier transforms, and Laplace transforms. Phasor analysis only allows us to find the steady-state response of relaxed systems to periodic signals. The Fourier transform, on the other hand, allows us to analyze relaxed systems with arbitrary inputs. The **Laplace transform** uses a complex frequency to extend the analysis both to a larger class of inputs and to systems with nonzero initial conditions. Different methods of system analysis allow different perspectives on both the system and the analysis results. Some are more suited to the time domain, others offer a perspective in the frequency domain, and yet others are more amenable to numerical computation.

1.3.6 Frequency-Domain Description of Discrete-Time Signals

It turns out that discrete-time sinusoids differ from their analog cousins in some fundamental ways. A discrete-time sinusoid is not periodic for any choice of frequency, but it has a *periodic spectrum* whose period equals the sampling frequency S. An important consequence of this result is that if the spectrum is periodic for a sampled sinusoid, it should also be periodic for a sampled combination of sinusoids. And since analog signals can be described as a combination of sinusoids (periodic ones by their Fourier series and others by their Fourier transform), their sampled combinations (and consequently any sampled signal) have a periodic spectrum in the frequency domain. The central period, from $-0.5S$ to $0.5S$, corresponds to the true spectrum of the analog signal if the sampling rate S exceeds the Nyquist rate, and to the aliased signal otherwise. The frequency-domain description of discrete-time signals is called the **discrete-time Fourier transform** (DTFT). Its periodicity is a consequence of the fundamental result that sampling a signal in one domain leads to periodicity in the other. Just as a periodic signal has a discrete spectrum, a discrete-time signal has a periodic spectrum. This duality also characterizes several other transforms. If the time signal is both discrete and periodic, its spectrum is also discrete and periodic and describes the **discrete Fourier transform** (DFT), whose computation is speeded up by the so-called **fast Fourier transform** (FFT) algorithms. The DFT is essentially the DTFT evaluated at a finite number of frequencies and is also periodic. The DFT can be used to approximate the spectrum of analog signals from their samples, provided the relations are understood in their proper context using the notion of implied periodicity. The DFT and FFT find extensive applications in fast convolution, signal interpolation, spectrum estimation, and transfer function estimation. Yet another transform for discrete signals is the ***z*-transform**, which may be regarded as a discrete version of the Laplace transform.

1.4 From Concept to Application

The term *filter* is often used to denote systems that process the input in a specified way, suppressing certain frequencies, for example. In practice, no measurement process is perfect or free from disturbances or **noise**. In a very broad sense, a signal may be viewed as something that is desirable, and noise may be regarded as an undesired characteristic that tends to degrade the signal. In this context, filtering describes a signal-processing operation that allows signal enhancement, noise reduction, or increased **signal-to-noise ratio**. Systems for the processing of discrete-time signals are also called **digital filters**. Analog filters are used both in analog signal processing and to band-limit the input to digital signal processors. Digital filters are used for tasks such as interpolation, extrapolation, smoothing, and prediction. Digital signal processing continues to play an increasingly important role in fields that range literally from A (astronomy) to Z (zeugmatography, or magnetic resonance imaging) and encompass applications such as compact disc players, speech recognition, echo cancellation in communication systems, image enhancement, geophysical exploration, and noninvasive medical imaging.

PGR2026-00033
Krisp Technologies EX1024 Page 21

Chapter 2

ANALOG SIGNALS

2.0 Scope and Objectives

Signals convey information and include physical quantities such as voltage, current, and intensity. This chapter deals with one-dimensional analog signals that are continuous functions, usually of time or frequency. It begins with signal classification in various forms, describes how signals can be manipulated by various operations, and quantifies the measures used to characterize signals. It is largely a compendium of facts, definitions, and concepts that are central to an understanding of the techniques of signal processing and system analysis described in subsequent chapters.

2.1 Signals

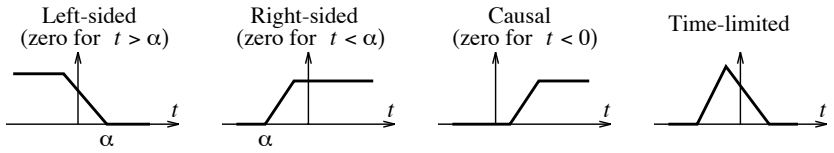
The study of signals allows us to assess how they might be processed to extract useful information. This is indeed what signal processing is all about. An analog signal may be described by a mathematical expression or graphically by a curve or even by a set of tabulated values. Real signals, alas, are not easy to describe quantitatively. They must often be approximated by idealized forms or *models* amenable to mathematical manipulation. It is these models that we concentrate on in this chapter.

2.1.1 Signal Classification by Duration and Area

Signals can be of finite or infinite duration. Finite duration signals are called **time-limited**. Signals of semi-infinite extent may be **right-sided** if they are zero for $t < \alpha$ (where α is finite) or **left-sided** if they are zero for $t > \alpha$. Signals that are zero for $t < 0$ are often called **causal**. The term *causal* is used in analogy with causal systems (discussed in Chapter 4).

REVIEW PANEL 2.1

Signals Can Be Left-Sided, Right-Sided, Causal, or Time-Limited



Piecewise continuous signals possess different expressions over different intervals. **Continuous signals**, such as $x(t) = \sin(t)$, are defined by a single expression for all time.

PGR2026-00033
Krisp Technologies EX1024 Page 22

Periodic signals are infinite-duration signals that repeat the same pattern endlessly. The *smallest* repetition interval is called the **period** T and leads to the formal definition

$$x_p(t) = x_p(t \pm nT) \qquad \text{(for integer } n\text{)} \tag{2.1}$$

One-sided or time-limited signals can never be periodic.

2.1.2 Absolute Area, Energy, and Power

The absolute area of a signal provides useful measures of its *size*. A signal $x(t)$ is called **absolutely integrable** if it possesses finite absolute area:

$$\int_{-\infty}^{\infty} |x(t)|\,dt < \infty \qquad \text{(for an absolutely integrable signal)} \tag{2.2}$$

All *time-limited* functions of finite amplitude have finite absolute area. The criterion of absolute integrability is often used to check for system stability or justify the existence of certain transforms.

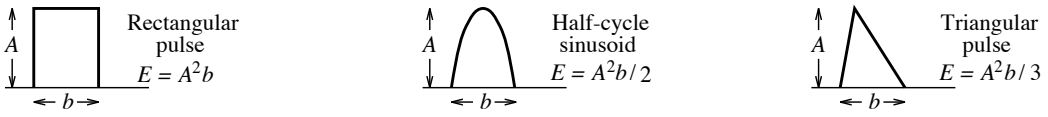
The area of $x^2(t)$ is tied to the power or energy delivered to a 1-Ω resistor. The **instantaneous power** $p_i(t)$ (in watts) delivered to a 1-Ω resistor may be expressed as $p_i(t) = x^2(t)$ where the signal $x(t)$ represents either the voltage across it or the current through it. The total energy E delivered to the 1-Ω resistor is called the **signal energy** (in joules) and is found by integrating the instantaneous power $p_i(t)$ for all time:

$$E = \int_{-\infty}^{\infty} p_i(t)\,dt = \int_{-\infty}^{\infty} |x^2(t)|\,dt \tag{2.3}$$

The absolute value $|x(t)|$ allows this relation to be used for complex-valued signals. The energy of some common signals is summarized in the following review panel.

REVIEW PANEL 2.2

Signal Energy for Three Useful Pulse Shapes (Height = A, Width = b)



The **signal power** P equals the *time average* of the signal energy *over all time*. If $x(t)$ is periodic with period T, the signal power is simply the *average energy per period*, and we have

$$P = \frac{1}{T}\int_T |x(t)|^2\,dt \qquad \text{(for periodic signals)} \tag{2.4}$$

Notation: We use $\int_T$ to mean integration over any convenient one-period duration.

Measures for Periodic Signals

Periodic signals are characterized by several measures. The **duty ratio** of a periodic signal $x_p(t)$ equals the ratio of its pulse width and period. Its **average value** x_{av} equals the average area per period. Its signal power P equals the average energy per period. Its **rms value** x_{rms} equals $\sqrt{P}$ and corresponds to a dc signal with the same power as $x_p(t)$.

$$x_{\text{av}} = \frac{1}{T}\int_T x(t)\,dt \qquad P = \frac{1}{T}\int_T |x(t)|^2\,dt \qquad x_{\text{rms}} = \sqrt{P} \tag{2.5}$$

PGR2026-00033
Krisp Technologies EX1024 Page 23

The average value can never exceed the rms value and thus $x_{\text{av}} \leq x_{\text{rms}}$. Two useful results pertaining to the power in sinusoids and complex exponentials are listed in the following review panel.

REVIEW PANEL 2.3

Signal Power in Analog Harmonics (Sinusoids and Complex Exponentials)

If $x(t) = A\cos(2\pi f_0 t + \theta)$, then $P = 0.5A^2$. If $x(t) = Ae^{\pm j(2\pi f_0 t + \theta)}$, then $P = A^2$.

If $x(t)$ is a nonperiodic power signal, we can compute the signal power (or average value) by averaging its energy (or area) over a finite stretch T_0, and letting $T_0 \to \infty$ to obtain the limiting form

$$P = \lim_{T_0 \to \infty} \frac{1}{T_0} \int_{T_0} |x(t)|^2 \, dt \qquad x_{\text{av}} = \lim_{T_0 \to \infty} \frac{1}{T_0} \int_{T_0} x(t) \, dt \quad \text{(for nonperiodic signals)} \tag{2.6}$$

We emphasize that these limiting forms are useful only for nonperiodic signals.

Energy Signals and Power Signals

The definitions for power and energy serve to classify signals as either energy signals or power signals. A signal with finite energy is called an **energy signal** or **square integrable**. Energy signals have zero signal power since we are averaging finite energy over all (infinite) time. All time-limited signals of finite amplitude are energy signals. Other examples of energy signals include one-sided or two-sided decaying and damped exponentials and damped sinusoids. If you must insist on a formal test, it turns out that $x(t)$ is an energy signal if it is finite valued (with no singularities) and $x^2(t)$ decays to zero *faster than* $1/|t|$ as $|t| \to \infty$.

Signals with finite power are called **power signals**. Power signals possess finite (nonzero) average power and infinite energy. The commonest examples of power signals are periodic signals and their combinations.

Remark: Power signals and energy signals are mutually exclusive because energy signals have zero power and power signals have infinite energy. A signal may of course be neither an energy signal nor a power signal if it has infinite energy and infinite power (examples are signals of polynomial or exponential growth, such as t^n) or infinite energy and zero power (such as $1/\sqrt{t},\ t \geq 1$).

REVIEW PANEL 2.4

Signal Energy E, Signal Power P (if Periodic), and rms Value (if Periodic)

$E = \displaystyle\int_{-\infty}^{\infty} |x(t)|^2 \, dt$ If periodic: $P = \dfrac{1}{T}\displaystyle\int_T |x(t)|^2 \, dt = \dfrac{\text{energy in one period}}{T}$ and $x_{\text{rms}} = \sqrt{P}$

Energy signals: Finite energy, zero power **Power signals:** Nonzero power, infinite energy

EXAMPLE 2.1 (Energy and Power)

(a) Find the signal energy for the signals shown in Figure E2.1A.

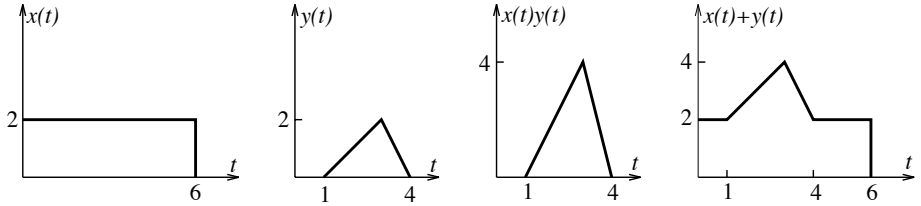


Figure E2.1A The signals for Example 2.1(a)

PGR2026-00033
Krisp Technologies EX1024 Page 24

Using the results of Review Panel 2.2, we find that
The energy in $x(t)$ is $E_x = (2)^2(6) = 24\,\text{J}$.
The energy in $y(t)$ is $E_y = \frac{1}{3}(2)^2(3) = 4\,\text{J}$.
The energy in $f(t) = x(t)y(t)$ is $E_f = \frac{1}{3}(4)^2(3) = 16\,\text{J}$.
The energy in $g(t) = x(t) + y(t)$ may be calculated by writing it as

$$E_g = \int_{\infty}^{\infty} \Big[x^2(t) + y^2(t) + 2x(t)y(t)\Big]\,dt = E_x + E_y + 2\int_{\infty}^{\infty} x(t)y(t)\,dt = 24 + 4 + 12 = 40\,\text{J}$$

Comment: The third term describes twice the area of $x(t)y(t)$ (and equals 12).

(b) The signal $x(t) = 2e^{-t} - 6e^{-2t},\ t > 0$ is an energy signal. Its energy is

$$E_x = \int_0^{\infty} x^2(t)\,dt = \int_0^{\infty} \Big(4e^{-2t} - 24e^{-3t} + 36e^{-4t}\Big)\,dt = 2 - 8 + 9 = 3\,\text{J} \quad \left(\text{Note: } \int_0^{\infty} e^{-\alpha t}\,dt = \frac{1}{\alpha}\right)$$

Comment: As a consistency check, ensure that the energy is always positive!

(c) Find the signal power for the periodic signals shown in Figure E2.1C.

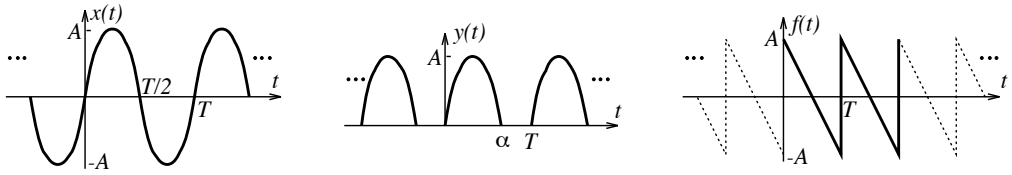


Figure E2.1C The signals for Example 2.1(c)

We use the results of Review Panel 2.2 to find the energy in one period.

- For $x(t)$: The energy E_x in one period is the sum of the energy in each half-cycle. We compute $E_x = \frac{1}{2}A^2(0.5T) + \frac{1}{2}(-A)^2(0.5T) = 0.5A^2T$.
  The power in $x(t)$ is thus $P_x = \dfrac{E_x}{T} = 0.5A^2$.
- For $y(t)$: The energy E_y in one period of $y(t)$ is $E_y = 0.5A^2\alpha$.
  Thus $P_y = \dfrac{E_y}{T} = 0.5A^2\dfrac{\alpha}{T} = 0.5A^2D$ where $D = \dfrac{\alpha}{T}$ is the duty ratio.
  For a half-wave rectified sine, $D = 0.5$ and the signal power equals $0.25A^2$.
  For a full-wave rectified sine, $D = 1$ and the signal power is $0.5A^2$.
- For $f(t)$: The energy E_f in one period is $E_f = \frac{1}{3}A^2(0.5T) + \frac{1}{3}(-A)^2(0.5T) = \dfrac{A^2T}{3}$.
  The signal power is thus $P_f = \dfrac{E_f}{T} = \dfrac{A^2}{3}$.

(d) Let $x(t) = Ae^{j\omega t}$. Since $x(t)$ is complex valued, we work with $|x(t)|$ (which equals A) to obtain

$$P_x = \frac{1}{T}\int_0^T |x(t)|^2\,dt = \frac{1}{T}\int_0^T A^2\,dt = A^2$$

PGR2026-00033
Krisp Technologies EX1024 Page 25

2.2 Operations on Signals

In addition to pointwise sums, differences, and products, the commonly used operations on signals include transformations of amplitude or time.

Amplitude scaling of $x(t)$ by C multiplies all signal values by C to generate $Cx(t)$.

An **amplitude shift** adds a constant K to $x(t)$ *everywhere* (even where it is zero) to form $K + x(t)$.

A **time shift** displaces a signal $x(t)$ in time without changing its shape. Consider the signal $y(t) = x(t-\alpha)$. The value of $y(t)$ at $t = \alpha$ corresponds to the value of $x(t)$ at $t = 0$. In other words, the signal $y(t)$ is a delayed (shifted right by α) replica of $x(t)$. Similarly, the signal $f(t) = x(t+\alpha)$ is an advanced (shifted left by α) replica of $x(t)$. To sketch $y(t) = x(t-\alpha)$, we simply shift the signal $x(t)$ to the right by α. This is equivalent to plotting the signal $x(t)$ on a new time axis t_n at the locations given by $t = t_n - \alpha$ or $t_n = t + \alpha$.

Time scaling speeds up or slows down time and results in signal compression or stretching. The signal $g(t) = x(t/2)$ describes a twofold *expansion* of $x(t)$ since t is slowed down to $t/2$. Similarly, the signal $g(t) = x(3t)$ describes a threefold *compression* of $x(t)$ since t is speeded up to $3t$. To sketch $y(t) = x(\alpha t)$, we compress the signal $x(t)$ by α. This is equivalent to plotting the signal $x(t)$ on a new time axis t_n at the locations given by $t = \alpha t_n$ or $t_n = t/\alpha$.

Reflection or **folding** is just a scaling operation with $\alpha = -1$. It creates the folded signal $x(-t)$ as a mirror image of $x(t)$ about the vertical axis passing through the origin $t = 0$.

Remark: Note that shifting or folding a signal $x(t)$ will not change its area or energy, but time scaling $x(t)$ to $x(\alpha t)$ will reduce both its area and energy by $|\alpha|$.

REVIEW PANEL 2.5

Time Delay $x(t) \Rightarrow x(t-\alpha),\ \alpha > 0$ **and Signal Compression** $x(t) \Rightarrow x(\alpha t),\ |\alpha| > 1$

Delay: $x(t) \Rightarrow x(t-\alpha)$, shift $x(t)$ *right* by α; $t \Rightarrow t_n - \alpha$. New axis: $t_n = t + \alpha$.
Scale: $x(t) \Rightarrow x(\alpha t)$, compress $x(t)$ by $|\alpha|$ (and fold if $\alpha < 0$); $t \Rightarrow \alpha t_n$. New axis: $t_n = t/\alpha$.

2.2.1 Operations in Combination

The signal $y(t) = x(\alpha t - \beta)$ may be generated from $x(t)$ by plotting $x(t)$ against a new time axis t_n where $t = \alpha t_n - \beta$. We may also use the shifting and scaling operations in succession. For example, we can generate the signal $x(2t-6)$ from $x(t)$ in one of two ways:

1. $x(t) \longrightarrow \boxed{\text{delay (shift right) by 6}} \longrightarrow x(t-6) \longrightarrow \boxed{\text{compress by 2}} \longrightarrow x(2t-6)$
2. $x(t) \longrightarrow \boxed{\text{compress by 2}} \longrightarrow x(2t) \longrightarrow \boxed{\text{delay (shift right) by 3}} \longrightarrow x(2t-6)$

In the second form, note that after compression the transformation $x(2t) \Rightarrow x(2t-6) = x[2(t-3)]$ implies a delay of only 3 (and not 6) units (because the signal $x(2t)$ is already compressed). In either case, as a consistency check for the sketch, the new time axis locations t_n are obtained from $t = 2t_n - 6$.

REVIEW PANEL 2.6

Operations in Combination: How to Sketch $x(\alpha t - \beta)$ **(Assuming** $\alpha > 1,\ \beta > 0$**)**

Method 1: Shift *right* by β: $[x(t) \Rightarrow x(t-\beta)]$. Then compress by α: $[x(t-\beta) \Rightarrow x(\alpha t - \beta)]$.
Method 2: Compress by α: $[x(t) \Rightarrow x(\alpha t)]$. Then shift *right* by $\frac{\beta}{\alpha}$: $[x(\alpha t) \Rightarrow x\{\alpha(t - \frac{\beta}{\alpha})\} = x(\alpha t - \beta)]$.
Check: Use $t \Rightarrow \alpha t_n - \beta$ to confirm new locations t_n for the origin $t = 0$ and the end points of $x(t)$.

PGR2026-00033
Krisp Technologies EX1024 Page 26

EXAMPLE 2.2 (Operations on Signals)

(a) Let $x(t) = 1.5t,\ 0 \le t \le 2$, and zero elsewhere. Sketch the following:

$$x(t), \quad f(t) = 1 + x(t-1), \quad g(t) = x(1-t), \quad h(t) = x(0.5t + 0.5), \quad w(t) = x(-2t+2)$$

Refer to Figure E2.2A for the sketches.

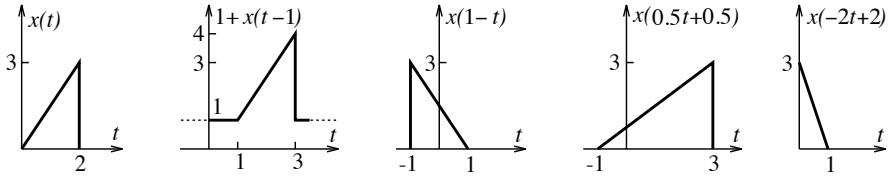


Figure E2.2A The signals for Example 2.2(a)

To generate $f(t) = 1 + x(t-1)$, we delay $x(t)$ by 1 and add a dc offset of 1 unit.

To generate $g(t) = x(1-t)$, we fold $x(t)$ and then shift *right* by 1.
Consistency check: With $t = 1 - t_n$, the edge of $x(t)$ at $t = 2$ translates to $t_n = 1 - t = -1$.

To generate $h(t) = x(0.5t + 0.5)$, first advance $x(t)$ by 0.5 and then stretch by 2 (or first stretch by 2 and then advance by 1).
Consistency check: With $t = 0.5t_n + 0.5$, the edge of $x(t)$ at $t = 2$ translates to $t_n = 2(t - 0.5) = 3$.

To generate $w(t) = x(-2t + 2)$, advance $x(t)$ by 2 units, then shrink by 2 and fold.
Consistency check: With $t = -2t_n + 2$, the edge of $x(t)$ at $t = 2$ translates to $t_n = -0.5(t - 2) = 0$.

(b) Express the signal $y(t)$ of Figure E2.2B in terms of the signal $x(t)$.

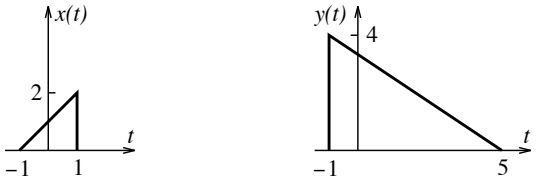


Figure E2.2B The signals $x(t)$ and $y(t)$ for Example 2.2(b)

We note that $y(t)$ is amplitude scaled by 2. It is also a folded, stretched, and shifted version of $x(t)$. If we fold $2x(t)$ and stretch by 3, the pulse edges are at $(-3,\ 3)$. We need a delay of 2 to get $y(t)$, and thus $y(t) = 2x[-(t-2)/3] = 2x(-\frac{t}{3} + \frac{2}{3})$.

Alternatively, with $y(t) = 2x(\alpha t + \beta)$, we use $t = \alpha t_n + \beta$ to solve for α and β by noting that $t = -1$ corresponds to $t_n = 5$ and $t = 1$ corresponds to $t_n = -1$. Then

$$\left.\begin{aligned} -1 &= \ \ 5\alpha + \beta \\ 1 &= -\alpha + \beta \end{aligned}\right\} \quad \Rightarrow \quad \alpha = -\frac{1}{3} \quad \beta = \frac{2}{3}$$

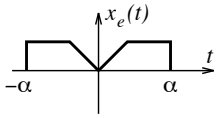
PGR2026-00033
Krisp Technologies EX1024 Page 27

2.3 Signal Symmetry

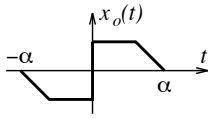
If a signal is identical to its folded version, with $x(t) = x(-t)$, it is called **even symmetric**. We see *mirror symmetry* about the vertical axis passing through the origin $t = 0$. If a signal and its folded version differ only in sign, with $x(t) = -x(-t)$, it is called **odd symmetric**. In either case, the signal extends over symmetric limits about the origin.

REVIEW PANEL 2.7

Symmetric Signals Cover a Symmetric Duration $(-\alpha,\ \alpha)$ About the Origin



Even symmetry: $x_e(t) = x_e(-t)$



Odd symmetry: $x_o(t) = -x_o(-t)$

For an even symmetric signal, the signal values at $t = \alpha$ and $t = -\alpha$ are *equal*. The area of an even symmetric signal is twice the area on either side of the origin. For an odd symmetric signal, the signal values at $t = \alpha$ and $t = -\alpha$ are *equal but opposite in sign* and the signal value at the origin equals zero. The area of an odd symmetric signal over symmetric limits $(-\alpha,\ \alpha)$ is always zero.

REVIEW PANEL 2.8

The Area of Symmetric Signals Over Symmetric Limits $(-\alpha,\ \alpha)$

Odd symmetry: $\displaystyle\int_{-\alpha}^{\alpha} x_o(t)\,dt = 0$ Even symmetry: $\displaystyle\int_{-\alpha}^{\alpha} x_e(t)\,dt = 2\int_{0}^{\alpha} x_e(t)\,dt$

Combinations (sums and products) of symmetric signals are also symmetric under certain conditions as summarized in the following review panel. These results are useful for problem solving.

REVIEW PANEL 2.9

Does the Sum or Product of Two Symmetric Signals Have Any Symmetry?

$x_e(t) + y_e(t)$:	Even symmetry	$x_o(t) + y_o(t)$:	Odd symmetry	$x_e(t) + y_o(t)$:	**No symmetry**
$x_e(t)y_e(t)$:	Even symmetry	$x_o(t)y_o(t)$:	Even symmetry	$x_e(t)y_o(t)$:	Odd symmetry

2.3.1 Even and Odd Parts of Signals

Even symmetry and odd symmetry are mutually exclusive. Consequently, if a signal $x(t)$ is formed by summing an even symmetric signal $x_e(t)$ and an odd symmetric signal $x_o(t)$, it will be devoid of either symmetry. Turning things around, any signal $x(t)$ may be expressed as the sum of an even symmetric part $x_e(t)$ and an odd symmetric part $x_o(t)$:

$$x(t) = x_e(t) + x_o(t) \tag{2.7}$$

To find $x_e(t)$ and $x_o(t)$ from $x(t)$, we fold $x(t)$ and invoke symmetry to get

$$x(-t) = x_e(-t) + x_o(-t) = x_e(t) - x_o(t) \tag{2.8}$$

PGR2026-00033
Krisp Technologies EX1024 Page 28

Adding and subtracting the two preceding equations, we obtain

$$x_e(t) = 0.5x(t) + 0.5x(-t) \qquad x_o(t) = 0.5x(t) - 0.5x(-t) \tag{2.9}$$

Naturally, if $x(t)$ has even symmetry, $x_o(t)$ will equal zero, and if $x(t)$ has odd symmetry, $x_e(t)$ will equal zero.

REVIEW PANEL 2.10

Any Signal Is the Sum of an Even Symmetric Part and an Odd Symmetric Part

$x(t) = x_e(t) + x_o(t)$ where $x_e(t) = 0.5x(t) + 0.5x(-t)$ and $x_o(t) = 0.5x(t) - 0.5x(-t)$

How to implement: Graphically, if possible. **How to check:** Does $x_e(t) + x_o(t)$ give $x(t)$?

2.3.2 Half-Wave Symmetry

Half-wave symmetry is defined only for *periodic signals*. If the value of a periodic signal $x_p(t)$ (with period T) at $t = \alpha$ and at $t = \alpha \pm 0.5T$, half a period away, differs only in sign, $x_p(t)$ is called half-wave symmetric. Thus

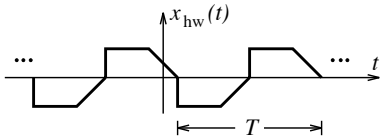
$$x_p(t) = -x_p(t \pm \tfrac{1}{2}T) = -x_p(t \pm nT \pm \tfrac{1}{2}T) \qquad \text{(for integer } n\text{)} \tag{2.10}$$

Half-wave symmetric signals always show *two half-cycles over one period with each half-cycle an inverted replica of the other* and the area of one period equals zero.

REVIEW PANEL 2.11

Half-Wave Symmetry Is Defined Only for Periodic Signals

There are two half-cycles per period.
Each is an inverted replica of the other.
$x_{\text{hw}}(t) = -x_{\text{hw}}(t \pm \frac{T}{2})$



EXAMPLE 2.3 (Even and Odd Parts of Signals)

(a) Find the even and odd parts of the signals $x(t)$ and $y(t)$ shown in Figure E2.3A(1).



Figure E2.3A(1) The signals for Example 2.3(a)

For $x(t)$, we create $0.5x(t)$ and $0.5x(-t)$, then add the two to give $x_e(t)$ and subtract to give $x_o(t)$ as shown in Figure E2.3A(2). Note how the components get added (or subtracted) when there is overlap.

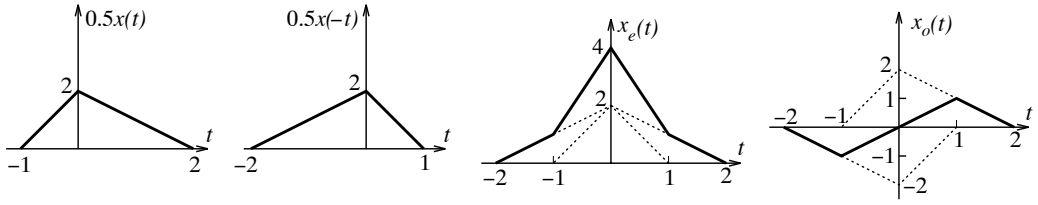


Figure E2.3A(2) The process for finding the even and odd parts of $x(t)$

PGR2026-00033
Krisp Technologies EX1024 Page 29

The process for finding the even and odd parts of $y(t)$ is identical and shown in Figure E2.3A(3).

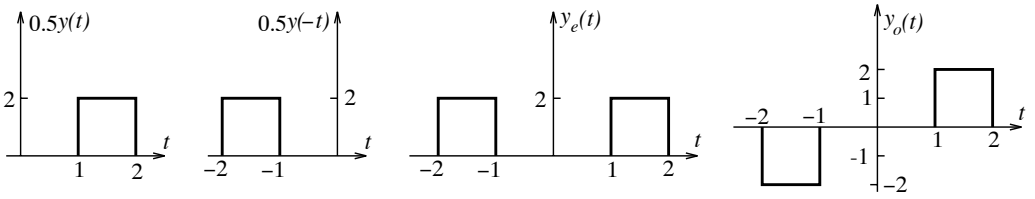


Figure E2.3A(3) The process for finding the even and odd parts of $y(t)$

In either case, as a consistency check, make sure that the even and odd parts display the appropriate symmetry and add up to the original signal.

(b) Let $x(t) = (\sin t + 1)^2$. To find its even and odd parts, we expand $x(t)$ to get

$$x(t) = (\sin t + 1)^2 = \sin^2 t + 2\sin t + 1$$

It is now easy to recognize the even part and odd part as

$$x_e(t) = \sin^2(t) + 1 \qquad x_o(t) = 2\sin(t)$$

2.4 Harmonic Signals and Sinusoids

Sinusoids and **harmonic signals** and are among the most useful periodic signals. They are described by the general forms

$$x_p(t) = A\cos(2\pi f_0 t + \theta) \qquad x(t) = Ae^{j(2\pi f_0 t + \theta)} \tag{2.11}$$

The two forms are related by Euler's identity as follows

$$x_p(t) = \text{ Re}\{Ae^{j(2\pi f_0 t+\theta)}\} = 0.5Ae^{j(2\pi f_0 t+\theta)} + 0.5Ae^{-j(2\pi f_0 t+\theta)} \tag{2.12}$$

The complex exponential form requires two separate plots (its real part and imaginary part, for example) for a graphical description.

REVIEW PANEL 2.12

Euler's Relation in Three Forms

$e^{\pm j\alpha} = 1\angle\pm\alpha = \cos\alpha \pm j\sin\alpha \qquad \cos\alpha = 0.5(e^{j\alpha} + e^{-j\alpha}) \qquad \sin\alpha = -j0.5(e^{j\alpha} - e^{-j\alpha})$

If we write $x_p(t) = A\cos(\omega_0 t + \theta) = A\cos[\omega_0(t - t_p)]$, the quantity $t_p = -\theta/\omega_0$ is called the **phase delay** and describes the time delay in the signal caused by a phase shift of θ.

The various time and frequency measures are related by

$$f_0 = \frac{1}{T} \qquad \omega_0 = \frac{2\pi}{T} = 2\pi f_0 \qquad \theta = \omega_0 t_p = 2\pi f_0 t_p = 2\pi\frac{t_p}{T} \tag{2.13}$$

We emphasize that an analog sinusoid or harmonic signal is *always periodic and unique* for any choice of period or frequency (quite in contrast to digital sinusoids, which we study later).

PGR2026-00033
Krisp Technologies EX1024 Page 30

REVIEW PANEL 2.13

An Analog Harmonic Signal Is Periodic for Any Choice of Frequency

$$x(t) \;=\; A\cos(2\pi f_0 t+\theta) \;=\; A\cos(\omega_0 t+\theta) \;=\; A\cos(2\pi\tfrac{t}{T}+\theta) \;=\; A\cos[2\pi f_0(t-t_p)]$$

Frequency: $\omega_0 = 2\pi f_0$ **Period:** $T = 1/f_0 = 2\pi/\omega_0$ **Phase delay:** $t_p = -\theta/\omega_0$

2.4.1 Combinations of Sinusoids

The **common period** or **time period** T of a combination of sinusoids is the *smallest duration* over which each sinusoid completes an integer number of cycles. It is given by the LCM (least common multiple) of the individual periods. The **fundamental frequency** f_0 is the reciprocal of T and equals the GCD (greatest common divisor) of the individual frequencies. We can find a common period or fundamental frequency only for a **commensurate** combination in which the *ratio* of any two periods (or frequencies) is a **rational fraction** (ratio of integers *with common factors canceled out*).

REVIEW PANEL 2.14

When Is a Sum of Harmonic Signals $y(t) = x_1(t) + x_2(t) + \cdots$ Periodic?

When the ratio of *each pair* of individual frequencies (or periods) is a rational fraction. If periodic:
The **fundamental frequency** f_0 is the GCD of the individual frequencies $f_0 = \text{GCD}(f_1,\ f_2, \ldots)$.
The **common period** is $T = 1/f_0$ *or* the LCM of the individual periods $T = \text{LCM}(T_1,\ T_2, \ldots)$.

For a combination of *sinusoids at different frequencies*, say $y(t) = x_1(t) + x_2(t) + \cdots$, the signal power P_y equals the sum of the individual powers and the rms value equals $\sqrt{P_y}$. The reason is that squaring $y(t)$ produces cross-terms such as $2x_1(t)x_2(t)$, *all of which integrate to zero.*

Almost Periodic Signals

For non-commensurate combinations such as $x(t) = 2\cos(\pi t) + 4\sin(3t)$, where the ratios of the periods (or frequencies) are not rational, we simply cannot find a common period (or frequency), and there is no repetition! Such combinations are called **almost periodic** or **quasiperiodic**.

REVIEW PANEL 2.15

The Signal Power Adds for a Sum of Sinusoids at Different Frequencies

$$\text{If}\quad y(t) = x_1(t) + x_2(t) + \cdots \quad \text{then} \quad P_y = P_{x_1} + P_{x_2} + \cdots \quad \text{and} \quad y_{\text{rms}} = \sqrt{P_y}.$$

EXAMPLE 2.4 (Periodic Combinations)

(a) Consider the signal $x(t) = 2\sin(\frac{2}{3}t) + 4\cos(\frac{1}{2}t) + 4\cos(\frac{1}{3}t - \frac{1}{5}\pi)$.

The periods (in seconds) of the individual components in $x(t)$ are 3π, 4π, and 6π, respectively. The common period of $x(t)$ is $T = \text{LCM}(3\pi, 4\pi, 6\pi) = 12\pi$ seconds. Thus, $\omega_0 = \frac{2\pi}{T} = \frac{1}{6}$ rad/s.

The frequencies (in rad/s) of the individual components are $\frac{2}{3}$, $\frac{1}{2}$, and $\frac{1}{3}$, respectively. The fundamental frequency is $\omega_0 = \text{GCD}(\frac{2}{3},\ \frac{1}{2},\ \frac{1}{3}) = \frac{1}{6}$ rad/s. Thus, $T = \frac{2\pi}{\omega_0} = 12\pi$ seconds.

The signal power is $P_x = 0.5\big(2^2 + 4^2 + 4^2\big) = 36$ W.

The rms value is $x_{\text{rms}} = \sqrt{P_x} = \sqrt{36} = 6$.

PGR2026-00033
Krisp Technologies EX1024 Page 31

(b) The signal $x(t) = \sin(t) + \sin(\pi t)$ is almost periodic because the frequencies $\omega_1 = 1\,\text{rad/s}$ and $\omega_2 = \pi\,\text{rad/s}$ of the two components are non-commensurate.
The signal power is $P_x = 0.5\big(1^2 + 1^2\big) = 1\,\text{W}$.

(c) Consider the signal $x(t) = \sin(t)\sin(\pi t)$.
We rewrite it as $x(t) = \sin(t)\sin(\pi t) = 0.5\cos[(1-\pi)t] - 0.5\cos[(1+\pi)t]$.
Since $\omega_1 = 1 - \pi$ rad/s and $\omega_2 = 1 + \pi$ rad/s are non-commensurate, $x(t)$ is almost periodic.
The signal power is $P_x = 0.5\Big[(0.5)^2 + (0.5)^2\Big] = 0.25\,\text{W}$.

Importance of Harmonic Signals

The importance of harmonic signals and sinusoids stems from the following aspects, which are discussed in detail in later chapters.

1. Any signal can be represented by a combination of harmonics—periodic ones by harmonics at discrete frequencies (Fourier series), and aperiodic ones by harmonics at all frequencies (Fourier transform).
2. The response of a *linear system* (defined in Chapter 4) to a harmonic input is also a harmonic signal at the input frequency. This forms the basis for system analysis in the frequency domain.

2.5 Commonly Encountered Signals

Besides sinusoids, several other signal models are of importance in signal processing. The **unit step** $u(t)$, **unit ramp** $r(t)$, and **signum function** $\text{sgn}(t)$ are piecewise linear signals and defined as

$$u(t) = \begin{cases} 0, & t < 0 \\ 1, & t > 0 \end{cases} \qquad r(t) = tu(t) = \begin{cases} 0, & t < 0 \\ t, & t > 0 \end{cases} \qquad \text{sgn}(t) = \begin{cases} -1, & t < 0 \\ 1, & t > 0 \end{cases} \tag{2.14}$$

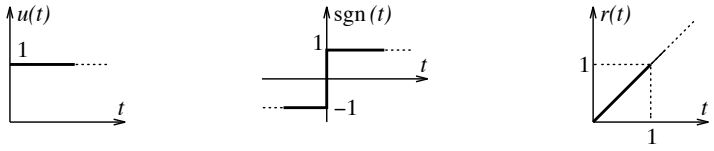
The unit step is discontinuous at $t = 0$, where its value is undefined. With $u(0) = 0.5$, $u(t)$ is called the **Heaviside unit step**. In its general form, $u[f(t)]$ equals 1 if $f(t) > 0$ and 0 if $f(t) < 0$. The signal $u(\alpha - t)$ describes a folded (left-sided) step that is zero past $t = \alpha$. The unit step $u(t)$ may also be regarded as the derivative of the unit ramp $r(t) = tu(t)$, and the unit ramp may be regarded as the running integral of the unit step.

$$u(t) = r'(t) \qquad tu(t) = r(t) = \int_0^t u(\tau)\,d\tau = \int_{-\infty}^t u(\tau)\,d\tau \tag{2.15}$$

The signum function is characterized by a *sign change* at $t = 0$. Its value at $t = 0$ is also undefined and chosen as zero. The step function (or its folded version) can be used to turn a signal $x(t)$ on (or off) while the signum function can be used to switch the polarity of a signal.

REVIEW PANEL 2.16

The Step, Ramp, and Signum Functions Are Piecewise Linear



PGR2026-00033
Krisp Technologies EX1024 Page 32

2.5.1 Pulse Signals

The rectangular pulse **rect**(t) and triangular pulse **tri**(t) are defined as

$$\text{rect}(t) = \begin{cases} 1, & |t| < 0.5 \\ 0, & \text{elsewhere} \end{cases} \quad (\text{width } = 1) \qquad \text{tri}(t) = \begin{cases} 1-|t|, & |t| \le 1 \\ 0, & \text{elsewhere} \end{cases} \quad (\text{width } = 2) \tag{2.16}$$

Both are even symmetric and possess unit area and unit height. The signal $f(t) = \text{rect}(\frac{t-\beta}{\alpha})$ describes a rectangular pulse of width α, centered at $t = \beta$. The signal $g(t) = \text{tri}(\frac{t-\beta}{\alpha})$ describes a triangular pulse of width 2α centered at $t = \beta$. These pulse signals serve as **windows** to limit and shape arbitrary signals. Thus, $h(t) = x(t)\text{rect}(t)$ equals $x(t)$ *abruptly truncated* past $|t| = 0.5$, whereas $x(t)\text{tri}(t)$ equals $x(t)$ *linearly tapered* about $t = 0$ and zero past $|t| = 1$.

REVIEW PANEL 2.17

The Signals rect(t) **and** tri(t) **Have Even Symmetry, Unit Area, and Unit Height**



An arbitrary signal may be represented in different forms each of which has its advantages, depending on the context. For example, we will find signal description by intervals quite useful in convolution, a description by a linear combination of shifted steps and ramps very useful in Laplace transforms and a description by linear combinations of shifted rect and tri functions extremely useful in Fourier transforms.

EXAMPLE 2.5 (Signal Representation)

(a) Sketch the following signals:

$f(t) = u(1-t), \quad g(\lambda) = u(t-\lambda)$ (vs. λ), $\quad x(t) = \text{rect}[0.5(t-1)], \quad y(t) = \text{tri}[\frac{1}{3}(t-1)]$

These signals are sketched in Figure E2.5A.

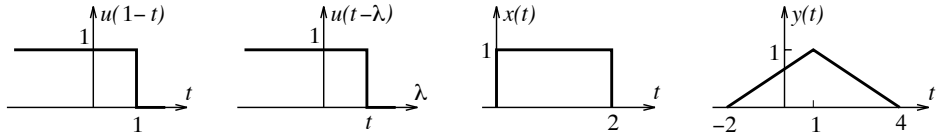


Figure E2.5A The signals for Example 2.5(a)

The signal $f(t) = u(1-t)$ is a folded step, delayed by 1 unit.

The signal $g(\lambda) = u(t-\lambda)$ versus λ is a folded step, shifted t units to the right. Pay particular attention to this form, which will rear its head again when we study convolution.

The signal $x(t) = \text{rect}[0.5(t-1)]$ is a stretched rectangular pulse (of width 2 units) delayed by (or centered at) 1 unit.

The signal $x(t) = \text{tri}[\frac{1}{3}(t-1)]$ is a stretched triangular pulse (of width 6 units) delayed by (or centered at) 1 unit.

PGR2026-00033
Krisp Technologies EX1024 Page 33

(b) Refer to Figure E2.5B. Describe $x(t)$ by a linear combination of rect and /or tri functions, $y(t)$ by a linear combination of steps and/or ramps, and both $x(t)$ and $y(t)$ by intervals.



Figure E2.5B The signals $x(t)$ and $y(t)$ for Example 2.5(b)

- The signal $x(t)$ may be described by a linear combination of shifted rect and tri functions as

$$x(t) = 3\,\text{rect}[\tfrac{1}{6}(t-3)] - 3\,\text{tri}[\tfrac{1}{3}(t-3)]$$

- The signal $y(t)$ may be described by a linear combination of shifted steps and ramps as

$$y(t) = r(t) - r(t-3) - 3u(t)$$

Caution: We could also write $y(t) = t\,\text{rect}[\frac{1}{3}(t-1.5)]$, but this is a product (not a linear combination) and not the preferred form.

- The signals $x(t)$ and $y(t)$ may be described by intervals as

$$x(t) = \begin{cases} 3-t, & 0 < t \le 3 \\ -3+t, & 3 \le t < 6 \\ 0, & \text{elsewhere} \end{cases} \qquad y(t) = \begin{cases} t, & 0 \le t < 3 \\ 0, & \text{elsewhere} \end{cases}$$

2.5.2 The Sinc Function

The sinc function will be encountered in many contexts and is defined as

$$\text{sinc}(t) = \frac{\sin(\pi t)}{\pi t} \tag{2.17}$$

Since the sine term oscillates while the factor $1/t$ decreases with time, $\text{sinc}(t)$ shows decaying oscillations. At $t = 0$, the sinc function produces the indeterminate form 0/0. Using the approximation $\sin(\alpha) \approx \alpha$ (or l'Hôpital's rule), we establish that $\text{sinc}(t) = 1$ in the limit as $t \to 0$:

$$\lim_{t\to 0} \frac{\sin(\pi t)}{\pi t} \approx \frac{\pi t}{\pi t} = 1 \qquad \text{or} \qquad \lim_{t\to 0} \frac{\sin(\pi t)}{\pi t} = \lim_{t\to 0} \frac{\pi\cos(\pi t)}{\pi} = 1$$

A sketch of $\text{sinc}(t)$ is shown in Figure 2.1. We see an even symmetric function with unit height at the origin, a central **mainlobe** between the first zero on either side of the origin and progressively decaying negative and positive **sidelobes**. The minimum peak equals -0.2178, and the next positive peak equals 0.1284. The factor π in the definition ensures both unit area and unit distance between zero crossings, which occur at $t = \pm 1, \pm 2, \pm 3, \ldots$. The signal $\text{sinc}^2(t)$ also has unit area with zeros at unit intervals but is entirely positive. The area of *any sinc or sinc-squared* function equals the area of the triangle inscribed within its central lobe as shown in Figure 2.1. Even though the sinc function is square integrable (an energy signal), it is not absolutely integrable (because it does not decay to zero *faster* than $1/|t|$).

PGR2026-00033
Krisp Technologies EX1024 Page 34

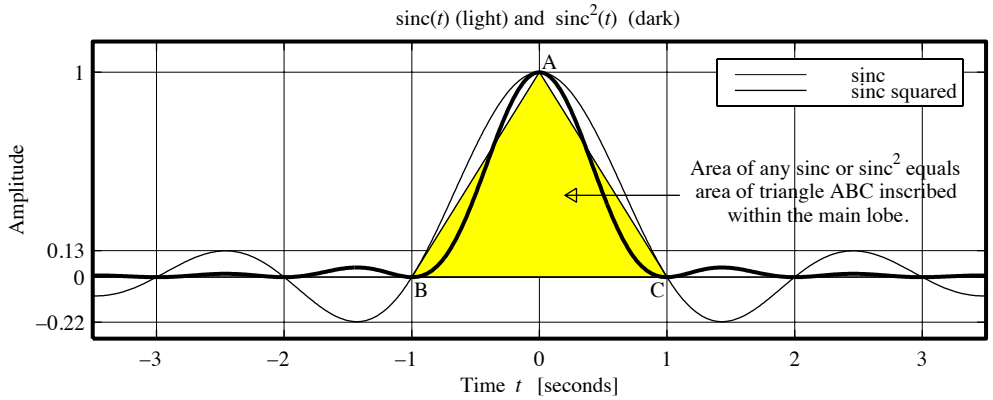


Figure 2.1 The sinc function and the sinc-squared function

REVIEW PANEL 2.18

The Signal $\text{sinc}(t)$ **Has Even Symmetry, Unit Area, and Unit-Spaced Zero Crossings**

$$\text{sinc}(t) = \frac{\sin(\pi t)}{\pi t} \qquad \text{sinc}(0) = 1 \qquad \text{sinc}(t) = 0 \ \text{ when } \ t = \pm 1,\ \pm 2,\ \ldots$$

The sinc function is square integrable (an energy signal) *but not absolutely integrable*.

2.6 The Impulse Function

Loosely speaking, an **impulse** is a tall, narrow spike with finite area. An informal definition of the **unit impulse function**, denoted by $\delta(t)$ and also called a **delta function** or **Dirac function**, is

$$\delta(t) = \begin{cases} 0, & t \neq 0 \\ \infty, & t = 0 \end{cases} \qquad \int_{-\infty}^{\infty} \delta(\tau)\, d\tau = 1 \tag{2.18}$$

It says that $\delta(t)$ is of *zero duration* but possesses *finite area*. To put the best face on this, we introduce a third, equally bizarre criterion that says $\delta(t)$ is unbounded (infinite or undefined) at $t = 0$ (all of which would make any mathematician wince).

REVIEW PANEL 2.19

An Impulse Is a Tall Narrow Spike with Finite Area and Infinite Energy

$$\int_{-\infty}^{\infty} \delta(\tau)\, d\tau = 1 \qquad \delta(t) = \begin{cases} 0, & t \neq 0 \\ \infty, & t = 0 \end{cases}$$

2.6.1 Impulses as Limiting Representations

A rigorous discussion of impulses involves distribution theory and the concept of **generalized functions**, which are beyond our scope. What emerges is that many sequences of ordinary functions behave like an impulse in their limiting form, in the sense of possessing the same *properties* as the impulse. As an example, consider the rectangular pulse $(1/\tau)\text{rect}(t/\tau)$ of width τ and height $1/\tau$ shown in Figure 2.2.

PGR2026-00033
Krisp Technologies EX1024 Page 35

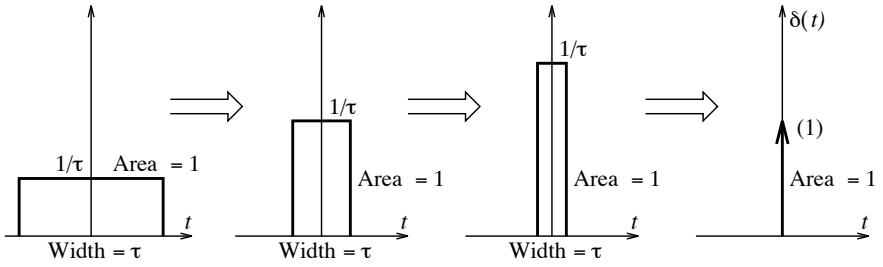


Figure 2.2 The genesis of the impulse function

As we decrease τ, its width shrinks and the height increases proportionately to maintain unit area. As $\tau \to 0$, we get a tall, narrow spike with unit area that satisfies all criteria associated with an impulse.

Signals such as the triangular pulse $\frac{1}{\tau}\,\mathrm{tri}(t/\tau)$, the exponentials $\frac{1}{\tau}\exp(-t/\tau)u(t)$ and $\frac{2}{\tau}\exp(-|t|/\tau)$, the sinc functions $\frac{1}{\tau}\,\mathrm{sinc}(t/\tau)$ and $\frac{1}{\tau}\,\mathrm{sinc}^2(t/\tau)$, the Gaussian $\frac{1}{\tau}\exp[-\pi(t/\tau)^2]$, and the Lorentzian $\tau/[\pi(\tau^2+t^2)]$ all possess unit area, and all are equivalent to the unit impulse $\delta(t)$ as $\tau \to 0$.

The signal $\delta(t-t_0)$ describes an impulse located at $t=t_0$. Its area may be evaluated using any lower and upper limits (say τ_1 and τ_2) that enclose its time of occurrence, t_0:

$$\int_{\tau_1}^{\tau_2} \delta(\tau - t_0)\, d\tau = \begin{cases} 1, & \tau_1 < t_0 < \tau_2 \\ 0, & \text{otherwise} \end{cases} \tag{2.19}$$

Notation: The area of the impulse $A\delta(t)$ equals A and is also called its **strength**. The function $A\delta(t)$ is shown as an arrow with its area A labeled next to the tip. For visual appeal, we make its height proportional to A. Remember, however, that its "height" at $t=0$ is infinite or undefined. An impulse with negative area is shown as an arrow directed downward.

2.6.2 Relating the Impulse and Step Function

If we consider the running integral of the unit impulse $\delta(t)$, it will equal zero until we enclose the impulse at $t=0$ and will equal the area of $\delta(t)$ (i.e. unity) thereafter. In other words, the running integral of $\delta(t)$ describes the unit step $u(t)$. The unit impulse may then be regarded as the derivative of $u(t)$.

REVIEW PANEL 2.20

The Impulse Function Is the Derivative of the Step Function

$$\delta(t) = \frac{d\,u(t)}{dt} \qquad\qquad u(t) = \int_{-\infty}^{t} \delta(t)\, dt$$

2.6.3 Derivative of Signals with Jumps

Since a signal with sudden jumps may be described by step functions, the derivative of such signals must give rise to impulses. For example, the derivative of $x(t) = Au(t) - Bu(t-\alpha)$ is given by $x'(t) = A\delta(t) - B\delta(t-\alpha)$. This describes two impulses whose strengths A and $-B$ correspond to the jumps (up and down) at $t=0$ and $t=\alpha$. In general, *the derivative at a jump results in an impulse whose strength equals the jump*. This leads to a simple rule for finding the *generalized derivative* of an arbitrary signal $x(t)$. For its piecewise continuous portions, we find the ordinary derivative $x'(t)$, and at each jump location in $x(t)$, we include impulses whose strength equals the jump.

PGR2026-00033
Krisp Technologies EX1024 Page 36

REVIEW PANEL 2.21

The Derivative of a Signal with Jumps Includes Impulses at the Jump Locations

To find the derivative: Sketch the signal and check for any jumps.
At each jump location, include impulses (of strength equal to the jump) in the ordinary derivative.

EXAMPLE 2.6 (Derivative of Signals with Jumps)

(a) Sketch the derivative of $y(t) = \text{rect}(t/2)$.

Since $y(t) = \text{rect}(t/2) = u(t+1) - u(t-1)$, its derivative is $y'(t) = \delta(t+1) - \delta(t-1)$.

The signals $y(t)$ and $y'(t)$ are sketched in Figure E2.6.

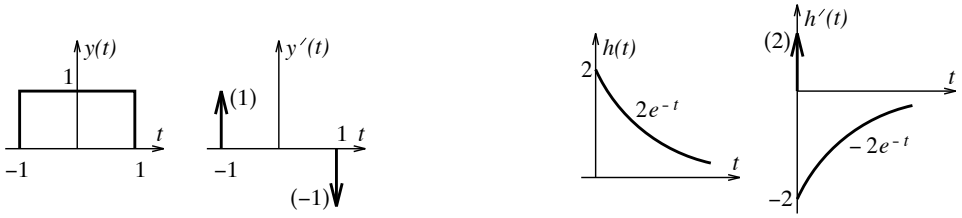


Figure E2.6 The signals for Example 2.6

(b) Find the derivative of $h(t) = 2e^{-t}u(t)$.

The derivative is best found by sketching $h(t)$ as shown. Its ordinary derivative is $-2e^{-t}$. We also include an impulse of strength 2 at the jump location $t = 0$. Thus, $h'(t) = 2\delta(t) - 2e^{-t}u(t)$. The signals $h(t)$ and $h'(t)$ are also sketched in Figure E2.6.

It is not such a good idea to use the chain rule and get $h'(t) = 2e^{-t}\delta(t) - 2e^{-t}u(t)$ because we must then use the product property (see the next section) to simplify the first term as $2e^{-t}\delta(t) = 2\delta(t)$.

2.6.4 Three Properties of the Impulse

Three useful properties of the impulse function are the scaling property, the product property, and the sifting property.

The Scaling Property

An impulse is a tall, narrow spike, and time scaling changes its area. Since $\delta(t)$ has unit area, the time-compressed impulse $\delta(\alpha t)$ should have an area of $\frac{1}{|\alpha|}$. Since the impulse $\delta(\alpha t)$ still occurs at $t = 0$, it may be regarded as an unscaled impulse $\frac{1}{|\alpha|}\delta(t)$. In other words, $\delta(\alpha t) = \frac{1}{|\alpha|}\delta(t)$. A formal proof is based on a change of variable in the defining integral. Since a time shift does not affect areas, we have the more general result:

$$\delta[\alpha(t-\beta)] = \left|\frac{1}{\alpha}\right| \delta(t-\beta) \tag{2.20}$$

PGR2026-00033
Krisp Technologies EX1024 Page 37

The Product Property

This property says that if you multiply a signal $x(t)$ by the impulse $\delta(t-\alpha)$, you get an *impulse* whose strength corresponds to the signal value at $t=\alpha$. The product of a signal $x(t)$ with a tall, narrow pulse of unit area centered at $t=\alpha$ is also a tall, narrow pulse whose height is scaled by the value of $x(t)$ at $t=\alpha$ or $x(\alpha)$. This describes an impulse whose area is $x(\alpha)$ and leads to the product property:

$$x(t)\delta(t-\alpha) = x(\alpha)\delta(t-\alpha) \qquad x(t)\delta(t) = x(0)\delta(t) \tag{2.21}$$

The Sifting Property

The product property immediately suggests that the area of the product $x(t)\delta(t-\alpha) = x(\alpha)\delta(t-\alpha)$ equals $x(\alpha)$. In other words, $\delta(t-\alpha)$ sifts out the value of $x(t)$ at the impulse location $t=\alpha$, and thus,

$$\int_{-\infty}^{\infty} x(t)\delta(t-\alpha)\,dt = x(\alpha) \qquad \int_{-\infty}^{\infty} x(t)\delta(t)\,dt = x(0) \tag{2.22}$$

This extremely important result is called the **sifting property**. It is the sifting action of an impulse (what it does) that purists actually regard as a formal definition of the impulse.

REVIEW PANEL 2.22

The Scaling, Product, and Sifting Properties of an Impulse

Scaling	Product	Sifting
$\delta(\alpha t) = \dfrac{1}{\lvert\alpha\rvert}\delta(t)$	$x(t)\delta(t-\alpha) = x(\alpha)\delta(t-\alpha)$	$\displaystyle\int_{-\infty}^{\infty} x(t)\delta(t-\alpha)\,dt = x(\alpha)$

EXAMPLE 2.7 (Properties of the Impulse Function)

(a) Consider the signal $x(t) = 2r(t) - 2r(t-2) - 4u(t-3)$. Sketch $x(t)$, $f(t) = x(t)\delta(t-1)$, and $g(t) = x'(t)$. Also evaluate $I = \int_{-\infty}^{\infty} x(t)\delta(t-2)\,dt$.

Refer to Figure E2.7A for the sketches.

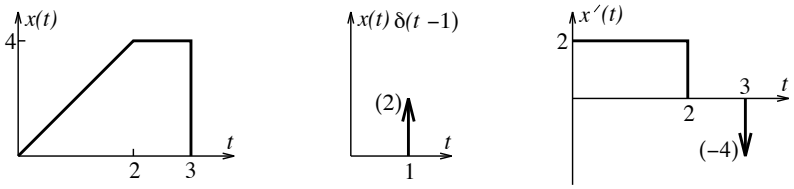


Figure E2.7A The signals for Example 2.7(a)

From the product property, $f(t) = x(t)\delta(t-1) = x(1)\delta(t-1)$. This is an impulse function with strength $x(1) = 2$.

The derivative $g(t) = x'(t)$ includes the ordinary derivative (slopes) of $x(t)$ and an impulse function of strength -4 at $t=3$.

By the sifting property, $I = \int_{-\infty}^{\infty} x(t)\delta(t-2)\,dt = x(2) = 4$.

(b) Evaluate $z(t) = 4t^2\delta(2t-4)$.

The scaling property gives $z(t) = (4t^2)(0.5)\delta(t-2) = 4(2)^2(0.5)\delta(t-2) = 8\delta(t-2)$.

PGR2026-00033
Krisp Technologies EX1024 Page 38

Thus, $z(t)$ is an impulse of strength 8 located at $t = 2$, a function we can sketch.

(c) Evaluate $I_1 = \int_0^\infty 4t^2\delta(t+1)\,dt$.

The result is $I_1 = 0$ because $\delta(t+1)$ (an impulse at $t = -1$) lies outside the limits of integration.

(d) Evaluate $I_2 = \int_{-4}^{2} \cos(2\pi t)\delta(2t+1)\,dt$.

Using the scaling and sifting properties of the impulse, we get

$$I_2 = \int_{-4}^{2} \cos(2\pi t)[0.5\delta(t+0.5)]\,dt = 0.5\cos(2\pi t)|_{t=-0.5} = -0.5$$

2.6.5 Signal Approximation by Impulses

A signal $x(t)$ multiplied by a periodic unit impulse train with period t_s yields the **ideally sampled signal**, $x_I(t)$, as shown in Figure 2.3. The ideally sampled signal is an impulse train described by

$$x_I(t) = x(t)\sum_{k=-\infty}^{\infty}\delta(t-kt_s) = \sum_{k=-\infty}^{\infty}x(kt_s)\delta(t-kt_s) \tag{2.23}$$

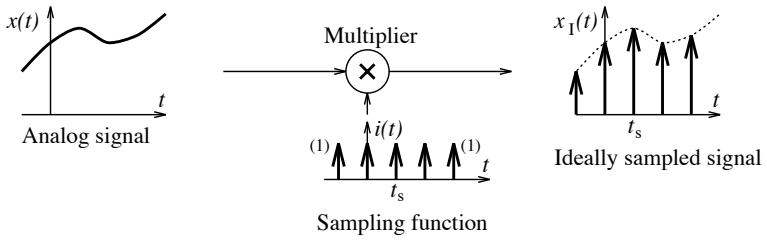


Figure 2.3 The ideally sampled signal is a (nonperiodic) impulse train

Note that even though $x_I(t)$ is an impulse train, *it is not periodic*. The strength of each impulse equals the signal value $x(kt_s)$. This form actually provides a link between analog and digital signals.

REVIEW PANEL 2.23

Ideally Sampled Signal: An Impulse Train Whose Strengths Are the Signal Values

$$x_I(t) = \sum_{k=-\infty}^{\infty}x(kt_s)\delta(t-kt_s)$$

To approximate a smooth signal $x(t)$ by impulses, we section it into narrow rectangular strips of width t_s as shown in Figure 2.4 and replace each strip at the location kt_s by an impulse $t_s x(kt_s)\delta(t-kt_s)$ whose strength equals the area $t_s x(kt_s)$ of the strip. This yields the impulse approximation

$$x(t) \approx \sum_{k=-\infty}^{\infty} t_s x(kt_s)\delta(t-kt_s) \tag{2.24}$$

PGR2026-00033
Krisp Technologies EX1024 Page 39

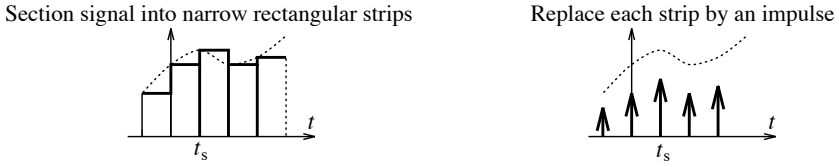


Figure 2.4 Signal approximation by impulses

Naturally, this approximation improves as t_s decreases. As $t_s \to d\lambda \to 0$, kt_s takes on a continuum of values λ, and the summation approaches an integral

$$x(t) = \lim_{t_s \to 0} \sum_{k=-\infty}^{\infty} t_s x(kt_s)\delta(t - kt_s) = \int_{-\infty}^{\infty} x(\lambda)\delta(t - \lambda)\, d\lambda \tag{2.25}$$

A signal $x(t)$ may thus be regarded as a weighted, infinite sum of shifted impulses.

REVIEW PANEL 2.24

We Can Approximate a Signal by a Sum of Impulses (Tall, Narrow Pulses)

The area of the kth impulse is the pulse width t_s times the signal height $x(kt_s)$.

$$x(t) \;\approx\; t_s \sum_{k=-\infty}^{\infty} x(kt_s)\delta(t - kt_s)$$

2.7 The Doublet

The derivative of an impulse is called a **doublet** and denoted by $\delta'(t)$. To see what it represents, consider the triangular pulse $x(t) = \frac{1}{\tau}\,\mathrm{tri}\left(\frac{t}{\tau}\right)$ of Figure 2.5. As $\tau \to 0$, $x(t)$ approaches $\delta(t)$. Its derivative $x'(t)$ should then correspond to $\delta'(t)$. Now, $x'(t)$ is odd and shows two pulses of height $1/\tau^2$ and $-1/\tau^2$ with zero area. As $\tau \to 0$, $x'(t)$ approaches $+\infty$ and $-\infty$ from below and above, respectively. Thus, $\delta'(t)$ is an odd function characterized by zero width, zero area, and amplitudes of $+\infty$ and $-\infty$ at $t = 0$. Formally, we write

$$\delta'(t) = \begin{cases} 0, & t \neq 0 \\ \text{undefined}, & t = 0 \end{cases} \qquad \int_{-\infty}^{\infty} \delta'(t)\, dt = 0 \qquad \delta'(-t) = -\delta'(t) \tag{2.26}$$

The two infinite spikes in $\delta'(t)$ are not impulses (their area is not constant), nor do they cancel. In fact, $\delta'(t)$ is indeterminate at $t = 0$. The signal $\delta'(t)$ is therefore sketched as a set of two spikes, which leads to the name *doublet*. Even though its area $\int \delta'(t)\, dt$ is zero, its absolute area $\int |\delta'(t)|\, dt$ is infinite.

2.7.1 Three Properties of the Doublet

The doublet also possesses three useful properties. The *scaling property comes* about if we write the form for the scaled impulse

$$\delta[\alpha(t - \beta)] = \frac{1}{|\alpha|}\delta(t - \beta) \tag{2.27}$$

and take the derivative of both sides to obtain

$$\alpha\delta'[\alpha(t - \beta)] = \frac{1}{|\alpha|}\delta'(t - \beta) \qquad \delta'[\alpha(t - \beta)] = \frac{1}{\alpha|\alpha|}\delta'(t - \beta) \tag{2.28}$$

PGR2026-00033
Krisp Technologies EX1024 Page 40

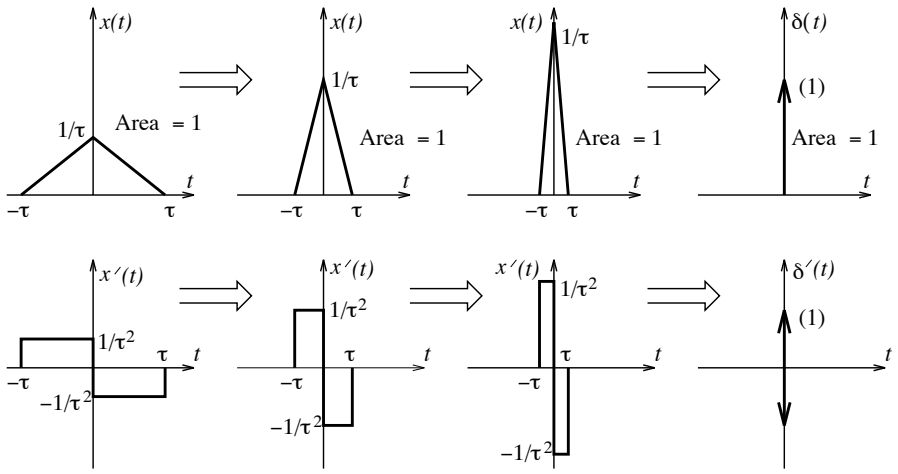


Figure 2.5 The genesis of the doublet

With $\alpha = -1$, we get $\delta'(-t) = -\delta'(t)$. This implies that $\delta'(t)$ is an odd function.

The product property of the doublet has a surprise in store. The derivative of $x(t)\delta(t-\alpha)$ may be described in one of two ways. First, using the rule for derivatives of products, we have

$$\frac{d}{dt}\left[x(t)\delta(t-\alpha)\right] = x'(t)\delta(t-\alpha) + x(t)\delta'(t-\alpha) = x'(\alpha)\delta(t-\alpha) + x(t)\delta'(t-\alpha) \tag{2.29}$$

Second, using the product property of impulses, we also have

$$\frac{d}{dt}\left[x(t)\delta(t-\alpha)\right] = \frac{d}{dt}\left[x(\alpha)\delta(t-\alpha)\right] = x(\alpha)\delta'(t-\alpha) \tag{2.30}$$

Comparing the two equations and rearranging, we get the rather unexpected result

$$x(t)\delta'(t-\alpha) = x(\alpha)\delta'(t-\alpha) - x'(\alpha)\delta(t-\alpha) \tag{2.31}$$

This is the **product property**. Unlike impulses, $x(t)\delta'(t-\alpha)$ does not just equal $x(\alpha)\delta'(t-\alpha)$!

Integrating the two sides that describe the product property, we obtain

$$\int_{-\infty}^{\infty} x(t)\delta'(t-\alpha)\,dt = \int_{-\infty}^{\infty} x(\alpha)\delta'(t-\alpha)\,dt - \int_{-\infty}^{\infty} x'(\alpha)\delta(t-\alpha)\,dt = -x'(\alpha) \tag{2.32}$$

This describes the **sifting property** of doublets. The doublet $\delta'(t-\alpha)$ sifts out the *negative derivative* of $x(t)$ at $t = \alpha$.

Remark: Higher derivatives of $\delta(t)$ obey $\delta^{(n)}(t) = (-1)^n\delta^{(n)}(t)$, are alternately odd and even, and possess zero area. All are limiting forms of the same sequences that generate impulses, provided their ordinary derivatives (up to the required order) exist. None are absolutely integrable. *The impulse is unique in being the only absolutely integrable function from among all its derivatives and integrals* (the step, ramp, etc.).

REVIEW PANEL 2.25

The Scaling, Product, and Sifting Properties of a Doublet

Scaling	Product	Sifting
$\delta'(\alpha t) = \dfrac{1}{\alpha\lvert\alpha\rvert}\delta'(t)$	$x(t)\delta'(t) = x(0)\delta'(t) - x'(0)\delta(t)$	$\displaystyle\int_{-\infty}^{\infty} x(t)\delta'(t)\,dt = -x'(0)$

PGR2026-00033
Krisp Technologies EX1024 Page 41

EXAMPLE 2.8 (Properties of the Doublet)

(a) Consider the signal $x(t) = 2r(t) - 2r(t-2) - 4u(t-3)$. Sketch $x(t)$ and its first and second derivative.

Refer to Figure E2.8A for sketches.

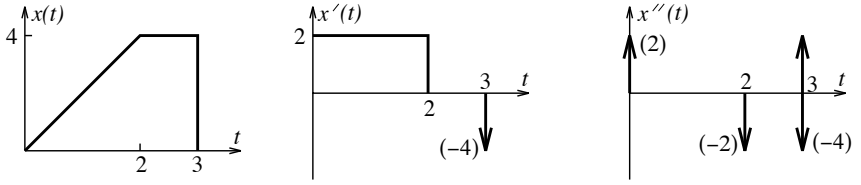


Figure E2.8A The signals for Example 2.8(a)

The first derivative $x'(t)$ results in a rectangular pulse (the ordinary derivative of $x(t)$) and an impulse (due to the jump) at $t = -3$.

The second derivative $x''(t)$ yields two impulses at $t = 0$ and $t = 2$ (the derivative of the rectangular pulse) and a doublet at $t = 3$ (the derivative of the impulse).

(b) What does the signal $x(t) = e^{-t}\delta'(t)$ describe?

By the product property, $x(t) = \delta'(t) - (-1)\delta(t) = \delta'(t) + \delta(t)$.

Thus, $x(t)$ is the sum of an impulse and a doublet!

(c) Evaluate $I = \displaystyle\int_{-2}^{2} [(t-3)\delta(2t+2) + 8\cos(\pi t)\delta'(t-0.5)]\,dt$.

With $\delta(2t+2) = 0.5\delta(t+1)$, the sifting property of impulses and doublets gives

$$I = 0.5(t-3)\Big|_{t=-1} - 8\,\frac{d}{dt}\left[\cos(\pi t)\right]\Big|_{t=0.5} = 0.5(-1-3) + 8\pi\sin 0.5\pi = -2 + 8\pi = 23.1327$$

2.8 Moments

Moments are general measures of signal "size" based on area. The nth **moment** is defined as

$$m_n = \int_{-\infty}^{\infty} t^n x(t)\,dt \tag{2.33}$$

The zeroth moment $m_0 = \int x(t)\,dt$ is just the area of $x(t)$. The normalized first moment $m_x = m_1/m_0$ is called the **mean**. Moments about the mean are called **central moments**. The nth central moment is denoted μ_n.

$$m_x = \frac{\int_{-\infty}^{\infty} tx(t)\,dt}{\int_{-\infty}^{\infty} x(t)\,dt} \qquad \mu_n = \int_{-\infty}^{\infty} (t - m_x)^n x(t)\,dt \tag{2.34}$$

PGR2026-00033
Krisp Technologies EX1024 Page 42

To account for complex valued signals or sign changes, it is often more useful to define moments in terms of the absolute quantities $|x(t)|$ or $|x(t)|^2$. The second central moment μ_2 is called the **variance**. It is often denoted by σ^2 and defined by

$$\mu_2 = \sigma^2 = \frac{m_2}{m_0} - m_x^2 \tag{2.35}$$

The first few moments find widespread application. In physics, if $x(t)$ represents the mass density, then m_x equals the centroid, σ^2 equals the moment of inertia, and σ equals the radius of gyration. In probability theory, if $x(t)$ represents the density function of a random variable, then m_x equals the mean, σ^2 equals the variance, and σ equals the *standard deviation*.

REVIEW PANEL 2.26

The Moment m_n and Central Moment μ_n Are Measures of Signal "Size"

$$m_n = \int_{-\infty}^{\infty} t^n x(t)\,dt \qquad m_x(\text{mean}) = \frac{m_1}{m_0} \qquad \mu_n = \int_{-\infty}^{\infty} (t - m_x)^n x(t)\,dt$$

For power signals, the normalized second moment, m_2/m_0, equals the total power, and σ^2 equals the ac power (the difference between the total power and the dc power). The variance σ^2 may thus be regarded as the power in a signal with its dc offset removed. For an energy signal, m_x is a measure of the effective signal **delay**, and σ is a measure of its effective width, or **duration**.

REVIEW PANEL 2.27

The Delay and Duration of Energy Signals

Delay: $m_x = \dfrac{m_1}{m_0}$ **Duration:** $\sigma = \left(\dfrac{m_2}{m_0} - m_x^2\right)^{1/2}$

EXAMPLE 2.9 (Moments of Energy Signals)

(a) Find the signal delay and duration for the signal $x(t) = \frac{1}{3}$, $0 \le t \le 3$.

The moments of $x(t)$ are $m_0 = \displaystyle\int_0^3 \tfrac{1}{3}\,dt = 1, \quad m_1 = \int_0^3 \tfrac{1}{3}t\,dt = \tfrac{3}{2}, \quad m_2 = \int_0^3 \tfrac{1}{3}t^2\,dt = 3.$

We find that delay $= m_x = \frac{m_1}{m_0} = 1.5,$ and duration $= \sigma = \left(\frac{m_2}{m_0} - m_x^2\right)^{1/2} = (3 - \frac{9}{4})^{1/2} = 0.866.$

(b) Find the signal delay and duration for the signal $x(t) = e^{-t}u(t)$.

The moments of $x(t)$ are $m_0 = \displaystyle\int_0^\infty e^{-t}\,dt = 1, \quad m_1 = \int_0^\infty te^{-t}\,dt = 1, \quad m_2 = \int_0^\infty t^2 e^{-t}\,dt = 2.$

We find that delay $= m_x = \frac{m_1}{m_0} = 1,$ and duration $= \sigma = \left(\frac{m_2}{m_0} - m_x^2\right)^{1/2} = (2 - 1)^{1/2} = 1.$

PGR2026-00033
Krisp Technologies EX1024 Page 43

CHAPTER 2 PROBLEMS

DRILL AND REINFORCEMENT

2.1 (Operations on Signals) For each signal $x(t)$ of Figure P2.1, sketch the following:

(a) $y(t) = x(-t)$ **(b)** $f(t) = x(t+3)$
(c) $g(t) = x(2t-2)$ **(d)** $h(t) = x(2-2t)$
(e) $p(t) = x[0.5(t-2)]$ **(f)** $s(t) = x(-0.5t-1)$
(g) $x_e(t)$ (its even part) **(h)** $x_o(t)$ (its odd part)

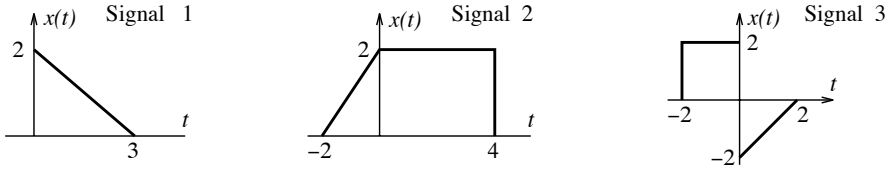


Figure P2.1 Signals for Problem 2.1

2.2 (Symmetry) Find the even and odd parts of each signal $x(t)$.

(a) $x(t) = e^{-t}u(t)$ **(b)** $x(t) = (1+t)^2$ **(c)** $x(t) = [\sin(\pi t) + \cos(\pi t)]^2$

2.3 (Symmetry) Evaluate the following integrals using the concepts of symmetry.

(a) $I = \displaystyle\int_{-3}^{3} (4 - t^2)\sin(5t)\,dt$ **(b)** $I = \displaystyle\int_{-2}^{2} \left[4 - t^3\cos(0.5\pi t)\right] dt$

2.4 (Classification) For each periodic signal shown in Figure P2.4, evaluate the average value x_{av}, the energy E in one period, the signal power P, and the rms value x_{rms}.

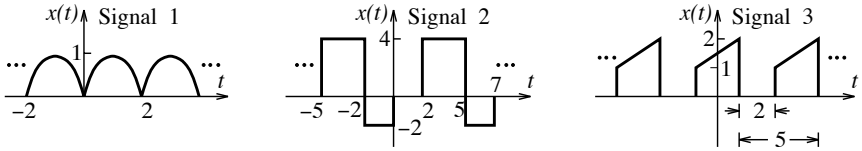


Figure P2.4 Periodic signals for Problem 2.4

2.5 (Signal Classification) Classify each signal as a power signal, energy signal, or neither and find its power or energy as appropriate.

(a) $te^{-t}u(t)$ **(b)** $e^{t}[u(t) - u(t-1)]$ **(c)** $te^{-|t|}$
(d) e^{-t} **(e)** $10e^{-t}\sin(t)u(t)$ **(f)** $\text{sinc}(t)u(t)$

2.6 (Periodic Signals) Classify each of the following signals as periodic, nonperiodic, or almost periodic and find the signal power where appropriate. For each periodic signal, also find the fundamental frequency and the common period.

(a) $x(t) = 4 - 3\sin(12\pi t) + \sin(30\pi t)$ **(b)** $x(t) = \cos(10\pi t)\cos(20\pi t)$
(c) $x(t) = \cos(10\pi t) - \cos(20t)$ **(d)** $x(t) = \cos(10\pi t)\cos(10t)$
(e) $x(t) = 2\cos(8\pi t) + \cos^2(6\pi t)$ **(f)** $x(t) = \cos(2t) - \sqrt{2}\cos(2t - \frac{\pi}{4})$

PGR2026-00033
Krisp Technologies EX1024 Page 44

2.7 (Periodic Signals) Classify each of the following signals as periodic, nonperiodic, or almost periodic and find the signal power where appropriate. For each periodic signal, also find the fundamental frequency and the common period.

(a) $x(t) = 4 - 3\sin^2(12\pi t)$ **(b)** $x(t) = \cos(\pi) + \cos(20\pi t)$ **(c)** $x(t) = \cos(\pi t) + \cos^2(t)$

2.8 (Signal Description) For each signal $x(t)$ shown in Figure P2.8,

(a) Express $x(t)$ by intervals.
(b) Express $x(t)$ as a linear combination of steps and/or ramps.
(c) Express $x(t)$ as a linear combination of rect and/or tri functions.
(d) Sketch the first derivative $x'(t)$.
(e) Find the signal energy in $x(t)$.

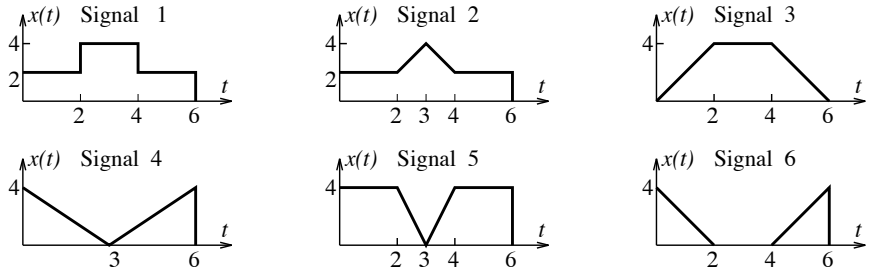


Figure P2.8 Signals for Problem 2.8

2.9 (Sketching Signals) Sketch each of the following signals.

(a) $x(t) = 3\sin(4\pi t - 30°)$ **(b)** $x(t) = 2\,\mathrm{tri}(t-1) + 2\,\mathrm{tri}\left(\frac{t-2}{2}\right)$ **(c)** $x(t) = 9te^{-3|t|}$
(d) $x(t) = t\,\mathrm{rect}(0.5t)$ **(e)** $x(t) = \mathrm{rect}(0.5t)\mathrm{sgn}(t)$ **(f)** $x(t) = e^{-2|t-1|}$
(g) $x(t) = \mathrm{sinc}(\frac{t-3}{3})$ **(h)** $x(t) = \mathrm{rect}(\frac{t}{6})\mathrm{tri}(\frac{t-2}{2})$ **(i)** $x(t) = u(1-|t|)$
(j) $x(t) = u(1-|t|)\mathrm{sgn}(t)$ **(k)** $x(t) = \mathrm{rect}(\frac{t}{6})\mathrm{rect}(\frac{t-2}{4})$ **(l)** $x(t) = u(t-1)u(3-t)$

2.10 (Impulses and Comb Functions) Sketch the following signals. Note that the comb function is a periodic train of unit impulses with unit spacing defined as $\mathrm{comb}(t) = \sum_{k=-\infty}^{\infty} \delta(t-k)$.

(a) $x(t) = 3\delta(t-2)$ **(b)** $x(t) = 3\delta(2t-2)$ **(c)** $x(t) = 3t\delta(t-2)$
(d) $x(t) = 3t\delta(2t-2)$ **(e)** $x(t) = \mathrm{comb}(t/2)$ **(f)** $x(t) = \mathrm{comb}(0.5t)\mathrm{rect}(\frac{t-5}{5})$

2.11 (The Sifting Property) Evaluate the following integrals.

(a) $\int_{-\infty}^{\infty} (4-t^2)\delta(t+3)\,dt$ **(b)** $\int_{-3}^{6} (4-t^2)\delta(t+4)\,dt$ **(c)** $\int_{-3}^{6} (6-t^2)[\delta(t+4) + 2\delta(2t+4)]\,dt$
(d) $\int_{-\infty}^{\infty} \delta(t-2)\delta(x-t)\,dt$ **(e)** $\int_{-3}^{6} (4-t^2)\delta'(t-4)\,dt$ **(f)** $\int_{-\infty}^{t} [\delta(t+2) - \delta(t-2)]\,dt$

2.12 (Generalized Derivatives) Sketch the signals $x(t)$, $x'(t)$, and $x''(t)$ for the following:

(a) $x(t) = 4\,\mathrm{tri}[\frac{1}{2}(t-2)]$ **(b)** $x(t) = e^{-t}u(t)$ **(c)** $x(t) = 2\,\mathrm{rect}(0.5t) + \mathrm{tri}(t)$
(d) $x(t) = e^{-|t|}$ **(e)** $x(t) = (1-e^{-t})u(t)$ **(f)** $x(t) = e^{-2t}\,\mathrm{rect}(\frac{t-1}{2})$

PGR2026-00033
Krisp Technologies EX1024 Page 45

2.13 (Ideally Sampled Signals) Sketch the ideally sampled signal and the impulse approximation for each of the following signals, assuming a sampling interval of $t_s = 0.5$ s.

(a) $x(t) = \text{rect}(t/4)$ **(b)** $x(t) = \text{tri}(t/2)$ **(c)** $x(t) = \sin(\pi t)$ **(d)** $x(t) = t\,\text{rect}(0.5t)$

REVIEW AND EXPLORATION

2.14 (Periodic Signals) Justify the following statements about periodic signals.

(a) The area of the product of two periodic signals is either infinite or zero.

(b) If a periodic signal $x_p(t)$ with period T is both even and half-wave symmetric, it must show even symmetry about $t = \frac{T}{2}$ and odd symmetry about $t = \frac{T}{4}$.

(c) If a periodic signal $x_p(t)$ with period T is both odd and half-wave symmetric, it must show odd symmetry about $t = \frac{T}{2}$ and even symmetry about $t = \frac{T}{4}$.

(d) If a periodic signal $x_p(t)$ is shifted to $x_p(t - \alpha)$ or scaled to $x_p(\alpha t)$, the average value, signal power, and rms value do not change.

2.15 (rms Value) Find the signal power and rms value for a periodic pulse train with peak value A and duty ratio D if the pulse shape is the following:

(a) Rectangular **(b)** Half-sinusoid **(c)** Sawtooth **(d)** Triangular

2.16 (Sketching Signals) Sketch the following signals. Which of these signals (if any) are identical?

(a) $x(t) = r(t-2)$ **(b)** $x(t) = tu(t) - 2u(t-2)$ **(c)** $x(t) = 2u(t) - (t-2)u(t-2)$
(d) $x(t) = tu(t-2) - 2u(t-2)$ **(e)** $x(t) = tu(t-2) - 2u(t)$ **(f)** $x(t) = (t-2)u(t) - u(t-2)$

2.17 (Signals and Derivatives) Sketch each signal $x(t)$ and represent it as a linear combination of step and/or ramp functions where possible.

(a) $x(t) = u(t+1)u(1-t)$ **(b)** $x(t) = \text{sgn}(t)\text{rect}(t)$ **(c)** $x(t) = t\,\text{rect}(t)$
(d) $x(t) = t\,\text{rect}(t-0.5)$ **(e)** $x(t) = t\,\text{rect}(t-2)$ **(f)** $x(t) = u(t+1)u(1-t)\text{tri}(t+1)$

2.18 (Areas) Use the signals $x(t) = \delta(t)$ and $x(t) = \text{sinc}(t)$ as examples to justify the following:

(a) If the area of $|x(t)|$ is finite, the area of $x^2(t)$ need not be finite.

(b) If the area of $x^2(t)$ is finite, the area of $|x(t)|$ need not be finite.

(c) If the area of $x^2(t)$ is finite, the area of $x(t)$ is also finite.

2.19 (Energy) Consider an energy signal $x(t)$, over the range $-3 \le t \le 3$, with energy $E = 12$ J. Find the range of the following signals and compute their signal energy.

(a) $x(3t)$ **(b)** $2x(t)$ **(c)** $x(t-4)$ **(d)** $x(-t)$ **(e)** $x(-\frac{t}{3} - 2)$

2.20 (Power) Consider a periodic signal $x(t)$ with time period $T = 6$ and power $P = 4$ W. Find the time period of the following signals and compute their signal power.

(a) $x(3t)$ **(b)** $2x(t)$ **(c)** $x(t-4)$ **(d)** $x(-t)$ **(e)** $x(-\frac{t}{3} - 2)$

PGR2026-00033
Krisp Technologies EX1024 Page 46

2.21 (Energy) If $y(t) = x_1(t) + x_2(t)$, the area of $y^2(t)$ equals

$$\int y^2(t)\,dt = \int x_1^2(t)\,dt + \int x_2^2(t)\,dt + 2\int x_1(t)x_2(t)\,dt$$

Use this result to show that for any energy signal, the signal energy equals the sum of the energy in its odd and even parts.

2.22 (Areas and Energy) The area of the signal $e^{-t}u(t)$ equals unity. Use this result and the notion of how the area changes upon time scaling to find the following (without formal integration).

(a) The area of $x(t) = e^{-2t}u(t)$.
(b) The energy of $x(t) = e^{-2t}u(t)$.
(c) The area of $y(t) = 2e^{-2t}u(t) - 6e^{-t}u(t)$.
(d) The energy of $y(t) = 2e^{-2t}u(t) - 6e^{-t}u(t)$.

2.23 (Power) Over one period, a periodic signal increases linearly from A to B in T_1 seconds, decreases linearly from B to A in T_2 seconds, and equals A for the rest of the period. What is the power of this periodic signal if its period T is given by $T = 2(T_1 + T_2)$?

2.24 (Power and Energy) Use simple signals such as $u(t)$, $u(t-1)$, $e^{-t}u(t)$ (and others) as examples to argue for or against the following statements.

(a) The sum of energy signals is an energy signal.
(b) The sum of a power and an energy signal is a power signal.
(c) The algebraic sum of two power signals can be an energy signal or a power signal or identically zero.
(d) The product of two energy signals is zero or an energy signal.
(e) The product of a power and energy signal is an energy signal or identically zero.
(f) The product of two power signals is a power signal or identically zero.

2.25 (Switched Periodic Signals) Let $x(t)$ be a periodic signal with power P_x. Show that the power P_y of the switched periodic signal $y(t) = x(t)u(t - t_0)$ is given by $P_y = 0.5P_x$. Use this result to compute the signal power for the following:

(a) $y(t) = u(t)$ **(b)** $y(t) = |\sin(t)|u(-t)$ **(c)** $y(t) = 2\sin(2t)u(t) + 2\sin(t)$
(d) $y(t) = 2 - u(t)$ **(e)** $y(t) = (1 - e^{-t})u(t)$ **(f)** $y(t) = 2\sin(\pi t)u(t) + 2\sin(\pi t)$

2.26 (Power and Energy) Compute the signal energy or signal power as appropriate for each $x(t)$.

(a) $x(t) = e^{2t}u(-t)$ **(b)** $x(t) = e^{t-1}u(-t)$ **(c)** $x(t) = e^{-(1-t)}u(1-t)$
(d) $x(t) = e^{1+2t}u(1-t)$ **(e)** $x(t) = e^{-(1-2t)}u(1-2t)$ **(f)** $x(t) = e^{t}u(-t-2)$
(g) $x(t) = e^{|1-t|}$ **(h)** $x(t) = \text{sinc}(3t-1)$ **(i)** $x(t) = e^{-t^2/2}$

2.27 (Power and Energy) Classify each signal as a power signal, energy signal, or neither, and compute the signal energy or signal power where appropriate.

(a) $x(t) = u(t)$ **(b)** $x(t) = 1 + u(t)$ **(c)** $x(t) = \dfrac{1}{1+|t|}$
(d) $x(t) = \dfrac{1}{1+t^2}$ **(e)** $x(t) = 1 + \cos(\pi t)u(t)$ **(f)** $x(t) = \dfrac{1}{\sqrt{t}},\ t \geq 1$
(g) $x(t) = \dfrac{1}{t},\ t \geq 1$ **(h)** $x(t) = \cos(\pi t)u(t)$ **(i)** $x(t) = \cos(\pi t)u(t) - \cos[\pi(t-4)]u(t-4)$

PGR2026-00033
Krisp Technologies EX1024 Page 47

2.28 (Periodicity) The sum of two periodic signals is periodic if their periods T_1 and T_2 are commensurate. Under what conditions will their product be periodic? Use sinusoids as examples to prove your point.

2.29 (Periodicity) Use Euler's identity to confirm that the signal $x(t) = e^{j2\pi f_0 t}$ is periodic with period $T = 1/f_0$ and use this result in the following:

(a) Is the signal $y(t) = x(2t) + 3x(0.5t)$ periodic? If so, what is its period?
(b) Is the signal $f(t) = 2e^{j16\pi t} + 3e^{-j7\pi t}$ periodic? If so, what is its period?
(c) Is the signal $g(t) = 4e^{j16\pi t} - 5e^{-7\pi t}$ periodic? If so, what is its period?
(d) Is the signal $h(t) = 3e^{j16\pi t} - 2e^{-7\pi}$ periodic? If so, what is its period?
(e) Is the signal $s(t) = \sum_{k=-\infty}^{\infty} X[k]e^{j2\pi k f_0 t}$ periodic? If so, what is its period?

2.30 (Periodicity) It is claimed that each of the following signals is periodic. Verify this claim by sketching each signal and finding its period. Find the signal power for those that are power signals.

(a) $u[\sin(\pi t)]$ **(b)** $\text{sgn}[\sin(\pi t)]$ **(c)** $\delta[\sin(\pi t)]$

2.31 (Periodicity) Let $x(t) = \sin(2\pi t)$.

(a) Sketch $x(t)$ and find its period and power.
(b) Is $y(t) = e^{x(t)}$ periodic? If so, find its period and sketch.
(c) Is $z(t) = e^{jx(t)}$ periodic? If so, find its period and power.

2.32 (Periodicity) It is claimed that the sum of an energy signal $x(t)$ and its shifted (by multiples of T) replicas is a periodic signal with period T. Verify this claim by sketching the following and, for each case, compare the area of one period of the periodic extension with the total area of $x(t)$.

(a) The sum of $x(t) = \text{tri}(t/2)$ and its replicas shifted by $T = 6$.
(b) The sum of $x(t) = \text{tri}(t/2)$ and its replicas shifted by $T = 4$.
(c) The sum of $x(t) = \text{tri}(t/2)$ and its replicas shifted by $T = 3$.

2.33 (Periodic Extension) The sum of an absolutely integrable signal $x(t)$ and its shifted (by multiples of T) replicas is called the *periodic extension* of $x(t)$ with period T. Show that the periodic extension of the signal $x(t) = e^{-\alpha t}u(t)$ with period T is $y(t) = x(t)/(1 - e^{-\alpha T})$. How does the area of one period of $y(t)$ compare with the total area of $x(t)$. Sketch $y(t)$ and find its signal power.

2.34 (Half-Wave Symmetry) Argue that if a half-wave symmetric signal $x(t)$ with period T is made up of several sinusoidal components, each component is also half-wave symmetric over one period T. Which of the following signals show half-wave symmetry?

(a) $x(t) = \cos(2\pi t) + \cos(6\pi t) + \cos(10\pi t)$
(b) $x(t) = 2 + \cos(2\pi t) + \sin(6\pi t) + \sin(10\pi t)$
(c) $x(t) = \cos(2\pi t) + \cos(4\pi t) + \sin(6\pi t)$

2.35 (Derivatives) Each of the following signals is zero outside the interval $-1 \le t \le 1$. Sketch the signals $x(t)$, $x'(t)$, and $x''(t)$.

(a) $x(t) = \cos(0.5\pi t)$ **(b)** $x(t) = 1 + \cos(\pi t)$ **(c)** $x(t) = \text{tri}(t)$ **(d)** $x(t) = 1 - t^2$

PGR2026-00033
Krisp Technologies EX1024 Page 48

2.36 (Practical Signals) Energy signals that are commonly encountered as the response of analog systems include the decaying exponential, the exponentially damped ramp, and the exponentially damped sine. Compute the signal energy for the following:

(a) $x(t) = e^{-t/\tau}u(t)$ **(b)** $x(t) = te^{-t/\tau}u(t)$ **(c)** $f(t) = e^{-t}\sin(2t)u(t)$

2.37 (Time Constant) For an exponential signal of the form $x(t) = Ae^{-t/\tau}u(t)$, the quantity τ is called the time constant and provides a measure of how rapidly the signal decays. A practical estimate of the time it takes for the signal to decay to less than 1% of its initial value is 5τ. What is the actual time it takes for $x(t)$ to decay to exactly 1% of its initial value? How well does the practical estimate compare with the exact result?

2.38 (Rise Time) The rise time is a measure of how fast a signal reaches a constant final value and is commonly defined as the time it takes to rise from 10% to 90% of the final value. Compute the rise time of the following signals.

(a) $x(t) = (1 - e^{-t})u(t)$ **(b)** $y(t) = \begin{cases} \sin(0.5\pi t), & 0 \le t \le 1 \\ 1, & t \ge 1 \end{cases}$

2.39 (Rise Time and Scaling) In practice the rise time t_R of the signal $x(t) = (1 - e^{-t/\tau})u(t)$ is often approximated as $t_R \approx 2.2\tau$.

(a) What is the actual rise time of $x(t)$, and how does it compare with the practical estimate?
(b) Compute the rise time of the signals $f(t) = x(3t)$ and $g(t) = x(t/3)$. How are these values related to the rise time of $x(t)$? Generalize this result to find the rise time of $h(t) = x(\alpha t)$.

2.40 (Settling Time) The settling time is another measure for signals that reach a nonzero final value. The 5% settling time is defined as the time it takes for a signal to settle to within 5% of its final value. Compute the 5% settling time of the following signals.

(a) $x(t) = (1 - e^{-t})u(t)$ **(b)** $y(t) = \begin{cases} \sin(0.5\pi t), & 0 \le t \le 1 \\ 1, & t \ge 1 \end{cases}$

2.41 (Signal Delay) The delay of an energy signal is a measure of how far the signal has been shifted from its mean position and is defined in one of two ways:

$$D_1 = \frac{\int_{-\infty}^{\infty} tx(t)\,dt}{\int_{-\infty}^{\infty} x(t)\,dt} \qquad D_2 = \frac{\int_{-\infty}^{\infty} tx^2(t)\,dt}{\int_{-\infty}^{\infty} x^2(t)\,dt}$$

(a) Verify that the delays D_1 and D_2 of $x(t) = \text{rect}(t)$ are both zero.
(b) Find and compare the delays D_1 and D_2 of the following signals.
(1) $x(t) = \text{rect}(t - 2)$ **(2)** $x(t) = e^{-\alpha t}u(t)$ **(3)** $x(t) = te^{-\alpha t}u(t)$

2.42 (Signal Models) Argue that each of the following models can describe the signal of Figure P2.42, and find the parameters A and α for each model.

(a) $x(t) = Ate^{-\alpha t}u(t)$ **(b)** $x(t) = A(e^{-\alpha t} - e^{-2\alpha t})$ **(c)** $x(t) = \dfrac{At}{\alpha + t^2}$

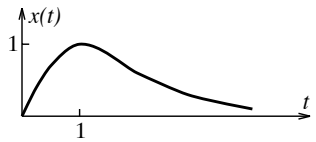


Figure P2.42 Signals for Problem 2.42

PGR2026-00033
Krisp Technologies EX1024 Page 49

2.43 (Instantaneous Frequency) The instantaneous phase of the sinusoid $x(t) = \cos[\phi(t)]$ is defined as its argument $\phi(t)$, and the instantaneous frequency $f_i(t)$ is then defined by the derivative of the instantaneous phase as $f_i(t) = \phi'(t)/2\pi$. Consider the signal $y(t) = \cos(2\pi f_0 t + \theta)$. Show that its instantaneous frequency is constant and equals f_0 Hz.

2.44 (Chirp Signals) Signals whose frequency varies linearly with time are called swept-frequency signals, or chirp signals. Consider the signal $x(t) = \cos[\phi(t)]$ where the time-varying phase $\phi(t)$ is also called the instantaneous phase. The instantaneous frequency $\omega_i(t) = \phi'(t)$ is defined as the derivative of the instantaneous phase (in rad/s).

(a) What is the expression for $\phi(t)$ and $x(t)$ if the instantaneous frequency is to be 10 Hz?

(b) What is the expression for $\phi(t)$ and $x(t)$ if the instantaneous frequency varies linearly from 0 to 100 Hz in 2 seconds?

(c) What is the expression for $\phi(t)$ and $x(t)$ if the instantaneous frequency varies linearly from 50 Hz to 100 Hz in 2 seconds?

(d) Set up a general expression for a chirp signal $x(t)$ whose frequency varies linearly from f_0 Hz to f_1 Hz in t_0 seconds.

2.45 (Chirp Signals) Chirp signals whose frequency varies linearly with time are often used in signal-processing applications (such as radar). Consider the signal $x(t) = \cos(\alpha t^2)$. How does the instantaneous frequency of $x(t)$ vary with time. What value of α will result in a signal whose frequency varies from dc to 10 Hz in 4 seconds?

2.46 (Impulses as Limiting Forms) Argue that the following signals describe the impulse $x(t) = A\delta(t)$ as $\alpha \to 0$. What is the constant A for each signal?

(a) $x(t) = \frac{1}{\alpha} e^{-\pi t^2/\alpha^2}$ **(b)** $x(t) = \dfrac{\alpha}{\alpha^2 + t^2}$ **(c)** $x(t) = \frac{1}{\alpha}\,\mathrm{sinc}(\frac{t}{\alpha})$ **(d)** $x(t) = \frac{1}{\alpha} e^{-|t|/\alpha}$

2.47 (Impulses) It is possible to show that the signal $\delta[f(t)]$ is a string of impulses at the roots t_k of $f(t) = 0$ whose strengths equal $1/\mid f'(t_k) \mid$. Use this result to sketch the following signals.

(a) $x(t) = \delta(t^2 - 3t + 2)$ **(b)** $x(t) = \delta[\sin(\pi t)]$

COMPUTATION AND DESIGN

ctsiggui A GUI for Signal Visualization

The graphical user interface `ctsiggui` allows you to plot a continuous-time signal $x(t)$.
You can display results of the operation $y(t) = A + Bx(Ct + D)$ for your choice of parameters.
You can also display the odd part, even part, derivative, and running integral of $x(t)$.
To explore this routine, type `ctsiggui` at the MATLAB prompt.

mathgui A GUI for Special Mathematical Functions

The graphical user interface `ctsiggui` allows you to plot and visualize the special mathematical functions encountered in this book.
They include the sine integral, Fresnel integrals, Chebyshev functions, and Bessel functions.
To explore this routine, type `mathgui` at the MATLAB prompt.

PGR2026-00033
Krisp Technologies EX1024 Page 50

2.48 (Periodicity) Use MATLAB to plot each signal over the range $0 \le t \le 3$, using a small time step (say, 0.01 s). If periodic, determine the period and compute the signal power (by hand if possible or using MATLAB otherwise).

(a) $x(t) = \sin(2\pi t)$ **(b)** $y(t) = e^{x(t)}$ **(c)** $z(t) = e^{jx(t)}$
(d) $f(t) = \cos[\pi x(t)]$ **(e)** $g(t) = \cos[\pi x^2(t)]$

2.49 (Curious Signals) Let $s(t) = \sin(\pi t)$. Use MATLAB to sketch each of the following signals over the range $-2 \le t \le 6$, using a small time step (say, 0.02 s). Confirm that each signal is periodic and find the period and power.

(a) $x(t) = u[s(t)]$ **(b)** $y(t) = \text{sgn}[s(t)]$ **(c)** $f(t) = \text{sgn}[s(t)] + \text{sgn}[s(t + 0.5)]$
(d) $g(t) = r[s(t)]$ **(e)** $h(t) = e^{s(t)}$

2.50 (Numerical Integration) A crude way to compute the definite integral $I = \int_\alpha^\beta x(t)\,dt$ is to approximate $x(t)$ by N rectangular strips of width t_s and find I as the sum of the areas under each strip:

$$I \approx t_s[x(t_1) + x(t_2) + \cdots + x(t_N)], \qquad t_k = kt_s$$

This is called the *rectangular rule*. The sum of the quantity in brackets can be computed using the MATLAB routine `sum` and then multiplied by t_s to approximate the integral I as the area.

(a) Use the rectangular rule to approximate the integrals of $x(t) = \text{tri}(t)$ and $y(t) = \sin^2(\pi t),\ 0 \le t \le 1$ with $N = 5$ and $N = 10$ and compare with the exact values. Does increasing the the number of strips N lead to more accurate results?

(b) The *trapezoidal rule* uses trapezoidal strips of width t_s to approximate the integral I. Show that this rule leads to the approximation

$$I \approx t_s[\tfrac{1}{2}x(t_1) + x(t_2) + \cdots + \tfrac{1}{2}x(t_N)], \qquad t_k = kt_s$$

(c) Use the trapezoidal rule to approximate the integral of the signals $y(t) = \sin^2(\pi t),\ 0 \le t \le 1$ and $x(t) = \text{tri}(t)$ with $N = 5$ and $N = 10$ and compare with the exact values. Are the results for a given N more accurate than those found by the rectangular rule?

2.51 (Numerical Integration) A crude way to find the running integral $y(t) = \int_0^t x(t)\,dt$ is to approximate $x(t)$ by rectangular strips and compute

$$y(t) \approx t_s \sum_{k=0}^{n} x(kt_s)$$

The cumulative sum can be computed using the MATLAB routine `cumsum` and then multiplied by t_s to obtain $y(t)$. Let $x(t) = 10e^{-t}\sin(2t),\ 0 \le t \le T_0$. Find an exact closed form result for $y(t)$.

(a) Let $T_0 = 2$. Plot the exact expression and the approximate running integral of $x(t)$ using a time step of $t_s = 0.1$ s and $t_s = 0.01$ s. Comment on the differences.

(b) Let $T_0 = 5$. Plot the approximate running integral of $x(t)$ with $t_s = 0.1$ s and $t_s = 0.01$ s. From the graph, can you predict the area of $x(t)$ as $T_0 \to \infty$? Does the error between this predicted value and the value from the exact result decrease as t_s decreases? Should it? Explain.

2.52 (Numerical Derivatives) The derivative $x'(t)$ can be numerically approximated by the slope

$$x'(t)|_{t=nt_s} \approx \frac{x[n] - x[n-1]}{t_s}$$

PGR2026-00033
Krisp Technologies EX1024 Page 51

where t_s is a small time step. The MATLAB routine `diff` yields the difference $x[n] - x[n-1]$ (whose length is 1 less than the length of $x[n]$). Use MATLAB to obtain the approximate derivative of $x(t)$ over $0 \le t \le 3$ with $t_s = 0.1$ s. Compute the exact derivative of $x(t) = 10e^{-t}\sin(2t)u(t)$ and plot both the exact and approximate results on the same plot. Also plot the error between the exact and approximate results. What happens to the error if t_s is halved?

2.53 (Signal Operations) The ADSP routine `operate` can be used to plot scaled and/or shifted versions of a signal $x(t)$. Let $x(t) = 2u(t+1) - r(t+1) + r(t-1)$. Use MATLAB to plot the signals $x(t)$, $y(t) = x(2t-1)$, and $f(t) = x(1-2t)$.

2.54 (Periodic Signals) The ADSP routines `lcm1` and `gcd1` allow us to find the LCM or GCD of an array of rational fractions. Use these routines to find the common period and fundamental frequency of the signal $x(t) = 2\cos(2.4\pi t) - 3\sin(5.4\pi t) + \cos(14.4\pi t - 0.2\pi)$.

2.55 (Energy and Power) The ADSP routine `enerpwr` also computes the energy (or power) if $x(t)$ is a *string* expression (it does not require you to specify a time step). Let $x(t) = 6\,\mathrm{sinc}(2t),\ -0.5 \le t \le 0.5$. Use the routine `enerpwr` to find the energy in this signal and compute the signal power if $x(t)$ describes one period of a periodic signal with period $T = 1.4$ s.

2.56 (Beats) Consider the amplitude modulated signal given by $x(t) = \cos(2\pi f_0 t)\cos[2\pi(f_0 + \Delta f)t]$.

(a) This signal can be expressed in the form $x(t) = A\cos(2\pi f_1 t) + B\cos(2\pi f_2 t)$. How are A, B, f_1, and f_2 related to the parameters of the original signal? Use MATLAB to plot $x(t)$ for $0 \le t \le 1$ s with a time step of $\frac{1}{8192}$ s, $f_0 = 400$ Hz, and $A = B = 1$ for the following values of Δf.

(1) $\Delta f = 0$ Hz **(2)** $\Delta f = 1$ Hz **(3)** $\Delta f = 4$ Hz **(4)** $\Delta f = 5$ Hz

Use the command `sound` to listen to each signal. Explain how you might determine the frequency Δf from the plots.

(b) If $\Delta f \ll f_0$, the frequencies f_1 and f_2 are very close together, and if we listen to the signal, it seems to fade and grow, causing beats. Explain why this happens. Can you determine the frequency of the beats? Would you be able to hear the beat signal even if f_0 is not in the audible range?

2.57 (Chirp Signals) Consider the chirp signal $x(t) = \cos(\pi t^2/6)$. Plot $x(t)$ over $0 \le t \le T$ using a small time step (say, 0.02 s) with $T = 2,\ 6,\ 10$ s. What do the plots reveal as T is increased? Is this signal periodic? Should it be? How does its instantaneous frequency vary with time?

2.58 (Simulating an Impulse) An impulse may be regarded as a tall, narrow spike that arises as a limiting form of many ordinary functions such as the sinc and Gaussian. Consider the Gaussian signal $x(t) = \frac{1}{t_0}e^{-\pi(t/t_0)^2}$. As we decrease t_0, its height increases to maintain a constant area.

(a) Plot $x(t)$ over $-2 \le t \le 2$ for $t_0 = 1, 0.5, 0.1, 0.05, 0.01$ using a time step of $0.1t_0$. What can you say about the symmetry in $x(t)$?

(b) Find the area of $x(t)$ for each t_0, using the MATLAB command `sum`. How does the area change with t_0?

(c) Does $x(t)$ approach an impulse as $t_0 \to 0$? If $x(t) \to A\delta(t)$, what is the value of A?

(d) Plot the (numerical) derivative $x'(t)$ for each t_0, using the MATLAB command `diff`. Find the area of $x'(t)$ for each t_0. How does the area change with t_0? What can you say about the nature of $x(t)$ as $t_0 \to 0$?

PGR2026-00033
Krisp Technologies EX1024 Page 52

Chapter 3

DISCRETE SIGNALS

3.0 Scope and Objectives

This chapter begins with an overview of discrete signals and how they arise. It develops various ways of signal classification, shows how discrete signals can be manipulated by various operations, and quantifies the measures used to characterize such signals. It explains how discrete-time (DT) sinusoids differ from their analog counterparts, introduces the sampling theorem that forms the key to sampling analog signals without loss of information, and concludes with a brief introduction to random signals.

3.1 Discrete Signals

Discrete signals (such as the annual population) may arise naturally or as a consequence of sampling continuous signals (typically at a uniform sampling interval t_s). A sampled or discrete signal $x[n]$ is just an *ordered* sequence of values corresponding to the integer index n that embodies the time history of the signal. It contains no direct information about the **sampling interval** t_s, except through the index n of the sample locations. A discrete signal $x[n]$ is plotted as lines against the index n. When comparing analog signals with their sampled versions, we shall assume that the origin $t = 0$ also corresponds to the sampling instant $n = 0$. We need information about t_s only in a few situations such as plotting the signal explicitly against time t (at $t = nt_s$) or approximating the area of the underlying analog signal from its samples (as $t_s \sum x[n]$).

REVIEW PANEL 3.1

Notation for a Numeric Sequence $x[n]$

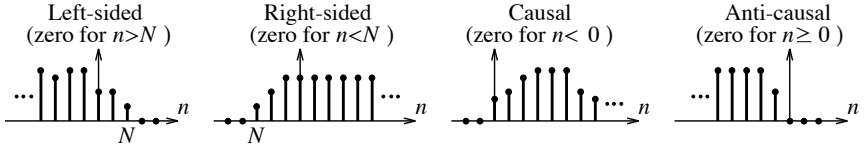
A marker ($\Downarrow$) indicates the origin $n = 0$. Example: $x[n] = \{1,\ 2,\ \overset{\Downarrow}{4},\ 8, \ldots\}$.

Ellipses ($\ldots$) denote infinite extent on either side. Example: $x[n] = \{2,\ 4,\ \overset{\Downarrow}{6},\ 8, \ldots\}$

A discrete signal $x[n]$ is called **right-sided** if it is zero for $n < N$ (where N is finite), **causal** if it is zero for $n < 0$, **left-sided** if it is zero for $n > N$, and **anti-causal** if it is zero for $n \geq 0$.

REVIEW PANEL 3.2

Discrete Signals Can Be Left-Sided, Right-Sided, Causal, or Anti-Causal



PGR2026-00033
Krisp Technologies EX1024 Page 53

A discrete periodic signal repeats every N samples and is described by

$$x[n] = x[n \pm kN], \qquad k = 0, 1, 2, 3, \ldots \tag{3.1}$$

The period N is the *smallest* number of samples that repeats. Unlike its analog counterpart, the period N of discrete signals is always an integer. The common period of a linear combination of periodic discrete signals is given by the least common multiple (LCM) of the individual periods.

REVIEW PANEL 3.3

The Period of Discrete Signals Is Measured as the Number of Samples per Period

The period N is *always an integer*. For combinations, N is the LCM of the individual periods.

3.1.1 Signal Measures

Signal measures for discrete signals are often based on summations. Summation is the discrete-time equivalent of integration. The **discrete sum** S_D, the *absolute sum* S_A, and the *cumulative sum* (running sum) $s_C[n]$ of a signal $x[n]$ are defined by

$$S_D = \sum_{n=-\infty}^{\infty} x[n] \qquad S_A = \sum_{n=-\infty}^{\infty} |x[n]| \qquad s_C[n] = \sum_{k=-\infty}^{n} x[k] \tag{3.2}$$

Signals for which the absolute sum $|x[n]|$ is finite are called **absolutely summable**. For nonperiodic signals, the **signal energy** E is a useful measure. It is defined as the sum of the squares of the signal values

$$E = \sum_{n=-\infty}^{\infty} |x[n]|^2 \tag{3.3}$$

The absolute value allows us to extend this relation to complex-valued signals. Measures for periodic signals are based on averages since their signal energy is infinite. The **average value** x_{av} and **signal power** P of a periodic signal $x[n]$ with period N are defined as the average sum per period and average energy per period, respectively:

$$x_{\text{av}} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] \qquad P = \frac{1}{N} \sum_{n=0}^{N-1} |x[n]|^2 \tag{3.4}$$

Note that the index runs from $n = 0$ to $n = N - 1$ and includes all N samples in one period. Only for nonperiodic signals is it useful to use the limiting forms

$$x_{\text{av}} = \lim_{M \to \infty} \frac{1}{2M+1} \sum_{n=-M}^{M} x[n] \qquad P = \lim_{M \to \infty} \frac{1}{2M+1} \sum_{n=-M}^{M} |x[n]|^2 \tag{3.5}$$

Signals with finite energy are called *energy signals* (or square summable). Signals with finite power are called *power signals*. All periodic signals are power signals.

REVIEW PANEL 3.4

Energy and Power in Discrete Signals

Energy: $E = \sum_{n=-\infty}^{\infty} |x[n]|^2$ **Power (if periodic with period N):** $P = \frac{1}{N} \sum_{n=0}^{N-1} |x[n]|^2$

PGR2026-00033
Krisp Technologies EX1024 Page 54

EXAMPLE 3.1 (Signal Energy and Power)

(a) Find the energy in the signal $x[n] = 3(0.5)^n,\ n \geq 0$.

This describes a one-sided decaying exponential. Its signal energy is

$$E = \sum_{n=-\infty}^{\infty} x^2[n] = \sum_{n=0}^{\infty} |3(0.5)^n|^2 = \sum_{n=0}^{\infty} 9(0.25)^n = \frac{9}{1-0.25} = 12\,\text{J} \quad \left(\text{Note: } \sum_{n=0}^{\infty} \alpha^n = \frac{1}{1-\alpha}\right)$$

(b) Consider the periodic signal $x[n] = 6\cos(2\pi n/4)$ whose period is $N = 4$.

One period of this signal is $x_1[n] = \{\overset{\Downarrow}{6},\ 0,\ -6,\ 0\}$. The average value and signal power of $x[n]$ is

$$x_{\text{av}} = \frac{1}{4}\sum_{n=0}^{3} x[n] = 0 \qquad P = \frac{1}{4}\sum_{n=0}^{3} x^2[n] = \frac{1}{4}\Big(36 + 36\Big) = 18\,\text{W}$$

(c) Consider the periodic signal $x[n] = 6e^{j2\pi n/4}$ whose period is $N = 4$.

This signal is complex-valued, with $|x[n]| = 6$. Its signal power is thus

$$P = \frac{1}{4}\sum_{n=0}^{3} |x[n]|^2 = \frac{1}{4}\Big(36 + 36 + 36 + 36\Big) = 36\,\text{W}$$

3.2 Operations on Discrete Signals

Common operations on discrete signals include elementwise addition and multiplication. In analogy with analog signals, two other useful operations are **time shift** and **folding** or time reversal.

Time Shift: The signal $y[n] = x[n-\alpha]$ describes a delayed version of $x[n]$ for $\alpha > 0$. In other words, if $x[n]$ starts at $n = N$, then its shifted version $y[n] = x[n-\alpha]$ starts at $n = N+\alpha$. Thus, the signal $y[n] = x[n-2]$ is a delayed (shifted right by 2) version of $x[n]$, and the signal $g[n] = x[n+2]$ is an advanced (shifted left by 2) version of $x[n]$. A useful consistency check for sketching shifted signals is based on the fact that if $y[n] = x[n-\alpha]$, a sample of $x[n]$ at the original index n gets relocated to the new index n_N based on the operation $n = n_N - \alpha$.

Folding: The signal $y[n] = x[-n]$ represents a folded version of $x[n]$, a *mirror image* of the signal $x[n]$ about the origin $n = 0$. The signal $y[n] = x[-n-\alpha]$ may be obtained from $x[n]$ in one of two ways:

(a) $x[n] \longrightarrow \boxed{\text{delay (shift right) by } \alpha} \longrightarrow x[n-\alpha] \longrightarrow \boxed{\text{fold}} \longrightarrow x[-n-\alpha]$.

(b) $x[n] \longrightarrow \boxed{\text{fold}} \longrightarrow x[-n] \longrightarrow \boxed{\text{advance (shift left) by } \alpha} \longrightarrow x[-n-\alpha]$.

In either case, a sample of $x[n]$ at the original index n will be plotted at a new index n_N given by $n = -n_N - \alpha$, and this can serve as a consistency check in sketches.

PGR2026-00033
Krisp Technologies EX1024 Page 55

REVIEW PANEL 3.5

Time delay means $x[n] \Rightarrow x[n-M], \quad M > 0$**, and folding means** $x[n] \Rightarrow x[-n]$

You can generate $x[-n-M]$ **from** $x[n]$ **in one of two ways:**

1. Shift *right* M units: $x[n] \Rightarrow x[n-M]$. Then fold: $x[n-M] \Rightarrow x[-n-M]$.

2. Fold: $x[n] \Rightarrow x[-n]$. Then shift *left* M units: $x[-n] \Rightarrow x[-(n+M)] = x[-n-M]$.

Check: Use $n \Rightarrow -n_N - M$ to confirm new locations n_N for the origin $n = 0$ and end points of $x[n]$.

EXAMPLE 3.2 (Operations on Discrete Signals)

Let $x[n] = \{2,\ 3,\ \overset{\Downarrow}{4},\ 5,\ 6,\ 7\}$. Find and sketch the following:

$$y[n] = x[n-3], \quad f[n] = x[n+2], \quad g[n] = x[-n], \quad h[n] = x[-n+1], \quad s[n] = x[-n-2]$$

Here is how we obtain the various signals:

$y[n] = x[n-3] = \{\overset{\Downarrow}{0},\ 2,\ 3,\ 4,\ 5,\ 6,\ 7\}$ (shift $x[n]$ right 3 units)

$f[n] = x[n+2] = \{2,\ 3,\ 4,\ 5,\ \overset{\Downarrow}{6},\ 7\}$ (shift $x[n]$ left 2 units)

$g[n] = x[-n] = \{7,\ 6,\ 5,\ \overset{\Downarrow}{4},\ 3,\ 2\}$ (fold $x[n]$ about $n = 0$)

$h[n] = x[-n+1] = \{7,\ 6,\ \overset{\Downarrow}{5},\ 4,\ 3,\ 2\}$ (fold $x[n]$, then delay by 1)

$s[n] = x[-n-2] = \{7,\ 6,\ 5,\ 4,\ 3,\ \overset{\Downarrow}{2}\}$ (fold $x[n]$, then advance by 2)

Refer to Figure E3.2 for the sketches.

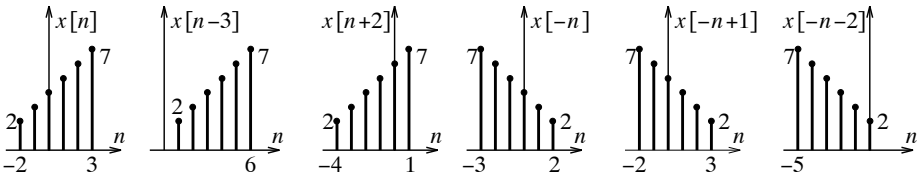


Figure E3.2 The signals for Example 3.2

3.2.1 Symmetry

If a signal $x[n]$ is identical to its mirror image $x[-n]$, it is called an **even symmetric signal**. If $x[n]$ differs from its mirror image $x[-n]$ only in sign, it is called an **odd symmetric** or **antisymmetric signal**. Mathematically,

$$x_e[n] = x_e[-n] \qquad x_o[n] = -x_o[-n] \tag{3.6}$$

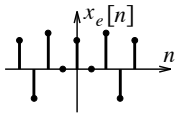
In either case, the signal extends over symmetric limits $-N \le n \le N$. For an odd symmetric signal, $x_o[0] = 0$ and the sum of $x_o[n]$ over symmetric limits $(-\alpha,\ \alpha)$ equals zero:

$$\sum_{k=-M}^{M} x_o[k] = 0 \tag{3.7}$$

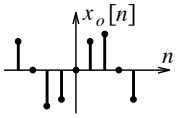
PGR2026-00033
Krisp Technologies EX1024 Page 56

REVIEW PANEL 3.6

Characteristics of Symmetric Signals



Even symmetry: $x_e[n] = x_e[-n]$



Odd symmetry: $x_o[n] = -x_o[-n]$ and $x_o[0] = 0$

3.2.2 Even and Odd Parts of Signals

Even symmetry and odd symmetry are mutually exclusive. Consequently, if a signal $x(t)$ is formed by summing an even symmetric signal $x_e[n]$ and an odd symmetric signal $x_o[n]$, it will be devoid of either symmetry. Turning things around, any signal $x[n]$ may be expressed as the sum of an even symmetric part $x_e[n]$ and an odd symmetric part $x_o[n]$:

$$x[n] = x_e[n] + x_o[n] \tag{3.8}$$

To find $x_e[n]$ and $x_o[n]$ from $x[n]$, we fold $x[n]$ and invoke symmetry to get

$$x[-n] = x_e[-n] + x_o[-n] = x_e[n] - x_o[n] \tag{3.9}$$

Adding and subtracting the two preceding equations, we obtain

$$x_e[n] = 0.5x[n] + 0.5x[-n] \qquad x_o[n] = 0.5x[n] - 0.5x[-n] \tag{3.10}$$

Naturally, if $x[n]$ has even symmetry, $x_o[n]$ will equal zero, and if $x[n]$ has odd symmetry, $x_e[n]$ will equal zero.

REVIEW PANEL 3.7

Any Discrete Signal Is the Sum of an Even Symmetric and an Odd Symmetric Part

$x[n] = x_e[n] + x_o[n]$ where $x_e[n] = 0.5x[n] + 0.5x[-n]$ and $x_o[n] = 0.5x[n] - 0.5x[-n]$

How to implement: Graphically, if possible. **How to check:** Does $x_e[n] + x_o[n]$ give $x[n]$?

EXAMPLE 3.3 (Signal Symmetry)

(a) Let $x[n] = \{4,\ -2,\ \overset{\Downarrow}{4},\ -6\}$. Find and sketch its odd and even parts.

We zero-pad the signal to $x[n] = \{4,\ -2,\ \overset{\Downarrow}{4},\ -6,\ 0\}$ so that it covers symmetric limits. Then

$$0.5x[n] = \{2,\ -1,\ \overset{\Downarrow}{2},\ -3,\ 0\} \qquad 0.5x[-n] = \{0,\ -3,\ \overset{\Downarrow}{2},\ -1,\ 2\}$$

Zero-padding, though not essential, allows us to perform elementwise addition or subtraction with ease to obtain

$$x_e[n] = 0.5x[n] + 0.5x[-n] = \{2,\ -4,\ \overset{\Downarrow}{4},\ -4,\ 2\}$$
$$x_o[n] = 0.5x[n] + 0.5x[-n] = \{2,\ 2,\ \overset{\Downarrow}{0},\ -2,\ -2\}$$

The various signals are sketched in Figure E3.3A. As a consistency check you should confirm that $x_o[0] = 0$, $\sum x_o[n] = 0$, and that the sum $x_e[n] + x_o[n]$ recovers $x[n]$.

PGR2026-00033
Krisp Technologies EX1024 Page 57

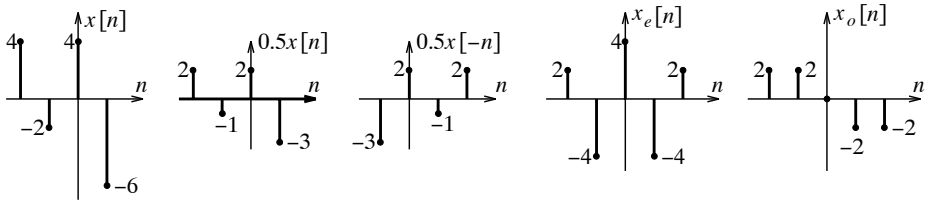


Figure E3.3A The signal $x[n]$ and its odd and even parts for Example 3.3(a)

(b) Let $x[n] = u[n] - u[n-5]$. Find and sketch its odd and even parts.

The signal $x[n]$ and the genesis of its odd and even parts are shown in Figure E3.3B. Note the value of $x_e[n]$ at $n = 0$ in the sketch.

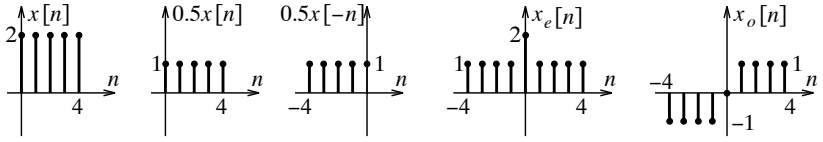


Figure E3.3B The signal $x[n]$ and its odd and even parts for Example 3.3(b)

3.3 Decimation and Interpolation

The time scaling of discrete-time signals must be viewed with care. For discrete-time signals, time scaling is equivalent to decreasing or increasing the signal length. The problems that arise in time scaling are not in what happens but *how* it happens.

3.3.1 Decimation

Suppose $x[n]$ corresponds to an analog signal $x(t)$ sampled at intervals t_s. The signal $y[n] = x[2n]$ then corresponds to the compressed signal $x(2t)$ sampled at t_s and contains only alternate samples of $x[n]$ (corresponding to $x[0], x[2], x[4], \ldots$). We can also obtain $y[n]$ directly from $x(t)$ (not its compressed version) if we sample it at intervals $2t_s$ (or at a sampling rate $S = 1/2t_s$). This means a twofold reduction in the sampling rate. Decimation by a factor of N is equivalent to sampling $x(t)$ at intervals Nt_s and implies an *N-fold reduction in the sampling rate*. The decimated signal $x[Nn]$ is generated from $x[n]$ by retaining every Nth sample corresponding to the indices $k = Nn$ and discarding all others.

3.3.2 Interpolation

If $x[n]$ corresponds to $x(t)$ sampled at intervals t_s, then $y[n] = x[n/2]$ corresponds to $x(t)$ sampled at $t_s/2$ and has twice the length of $x[n]$ with one new sample between adjacent samples of $x[n]$. If an expression for $x[n]$ (or the underlying analog signal) were known, it would be no problem to determine these new sample values. If we are only given the sample values of $x[n]$ (without its analytical form), the best we can do is interpolate between samples. For example, we may choose each new sample value as zero (*zero interpolation*), a constant equal to the previous sample value (*step interpolation*), or the average of adjacent sample values (*linear interpolation*). Zero interpolation is referred to as **up-sampling** and plays an important role in practical interpolation schemes. Interpolation by a factor of N is equivalent to sampling $x(t)$ at intervals t_s/N and implies an *N-fold increase in both the sampling rate and the signal length*.

PGR2026-00033
Krisp Technologies EX1024 Page 58

REVIEW PANEL 3.8

How to Decimate or Interpolate by N

Decimate: Keep every Nth sample (at $n = kN$). This leads to potential loss of information.
Interpolate: Insert $N-1$ new values after each sample. The new sample values may equal zero (zero interpolation), or the previous value (step interpolation), or linearly interpolated values.

Some Caveats

Consider the two sets of operations shown below:

$$x[n] \longrightarrow \boxed{\text{decimate by 2}} \longrightarrow x[2n] \longrightarrow \boxed{\text{interpolate by 2}} \longrightarrow x[n]$$

$$x[n] \longrightarrow \boxed{\text{interpolate by 2}} \longrightarrow x[n/2] \longrightarrow \boxed{\text{decimate by 2}} \longrightarrow x[n]$$

On the face of it, both sets of operations start with $x[n]$ and appear to recover $x[n]$, suggesting that interpolation and decimation are inverse operations. In fact, only the second sequence of operations (interpolation followed by decimation) recovers $x[n]$ exactly. To see why, let $x[n] = \{\overset{\Downarrow}{1},\ 2,\ 6,\ 4,\ 8\}$. Using step interpolation, for example, the two sequences of operations result in

$$\{\overset{\Downarrow}{1},\ 2,\ 6,\ 4,\ 8\} \xrightarrow[n \to 2n]{\text{decimate}} \{\overset{\Downarrow}{1},\ 6,\ 8\} \xrightarrow[n \to n/2]{\text{interpolate}} \{\overset{\Downarrow}{1},\ 1,\ 6,\ 6,\ 8,\ 8\}$$

$$\{\overset{\Downarrow}{1},\ 2,\ 6,\ 4,\ 8\} \xrightarrow[n \to n/2]{\text{interpolate}} \{\overset{\Downarrow}{1},\ 1,\ 2,\ 2,\ 6,\ 6,\ 4,\ 4,\ 8,\ 8\} \xrightarrow[n \to 2n]{\text{decimate}} \{\overset{\Downarrow}{1},\ 2,\ 6,\ 4,\ 8\}$$

We see that decimation is indeed the inverse of interpolation, but the converse is not necessarily true. After all, it is highly unlikely for any interpolation scheme to recover or predict the exact value of the samples that were discarded during decimation. In situations where both interpolation and decimation are to be performed in succession, it is therefore best to interpolate first. In practice, of course, interpolation or decimation should preserve the information content of the original signal, and this imposes constraints on the rate at which the original samples were acquired.

REVIEW PANEL 3.9

Decimation Is the Inverse of Interpolation but Not the Other Way Around

If a signal $x[n]$ is interpolated by N and then decimated by N, we recover the original signal $x[n]$.
If a signal $x[n]$ is first decimated by N and then interpolated by N, we *may not recover* $x[n]$.

Remember: If both interpolation and decimation are required, it is better to interpolate first.

EXAMPLE 3.4 (Decimation and Interpolation)

(a) Let $x[n] = \{1,\ \overset{\Downarrow}{2},\ 5,\ -1\}$. Generate $x[2n]$ and various interpolated versions of $x[n/3]$.

To generate $y[n] = x[2n]$, we remove samples at the odd indices to obtain $x[2n] = \{\overset{\Downarrow}{2},\ -1\}$.

The zero-interpolated signal is $g[n] = x[\frac{n}{3}] = \{1,\ 0,\ 0,\ \overset{\Downarrow}{2},\ 0,\ 0,\ 5,\ 0,\ 0,\ -1,\ 0,\ 0\}$.

The step-interpolated signal is $h[n] = x[\frac{n}{3}] = \{1,\ 1,\ 1,\ \overset{\Downarrow}{2},\ 2,\ 2,\ 5,\ 5,\ 5,\ -1,\ -1,\ -1\}$.

PGR2026-00033
Krisp Technologies EX1024 Page 59

The linearly interpolated signal is $s[n] = x[\frac{n}{3}] = \{1,\ \frac{4}{3},\ \frac{5}{3},\ \underset{\Downarrow}{2},\ 3,\ 4,\ 5,\ 3,\ 1,\ -1,\ -\frac{2}{3},\ -\frac{1}{3}\}$.

In linear interpolation, note that we interpolated the last two values toward zero.

(b) Let $x[n] = \{3,\ 4,\ \underset{\Downarrow}{5},\ 6\}$. Find $g[n] = x[2n-1]$ and the step-interpolated signal $h[n] = x[0.5n-1]$.

In either case, we first find $y[n] = x[n-1] = \{3,\ \underset{\Downarrow}{4},\ 5,\ 6\}$. Then

$g[n] = y[2n] = x[2n-1] = \{\underset{\Downarrow}{4},\ 6\}$.

$h[n] = y[\frac{n}{2}] = x[0.5n-1] = \{3,\ 3,\ \underset{\Downarrow}{4},\ 4,\ 5,\ 5,\ 6,\ 6\}$.

(c) Let $x[n] = \{3,\ 4,\ \underset{\Downarrow}{5},\ 6\}$. Find $y[n] = x[2n/3]$ assuming step interpolation where needed.

Since we require both interpolation and decimation, we first interpolate and then decimate to get

After interpolation: $g[n] = x[\frac{n}{3}] = \{3,\ 3,\ 3,\ 4,\ 4,\ 4,\ \underset{\Downarrow}{5},\ 5,\ 5,\ 6,\ 6,\ 6\}$.

After decimation: $y[n] = g[2n] = x[\frac{2}{3}n] = \{3,\ 3,\ 4,\ \underset{\Downarrow}{5},\ 5,\ 6\}$.

3.3.3 Fractional Delays

Fractional (typically half-sample) delays are sometimes required in practice and can be implemented using interpolation and decimation. If we require that interpolation be followed by decimation and integer shifts, the only sequence of operations that will yield the correct result is interpolation followed by an integer shift and decimation, as summarized in the following review panel.

REVIEW PANEL 3.10

Fractional Delay of $x[n]$ Requires Interpolation, Shift, and Decimation (in That Order)

The idea is to ensure that each operation (interpolation, shift, and decimation) involves integers.
For $x[n] \Rightarrow x[n - \frac{M}{N}] = x[\frac{Nn-M}{N}]$: Interpolate $x[n]$ by N, delay by M, then decimate by N.

EXAMPLE 3.5 (Fractional Delay)

Let $x[n] = \{2,\ 4,\ \underset{\Downarrow}{6},\ 8\}$. Find the signal $y[n] = x[n-0.5]$ assuming linear interpolation where needed.

We first interpolate by 2, then delay by 1, and then decimate by 2 to get

After interpolation: $g[n] = x[\frac{n}{2}] = \{2,\ 3,\ 4,\ 5,\ \underset{\Downarrow}{6},\ 7,\ 8,\ 4\}$ (last sample interpolated to zero).

After delay: $h[n] = g[n-1] = x[\frac{n-1}{2}] = \{2,\ 3,\ 4,\ \underset{\Downarrow}{5},\ 6,\ 7,\ 8,\ 4\}$.

After decimation: $y[n] = h[2n] = x[\frac{2n-1}{2}] = x[n-0.5] = \{\ 3,\ \underset{\Downarrow}{5},\ 7,\ 4\}$.

PGR2026-00033
Krisp Technologies EX1024 Page 60

3.4 Common Discrete Signals

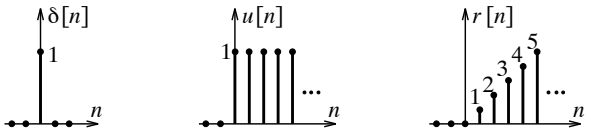
The discrete versions of the **unit impulse** (or **sample**) $\delta[n]$, **unit step** $u[n]$, and **unit ramp** are defined as

$$\delta[n] = \begin{cases} 0, & n \neq 0 \\ 1, & n = 0 \end{cases} \qquad u[n] = \begin{cases} 0, & n < 0 \\ 1, & n \geq 0 \end{cases} \qquad r[n] = nu[n] = \begin{cases} 0, & n < 0 \\ n, & n \geq 0 \end{cases} \tag{3.11}$$

The discrete impulse is just a unit sample at $n = 0$. It is completely free of the kind of ambiguities associated with the analog impulse $\delta(t)$ at $t = 0$. The discrete unit step $u[n]$ also has a well-defined, unique value of $u[0] = 1$ (unlike its analog counterpart $u(t)$). The signal $x[n] = Anu[n] = Ar[n]$ describes a discrete ramp whose "slope" A is given by $x[k] - x[k-1]$, the difference between adjacent sample values.

REVIEW PANEL 3.11

The Discrete Impulse, Step, and Ramp Are Well Defined at $n = 0$



3.4.1 Properties of the Discrete Impulse

The product of a signal $x[n]$ with the impulse $\delta[n-k]$ results in

$$x[n]\delta[n-k] = x[k]\delta[n-k] \tag{3.12}$$

This is just an impulse with strength $x[k]$. The product property leads directly to

$$\sum_{n=-\infty}^{\infty} x[n]\delta[n-k] = x[k] \tag{3.13}$$

This is the sifting property. The impulse extracts the value $x[k]$ from $x[n]$ at the impulse location $n = k$. The product and sifting properties are analogous to their analog counterparts.

3.4.2 Signal Representation by Impulses

Any discrete signal $x[n]$ may be expressed as a sum of shifted impulses $\delta[n-k]$ whose strengths $x[k]$ correspond to the signal values at $n = k$. Thus,

$$x[n] = \sum_{k=-\infty}^{\infty} x[k]\delta[n-k] \tag{3.14}$$

For example, the signals $u[n]$ and $r[n]$ may be expressed as a train of shifted impulses:

$$u[n] = \sum_{k=0}^{\infty} \delta[n-k] \qquad r[n] = \sum_{k=0}^{\infty} k\delta[n-k] \tag{3.15}$$

The signal $u[n]$ may also be expressed as the cumulative sum of $\delta[n]$, and the signal $r[n]$ may be described as the cumulative sum of $u[n]$:

$$u[n] = \sum_{k=-\infty}^{n} \delta[k] \qquad r[n] = \sum_{k=-\infty}^{n} u[k] \tag{3.16}$$

PGR2026-00033
Krisp Technologies EX1024 Page 61

3.4.3 Discrete Pulse Signals

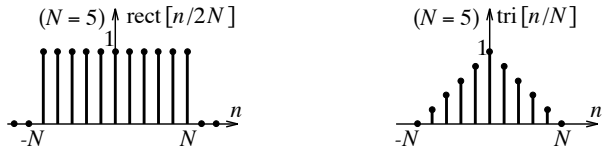
The discrete versions of the rectangular pulse **rect**$(n/2N)$ and triangular pulse **tri**(n/N) are given by

$$\text{rect}\Big(\frac{n}{2N}\Big) = \begin{cases} 1, & |n| \le N \\ 0, & \text{elsewhere} \end{cases} \qquad \text{tri}\Big(\frac{n}{N}\Big) = \begin{cases} 1 - \frac{|n|}{N}, & |n| \le N \\ 0, & \text{elsewhere} \end{cases} \tag{3.17}$$

The signal $\text{rect}(\frac{n}{2N})$ has $2N+1$ unit samples over $-N \le n \le N$. The factor $2N$ in $\text{rect}(\frac{n}{2N})$ gets around the problem of having to deal with half-integer values of n when N is odd. The signal $x[n] = \text{tri}(n/N)$ also has $2N+1$ samples over $-N \le n \le N$, with the end samples $x[N]$ and $x[-N]$ being zero.

REVIEW PANEL 3.12

The Discrete rect and tri Functions



EXAMPLE 3.6 (Describing Sequences and Signals)

(a) Let $x[n] = (2)^n$ and $y[n] = \delta[n-3]$. Find $z[n] = x[n]y[n]$ and evaluate the sum $A = \sum z[n]$.

The product, $z[n] = x[n]y[n] = (2)^3\delta[n-3] = 8\delta[n-3]$, is an impulse.

The sum, $A = \sum z[n]$, is given by $\sum (2)^n\delta[n-3] = (2)^3 = 8$.

(b) Mathematically describe the signals of Figure E3.6B in at least two different ways.

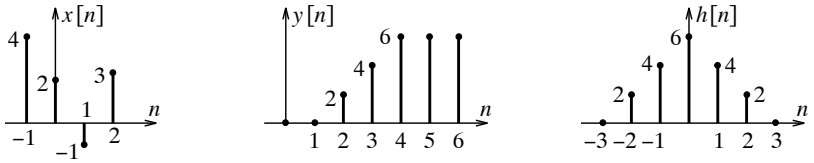


Figure E3.6B The signals for Example 3.6(b)

1. The signal $x[n]$ may be described as the sequence $x[n] = \{4,\ \underset{\Downarrow}{2},\ -1,\ 3\}$.
   It may also be written as $x[n] = 4\delta[n+1] + 2\delta[n] - \delta[n-1] + 3\delta[n-2]$.
2. The signal $y[n]$ may be represented variously as
   A numeric sequence: $y[n] = \{0,\ 0,\ 2,\ 4,\ 6,\ 6,\ 6\}$.
   A sum of shifted impulses: $y[n] = 2\delta[n-2] + 4\delta[n-3] + 6\delta[n-4] + 6\delta[n-5] + 6\delta[n-6]$.
   A sum of steps and ramps: $y[n] = 2r[n-1] - 2r[n-4] - 6u[n-7]$.
   Note carefully that the argument of the step function is $[n-7]$ (and not $[n-6]$).
3. The signal $h[n]$ may be described as $h[n] = 6\,\text{tri}(n/3)$ or variously as
   A numeric sequence: $h[n] = \{0,\ 2,\ 4,\ \underset{\Downarrow}{6},\ 4,\ 2,\ 0\}$.
   A sum of impulses: $h[n] = 2\delta[n+2] + 4\delta[n+1] + 6\delta[n] + 4\delta[n-1] + 2\delta[n-2]$.
   A sum of steps and ramps: $h[n] = 2r[n+3] - 4r[n] + 2r[n-3]$.

PGR2026-00033
Krisp Technologies EX1024 Page 62

3.4.4 The Discrete Sinc Function

The discrete sinc function is defined by

$$\text{sinc}\Big(\frac{n}{N}\Big) = \frac{\sin(n\pi/N)}{(n\pi/N)}, \qquad \text{sinc}(0) = 1 \tag{3.18}$$

The signal $\text{sinc}(n/N)$ equals zero at $n = kN$, $k = \pm 1, \pm 2, \ldots$. At $n = 0$, $\text{sinc}(0) = 0/0$ and cannot be evaluated in the limit since n can take on only integer values. *We therefore define* $\text{sinc}(0) = 1$. The *envelope* of the sinc function shows a mainlobe and gradually decaying sidelobes. The definition of $\text{sinc}(n/N)$ also implies that $\text{sinc}(n) = \delta[n]$.

3.4.5 Discrete Exponentials

Discrete exponentials are often described using a rational base. For example, the signal $x[n] = 2^n u[n]$ shows exponential growth while $y[n] = (0.5)^n u[n]$ is a decaying exponential. The signal $f[n] = (-0.5)^n u[n]$ shows values that alternate in sign. The exponential $x[n] = \alpha^n u[n]$, where $\alpha = re^{j\theta}$ is complex, may be described using the various formulations of a complex number as

$$x[n] = \alpha^n u[n] = \left(re^{j\theta}\right)^n u[n] = r^n e^{jn\theta} u[n] = r^n[\cos(n\theta) + j\sin(n\theta)]u[n] \tag{3.19}$$

This complex-valued signal requires two separate plots (the real and imaginary parts, for example) for a graphical description. If $0 < r < 1$, $x[n]$ describes a signal whose real and imaginary parts are exponentially decaying cosines and sines. If $r = 1$, the real and imaginary parts are pure cosines and sines with a peak value of unity. If $r > 1$, we obtain exponentially growing sinusoids.

3.5 Discrete-Time Harmonics and Sinusoids

If we sample an analog sinusoid $x(t) = \cos(2\pi f_0 t)$ at intervals of t_s corresponding to a sampling rate of $S = 1/t_s$ samples/s (or S Hz), we obtain the sampled sinusoid

$$x[n] = \cos(2\pi f n t_s + \theta) = \cos(2\pi n \tfrac{f}{S} + \theta) = \cos(2\pi n F + \theta) \tag{3.20}$$

The quantities f and $\omega = 2\pi f$ describe analog frequencies. The normalized frequency $F = f/S$ is called the **digital frequency** and has units of cycles/sample. The frequency $\Omega = 2\pi F$ is the digital *radian* frequency with units of radians/sample. The various analog and digital frequencies are compared in Figure 3.1. Note that the analog frequency $f = S$ (or $\omega = 2\pi S$) corresponds to the digital frequency $F = 1$ (or $\Omega = 2\pi$).

REVIEW PANEL 3.13

The Digital Frequency Is the Analog Frequency Normalized by Sampling Rate S

$$F(\text{cycles/sample}) = \frac{f(\text{cycles/sec})}{S(\text{samples/sec})} \qquad \Omega(\text{radians/sample}) = \frac{\omega(\text{radians/sec})}{S(\text{samples/sec})} = 2\pi F$$

3.5.1 Discrete-Time Sinusoids Are Not Always Periodic in Time

An analog sinusoid $x(t) = \cos(2\pi f t + \theta)$ has two remarkable properties. It is unique for every frequency. And it is periodic in time for every choice of the frequency f. Its sampled version, however, is a beast of a different kind.

PGR2026-00033
Krisp Technologies EX1024 Page 63

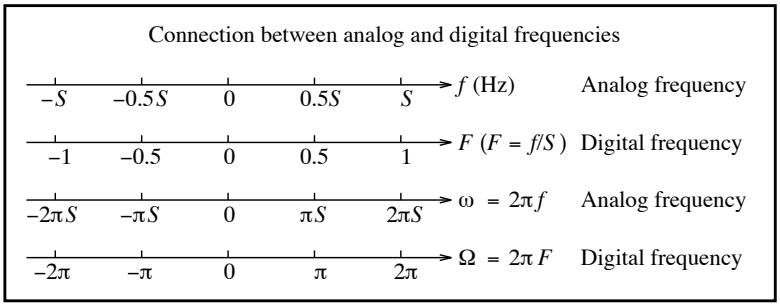


Figure 3.1 Comparison of analog and digital frequencies

Are all discrete-time sinusoids and harmonics periodic in time? Not always! To understand this idea, suppose $x[n]$ is periodic with period N such that $x[n] = x[n + N]$. This leads to

$$\cos(2\pi nF + \theta) = \cos[2\pi(n + N)F + \theta] = \cos(2\pi nF + \theta + 2\pi NF) \tag{3.21}$$

The two sides are equal provided NF equals an integer k. In other words, F must be a *rational fraction* (ratio of integers) of the form k/N. What we are really saying is that a DT sinusoid is *not* always periodic but only if its digital frequency is a ratio of integers or a *rational fraction*. The period N equals the denominator of k/N, provided common factors have been canceled from its numerator and denominator. The significance of k is that it takes k full periods of the analog sinusoid to yield one full period of the sampled sinusoid. The common period of a combination of periodic DT sinusoids equals the least common multiple (LCM) of their individual periods. If F is not a rational fraction, there is no periodicity, and the DT sinusoid is classified as nonperiodic or *almost periodic*. Examples of periodic and nonperiodic DT sinusoids appear in Figure 3.2. Even though a DT sinusoid may not always be periodic, it will always have a periodic *envelope*.

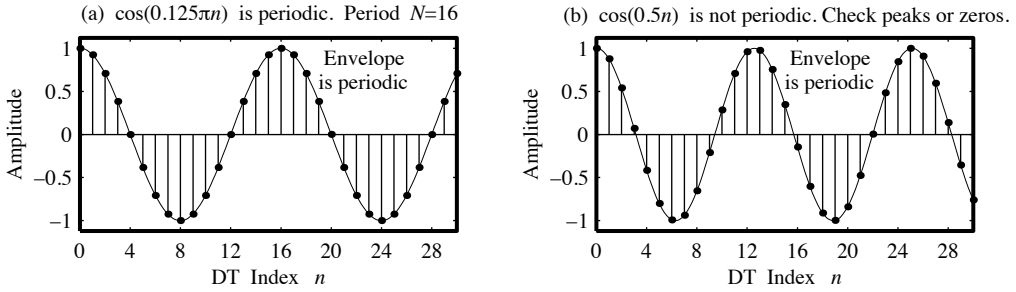


Figure 3.2 Discrete-time sinusoids are not always periodic

REVIEW PANEL 3.14

A DT Harmonic $\cos(2\pi nF_0 + \theta)$ or $e^{j(2\pi nF_0+\theta)}$ Is Not Always Periodic in Time

It is periodic only if its digital frequency $F_0 = k/N$ can be expressed as a ratio of integers.
Its period equals N if common factors have been canceled in k/N.
One period of the sampled sinusoid is obtained from k full periods of the analog sinusoid.

PGR2026-00033
Krisp Technologies EX1024 Page 64

EXAMPLE 3.7 (Discrete-Time Harmonics and Periodicity)

(a) Is $x[n] = \cos(2\pi Fn)$ periodic if $F = 0.32$? If $F = \sqrt{3}$? If periodic, what is its period N?

If $F = 0.32$, $x[n]$ is periodic because $F = 0.32 = \frac{32}{100} = \frac{8}{25} = \frac{k}{N}$. The period is $N = 25$.

If $F = \sqrt{3}$, $x[n]$ is *not* periodic because F is irrational and cannot be expressed as a ratio of integers.

(b) What is the period of the harmonic signal $x[n] = e^{j0.2n\pi} + e^{-j0.3n\pi}$?

The digital frequencies in $x[n]$ are $F_1 = 0.1 = \frac{1}{10} = \frac{k_1}{N_1}$ and $F_2 = 0.15 = \frac{3}{20} = \frac{k_2}{N_2}$.

Their periods are $N_1 = 10$ and $N_2 = 20$.

The common period is thus $N = \text{LCM}(N_1,\ N_2) = \text{LCM}(10,\ 20) = 20$.

(c) The signal $x(t) = 2\cos(40\pi t) + \sin(60\pi t)$ is sampled at 75 Hz. What is the common period of the sampled signal $x[n]$, and how many full periods of $x(t)$ does it take to obtain one period of $x[n]$?

The frequencies in $x(t)$ are $f_1 = 20$ Hz and $f_2 = 30$ Hz. The digital frequencies of the individual components are $F_1 = \frac{20}{75} = \frac{4}{15} = \frac{k_1}{N_1}$ and $F_2 = \frac{30}{75} = \frac{2}{5} = \frac{k_2}{N_2}$. Their periods are $N_1 = 15$ and $N_2 = 5$. The common period is thus $N = \text{LCM}(N_1,\ N_2) = \text{LCM}(15, 5) = 15$.

The fundamental frequency of $x(t)$ is $f_0 = \text{GCD}(20,\ 30) = 10$ Hz. One period of $x(t)$ is $T = \frac{1}{f_0} = 0.1\,\text{s}$. Since $N = 15$ corresponds to a duration of $Nt_s = \frac{N}{S} = 0.2$ s, it takes two full periods of $x(t)$ to obtain one period of $x[n]$. We also get the same result by computing $\text{GCD}(k_1,\ k_2) = \text{GCD}(4,\ 2) = 2$.

3.5.2 Discrete-Time Sinusoids Are Always Periodic in Frequency

Unlike analog sinusoids, discrete-time sinusoids and harmonics are *always* periodic in frequency. If we start with the sinusoid $x[n] = \cos(2\pi nF + \theta)$ and add an integer m to F, we get

$$\cos[2\pi n(F + m) + \theta] = \cos(2\pi nF + \theta + 2\pi nm) = \cos(2\pi nF + \theta) = x[n]$$

This result says that discrete-time (DT) sinusoids at the frequencies $F \pm m$ are *identical*. Put another way, a DT sinusoid is *periodic* in frequency (has a periodic spectrum) with period $F = 1$. The range $-0.5 \le F \le 0.5$ defines the **principal period** or **principal range**. A DT sinusoid with a frequency outside this range can always be expressed as a DT sinusoid whose frequency is in the principal period by subtracting out an integer. A DT sinusoid can be uniquely identified only if its frequency falls in the principal period.

To summarize, *a discrete-time sinusoid is periodic in time only if its digital frequency F is a rational fraction, but it is always periodic in frequency with period $F = 1$.*

REVIEW PANEL 3.15

A DT Harmonic $\cos(2\pi nF_0 + \theta)$ **Is Always Periodic in Frequency but Not Always Unique**

Its frequency period is unity (harmonics at F_0 and $F_0 \pm M$ are identical for *integer* M).
It is **unique** only if F_0 lies in the *principal period* $-0.5 < F_0 \le 0.5$.
If $F_0 > 0.5$, the unique frequency is $F_a = F_0 - M$, where M is chosen to ensure $-0.5 < F_a \le 0.5$.

PGR2026-00033
Krisp Technologies EX1024 Page 65

3.6 Aliasing and the Sampling Theorem

The central concept in the digital processing of analog signals is that the sampled signal must be a unique representation of the analog signal. For a sinusoid $x(t) = \cos(2\pi f_0 t + \theta)$, a sampling rate $S > 2f_0$ ensures that the digital frequency of the sampled signal falls in the principal period $-0.5 \le F \le 0.5$ and permits a unique correspondence with the underlying analog signal. More generally, the **sampling theorem** says that for a unique correspondence between an analog signal and the version reconstructed from its samples (using the *same* sampling rate), the sampling rate must *exceed* **twice** the highest signal frequency $f_{\max}$. The critical rate $S = 2f_{\max}$ is called the **Nyquist rate** or **Nyquist frequency**, and $t_s = \frac{1}{2f_{\max}}$ is called the **Nyquist interval**. For the sinusoid $x(t) = \cos(2\pi f_0 t + \theta)$, the Nyquist rate is $S_N = 2f_0 = \frac{2}{T}$ and corresponds to taking *two samples per period* (because the sampling interval is $t_s = \frac{T}{2}$).

REVIEW PANEL 3.16

The Sampling Theorem: How to Sample an Analog Signal Without Loss of Information

For an analog signal band-limited to $f_{\max}$ Hz, the sampling rate S must *exceed* $2f_{\max}$.
$S = 2f_{\max}$ defines the **Nyquist rate**. $t_s = \frac{1}{2f_{\max}}$ defines the **Nyquist interval**.
For an analog sinusoid: The Nyquist rate corresponds to taking *two samples per period.*

Consider an analog signal $x(t) = \cos(2\pi f_0 t + \theta)$ and its sampled version $x[n] = \cos(2\pi n F_0 + \theta)$, where $F_0 = f_0/S$. If $x[n]$ is to be a unique representation of $x(t)$, we must be able to reconstruct $x(t)$ from $x[n]$. In practice, reconstruction uses only the copy or **image** of the periodic spectrum of $x[n]$ in the **principal period** $-0.5 \le F \le 0.5$, which corresponds to the analog frequency range $-0.5S \le f \le 0.5S$. We use a lowpass filter to remove all other replicas or images, and the output of the lowpass filter corresponds to the reconstructed analog signal. As a result, the highest frequency f_H we can identify in the signal reconstructed from its samples is $f_H = 0.5S$.

REVIEW PANEL 3.17

The Highest Frequency Cannot Exceed Half the Sampling or Reconstruction Rate

Upon sampling or reconstruction, all frequencies must be brought into the principal period.

Whether the frequency of the reconstructed analog signal matches $x(t)$ or not depends on the sampling rate S. If $S > 2f_0$, the digital frequency $F_0 = f_0/S$ is always in the principal range $-0.5 \le F \le 0.5$, and the reconstructed analog signal is *identical* to $x(t)$. If $S < 2f_0$, the digital frequency exceeds 0.5. Its image in the principal range appears at the *lower* digital frequency $F_a = F_0 - M$ (corresponding to the *lower* analog frequency $f_a = f_0 - MS$), where M is an integer that places the digital frequency F_a between -0.5 and 0.5 (or the analog frequency f_a between $-0.5S$ and $0.5S$). The reconstructed analog signal $x_a(t) = \cos(2\pi f_a t + \theta)$ is at a *lower* frequency $f_a = SF_a$ than f_0 and is no longer a replica of $x(t)$. This phenomenon, where a reconstructed sinusoid appears at a lower frequency than the original, is called **aliasing**. The real problem is that the original signal $x(t)$ and the aliased signal $x_a(t)$ yield *identical* sampled representations at the sampling frequency S and prevent unique identification of $x(t)$ from its samples!

REVIEW PANEL 3.18

Aliasing Occurs if the Analog Signal $\cos(2\pi f_0 t + \theta)$ **Is Sampled Below the Nyquist Rate**

If $S < 2f_0$, the *reconstructed* analog signal is aliased to a *lower* frequency $|f_a| < 0.5S$. We find f_a as $f_a = f_0 - MS$, where M is an integer that places f_a in the *principal period* $(-0.5S < f_a \le 0.5S)$.

PGR2026-00033
Krisp Technologies EX1024 Page 66

EXAMPLE 3.8 (Aliasing and Its Effects)

(a) A 100-Hz sinusoid $x(t)$ is sampled at 240 Hz. Has aliasing occurred? How many full periods of $x(t)$ are required to obtain one period of the sampled signal?

The sampling rate exceeds 200 Hz, so there is no aliasing. The digital frequency is $F = \frac{100}{240} = \frac{5}{12}$. Thus, five periods of $x(t)$ yield 12 samples (one period) of the sampled signal.

(b) A 100-Hz sinusoid is sampled at rates of 240 Hz, 140 Hz, 90 Hz, and 35 Hz. In each case, has aliasing occurred, and if so, what is the aliased frequency?

To avoid aliasing, the sampling rate must exceed 200 Hz. If $S = 240$ Hz, there is no aliasing, and the reconstructed signal (from its samples) appears at the original frequency of 100 Hz. For all other choices of S, the sampling rate is too low and leads to aliasing. The aliased signal shows up at a lower frequency. The aliased frequencies corresponding to each sampling rate S are found by subtracting out multiples of S from 100 Hz to place the result in the range $-0.5S \leq f \leq 0.5S$. If the original signal has the form $x(t) = \cos(200\pi t + \theta)$, we obtain the following aliased frequencies and aliased signals:

1. $S = 140$ Hz, $\quad f_a = 100 - 140 = -40$ Hz, $\quad x_a(t) = \cos(-80\pi t + \theta) = \cos(80\pi t - \theta)$
2. $S = 90$ Hz, $\quad f_a = 100 - 90 = 10$ Hz, $\quad x_a(t) = \cos(20\pi t + \theta)$
3. $S = 35$ Hz, $\quad f_a = 100 - 3(35) = -5$ Hz, $\quad x_a(t) = \cos(-10\pi t + \theta) = \cos(10\pi t - \theta)$

We thus obtain a 40-Hz sinusoid (with reversed phase), a 10-Hz sinusoid, and a 5-Hz sinusoid (with reversed phase), respectively. Notice that negative aliased frequencies simply lead to a phase reversal and do not represent any new information. Finally, had we used a sampling rate exceeding the Nyquist rate of 200 Hz, we would have recovered the original 100-Hz signal every time. Yes, it pays to play by the rules of the sampling theorem!

(c) Two analog sinusoids $x_1(t)$ (shown light) and $x_2(t)$ (shown dark) lead to an identical sampled version as illustrated in Figure E3.8C. Has aliasing occurred? Identify the original and aliased signal. Identify the digital frequency of the sampled signal corresponding to each sinusoid. What is the analog frequency of each sinusoid if $S = 50$ Hz? Can you provide exact expressions for each sinusoid?

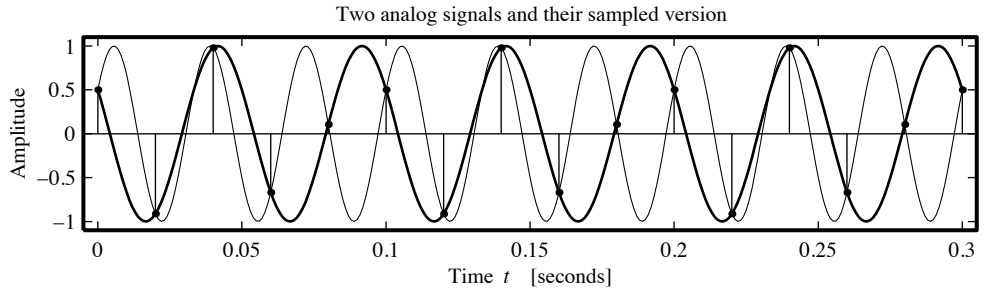


Figure E3.8C The sinusoids for Example 3.8(c)

Look at the interval (0, 0.1) s. The sampled signal shows five samples per period. This covers three full periods of $x_1(t)$ and so $F_1 = \frac{3}{5}$. This also covers two full periods of $x_2(t)$, and so $F_2 = \frac{2}{5}$. Clearly, $x_1(t)$ (with $|F_1| > 0.5$) is the original signal that is aliased to $x_2(t)$. The sampling interval is 0.02 s.

PGR2026-00033
Krisp Technologies EX1024 Page 67

So, the sampling rate is $S = 50$ Hz. The original and aliased frequencies are $f_1 = SF_1 = 30$ Hz and $f_2 = SF_2 = 20$ Hz.

From the figure, we can identify exact expressions for $x_1(t)$ and $x_2(t)$ as follows. Since $x_1(t)$ is a delayed cosine with $x_1(0) = 0.5$, we have $x_1(t) = \cos(60\pi t - \frac{\pi}{3})$. With $S = 50$ Hz, the frequency $f_1 = 30$ Hz actually aliases to $f_2 = -20$ Hz, and thus $x_2(t) = \cos(-40\pi t - \frac{\pi}{3}) = \cos(40\pi t + \frac{\pi}{3})$. With $F = \frac{30}{50} = 0.6$ (or $F = -0.4$), the expression for the sampled signal is $x[n] = \cos(2\pi nF - \frac{\pi}{3})$.

(d) A 100-Hz sinusoid is sampled, and the reconstructed signal (from its samples) shows up at 10 Hz. What was the sampling rate S?

If you said 90 Hz ($100 - S = 10$), you are not wrong. But you could also have said 110 Hz ($100 - S = -10$). In fact, we can also subtract out integer multiples of S from 100 Hz, and S is then found from the following expressions (as long as we ensure that $S > 20$ Hz):

1. $100 - MS = 10$
2. $100 - MS = -10$

Solving the first expression for S, we find, for example, $S = 45$ Hz (with $M = 2$) or $S = 30$ Hz (with $M = 3$). Similarly, the second expression gives $S = 55$ Hz (with $M = 2$). Which of these sampling rates was actually used? We have no way of knowing!

3.6.1 Reconstruction Using a Different Sampling Rate

There are situations when we sample a signal using one sampling rate S_1 but reconstruct the analog signal from samples using a different sampling rate S_2. In such situations, a frequency f_0 in the original signal will result in a recovered frequency of $f_r = f_0(S_2/S_1)$ if $S_1 > 2f_0$ (no aliasing) or $f_r = f_a(S_2/S_1)$ if $S_1 < 2f_0$, where f_a is the aliased frequency. In other words, all frequencies, aliased or recovered, are identified by their principal period.

REVIEW PANEL 3.19

Aliased or Reconstructed Frequencies Are Always Identified by Their Principal Period

Sampling: Unique digital frequencies always lie in the principal period $-0.5 < F_a \leq 0.5$.
Reconstruction at S_R: Analog frequencies lie in the principal period $-0.5S_R < f_a \leq 0.5S_R$.
The reconstructed frequency is $f_R = S_R F_a = S_R \frac{f_a}{S}$.

EXAMPLE 3.9 (Signal Reconstruction at Different Sampling Rates)
A 100-Hz sinusoid is sampled at S Hz, and the sampled signal is then reconstructed at 540 Hz. What is the frequency of the reconstructed signal if $S = 270$ Hz? If $S = 70$ Hz?

1. If $S = 270$ Hz, the digital frequency of the sampled signal is $F = \frac{100}{270} = \frac{10}{27}$, which lies in the principal period. The frequency f_r of the reconstructed signal is then $f_r = SF = 540F = 200$ Hz.
2. If $S = 70$ Hz, the digital frequency of the sampled signal is $F = \frac{100}{70} = \frac{10}{7}$, which does not lie in the principal period. The frequency of the principal period is $F = \frac{10}{7} - 1 = -\frac{3}{7}$, and the frequency f_r of reconstructed signal is then $f_r = 70F = SF = -30$ Hz. The negative sign simply translates to a phase reversal in the reconstructed signal.

PGR2026-00033
Krisp Technologies EX1024 Page 68

3.7 Random Signals

The signals we have studied so far are called **deterministic** or predictable. They are governed by a unique mathematical representation that, once established, allows us to completely characterize the signal for all time, past, present, or future. In contrast to this is the class of signals known as **random** or **stochastic** whose precise value cannot be predicted in advance. We stress that only the future values of a random signal pose a problem since past values of any signal, random or otherwise, are exactly known once they have occurred. Randomness or uncertainty about future signal values is inherent in many practical situations. In fact, a degree of uncertainty is essential for communicating information. The longer we observe a random signal, the more the additional information we gain and the less the uncertainty. To fully understand the nature of random signals requires the use of probability theory, random variables, and statistics. Even with such tools, the best we can do is to characterize random signals only *on the average* based on their past behavior.

3.7.1 Probability

Figure 3.3 shows the results of two experiments, each repeated under identical conditions. The first experiment always yields identical results no matter how many times it is run and yields a deterministic signal. We need to run the experiment only once to predict what the next, or any other run, will yield.

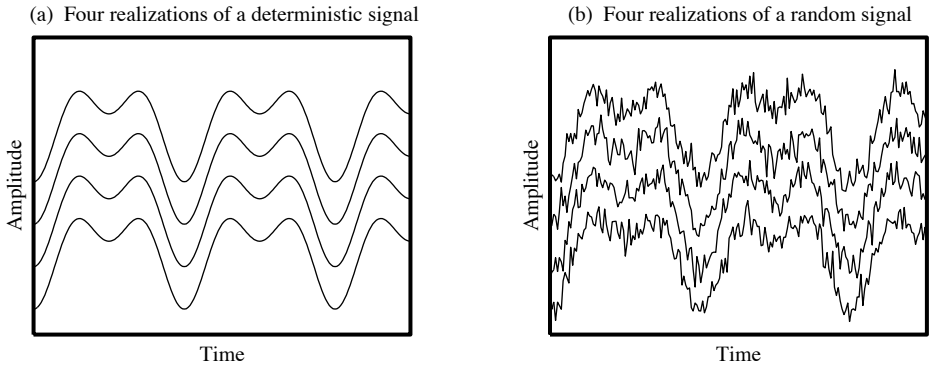


Figure 3.3 Realizations of a deterministic and stochastic process

The second experiment gives a different result or **realization** $x(t)$ every time the experiment is repeated and describes a **stochastic** or random system. A **random signal** or **random process** $X(t)$ comprises the family or **ensemble** of all such realizations obtained by repeating the experiment many times. Each realization $x(t)$, once obtained, ceases to be random and can be subjected to the same operations as we use for deterministic signals (such as derivatives, integrals, and the like). The randomness of the signal stems from the fact that one realization provides no clue as to what the next, or any other, realization might yield. At a given instant t, each realization of a random signal can assume a different value, and the collection of all such values defines a **random variable**. Some values are more likely to occur, or more probable, than others. The concept of probability is tied to the idea of repeating an experiment a large number of times in order to estimate this probability. Thus, if the value 2 V occurs 600 times in 1000 runs, we say that the probability of occurrence of 2 V is 0.6.

The **probability of an event** A, denoted $\Pr(A)$, is the proportion of successful outcomes to the (very large) number of times the experiment is run and is a fraction between 0 and 1 since the number of successful

PGR2026-00033
Krisp Technologies EX1024 Page 69

runs cannot exceed the total number of runs. The larger the probability $\Pr(A)$, the more the chance of event A occurring. To fully characterize a random variable, we must answer two questions:

1. What is the range of all possible (nonrandom) values it can acquire? This defines an **ensemble space**, which may be finite or infinite.
2. What are the probabilities for all the possible values in this range? This defines the **probability distribution function** $F(x)$. Clearly, $F(x)$ must always lie between 0 and 1.

It is common to work with the derivative of the probability distribution function called the **probability density function** $f(x)$. The distribution function $F(x)$ is simply the running integral of the density $f(x)$:

$$f(x) = \frac{d\,F(x)}{dx} \qquad \text{or} \qquad F(x) = \int_{-\infty}^{x} f(\lambda)\,d\lambda \tag{3.22}$$

The probability $F(x_1) = \Pr[X \le x_1]$ that X is less than x_1 is given by

$$\Pr[X \le x_1] = \int_{-\infty}^{x_1} f(x)\,dx \tag{3.23}$$

The probability that X lies between x_1 and x_2 is $\Pr[x_1 < X \le x_2] = F(x_2) - F(x_1)$. The area of $f(x)$ is 1.

3.7.2 Measures for Random Variables

Measures or features of a random variable X are based on its distribution. Two commonly used features are the **mean** and **variance**, defined by

$$\hat{E}(x) = m_x = \int_{-\infty}^{\infty} x f(x)\,dx \qquad\qquad \sigma_x^2 = \hat{E}[(x - m_x)^2] = \int_{-\infty}^{\infty} (x - m_x)^2 f(x)\,dx \tag{3.24}$$

The mean, or **expectation**, is a measure of where the distribution is *centered*. The **variance** measures the spread of the distribution about its mean. The less the spread, the smaller is the variance. The variance is also a measure of the ac power in a signal. The quantity σ is known as the **standard deviation** and provides a measure of the uncertainty in a physical measurement.

3.7.3 Probability Distributions

Two of the most commonly encountered probability distributions are the uniform distribution and normal (or Gaussian) distribution. These are illustrated in Figure 3.4.

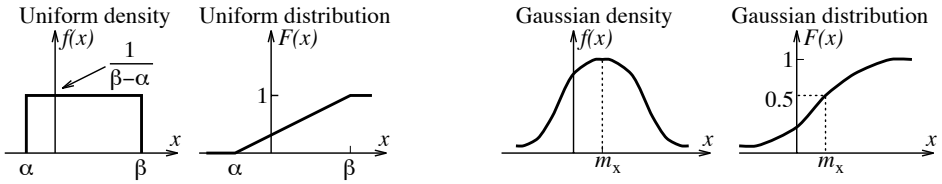


Figure 3.4 The uniform and normal probability distributions

In a uniform distribution, every value is equally likely, since the random variable shows no preference for a particular value. The density function $f(x)$ is just a rectangular pulse defined by

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha}, & \alpha \le x \le \beta \\ 0, & \text{otherwise} \end{cases} \qquad \text{(uniform distribution)} \tag{3.25}$$

PGR2026-00033
Krisp Technologies EX1024 Page 70

and the distribution $F(x)$ is a ramp that flattens out. When quantizing signals in uniform steps, the error in representing a signal value is assumed to be uniformly distributed between -0.5Δ and 0.5Δ, where Δ is the quantization step. The density function of the phase of a sinusoid with random phase is also uniformly distributed between $-\pi$ and π.

The bell-shaped Gaussian probability density is also referred to as **normal** and defined by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-m_x)^2}{2\sigma^2}\right] \qquad \text{(normal distribution)} \tag{3.26}$$

The mean (or variance) of the sum of Gaussian distributions equals the sum of the individual means (or variances). The probability distribution of combinations of statistically independent, random phenomena often tends to a Gaussian. This is the **central limit theorem**.

The idea of distributions also applies to *deterministic* periodic signals for which they can be found as exact analytical expressions. Consider the periodic signal $x(t)$ of Figure 3.5. The probability $\Pr[X < 0]$ that $x(t) < 0$ is zero. The probability $\Pr[X < 3]$ that $x(t)$ is less than 3 is 1. Since $x(t)$ is linear over one period ($T = 3$), all values in this range are equally likely, and $F(x)$ must vary linearly from 0 to 1 over this range. This yields the distribution $F(x)$ and density $f(x)$ as shown. Note that the area of $f(x)$ equals unity.

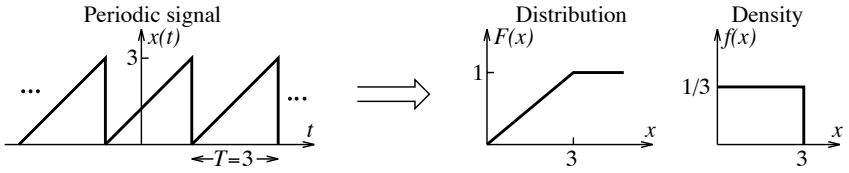


Figure 3.5 A periodic signal and its density and distribution functions

3.7.4 Stationary, Ergodic, and Pseudorandom Signals

A random signal is called **stationary** if its statistical features do not change over time. Stationarity suggests a state of statistical equilibrium akin to the steady-state for deterministic situations. A stationary process is typically characterized by a constant mean and constant variance. The statistical properties of a stationary random signal may be found as **ensemble averages** *across* the process by averaging over all realizations at one specific time instant or as **time averages** *along* the process by averaging a single realization over time. The two are not always equal. If they are, a stationary process is said to be **ergodic**. The biggest advantage of ergodicity is that we can use features from one realization to describe the whole process. Ergodicity is assumed in most practical situations! For an ergodic signal, the mean equals the time average, and the variance equals the ac power (the power in the signal with its dc component removed).

3.7.5 Statistical Estimates

Probability theory allows us to fully characterize a random signal from an a priori knowledge of its probability distribution. This yields features like the mean and variance of the random variable. In practice, we are faced with exactly the opposite problem of finding such features from a set of discrete data, often in the absence of a probability distribution. The best we can do is get an estimate of such features and perhaps even the distribution itself. This is what statistical estimation achieves. The mean and variance are typically estimated directly from the observations $x_k,\ k = 0, 1, 2, \ldots, N-1$ as

$$m_x = \frac{1}{N}\sum_{k=0}^{N-1} x_k \qquad \sigma_x^2 = \frac{1}{N-1}\sum_{k=0}^{N-1} (x_k - m_x)^2 \tag{3.27}$$

PGR2026-00033
Krisp Technologies EX1024 Page 71

In many situations, we use artificially generated signals (which can never be truly random) with prescribed statistical features called **pseudorandom signals**. Such signals are actually periodic (with a *very* long period), but over one period their statistical features approximate those of random signals.

Histograms: The estimates f_k of a probability distribution are obtained by constructing a **histogram** from a large number of observations. A histogram is a bar graph of the number of observations falling within specified amplitude levels, or *bins*, as illustrated in Figure 3.6.

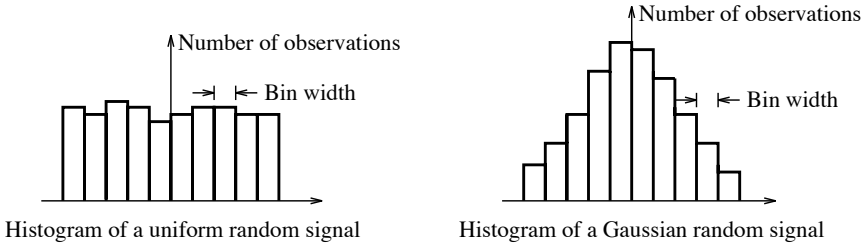


Figure 3.6 Histograms of a uniformly distributed and a normally distributed random signal

Signal-to-Noise Ratio: For a noisy signal $x(t) = s(t) + An(t)$ with a signal component $s(t)$ and a noise component $An(t)$ (with noise amplitude A), the **signal-to-noise ratio** (SNR) is the ratio of the signal power σ_s^2 and noise power $A^2\sigma_n^2$ and usually defined in **decibels** (dB) as

$$\text{SNR} = 10\log\left(\frac{\sigma_s^2}{A^2\sigma_n^2}\right) \text{ dB} \tag{3.28}$$

3.7.6 Application: Coherent Signal Averaging

Coherent signal averaging is a method of extracting signals from noise and assumes that the experiment can be repeated and the noise corrupting the signal is random (and uncorrelated). Averaging the results of many runs tends to average out the noise to zero, and the signal quality (or signal-to-noise ratio) improves. The more the number of runs, the smoother and less noisy the averaged signal. We often remove the mean or any linear trend before averaging. Figure 3.7 shows one realization of a noisy sine wave and the much smoother results of averaging 8 and 48 such realizations. This method is called **coherent** because it requires *time coherence* (time alignment of the signal for each run). It relies, for its success, on perfect synchronization of each run and on the statistical independence of the contaminating noise.

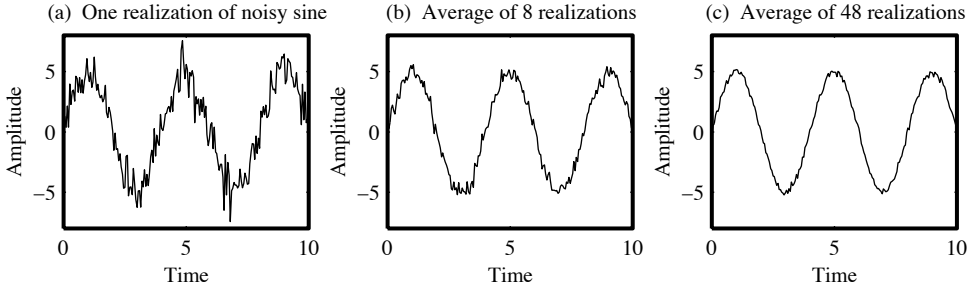


Figure 3.7 Coherent averaging of a noisy sine wave

PGR2026-00033
Krisp Technologies EX1024 Page 72

CHAPTER 3 PROBLEMS

DRILL AND REINFORCEMENT

3.1 (Discrete Signals) Sketch each signal and find its energy or power as appropriate.

(a) $x[n] = \{\overset{\Downarrow}{6},\ 4,\ 2,\ 2\}$ **(b)** $x[n] = \{-3,\ -2,\ \overset{\Downarrow}{-1},\ 0,\ 1\}$
(c) $x[n] = \{\overset{\Downarrow}{0},\ 2,\ 4,\ 6\}$ **(d)** $x[n] = u[n] - u[n-4]$
(e) $x[n] = \cos(n\pi/2)$ **(f)** $x[n] = 8(0.5)^n u[n]$

3.2 (Operations) Let $x[n] = \{\overset{\Downarrow}{6},\ 4,\ 2,\ 2\}$. Sketch the following signals and find their signal energy.

(a) $y[n] = x[n-2]$ **(b)** $f[n] = x[n+2]$ **(c)** $g[n] = x[-n+2]$ **(d)** $h[n] = x[-n-2]$

3.3 (Operations) Let $x[n] = 8(0.5)^n(u[n+1] - u[n-4])$. Sketch the following signals.

(a) $y[n] = x[n-3]$ **(b)** $f[n] = x[n+1]$ **(c)** $g[n] = x[-n+4]$ **(d)** $h[n] = x[-n-2]$

3.4 (Decimation and Interpolation) Let $x[n] = \{4,\ 0,\ \overset{\Downarrow}{2},\ -1,\ 3\}$. Find and sketch the following signals and compare their signal energy with the energy in $x[n]$.

(a) The decimated signal $d[n] = x[2n]$
(b) The *zero-interpolated* signal $f[n] = x[\frac{n}{2}]$
(c) The *step-interpolated* signal $g[n] = x[\frac{n}{2}]$
(d) The *linearly interpolated* signal $h[n] = x[\frac{n}{2}]$

3.5 (Symmetry) Sketch each signal and its even and odd parts.

(a) $x[n] = 8(0.5)^n u[n]$ **(b)** $x[n] = u[n]$ **(c)** $x[n] = 1 + u[n]$
(d) $x[n] = u[n] - u[n-4]$ **(e)** $x[n] = \text{tri}(\frac{n-3}{3})$ **(f)** $x[n] = \{6,\ \overset{\Downarrow}{4},\ 2,\ 2\}$

3.6 (Sketching Discrete Signals) Sketch each of the following signals:

(a) $x[n] = r[n+2] - r[n-2] - 4u[n-6]$ **(b)** $x[n] = \text{rect}(\frac{n}{6})$
(c) $x[n] = \text{rect}(\frac{n-2}{4})$ **(d)** $x[n] = 6\,\text{tri}(\frac{n-4}{3})$

3.7 (Signal Description) For each signal shown in Figure P3.7,

(a) Write out the numeric sequence, and mark the index $n = 0$ by an arrow.
(b) Write an expression for each signal using impulse functions.
(c) Write an expression for each signal using steps and/or ramps.
(d) Find the signal energy.
(e) Find the signal power, assuming that the sequence shown repeats itself.

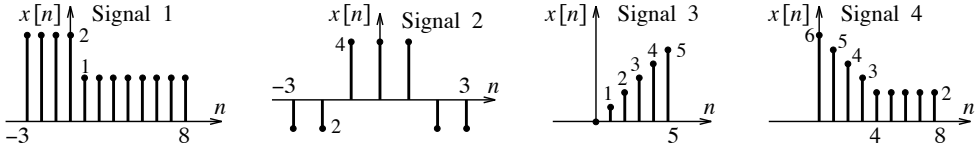


Figure P3.7 Signals for Problem 3.7

PGR2026-00033
Krisp Technologies EX1024 Page 73

3.8 (Discrete-Time Harmonics) Check for the periodicity of the following signals, and compute the common period N if periodic.

(a) $x[n] = \cos(\frac{n\pi}{2})$ **(b)** $x[n] = \cos(\frac{n}{2})$
(c) $x[n] = \sin(\frac{n\pi}{4}) - 2\cos(\frac{n\pi}{6})$ **(d)** $x[n] = 2\cos(\frac{n\pi}{4}) + \cos^2(\frac{n\pi}{4})$
(e) $x[n] = 4 - 3\sin(\frac{7n\pi}{4})$ **(f)** $x[n] = \cos(\frac{5n\pi}{12}) + \cos(\frac{4n\pi}{9})$
(g) $x[n] = \cos(\frac{8}{3}n\pi) + \cos(\frac{8}{3}n)$ **(h)** $x[n] = \cos(\frac{8n\pi}{3})\cos(\frac{n\pi}{2})$
(i) $x[n] = e^{j0.3n\pi}$ **(j)** $x[n] = 2e^{j0.3n\pi} + 3e^{j0.4n\pi}$
(k) $x[n] = e^{j0.3n}$ **(l)** $x[n] = (j)^{n/2}$

3.9 (Digital Frequency) Set up an expression for each signal, using a digital frequency $|F| < 0.5$, and another expression using a digital frequency in the range $4 < F < 5$.

(a) $x[n] = \cos(\frac{4n\pi}{3})$ **(b)** $x[n] = \sin(\frac{4n\pi}{3}) + 3\sin(\frac{8n\pi}{3})$

3.10 (Sampling and Aliasing) Each of the following sinusoids is sampled at $S = 100$ Hz. Determine if aliasing has occurred and set up an expression for each sampled signal using a digital frequency in the principal range ($|F| < 0.5$).

(a) $x(t) = \cos(320\pi t + \frac{\pi}{4})$ **(b)** $x(t) = \cos(140\pi t - \frac{\pi}{4})$ **(c)** $x(t) = \sin(60\pi t)$

REVIEW AND EXPLORATION

3.11 (Signal Duration) Use examples to argue that the product of a right-sided and a left-sided discrete-time signal is always time-limited or identically zero.

3.12 (Signal Representation) The two signals shown in Figure P3.12 may be expressed as

(a) $x[n] = A\alpha^n(u[n] - u[n-N])$ **(b)** $y[n] = A\cos(2\pi Fn + \theta)$

Find the constants in each expression and then find the signal energy or power as appropriate.

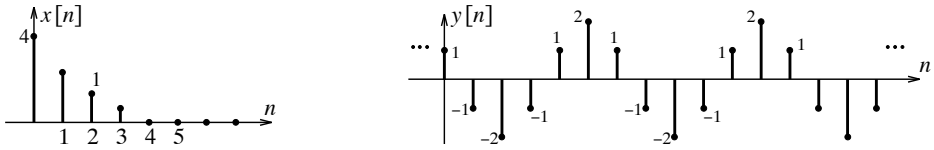


Figure P3.12 Signals for Problem 3.12

3.13 (Energy and Power) Classify the following as energy signals, power signals, or neither and find the energy or power as appropriate.

(a) $x[n] = 2^n u[-n]$ **(b)** $x[n] = 2^n u[-n-1]$ **(c)** $x[n] = \cos(n\pi)$
(d) $x[n] = \cos(n\pi/2)$ **(e)** $x[n] = \frac{1}{n}u[n-1]$ **(f)** $x[n] = \frac{1}{\sqrt{n}}u[n-1]$
(g) $x[n] = \frac{1}{n^2}u[n-1]$ **(h)** $x[n] = e^{jn\pi}$ **(i)** $x[n] = e^{jn\pi/2}$
(j) $x[n] = e^{(j+1)n\pi/4}$ **(k)** $x[n] = j^{n/4}$ **(l)** $x[n] = (\sqrt{j})^n + (\sqrt{j})^{-n}$

3.14 (Energy and Power) Sketch each of the following signals, classify as an energy signal or power signal, and find the energy or power as appropriate.

PGR2026-00033
Krisp Technologies EX1024 Page 74

(a) $x[n] = \sum_{k=-\infty}^{\infty} y[n-kN]$, where $y[n] = u[n] - u[n-3]$ and $N = 6$

(b) $x[n] = \sum_{k=-\infty}^{\infty} (2)^{n-5k}(u[n-5k] - u[n-5k-4])$

3.15 (Sketching Signals) Sketch the following signals and describe how they are related.

(a) $x[n] = \delta[n]$ **(b)** $f[n] = \text{rect}(n)$ **(c)** $g[n] = \text{tri}(n)$ **(d)** $h[n] = \text{sinc}(n)$

3.16 (Discrete Exponentials) A causal discrete exponential has the form $x[n] = \alpha^n u[n]$.

(a) Assume that α is real and positive. Pick convenient values for $\alpha > 1,\ \alpha = 1$, and $\alpha < 1$; sketch $x[n]$; and describe the nature of the sketch for each choice of α.

(b) Assume that α is real and negative. Pick convenient values for $\alpha < -1,\ \alpha = -1$, and $\alpha > -1$; sketch $x[n]$; and describe the nature of the sketch for each choice of α.

(c) Assume that α is complex and of the form $\alpha = Ae^{j\theta}$, where A is a positive constant. Pick convenient values for θ and for $A < 1,\ A = 1$, and $A > 1$; sketch the real part and imaginary part of $x[n]$ for each choice of A; and describe the nature of each sketch.

(d) Assume that α is complex and of the form $\alpha = Ae^{j\theta}$, where A is a positive constant. Pick convenient values for θ and for $A < 1,\ A = 1$, and $A > 1$; sketch the magnitude and imaginary phase of $x[n]$ for each choice of A; and describe the nature of each sketch.

3.17 (Interpolation and Decimation) Let $x[n] = 4\,\text{tri}(n/4)$. Sketch the following signals and describe how they differ.

(a) $x[\frac{2}{3}n]$, using zero interpolation followed by decimation
(b) $x[\frac{2}{3}n]$, using step interpolation followed by decimation
(c) $x[\frac{2}{3}n]$, using decimation followed by zero interpolation
(d) $x[\frac{2}{3}n]$, using decimation followed by step interpolation

3.18 (Fractional Delay) Starting with $x[n]$, we can generate the signal $x[n-2]$ (using a delay of 2) or $x[2n-3]$ (using a delay of 3 followed by decimation). However, to generate a *fractional delay* of the form $x[n - \frac{M}{N}]$ requires a delay, interpolation, and decimation!

(a) Describe the sequence of operations required to generate $x[n - \frac{2}{3}]$ from $x[n]$.

(b) Let $x[n] = \{\overset{\Downarrow}{1},\ 4,\ 7,\ 10,\ 13\}$. Sketch $x[n]$ and $x[n - \frac{2}{3}]$. Use *linear* interpolation where required.

(c) Generalize the results of part (a) to generate $x[n - \frac{M}{N}]$ from $x[n]$. Any restrictions on M and N?

3.19 (The Roots of Unity) The N roots of the equation $z^N = 1$ can be found by writing it as $z^N = e^{j2k\pi}$ to give $z = e^{j2k\pi/N},\ k = 0, 1, \ldots, N-1$. What is the magnitude of each root? The roots can be displayed as vectors directed from the origin whose tips lie on a circle.

(a) What is the length of each vector and the angular spacing between adjacent vectors? Sketch for $N = 5$ and $N = 6$.

(b) Extend this concept to find the roots of $z^N = -1$ and sketch for $N = 5$ and $N = 6$.

3.20 (Digital Sinusoids) Find the period N of each signal if periodic. Express each signal using a digital frequency in the principal range ($|F| < 0.5$) and in the range $3 \le F \le 4$.

(a) $x[n] = \cos(\frac{7n\pi}{3})$ **(b)** $x[n] = \cos(\frac{7n\pi}{3}) + \sin(0.5n\pi)$ **(c)** $x[n] = \cos(n)$

PGR2026-00033
Krisp Technologies EX1024 Page 75

3.21 (Aliasing and Signal Reconstruction) The signal $x(t) = \cos(320\pi t + \frac{\pi}{4})$ is sampled at 100 Hz, and the sampled signal $x[n]$ is reconstructed at 200 Hz to recover the analog signal $x_r(t)$.

(a) Has aliasing occurred? What is the period N and the digital frequency F of $x[n]$?
(b) How many full periods of $x(t)$ are required to generate one period of $x[n]$?
(c) What is the analog frequency of the recovered signal $x_r(t)$?
(d) Write expressions for $x[n]$ (using $|F| < 0.5$) and for $x_r(t)$.

3.22 (Digital Pitch Shifting) One way to accomplish *pitch shifting* is to play back (or reconstruct) a sampled signal at a *different* sampling rate. Let the analog signal $x(t) = \sin(15800\pi t + 0.25\pi)$ be sampled at a sampling rate of 8 kHz.

(a) Find its sampled representation with digital frequency $|F| < 0.5$.
(b) What frequencies are heard if the signal is reconstructed at a rate of 4 kHz?
(c) What frequencies are heard if the signal is reconstructed at a rate of 8 kHz?
(d) What frequencies are heard if the signal is reconstructed at a rate of 20 kHz?

3.23 (Discrete-Time Chirp Signals) Consider the signal $x(t) = \cos[\phi(t)]$, where $\phi(t) = \alpha t^2$. Show that its instantaneous frequency $f_i(t) = \frac{1}{2\pi}\phi'(t)$ varies linearly with time.

(a) Choose α such that the frequency varies from 0 Hz to 2 Hz in 10 seconds, and generate the sampled signal $x[n]$ from $x(t)$, using a sampling rate of $S = 4$ Hz.
(b) It is claimed that, unlike $x(t)$, the signal $x[n]$ is periodic. Verify this claim, using the condition for periodicity ($x[n] = x[n + N]$), and determine the period N of $x[n]$.
(c) The signal $y[n] = \cos(\pi F_0 n^2/M),\ n = 0, 1, \ldots, M-1$, describes an M-sample chirp whose digital frequency varies linearly from 0 to F_0. What is the period of $y[n]$ if $F_0 = 0.25$ and $M = 8$?

3.24 (Time Constant) For exponentially decaying discrete signals, the **time constant** is a measure of how fast a signal decays. The 60-dB time constant describes the (integer) number of samples it takes for the signal level to decay by a factor of 1000 (or $20 \log 1000 = 60$ dB).

(a) Let $x[n] = (0.5)^n u[n]$. Compute its 60-dB time constant and 40-dB time constant.
(b) Compute the time constant in seconds if the discrete-time signal is derived from an analog signal sampled at 1 kHz.

3.25 (Signal Delay) The delay D of a discrete-time energy signal $x[n]$ is defined by

$$D = \frac{\sum_{k=-\infty}^{\infty} k x^2[k]}{\sum_{k=-\infty}^{\infty} x^2[k]}$$

(a) Verify that the delay of the symmetric sequence $x[n] = \{4, 3, 2, 1, \underset{\Downarrow}{0}, 1, 2, 3, 4\}$ is zero.
(b) Compute the delay of the signals $g[n] = x[n-1]$ and $h[n] = x[n-2]$.
(c) What is the delay of the signal $y[n] = 1.5(0.5)^n u[n] - 2\delta[n]$?

3.26 (Periodicity) It is claimed that the sum of an absolutely summable signal $x[n]$ and its shifted (by multiples of N) replicas is a periodic signal $x_p[n]$ with period N. Verify this claim by sketching the following and, for each case, compute the power in the resulting periodic signal $x_p[n]$ and compare the sum and energy of one period of $x_p[n]$ with the sum and energy of $x[n]$.

PGR2026-00033
Krisp Technologies EX1024 Page 76

(a) The sum of $x[n] = \text{tri}(n/3)$ and its replicas shifted by $N = 7$
(b) The sum of $x[n] = \text{tri}(n/3)$ and its replicas shifted by $N = 6$
(c) The sum of $x[n] = \text{tri}(n/3)$ and its replicas shifted by $N = 5$
(d) The sum of $x[n] = \text{tri}(n/3)$ and its replicas shifted by $N = 4$
(e) The sum of $x[n] = \text{tri}(n/3)$ and its replicas shifted by $N = 3$

3.27 (Periodic Extension) The sum of an absolutely summable signal $x[n]$ and its shifted (by multiples of N) replicas is called the *periodic extension* of $x[n]$ with period N. Show that one period of the periodic extension of the signal $x[n] = \alpha^n u[n]$ with period N is $y[n] = \dfrac{x[n]}{1-\alpha^N},\ 0 \le n \le N-1$. How does the one-period sum of $y[n]$ compare with the sum of $x[n]$? What is the signal power in $x[n]$ and $y[n]$?

3.28 (Signal Norms) Norms provide a measure of the *size* of a signal. The **p-norm**, or **Hölder norm**, $\|x\|_p$ for discrete signals is defined by $\|x\|_p = (\sum |x|^p)^{1/p}$, where $0 < p < \infty$ is a positive integer. For $p = \infty$, we also define $\|x\|_\infty$ as the peak absolute value $|x|_{\max}$.
(a) Let $x[n] = \{3,\ -j4,\ 3 + j4\}$. Find $\|x\|_1$, $\|x\|_2$, and $\|x\|_\infty$.
(b) What is the significance of each of these norms?

COMPUTATION AND DESIGN

dtsiggui A GUI for Signal Visualization

The graphical user interface `ctsiggui` allows you to plot a discrete-time signal $x[n]$.
You can display results of the operation $y[n] = A + Bx[Cn + D]$ for your choice of parameters.
You can also display the odd part, even part, first difference, and running sum of $x[n]$.
To explore this routine, type `dtsiggui` at the MATLAB prompt.

randgui A GUI for Random Signal Visualization

The graphical user interface `randgui` allows you to plot a random signal and its histogram.
You can select the distribution and display its histogram using various bin sizes.
You can also display the effect of adding several realizations of a random signal.
To explore this routine, type `randgui` at the MATLAB prompt.

3.29 (Discrete Signals) Plot each signal $x[n]$ over $-10 \le n \le 10$. Then, using the ADSP routine `operate` (or otherwise), plot each signal $y[n]$ and compare with the original.

(a) $x[n] = u[n+4] - u[n-4] + 2\delta[n+6] - \delta[n-3]$ $\quad y[n] = x[-n-4]$
(b) $x[n] = r[n+6] - r[n+3] - r[n-3] + r[n-6]$ $\quad y[n] = x[n-4]$
(c) $x[n] = \text{rect}(\frac{n}{10}) - \text{rect}(\frac{n-3}{6})$ $\quad y[n] = x[n+4]$
(d) $x[n] = 6\,\text{tri}(\frac{n}{6}) - 3\,\text{tri}(\frac{n}{3})$ $\quad y[n] = x[-n+4]$

3.30 (Signal Interpolation) Let $h[n] = \sin(n\pi/3),\ \ 0 \le n \le 10$. Using the ADSP routine `interpol` (or otherwise), plot $h[n]$, the *zero-interpolated*, *step-interpolated*, and *linearly interpolated* signals using interpolation by 3.

3.31 (Discrete Exponentials) A causal discrete exponential may be expressed as $x[n] = \alpha^n u[n]$, where the nature of α dictates the form of $x[n]$. Plot the following over $0 \le n \le 40$ and comment on the nature of each plot.

PGR2026-00033
Krisp Technologies EX1024 Page 77

(a) The signal $x[n]$ for $\alpha = 1.2,\ \alpha = 1$, and $\alpha = 0.8$.
(b) The signal $x[n]$ for $\alpha = -1.2,\ \alpha = -1$, and $\alpha = -0.8$.
(c) The real part and imaginary parts of $x[n]$ for $\alpha = Ae^{j\pi/4}$, with $A = 1.2,\ A = 1$, and $A = 0.8$.
(d) The magnitude and phase of $x[n]$ for $\alpha = Ae^{j\pi/4}$, with $A = 1.2,\ A = 1$, and $A = 0.8$.

3.32 (Discrete-Time Sinusoids) Which of the following signals are periodic and with what period? Plot each signal over $-10 \le n \le 30$. Do the plots confirm your expectations?

(a) $x[n] = 2\cos(\frac{n\pi}{2}) + 5\sin(\frac{n\pi}{5})$ **(b)** $x[n] = 2\cos(\frac{n\pi}{2})\sin(\frac{n\pi}{3})$
(c) $x[n] = \cos(0.5n)$ **(d)** $x[n] = 5\sin(\frac{n\pi}{8} + \frac{\pi}{4}) - 5\cos(\frac{n\pi}{8} - \frac{\pi}{4})$

3.33 (Complex-Valued Signals) A complex-valued signal $x[n]$ requires *two* plots for a complete description in one of two forms—the *magnitude* and *phase* vs. n or the *real part* vs. n and *imaginary part* vs. n.

(a) Let $x[n] = \{\overset{\Downarrow}{2},\ \ 1+j,\ \ -j2,\ \ 2-j2,\ \ -4\}$. Sketch each form for $x[n]$ by hand.
(b) Let $x[n] = e^{-j0.3n\pi}$. Use MATLAB to plot each form over $-30 \le n \le 30$. Is $x[n]$ periodic? If so, can you identify its period from the MATLAB plots? From which form, and how?

3.34 (Complex Exponentials) Let $x[n] = 5\sqrt{2}e^{j(\frac{n\pi}{9} - \frac{\pi}{4})}$. Plot the following signals and, for each case, derive analytic expressions for the signals plotted and compare with your plots. Is the signal $x[n]$ periodic? What is the period N? Which plots allow you determine the period of $x[n]$?

(a) The real part and imaginary part of $x[n]$ over $-20 \le n \le 20$
(b) The magnitude and phase of $x[n]$ over $-20 \le n \le 20$
(c) The sum of the real and imaginary parts over $-20 \le n \le 20$
(d) The difference of the real and imaginary parts over $-20 \le n \le 20$

3.35 (Complex Exponentials) Let $x[n] = (\sqrt{j})^n + (\sqrt{j})^{-n}$. Plot the following signals and, for each case, derive analytic expressions for the sequences plotted and compare with your plots. Is the signal $x[n]$ periodic? What is the period N? Which plots allow you determine the period of $x[n]$?

(a) The real part and imaginary part of $x[n]$ over $-20 \le n \le 20$
(b) The magnitude and phase of $x[n]$ over $-20 \le n \le 20$

3.36 (Discrete-Time Chirp Signals) An N-sample *chirp* signal $x[n]$ whose digital frequency *varies linearly* from F_0 to F_1 is described by

$$x[n] = \cos\left[2\pi\left(F_0 n + \frac{F_1 - F_0}{2N}n^2\right)\right], \qquad n = 0, 1, \ldots, N-1$$

(a) Generate and plot 800 samples of a chirp signal `x` whose digital frequency varies from $F = 0$ to $F = 0.5$. Observe how the frequency of `x` varies linearly with time, using the ADSP command `timefreq(x)`.
(b) Generate and plot 800 samples of a chirp signal `x` whose digital frequency varies from $F = 0$ to $F = 1$. Is the frequency always increasing? If not, what is the likely explanation?

3.37 (Chirp Signals) It is claimed that the chirp signal $x[n] = \cos(\pi n^2/6)$ is periodic (unlike the analog chirp signal $x(t) = \cos(\pi t^2/6)$). Plot $x[n]$ over $0 \le n \le 20$. Does $x[n]$ appear periodic? If so, can you identify the period N? Justify your results by trying to find an integer N such that $x[n] = x[n+N]$ (the basis for periodicity).

PGR2026-00033
Krisp Technologies EX1024 Page 78

3.38 (Signal Averaging) Extraction of signals from noise is an important signal-processing application. Signal averaging relies on averaging the results of many runs. The noise tends to average out to zero, and the signal quality or **signal-to-noise ratio** (SNR) improves.

(a) Generate samples of the sinusoid $x(t) = \sin(800\pi t)$ sampled at $S = 8192$ Hz for 2 seconds. The sampling rate is chosen so that you may also listen to the signal if your machine allows.

(b) Create a noisy signal $s[n]$ by adding $x[n]$ to samples of uniformly distributed noise such that $s[n]$ has an SNR of 10 dB. Compare the noisy signal with the original and compute the actual SNR of the noisy signal.

(c) Sum the signal $s[n]$ 64 times and average the result to obtain the signal $s_a[n]$. Compare the averaged signal $s_a[n]$, the noisy signal $s[n]$, and the original signal $x[n]$. Compute the SNR of the averaged signal $x_a[n]$. Is there an improvement in the SNR? Do you notice any (visual and audible) improvement? Should you?

(d) Create the averaged result $x_b[n]$ of 64 *different* noisy signals and compare the averaged signal $x_b[n]$ with the original signal $x[n]$. Compute the SNR of the averaged signal $x_b[n]$. Is there an improvement in the SNR? Do you notice any (visual and/or audible) improvement? Explain how the signal $x_b[n]$ differs from $x_a[n]$.

(e) The reduction in SNR is a function of the noise distribution. Generate averaged signals, using different noise distributions (such as Gaussian noise) and comment on the results.

3.39 (The Central Limit Theorem) The central limit theorem asserts that the sum of independent noise distributions tends to a Gaussian distribution as the number N of distributions in the sum increases. In fact, one way to generate a random signal with a Gaussian distribution is to add many (typically 6 to 12) uniformly distributed signals.

(a) Generate the sum of uniformly distributed random signals using $N = 2$, $N = 6$, and $N = 12$ and plot the histograms of each sum. Does the histogram begin to take on a Gaussian shape as N increases? Comment on the shape of the histogram for $N = 2$.

(b) Generate the sum of random signals with different distributions using $N = 6$ and $N = 12$. Does the central limit theorem appear to hold even when the distributions are not identical (as long as you select a large enough N)? Comment on the physical significance of this result.

3.40 (Music Synthesis I) A musical composition is a combination of *notes*, or signals, at various frequencies. An *octave* covers a range of frequencies from f_0 to $2f_0$. In the western musical scale, there are 12 notes per octave, *logarithmically equispaced*. The frequencies of the notes from f_0 to $2f_0$ correspond to

$$f = 2^{k/12} f_0 \qquad k = 0, 1, 2, \ldots, 11$$

The 12 notes are as follows (the $\sharp$ and $\flat$ stand for *sharp* and *flat*, and each pair of notes in parentheses has the same frequency):

A (A$^\sharp$ or B$^\flat$) B C (C$^\sharp$ or D$^\flat$) D (D$^\sharp$ or E$^\flat$) E F (F$^\sharp$ or G$^\flat$) G (G$^\sharp$ or A$^\flat$)

An Example: Raga Malkauns: In Indian classical music, a *raga* is a musical composition based on an ascending and descending scale. The notes and their order form the musical alphabet and grammar from which the performer constructs musical passages, using only the notes allowed. The performance of a raga can last from a few minutes to an hour or more! Raga *malkauns* is a pentatonic raga (with five notes) and the following scales:

Ascending: D F G B$^\flat$ C D Descending: C B$^\flat$ G F D

PGR2026-00033
Krisp Technologies EX1024 Page 79

The final note in each scale is held twice as long as the rest. To synthesize this scale in MATLAB, we start with a frequency f_0 corresponding to the first note D and go up in frequency to get the notes in the ascending scale; when we reach the note D, which is an octave higher, we go down in frequency to get the notes in the descending scale. Here is a MATLAB code fragment.

```
f0=340; d=f0;                          % Pick a frequency and the note D
f=f0*(2^(3/12)); g=f0*(2^(5/12));      % The notes F and G
bf=f0*(2^(8/12)); c=f0*(2^(10/12));    % The notes B(flat) and C
d2=2*d;                                % The note D (an octave higher)
```

Generate sampled sinusoids at these frequencies, using an appropriate sampling rate (say, 8192 Hz); concatenate them, assuming silent passages between each note; and play the resulting signal, using the MATLAB command `sound`. Use the following MATLAB code fragment as a guide:

```
ts=1/8192;                                    % Sampling interval
t=0:ts:0.4;                                   % Time for each note (0.4 s)
s1=0*(0:ts:0.1);                              % Silent period (0.1 s)
s2=0*(0:ts:0.05);                             % Shorter silent period (0.05 s)
tl=0:ts:1;                                    % Time for last note of each scale
d1=sin(2*pi*d*t);                             % Start generating the notes
f1=sin(2*pi*f*t); g1=sin(2*pi*g*t);
bf1=sin(2*pi*bf*t); c1=sin(2*pi*c*t);
dl1=sin(2*pi*d2*tl); dl2=sin(2*pi*d*tl);
asc=[d1 s1 f1 s1 g1 s1 bf1 s1 c1 s2 dl1]; % Create ascending scale
dsc=[c1 s1 bf1 s1 g1 s1 f1 s1 dl2];       % Create descending scale
y=[asc s1 dsc s1]; sound(y)                   % Malkauns scale (y)
```

3.41 (Music Synthesis II) The raw scale of raga *malkauns* will sound pretty dry! The reason for this is the manner in which the sound from a musical instrument is generated. Musical instruments produce sounds by the vibrations of a string (in string instruments) or a column of air (in woodwind instruments). Each instrument has its characteristic sound. In a guitar, for example, the strings are plucked, held, and then released to sound the notes. Once plucked, the sound dies out and decays. Furthermore, the notes are never pure but contain overtones (harmonics). For a realistic sound, we must include the overtones and the attack, sustain, and release (decay) characteristics. The sound signal may be considered to have the form $x(t) = \alpha(t)\cos(2\pi f_0 t + \theta)$, where f_0 is the pitch and $\alpha(t)$ is the envelope that describes the attack-sustain-release characteristics of the instrument played. A crude representation of some envelopes is shown in Figure P3.41 (the piecewise linear approximations will work just as well for our purposes). Woodwind instruments have a much longer sustain time and a much shorter release time than do plucked string and keyboard instruments.

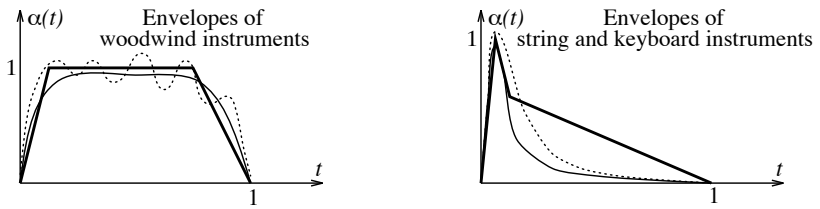


Figure P3.41 Envelopes and their piecewise linear approximations (dark) for Problem 3.41

Experiment with the scale of raga *malkauns* and try to produce a guitar-like sound, using the appropriate envelope form. You should be able to discern an audible improvement.

PGR2026-00033
Krisp Technologies EX1024 Page 80

3.42 (Music Synthesis III) Synthesize the following notes, using a woodwind envelope, and synthesize the same notes using a plucked string envelope.

$$\mathrm{F}^\sharp(0.3) \quad \mathrm{D}(0.4) \quad \mathrm{E}(0.4) \quad \mathrm{A}(1) \quad \mathrm{A}(0.4) \quad \mathrm{E}(0.4) \quad \mathrm{F}^\sharp(0.3) \quad \mathrm{D}(1)$$

All the notes cover one octave, and the numbers in parentheses give a rough indication of their relative duration. Can you identify the music? (It is *Big Ben*.)

3.43 (Music Synthesis IV) Synthesize the first bar of *Pictures at an Exhibition* by Mussorgsky, which has the following notes:

$$\mathrm{A}(3) \quad \mathrm{G}(3) \quad \mathrm{C}(3) \quad \mathrm{D}(2) \quad \mathrm{G}^\star(1) \quad \mathrm{E}(3) \quad \mathrm{D}(2) \quad \mathrm{G}^\star(1) \quad \mathrm{E}(3) \quad \mathrm{C}(3) \quad \mathrm{D}(3) \quad \mathrm{A}(3) \quad \mathrm{G}(3)$$

All the notes cover one octave except the note $\mathrm{G}^\star$, which is an octave above G. The numbers in parentheses give a rough indication of the relative duration of the notes (for more details, you may want to listen to an actual recording). Assume that a keyboard instrument (such as a piano) is played.

3.44 (DTMF Tones) In dual-tone multi-frequency (DTMF) or touch-tone telephone dialing, each number is represented by a dual-frequency tone. The frequencies for each digit are listed in Chapter 18.

(a) Generate DTMF tones corresponding to the telephone number 487-2550, by sampling the sum of two sinusoids at the required frequencies at $S = 8192$ Hz for each digit. Concatenate the signals by putting 50 zeros between each signal (to represent silence) and listen to the signal using the MATLAB command `sound`.

(b) Write a MATLAB program that generates DTMF signals corresponding to a vector input representing the digits in a phone number. Use a sampling frequency of $S = 8192$ Hz.

PGR2026-00033
Krisp Technologies EX1024 Page 81

Chapter 4

ANALOG SYSTEMS

4.0 Scope and Objectives

Analog systems are used to process analog signals. A description of systems relies heavily on how they respond to arbitrary or specific signals. In the time domain, many analog systems can be described by their response to arbitrary signals using differential equations. The class of *linear, time-invariant* systems can also be described by their impulse response, the response to an impulse input. This chapter deals with continuous-time systems and their classification, time-domain representation, and analysis based on the solution of differential equations. It also introduces the all-important concept of the impulse response, which forms a key ingredient in both system description and system analysis.

4.1 Introduction

In its broadest sense, a physical system is an interconnection of devices and elements subject to physical laws. A system that processes analog signals is referred to as an **analog system** or **continuous-time** (CT) **system**. The signal to be processed forms the **excitation** or **input** to the system. The processed signal is termed the **response** or **output**.

The response of any system is governed by the input and the system details. A system may of course be excited by more than one input, and this leads to the more general idea of *multiple-input* systems. We address only *single-input, single-output* systems in this text. The study of systems involves the input, the output, and the system specifications. Conceptually, we can determine any one of these in terms of the other two. **System analysis** implies a study of the response subject to known inputs and system formulations. Known input-output specifications, on the other hand, usually allow us to identify, or synthesize, the system. **System identification** or **synthesis** is much more difficult because many system formulations are possible for the same input-output relationship.

Most real-world systems are quite complex and almost impossible to analyze quantitatively. Of necessity, we are forced to use **models** or abstractions that retain the essential features of the system and simplify the analysis, while still providing meaningful results. The analysis of systems refers to the analysis of the models that in fact describe such systems, and it is customary to treat the system and its associated models synonymously. In the context of signal processing, a system that processes the input signal in some fashion is also called a **filter**.

4.1.1 Terminology of Systems

A system requires two separate descriptions for a complete specification, one in terms of its components or external structure and the other in terms of its energy level or internal **state**. The *state* of a system is described by a set of **state variables** that allow us to establish the energy level of the system at any instant.

PGR2026-00033
Krisp Technologies EX1024 Page 82

Such variables may represent physical quantities or may have no physical significance whatever. Their choice is governed primarily by what the analysis requires. For example, capacitor voltages and inductor currents are often used as state variables since they provide an instant measure of the system energy. Any inputs applied to the system result in a change in the energy or state of the system. All physical systems are, by convention, referenced to a zero-energy state (variously called the **ground state**, the **rest state**, the **relaxed state**, or the **zero state**) at $t = -\infty$.

The behavior of a system is governed not only by the input but also by the state of the system at the instant at which the input is applied. The initial values of the state variables define the **initial conditions** or **initial state**. This initial state, which must be known before we can establish the complete system response, embodies the past history of the system. It allows us to predict the future response due to any input *regardless of how the initial state was arrived at.*

4.1.2 Operators

Any equation is based on a set of operations. An **operator** is a rule or a set of directions—a recipe if you will—that shows us how to transform one function to another. For example, the derivative operator $s \equiv \frac{d}{dt}$ transforms a function $x(t)$ to $y(t) = s\{x(t)\}$ or $\frac{d\,x(t)}{dt}$. If an operator or a rule of operation is represented by the symbol $\mathcal{O}$, the equation

$$\mathcal{O}\{x(t)\} = y(t) \tag{4.1}$$

implies that if the function $x(t)$ is treated exactly as the operator $\mathcal{O}$ requires, we obtain the function $y(t)$. For example, the operation $\mathcal{O}\{\ \} = 4\,\frac{d}{dt}\{\ \} + 6$ says that to get $y(t)$, we must take the derivative of $x(t)$, multiply by 4 and then add 6 to the result $4\,\frac{d}{dt}\{x(t)\} + 6 = 4\,\frac{dx}{dt} + 6 = y(t)$.

If an operation on the sum of two functions is equivalent to the sum of operations applied to each separately, the operator is said to be **additive**. In other words,

$$\mathcal{O}\{x_1(t) + x_2(t)\} = \mathcal{O}\{x_1(t)\} + \mathcal{O}\{x_2(t)\} \qquad \text{(for an additive operation)} \tag{4.2}$$

If an operation on $Kx(t)$ is equivalent to K times the linear operation on $x(t)$ where K is a scalar, the operator is said to be **homogeneous**. In other words,

$$\mathcal{O}\{Kx(t)\} = K\mathcal{O}\{x(t)\} \qquad \text{(for a homogeneous operation)} \tag{4.3}$$

Together, the two describe the **principle of superposition**. An operator $\mathcal{O}$ is termed a **linear operator** if it is both additive and homogeneous. In other words,

$$\mathcal{O}\{Ax_1(t) + Bx_2(t)\} = A\mathcal{O}\{x_1(t)\} + B\mathcal{O}\{x_2(t)\} \qquad \text{(for a linear operation)} \tag{4.4}$$

If an operation performed on a linear combination of $x_1(t)$ and $x_2(t)$ produces the same results as a linear combination of operations on $x_1(t)$ and $x_2(t)$ separately, the operation is linear. If not, it is nonlinear. Linearity thus implies superposition. An important concept that forms the basis for the study of linear systems is that *the superposition of linear operators is also linear.*

REVIEW PANEL 4.1

A Linear Operator Obeys Superposition: $a\mathcal{O}\{x_1(t)\} + b\mathcal{O}\{x_2(t)\} = \mathcal{O}\{ax_1(t) + bx_2(t)\}$

Superposition implies both **homogeneity** and **additivity.**

Homogeneity: $\mathcal{O}\{ax(t)\} = a\mathcal{O}\{x(t)\}$ **Additivity:** $\mathcal{O}\{x_1(t)\} + \mathcal{O}\{x_2(t)\} = \mathcal{O}\{x_1(t) + x_2(t)\}$

Testing an Operator for Linearity: If an operator fails either the additive or the homogeneity test, it is nonlinear. In all but a few (usually contrived) cases, if an operator passes *either* the additive *or* the homogeneity test, it is linear (meaning that it will also pass the other). In other words, only one test, additivity or homogeneity, suffices to confirm linearity (or lack thereof) in most cases.

PGR2026-00033
Krisp Technologies EX1024 Page 83

EXAMPLE 4.1 (Testing for Linear Operations)

(a) Consider the operator $\mathcal{O}\{\ \} = \log\{\ \}$.

Since $\log(Kx) \neq K \log x$, the log operator is nonlinear because it is not homogeneous.

(b) Consider the operator $\mathcal{O}\{x\} = C\{x\} + D$.

If we test for homogeneity, we find that $\mathcal{O}\{Kx\} = KCx + D$ but $K\mathcal{O}\{x\} = KCx + KD$.

The operator is thus nonlinear. Only if $D = 0$ is the operation linear.

(c) Consider the squaring operator $\mathcal{O}\{\ \} = \{\ \}^2$, which transforms $x(t)$ to $x^2(t)$.

We find $A\mathcal{O}\{x(t)\} = Ax^2(t)$ but $\mathcal{O}\{Ax(t)\} = [Ax(t)]^2 = A^2x^2(t)$.

Since the two are not equal, the squaring operator is also nonlinear.

(d) Consider the derivative operator $\mathcal{O}\{\ \} = \frac{d\{\ \}}{dt}$, which transforms $x(t)$ to $x'(t)$.

We find that $A\mathcal{O}\{x(t)\} = Ax'(t)$ and $\mathcal{O}\{Ax(t)\} = x'(At) = Ax'(t)$.

The two are equal, and the derivative operator is homogeneous and thus linear.

Of course, to be absolutely certain, we could use the full force of the linearity relation to obtain $\mathcal{O}\{Ax_1(t) + Bx_2(t)\} = \frac{d}{dt}\left[Ax_1(t) + Bx_2(t)\right]$ and $A\mathcal{O}\{x_1(t)\} + A\mathcal{O}\{x_2(t)\} = Ax_1'(t) + Bx_2'(t)$.

The two results are equal, and we thus confirm the linearity of the derivative operator.

4.2 System Classification

Systems may be classified in several ways. Such classifications allow us to make informed decisions on the choice of a given method of analysis over others, depending on the context in which the system is considered. At the quantitative level, analog systems are usually modeled by differential equations that relate the output $y(t)$ and input $x(t)$ through the system parameters and the independent variable t. Using the notation $y^{(n)}(t) \equiv \frac{d^n y(t)}{dt^n}$, the general form of a differential equation may be written as

$$y^{(n)}(t) + a_1 y^{(n-1)}(t) + \cdots + a_{n-1}y^{(1)}(t) + a_n y(t) = b_0 x^{(m)}(t) + b_1 x^{(m-1)}(t) + \cdots + b_{m-1}x^{(1)}(t) + b_m x(t) \quad (4.5)$$

The **order** n of the differential equation refers to the order of the *highest derivative of the output* $y(t)$. It is customary to normalize the coefficient of the highest derivative of $y(t)$ to 1. The coefficients a_k and b_k may be functions of $x(t)$ and/or $y(t)$ and/or t. Using the *derivative operator* $s^k \equiv \frac{d^k}{dt^k}$ with $s^0 \equiv 1$, we may recast this equation in **operator notation** as

$$\{s^n + a_1 s^{n-1} + \cdots + a_{n-1}s + a_n\}y(t) = \{b_0 s^m + b_1 s^{m-1} + \cdots + b_{m-1}s + b_m\}x(t) \quad (4.6)$$

Notation: For low-order systems, we will also use the notation $y'(t) \equiv \frac{d\,y(t)}{dt}$, $y''(t) \equiv \frac{d^2\,y(t)}{dt^2}$ etc.

PGR2026-00033
Krisp Technologies EX1024 Page 84

4.2.1 Linearity and Time Invariance

A linear system is one for which superposition applies and implies three constraints:

1. The system equation must involve only linear operators.
2. The system must contain no internal independent sources.
3. The system must be relaxed (with zero initial conditions).

REVIEW PANEL 4.2

What Makes a System Nonlinear?

(1) nonlinear elements *or* **(2)** nonzero initial conditions *or* **(3)** internal sources

For a linear system, scaling the input leads to an identical scaling of the output. In particular, this means zero output for zero input and a linear input-output relation *passing through the origin*. This is possible only if *every* system element obeys a similar relationship at its own terminals. Since independent sources have terminal characteristics that are constant and do not pass through the origin, a system that includes such sources is therefore nonlinear. Formally, a linear system must also be relaxed (with zero initial conditions) if superposition is to hold. We can, however, use superposition even for a system with nonzero initial conditions (or internal sources) that is otherwise linear. We treat it as a multiple-input system by including the initial conditions (or internal sources) as additional inputs. The output then equals the superposition of the outputs due to each input acting alone, and any *changes* in the input are related linearly to *changes* in the response. As a result, the response can be written as the sum of a *zero-input response* (due to the initial conditions alone) and the *zero-state response* (due to the input alone). This is the **principle of decomposition**, which allows us to analyze linear systems in the presence of nonzero initial conditions. Both the zero-input response and the zero-state response obey superposition individually.

REVIEW PANEL 4.3

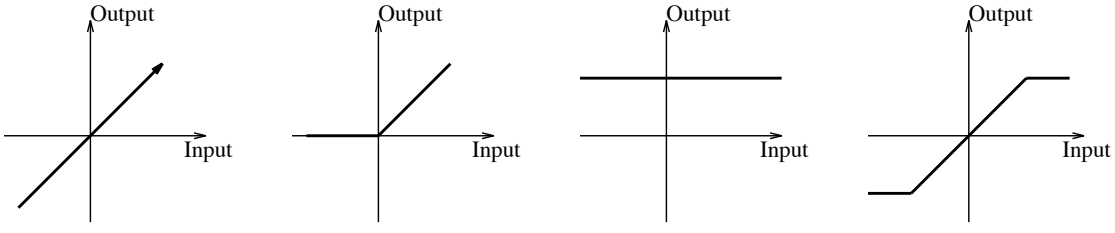
Linearity from the Input-Output Relation

The input-output relation is a **straight line** *passing through the origin*.

Examples: Input $= v(t)$ Output $= C\dfrac{dv(t)}{dt}$ Input $= v(t)$ Output $= \dfrac{1}{L}\displaystyle\int_{-\infty}^{t} v(t)\,dt$

EXAMPLE 4.2 (Linearity from the Input-Output Relation)

The input-output relations for four systems are shown in Figure E4.2. Which systems are linear?



(a) Linear (b) Nonlinear (c) Nonlinear (d) Nonlinear

Figure E4.2 Input-output relations of the systems for Example 4.2

Only the first one is linear because the input-output relation is a straight line that passes through the origin. All other systems are nonlinear. The second system describes a half-wave rectifier, the third describes an internal source, and the fourth describes an operational amplifier.

PGR2026-00033
Krisp Technologies EX1024 Page 85

4.2.2 Time-Invariant (Shift-Invariant) Systems

Time invariance (also called shift invariance) implies that the shape of the response $y(t)$ depends only on the shape of the input $x(t)$ and not on the time when it is applied. If the input is shifted to $x(t-\alpha)$, the response equals $y(t-\alpha)$ and is shifted by the same amount. In other words, the system does not change with time. Such a system is also called **stationary** or **fixed.**

Every element of a time-invariant system must itself be time invariant with a value that is constant *with respect to time*. If the element value depends on the input or output, this only makes the system nonlinear. Coefficients in a system equation that depend on the element values cannot show explicit dependence on time for time-invariant systems. In a time-varying system, the value of at least one system element is a function of time. As a result, the system equation contains time-dependent coefficients. An example is a system containing a time-varying resistor. In physical systems, aging of components often contributes to their time-varying nature.

Formally, if the operator $\mathcal{O}$ transforms the input $x(t)$ to the output $y(t)$ such that $\mathcal{O}\{x(t)\} = y(t)$, a time-invariant system requires that

$$\text{If} \quad \mathcal{O}\{x(t)\} = y(t) \quad \text{then} \quad \mathcal{O}\{x(t-t_0)\} = y(t-t_0) \quad \text{(for a time-invariant operation)} \tag{4.7}$$

REVIEW PANEL 4.4

Time Invariance from the Operational Relation

If $\mathcal{O}\{x(t)\} = y(t)$ then $\mathcal{O}\{x(t-t_0)\} = y(t-t_0)$ (shift input by $\alpha \Rightarrow$ shift output by α).

EXAMPLE 4.3 (Linearity and Time Invariance of Operators)

(a) $y(t) = x(t)x'(t)$ is nonlinear but time invariant.

The operation is $\mathcal{O}\{\ \} = (\{\ \})(\frac{d\{\ \}}{dt})$. We find that

$A\mathcal{O}\{x(t)\} = A[x(t)x'(t)]$ but $\mathcal{O}\{Ax(t)\} = [Ax(t)][Ax'(t)] = A^2x(t)x'(t)$. The two are not equal.

$\mathcal{O}\{x(t-t_0)\} = x(t-t_0)x'(t-t_0)$ and $y(t-t_0) = x(t-t_0)x'(t-t_0)$. The two are equal.

(b) $y(t) = tx(t)$ is linear but time varying.

The operation is $\mathcal{O}\{\ \} = t\{\ \}$. We find that

$A\mathcal{O}\{x(t)\} = A[tx(t)]$ and $\mathcal{O}\{Ax(t)\} = t[Ax(t)]$. The two are equal.

$\mathcal{O}\{x(t-t_0)\} = t[x(t-t_0)]$ but $y(t-t_0) = (t-t_0)x(t-t_0)$. The two are not equal.

(c) $y(t) = x(\alpha t)$ is linear but time varying. With $t \Rightarrow \alpha t$, we see that $A\mathcal{O}\{x(t)\} = A[x(\alpha t)]$ and $\mathcal{O}\{Ax(t)\} = Ax(\alpha t)$. The two are equal.

To test for time invariance, we find that $\mathcal{O}\{x(t-t_0)\} = x(\alpha t - t_0)$ but $y(t-t_0) = x[\alpha(t-t_0)]$. The two are not equal, and the time-scaling operation is time varying. Figure E4.3C illustrates this for $y(t) = x(2t)$, using a shift of $t_0 = 2$.

PGR2026-00033
Krisp Technologies EX1024 Page 86

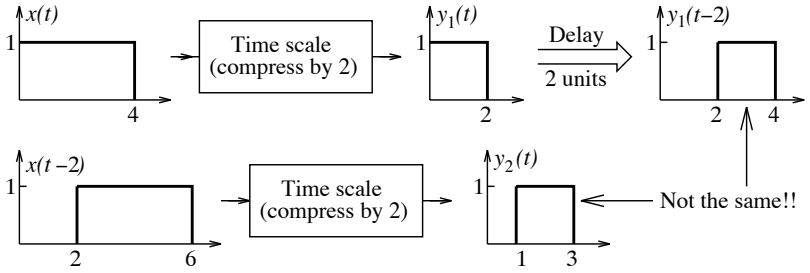


Figure E4.3C Illustrating time variance of the system for Example 4.3(c)

(d) $y(t) = x(t-2)$ is linear and time invariant. The operation $t \Rightarrow t-2$ reveals that
$A\mathcal{O}\{x(t)\} = A[x(t-2)]$ and $\mathcal{O}\{Ax(t)\} = Ax(t-2)$. The two are equal.
$\mathcal{O}\{x(t-t_0)\} = x(t-t_0-2)$ and $y(t-t_0) = x(t-t_0-2)$. The two are equal.

(e) $y(t) = e^{x(t)}x(t)$ is nonlinear but time invariant.
The operation is $\mathcal{O}\{\ \} = e^{\{\ \}}\{\ \}$ and reveals that
$A\mathcal{O}\{x(t)\} = Ae^{x(t)}x(t)$ but $\mathcal{O}\{Ax(t)\} = e^{Ax(t)}[Ax(t)]$. The two are not equal.
$\mathcal{O}\{x(t-t_0)\} = e^{x(t-t_0)}x(t-t_0)$ and $y(t-t_0) = e^{x(t-t_0)}x(t-t_0)$. The two are equal.

4.2.3 Linear Time-Invariant Systems

The important class of **linear time-invariant** (LTI) **systems** are described by differential equations with *constant coefficients*. To test for linearity or time invariance of systems described by differential equations, we can formally apply the linearity or time-invariance test to each operation or recognize the nonlinear or time-varying operations by generalizing the results of the previous examples as follows.

1. *Terms* containing products of the input and/or output make a system equation nonlinear. A constant *term* also makes a system equation nonlinear.
2. *Coefficients* of the input or output that are explicit functions of t make a system equation time varying. Time-scaled inputs or outputs such as $y(2t)$ also make a system equation time varying.

REVIEW PANEL 4.5

An LTI System Is Described by a Linear Constant-Coefficient Differential Equation (LCCDE)

$y^{(n)}(t) + A_{n-1}y^{(n-1)}(t) + \cdots + A_0 y(t) = B_m x^{(m)}(t) + B_{m-1}x^{(m-1)}(t) + \cdots + B_0 x(t)$

All *terms* contain $x(t)$ or $y(t)$. All *coefficients* are constant (not functions of $x(t)$ or $y(t)$ or t).

Notation: $y^{(n)}(t) = \frac{d^n\, y(t)}{dt^n}$, $\quad y^{(0)}(t) = y(t)$ $\qquad$ Also, $\quad y'(t) = \frac{d\, y(t)}{dt}$, $\quad y''(t) = \frac{d^2\, y(t)}{dt^2}$

REVIEW PANEL 4.6

What Makes a System Differential Equation Nonlinear or Time Varying?

It is **nonlinear** if *any term* is a constant or a *nonlinear function* of $x(t)$ or $y(t)$.
It is **time varying** if the *coefficient* of any term in $x(t)$ or $y(t)$ is an *explicit* function of t.

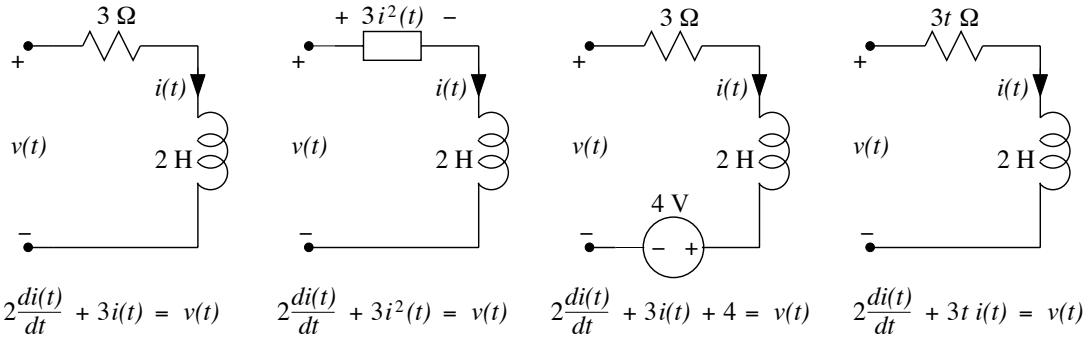
PGR2026-00033
Krisp Technologies EX1024 Page 87

EXAMPLE 4.4 (Linearity and Time Invariance of Systems)

(a) We test for the linearity and time invariance of the following systems:

1. $y'(t) - 2y(t) = 4x(t)$. This is LTI.
2. $y''(t) - 2ty'(t) = x(t)$. This is linear but time varying.
3. $y'(t) + 2y^2(t) = 2x'(t) - x(t)$. This is nonlinear but time invariant.
4. $y'(t) - 2y(t) = e^{x(t)}x(t)$. This is nonlinear but time invariant.
5. $y'(t) - 4y(t)y(2t) = x(t)$. This is nonlinear and time varying.

(b) What can you say about the linearity and time-invariance of the four circuits and their governing differential equations shown in the Figure E4.4B.



(a) LTI (b) Nonlinear (c) Nonlinear (d) Time varying

Figure E4.4B The circuits for Example 4.4(b)

For (a), $2i'(t) + 3i(t) = v(t)$. This is LTI because all the element values are constants.

For (b), $2i'(t) + 3i^2(t) = v(t)$. This is nonlinear due to the nonlinear element.

For (c), $2i'(t) + 3i(t) + 4 = v(t)$. This is nonlinear due to the 4-V internal source.

For (d), $2i'(t) + 3ti(t) = v(t)$. This is time varying due to the time-varying resistor.

Implications of Linearity and Time Invariance

In most practical cases, if an input $x(t)$ to a relaxed LTI system undergoes a linear operation, the output $y(t)$ undergoes the same linear operation. For example, the input $x'(t)$ results in the response $y'(t)$. The superposition property is what makes the analysis of linear systems so much more tractable. Often, an arbitrary function may be decomposed into its simpler constituents, the response due to each analyzed separately and more effectively, and the total response found using superposition. This approach forms the basis for several methods of system analysis. Representing a signal $x(t)$ as a weighted sum of impulses is the basis for the convolution method (Chapter 6), representing a periodic signal $x(t)$ as a linear combination of harmonic signals is the basis for the Fourier series (Chapter 8), and representing a signal $x(t)$ as a weighted sum of complex exponentials is the basis for Fourier and Laplace transforms (Chapters 9 and 11).

PGR2026-00033
Krisp Technologies EX1024 Page 88

4.2.4 Causal and Dynamic Systems

A **causal** or **non-anticipating** system is one whose present response does not depend on future values of the input. The system cannot anticipate future inputs in order to *generate or alter* the present response. Systems whose present response is affected by future inputs are termed **noncausal** or **anticipating**. A formal definition of causality requires identical inputs up to a given duration to produce identical responses over the same duration. Ideal systems (such as ideal filters) often turn out to be noncausal but also form the yardstick by which the performance of many practical systems, implemented to perform the same task, is usually assessed. A differential equation describes a noncausal system if, for example, the output terms have an argument of the form $y(t)$ and an input term has the argument $x(t+\alpha),\ \alpha > 0$.

A **dynamic system**, or a system with **memory**, is characterized by *differential* equations. Its present response depends on both present and past inputs. The memory or past history is due to energy-storage elements that lead to the differential form. In contrast, the response of resistive circuits or circuits operating in the steady state depends only on the instantaneous value of the input, not on past or future values. Such systems are also termed **instantaneous**, **memoryless**, or **static**. All instantaneous systems are also causal.

The system equation of an instantaneous system is *algebraic*, and the input and output are each of the form $x(t)$ and $y(t)$ (with identical scaling or shift, if any). Dynamic systems include (but are not limited to) systems described by differential equations.

REVIEW PANEL 4.7

Causal, Static, and Dynamic Systems

It is **causal** if the output $y(t)$ does not depend on *future inputs* such as $x(t+1)$.
It is **static** if the output $y(t_0)$ depends only on the *instantaneous* value of the input $x(t_0)$.
It is **dynamic** if energy storage is present, and $y(t_0)$ also depends on its own *past* history.

REVIEW PANEL 4.8

Noncausal and Static Systems from the System Differential Equation

It is **noncausal** if output terms have the form $y(t)$ and *any input term* contains $x(t+\alpha),\ \alpha > 0$.
It is **static** if *no derivatives* are present, and *every term* in x and y has *identical arguments*.

EXAMPLE 4.5 (Causal and Dynamic Systems)
We investigate the following systems for causality and memory.

(a) $y''(t) - 2ty'(t) = x(t)$. This is causal and dynamic.

(b) $y(t) = x(t) + 3$. This is causal and instantaneous (but nonlinear).

(c) $y(t) = 2(t+1)x(t)$. This is causal and instantaneous (but time varying).

(d) $y'(t) + 2y(t) = x(t+5)$. This is noncausal and dynamic.

(e) $y'(t+4) + 2y(t) = x(t+2)$. This is causal and dynamic.

(f) $y(t) = x(t+2)$. This is noncausal and dynamic (the arguments of x and y differ).

(g) $y(t) = 2x(\alpha t)$ is causal and instantaneous for $\alpha = 1$, causal and dynamic for $\alpha < 1$, and noncausal and dynamic for $\alpha > 1$. It is also time varying if $\alpha \neq 1$.

PGR2026-00033
Krisp Technologies EX1024 Page 89

4.3 Analysis of LTI Systems

LTI systems can be analyzed in the time domain using any of the following models:

The differential equation representation is quite general and applies even to nonlinear and time-varying systems. For LTI systems it allows the computation of the response using superposition even in the presence of initial conditions. Its major disadvantages are that, as the system order and complexity increases, both the formulation of the differential equations and the evaluation of the initial conditions become quite difficult.

The state variable representation describes an nth-order system by n simultaneous first-order equations, called **state equations**, in terms of n state variables. It is quite useful for complex or nonlinear systems and those with multiple inputs and outputs. For linear systems, the state equations can be solved using matrix methods. We do not discuss this method in this book.

The impulse response representation describes a *relaxed* LTI system by its impulse response $h(t)$. Unlike differential or state representations, the system response $y(t)$ appears explicitly in the governing relation called the *convolution integral*. It also provides the formal ties that allow us to relate time-domain and transformed-domain methods of system analysis. We discuss this method in detail in Chapter 6.

4.4 LTI Systems Described by Differential Equations

An nth-order differential equation requires n initial conditions for its complete solution. Typically, the initial conditions (ICs) are specified as $y(0),\ y^{(1)}(0), \ldots, y^{(n-1)}(0)$ (the response and its $n-1$ successive derivatives at $t = 0$), and the resulting solution is valid for $t \geq 0$. A convenient technique for solving a linear constant-coefficient differential equation (LCCDE) is the **method of undetermined coefficients**, which yields the total response as the sum of the **natural response** $y_N(t)$ and the **forced response** $y_F(t)$.

4.4.1 The Single-Input Case

We start with an nth-order differential equation with a single input $x(t)$ *whose coefficient is unity*:

$$y^{(n)}(t) + a_1 y^{(n-1)}(t) + \cdots + a_{n-1} y^{(1)}(t) + a_n y(t) = x(t) \tag{4.8}$$

$$\{a_0 s^n + a_1 s^{n-1} + \cdots + a_{n-1} s + a_n\} y(t) = x(t) \tag{4.9}$$

The Natural, Forced, and Total Response

The form of the natural response depends only on the system details and is independent of the nature of the input. It is a sum of exponentials whose exponents are the roots (real or complex) of the so-called *characteristic equation* or *characteristic polynomial* defined by

$$a_0 s^n + a_1 s^{n-1} + a_2 s^{n-2} + \cdots + a_{n-2} s^2 + a_{n-1} s + a_n = 0 \tag{4.10}$$

Its n roots, $s_1, s_2, \ldots, s_n$, define the form of the natural response as

$$y_N(t) = K_1 e^{s_1 t} + K_2 e^{s_2 t} + \cdots + K_n e^{s_n t} \tag{4.11}$$

Here, $K_1, K_2, \ldots, K_n$ are as yet undetermined constants (which are evaluated by using the specified initial conditions but *only after setting up the total response*). The form for $y_N(t)$ must be modified for multiple roots. Table 4.1 summarizes the various forms.

Since complex roots occur in conjugate pairs, their associated constants also form conjugate pairs to ensure that $y_N(t)$ is real. Algebraic details lead to the preferred form with two real constants, listed as the second and fourth entries in Table 4.1.

PGR2026-00033
Krisp Technologies EX1024 Page 90

Table 4.1 Form of the Natural Response for Analog LTI Systems

Entry	**Root of Characteristic Equation**	**Form of Natural Response**
1	Real and distinct: r	Ke^{rt}
2	Complex conjugate: $\beta \pm j\omega$	$e^{\beta t}[K_1 \cos(\omega t) + K_2 \sin(\omega t)]$
3	Real, repeated: r^{p+1}	$e^{rt}(K_0 + K_1 t + K_2 t^2 + \cdots + K_p t^p)$
4	Complex, repeated: $(\beta \pm j\omega)^{p+1}$	$e^{\beta t}\cos(\omega t)(A_0 + A_1 t + A_2 t^2 + \cdots + A_p t^p)$ $+ e^{\beta t}\sin(\omega t)(B_0 + B_1 t + B_2 t^2 + \cdots + B_p t^p)$

Table 4.2 Form of the Forced Response for Analog LTI Systems

Note: If the right-hand side (RHS) is $e^{\alpha t}$, where α is also a root of the characteristic equation repeated r times, the forced response form must be multiplied by t^r.

Entry	**Forcing Function (RHS)**	**Form of Forced Response**
1	C_0 (constant)	C_1 (another constant)
2	$e^{\alpha t}$ (see note above)	$Ce^{\alpha t}$
3	$\cos(\omega t + \beta)$	$C_1 \cos(\omega t) + C_2 \sin(\omega t)$ or $C\cos(\omega t + \theta)$
4	$e^{\alpha t}\cos(\omega t + \beta)$ (see note above)	$e^{\alpha t}[C_1 \cos(\omega t) + C_2 \sin(\omega t)]$
5	t	$C_0 + C_1 t$
6	t^p	$C_0 + C_1 t + C_2 t^2 + \cdots + C_p t^p$
7	$te^{\alpha t}$ (see note above)	$e^{\alpha t}(C_0 + C_1 t)$
8	$t^p e^{\alpha t}$ (see note above)	$e^{\alpha t}(C_0 + C_1 t + C_2 t^2 + \cdots + C_p t^p)$
9	$t\cos(\omega t + \beta)$	$(C_1 + C_2 t)\cos(\omega t) + (C_3 + C_4 t)\sin(\omega t)$

REVIEW PANEL 4.9

The Response of LTI Systems Described by Differential Equations

Total Response = Natural Response + Forced Response

The roots of the characteristic equation determine *only the form* of the natural response.
The input terms (RHS) of the differential equation *completely determine* the forced response.
Initial conditions *satisfy the total response* to yield the constants in the natural response.

PGR2026-00033
Krisp Technologies EX1024 Page 91

The **forced response** arises due to the interaction of the system with the input and thus depends on both the input and the system details. It satisfies the given differential equation and has the same form as the input. Table 4.2 summarizes these forms for various types of inputs. The constants in the forced response can be found uniquely and independently of the natural response or initial conditions simply by satisfying the given differential equation.

The **total response** is found by *first* adding the forced and natural response and then evaluating the undetermined constants (in the natural component) using the prescribed initial conditions.

Remarks: For stable systems, the natural response is also called the **transient response**, since it decays to zero with time. For systems with harmonic or switched harmonic inputs, the forced response is a harmonic at the input frequency and is termed the **steady-state response**.

EXAMPLE 4.6 (Natural and Forced Response)

(a) (A First-Order System)

Consider the first-order system $y'(t) + 2y(t) = x(t)$. Find its response if

1. $x(t) = 6,\ y(0) = 8$
2. $x(t) = \cos(2t),\ y(0) = 2$
3. $x(t) = e^{-2t},\ y(0) = 3$

The characteristic equation $s + 2 = 0$ has the root $s = -2$.

The natural response is thus $y_N(t) = Ke^{-2t}$.

1. Since $x(t) = 6$ is a constant, we choose $y_F(t) = C$.
   Then $y_F'(t) = 0$ and $y_F'(t) + 2y_F(t) = 2C = 6$, and thus $y_F(t) = C = 3$.
   The total response is $y(t) = y_N(t) + y_F(t) = Ke^{-2t} + 3$.
   With $y(0) = 8$, we find $8 = K + 3$ (or $K = 5$) and $y(t) = 5e^{-2t} + 3,\ t \geq 0$ or $y(t) = (5e^{-2t} + 3)u(t)$.
2. Since $x(t) = \cos(2t)$, we choose $y_F(t) = A\cos(2t) + B\sin(2t)$.
   Then $y_F'(t) = -2A\sin(2t) + 2B\cos(2t)$, and
   $y_F'(t) + 2y_F(t) = (2A + 2B)\cos(2t) + (2B - 2A)\sin(2t) = \cos(2t)$.
   Comparing the coefficients of the sine and cosine terms on either side, we obtain
   $2A + 2B = 1,\ \ 2B - 2A = 0$ or $A = 0.25,\ B = 0.25$. This gives
   $y_F(t) = 0.25\cos(2t) + 0.25\sin(2t)$.
   The total response is $y(t) = y_N(t) + y_F(t) = Ke^{-2t} + 0.25\cos(2t) + 0.25\sin(2t)$.
   With $y(0) = 2$, we find $2 = K + 0.25$ (or $K = 1.75$) and
   $$y(t) = [1.75e^{-2t} + 0.25\cos(2t) + 0.25\sin(2t)]u(t)$$
   The steady-state response is $y_{ss}(t) = 0.25\cos(2t) + 0.25\sin(2t)$, a sinusoid at the input frequency.
3. Since $x(t) = e^{-2t}$ has the same form as $y_N(t)$, we must choose $y_F(t) = Cte^{-2t}$ (see Table 4.2).
   Then $y_F'(t) = Ce^{-2t} - 2Cte^{-2t}$, and $y_F'(t) + 2y_F(t) = Ce^{-2t} - 2Cte^{-2t} + 2Cte^{-2t} = e^{-2t}$.
   This gives $C = 1$, and thus $y_F(t) = te^{-2t}$ and $y(t) = y_N(t) + y_F(t) = Ke^{-2t} + te^{-2t}$.
   With $y(0) = 3$, we find $3 = K + 0$ and $y(t) = (3e^{-2t} + te^{-2t})u(t)$.

PGR2026-00033
Krisp Technologies EX1024 Page 92

(b) (A Second-Order System)

Let $y''t + 3y'(t) + 2y(t) = x(t),\ t \geq 0$ with $y(0) = 3,\ y'(0) = 4$. Find its response if

1. $x(t) = 4e^{-3t}$
2. $x(t) = 4e^{-2t}$

The characteristic equation is $s^2 + 3s + 2 = 0$ with roots $s_1 = -1$ and $s_2 = -2$.

The natural response is $y_N(t) = K_1e^{-t} + K_2e^{-2t}$.

1. Since $x(t) = 4e^{-3t}$, we select the forced response as $y_F(t) = Ce^{-3t}$. Then $y'_F(t) = -3Ce^{-3t},\ y''_F(t) = 9Ce^{-3t}$, and $y''_F(t) + 3y'_F(t) + 2y_F(t) = (9C - 9C + 2C)e^{-3t} = 4e^{-3t}$. Thus, $C = 2,\ y_F(t) = 2e^{-3t}$, and $y(t) = y_N(t) + y_F(t) = K_1e^{-t} + K_2e^{-2t} + 2e^{-3t}$. Using initial conditions, we get $y(0) = K_1 + K_2 + 2 = 3$ and $y'(0) = -K_1 - 2K_2 - 6 = 4$. This gives $K_2 = -11,\ K_1 = 12$, and $y(t) = (12e^{-t} - 11e^{-2t} + 2e^{-3t})u(t)$.
2. Since $x(t) = 4e^{-2t}$ has the same form as a term of $y_N(t)$, we must choose $y_F(t) = Cte^{-2t}$. Then $y'_F(t) = -2Cte^{-2t} + Ce^{-2t}$, and $y''_F(t) = -2C(1 - 2t)e^{-2t} - 2Ce^{-2t}$. Thus, $y''_F(t) + 3y'_F(t) + 2y_F(t) = (-2C + 4Ct - 2C - 6Ct + 3C + 2Ct)e^{-2t} = 4e^{-2t}$. This gives $C = -4$. Thus, $y_F(t) = -4te^{-2t}$, and $y(t) = y_N(t) + y_F(t) = K_1e^{-t} + K_2e^{-2t} - 4te^{-2t}$. Using initial conditions, we get $y(0) = K_1 + K_2 = 3$ and $y'(0) = -K_1 - 2K_2 - 4 = 4$. Thus, $K_2 = -11,\ K_1 = 14$, and $y(t) = (14e^{-t} - 11e^{-2t} - 4te^{-2t})u(t)$.

4.4.2 The Zero-State Response and Zero-Input Response

It is often more convenient to describe the response $y(t)$ of an LTI system as the sum of its zero-state response (ZSR) $y_{zs}(t)$ (assuming zero initial conditions) and zero-input response (ZIR) $y_{zi}(t)$ (assuming zero input). Each component is found using the method of undetermined coefficients. Note that the natural and forced components $y_N(t)$ and $y_F(t)$ do not, in general, correspond to the zero-input and zero-state response, respectively, even though each pair adds up to the total response. Remember also that the zero-state response obeys superposition (as does the zero-input response).

REVIEW PANEL 4.10

The Zero-Input Response (ZIR) and Zero-State Response (ZSR) of LTI Systems

Total response = ZIR (assume zero input, use given ICs) + **ZSR** (assume zero ICs).
Each obeys superposition. Each is found from its own *natural* and *forced* components.

REVIEW PANEL 4.11

Solving $y^{(n)}(t) + a_1y^{(n-1)}(t) + \cdots + a_ny(t) = x(t)$

1. Find the ZSR from $y_{zs}^{(n)}(t) + a_1y_{zs}^{(n-1)}(t) + \cdots + a_ny_{zs}(t) = x(t)$ **with zero initial conditions**.

2. Find the ZIR from $y_{zi}^{(n)}(t) + a_1y_{zi}^{(n-1)}(t) + \cdots + a_ny_{zi}(t) = 0$ **with given initial conditions**.

3. Find complete response as $y(t) = y_{zs}(t) + y_{zi}(t)$

PGR2026-00033
Krisp Technologies EX1024 Page 93

EXAMPLE 4.7 (Zero-Input and Zero-State Response for the Single-Input Case)
Let $y''(t) + 3y'(t) + 2y(t) = x(t)$ with $x(t) = 4e^{-3t}$ and initial conditions $y(0) = 3$ and $y'(0) = 4$.
Find its zero-input response and zero-state response.

The characteristic equation is $s^2 + 3s + 2 = 0$ with roots $s_1 = -1$ and $s_2 = -2$.
Its natural response is $y_N(t) = K_1 e^{s_1 t} + K_2 e^{s_2 t} = K_1 e^{-t} + K_2 e^{-2t}$.

1. The zero-input response is found from $y_N(t)$ and the prescribed initial conditions:

$$y_{\text{zi}}(t) = K_1 e^{-t} + K_2 e^{-2t} \qquad y_{\text{zi}}(0) = K_1 + K_2 = 3 \qquad y'_{\text{zi}}(0) = -K_1 - 2K_2 = 4$$

   This yields $K_2 = -7$, $K_1 = 10$, and $y_{\text{zi}}(t) = 10e^{-t} - 7e^{-2t}$.

2. Similarly, $y_{\text{zs}}(t)$ is found from the general form of $y(t)$ but with zero initial conditions.
   Since $x(t) = 4e^{-3t}$, we select the forced response as $y_F(t) = Ce^{-3t}$.
   Then, $y'_F(t) = -3Ce^{-3t}$, $y''_F(t) = 9Ce^{-3t}$, and $y''_F(t) + 3y'_F(t) + 2y_F(t) = (9C - 9C + 2C)e^{-3t} = 4e^{-3t}$.
   Thus, $C = 2$, $y_F(t) = 2e^{-3t}$, and $y_{\text{zs}}(t) = K_1 e^{-t} + K_2 e^{-2t} + 2e^{-3t}$.
   With zero initial conditions, we obtain

$$y_{\text{zs}}(0) = K_1 + K_2 + 2 = 0 \qquad y'_{\text{zs}}(0) = -K_1 - 2K_2 - 6 = 0$$

   This yields $K_2 = -4$, $K_1 = 2$, and $y_{\text{zs}}(t) = 2e^{-t} - 4e^{-2t} + 2e^{-3t}$.

3. The total response is the sum of $y_{\text{zs}}(t)$ and $y_{\text{zi}}(t)$:

$$y(t) = y_{\text{zi}}(t) + y_{\text{zs}}(t) = 12e^{-t} - 11e^{-2t} + 2e^{-3t}, \qquad t \geq 0$$

4.4.3 Solution of the General Differential Equation

The solution to the general nth-order differential equation described by

$$y^{(n)}(t) + a_1 y^{(n-1)}(t) + \cdots + a_n y(t) = b_0 x^{(m)}(t) + b_1 x^{(m-1)}(t) + \cdots + b_m x(t) \tag{4.12}$$

is found by invoking linearity and superposition as follows:

1. Compute the zero-state response $y_0(t)$ to the single-input system

$$y_0^{(n)}(t) + a_1 y_0^{(n-1)}(t) + \cdots + a_n y_0(t) = x(t) \tag{4.13}$$

2. Use linearity and superposition to find $y_{\text{zs}}(t)$ as

$$y_{\text{zs}}(t) = b_0 y_0^{(m)}(t) + b_1 y_0^{(m-1)}(t) + \cdots + b_m y_0(t) \tag{4.14}$$

3. Find the zero-input response $y_{\text{zi}}(t)$.
4. Find the total response as $y(t) = y_{\text{zs}}(t) + y_{\text{zi}}(t)$.

Note that the zero-input response is computed and included just once.

PGR2026-00033
Krisp Technologies EX1024 Page 94

REVIEW PANEL 4.12

Solving $y^{(n)}(t) + a_1 y^{(n-1)}(t) + \cdots + a_n y(t) = b_0 x^{(m)}(t) + b_1 x^{(m-1)}(t) + \cdots + b_m x(t)$

1. Find $y_0(t)$ from $y_0^{(n)}(t) + a_1 y_0^{(n-1)}(t) + \cdots + a_n y_0(t) = x(t)$ **using zero initial conditions**.

2. Find ZSR (using superposition) as $y_{zs}(t) = b_0 y_0^{(m)}(t) + b_1 y_0^{(m-1)}(t) + \cdots + b_m y_0(t)$

3. Find ZIR from $y_{zi}^{(n)}(t) + a_1 y_{zi}^{(n-1)}(t) + \cdots + a_n y_{zi}(t) = 0$ **using given initial conditions**.

4. Find complete response as $y(t) = y_{zs}(t) + y_{zi}(t)$

EXAMPLE 4.8 (Solving a Differential Equation: The General Case)
Find the response of $y''(t) + 3y'(t) + 2y(t) = 2x'(t) + x(t)$, with $x(t) = 4e^{-3t}$, $y(0) = 0$, and $y'(0) = 1$.

1. The characteristic equation is $s^2 + 3s + 2 = 0$ with roots $s_1 = -1$ and $s_2 = -2$.
   Its natural response is $y_N(t)K_1 e^{-t} + K_2 e^{-2t}$. Since $x(t) = 4e^{-3t}$, choose $y_F(t) = Ce^{-3t}$.
   Then, $y''(t) + 3y'(t) + 2y(t) = 9Ce^{-3t} - 9Ce^{-3t} + 2Ce^{-3t} = 4e^{-3t}$, and $C = 2$, and $y_F(t) = 2e^{-3t}$
   Thus, $y_0(t)$ has the form $y_0(t) = K_1 e^{-t} + K_2 e^{-2t} + 2e^{-3t}$.
   Using zero initial conditions, $K_1 + K_2 + 2 = 0$ and $-K_1 - 2K_2 - 6 = 0$. Then $K_1 = 2$, $K_2 = -4$.
   Thus, $y_0(t) = 2e^{-t} - 4e^{-2t} + 2e^{-3t}$.

2. Since $y_0'(t) = -2e^{-t} + 8e^{-2t} - 6e^{-3t}$, we find the ZSR as
$$y_{zs}(t) = 2y_0'(t) + y_0(t) = -2e^{-t} + 12e^{-2t} - 10e^{-3t}, \qquad t \geq 0$$

3. The ZIR is found from $y_{zi}(t) = C_1 e^{-t} + C_2 e^{-2t}$, with $y(0) = 0$ and $y'(0) = 1$. This yields $y_{zi}(0) = C_1 + C_2 = 0$ and $y_{zi}'(0) - C_1 - 2C_2 = 1$. We find $C_1 = 1$ and $C_2 = -1$. Then,
$$y_{zi}(t) = e^{-t} - e^{-2t}$$

4. Finally, the total response is $y(t) = y_{zs}(t) + y_{zi}(t) = -e^{-t} + 11e^{-2t} - 10e^{-3t},\ t \geq 0$.

4.5 The Impulse Response of LTI Systems

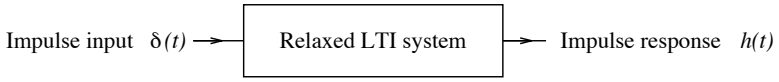
The **impulse response**, denoted by $h(t)$, is simply the response of a *relaxed* LTI system to a unit impulse $\delta(t)$. The impulse response provides us with a powerful method for evaluating the *zero-state response* of LTI systems to arbitrary inputs using superposition (as described in the next chapter). The impulse response is also related to the *transfer function*, which is used extensively for system analysis in the transformed domain (Chapters 8–12). The impulse response and step response are often used to characterize the time-domain performance of systems and filters.

Methods for finding the impulse response in the time domain are often tedious. It is often easier to find the impulse response of a system by resorting to the transformed domain. Here we consider approaches that rely only on time-domain methods and on the concepts of linearity.

PGR2026-00033
Krisp Technologies EX1024 Page 95

REVIEW PANEL 4.13

The Impulse Response and Step Response Is Defined Only for Relaxed LTI Systems



Impulse response $h(t)$: The output of a **relaxed** LTI system if the input is a unit impulse $\delta(t)$.
Step response $s(t)$: The output of a **relaxed** LTI system if the input is a unit step $u(t)$.

4.5.1 Impulse Response from the Step Response

For a linear system, if a unit step input $u(t)$ leads to the step response $s(t)$, the impulse $\delta(t) = u'(t)$ leads to the impulse response $h(t) = s'(t)$. This result is the basis for the assertion that the impulse response $h(t)$ equals the derivative of the step response $s(t)$, a concept that suggests a simple way of finding $h(t)$ for a relaxed system. We evaluate the step response $s(t)$ and then let $h(t) = s'(t)$.

REVIEW PANEL 4.14

The Impulse Response $h(t)$ Is the Derivative of the Step Response $s(t)$

$$h(t) = s'(t) = \frac{d\,s(t)}{dt} \qquad\qquad s(t) = \int_{-\infty}^{t} h(t)\,dt$$

4.5.2 Impulse Response of a First-Order System

Consider the first-order system $y'(t) + ay(t) = x(t)$, where $x(t) = u(t)$. Its characteristic equation $(s + a = 0)$ yields the natural response $s_N(t) = K\exp(-at),\ t \geq 0$. The forced response is a constant of the form $s_F(t) = B$. Substituting into the governing differential equation, we obtain $0 + aB = 1$ or $B = 1/a$. The total response is then given by

$$s(t) = s_F(t) + s_N(t) = \frac{1}{a} + Ke^{-at}, \quad t \geq 0$$

Since $s(0) = 0$, we get $0 = \frac{1}{a} + K$ or $K = -1/a$. The step response is thus

$$s(t) = \frac{1 - e^{-at}}{a}u(t) \tag{4.15}$$

The impulse response $h(t)$ equals the derivative of the step response. Thus,

$$h(t) = s'(t) = \frac{d}{dt}\left[\frac{1 - e^{-at}}{a}u(t)\right] = e^{-at}u(t) \tag{4.16}$$

4.5.3 Impulse Response for the Single-Input Case

To find the impulse response directly, we must apparently (and without success) solve $y'(t) + ay(t) = 0$, a homogeneous differential equation (because $\delta(t) = 0,\ t > 0$) subject to zero initial conditions. Of course, what is missing is that the impulse produces a sudden change in the initial conditions and then disappears. In other words, the impulse response $h(t)$ is equivalent to solving a homogeneous differential equation, but subject to the *nonzero initial conditions* produced by the impulse. For the impulse response of the first-order circuit considered previously, we note that $h(0) = 1$. Finding the impulse response of $y'(t) + ay(t) = x(t)$ is then equivalent to solving the homogeneous equation $y'(t) + ay(t) = 0$ with $y(0) = 1$.

PGR2026-00033
Krisp Technologies EX1024 Page 96

REVIEW PANEL 4.15

An Impulse Produces a Sudden Change in State (Initial Conditions)

Solving $y'(t) + \alpha y(t) = \delta(t), \quad y(0) = 0$ is equivalent to solving $y'(t) + \alpha y(t) = 0, \quad y(0) = 1.$

Similarly, it turns out that the impulse response of the second-order system $y''(t)+a_1y'(t)+a_2y(t) = x(t)$ can be found as the solution to the homogeneous equation $y''(t)+a_1y'(t)+a_2y(t) = 0$, with initial conditions $y(0) = 0$ and $y'(0) = 1$. These results can be generalized to higher-order systems. For the nth-order, single-input system given by

$$y^{(n)}(t) + a_1y^{(n-1)}(t) + \cdots + a_ny(t) = x(t) \tag{4.17}$$

the impulse response $h(t)$ is found as the solution to the homogeneous equation

$$h^{(n)}(t) + a_1h^{(n-1)}(t) + \cdots + a_nh(t) = 0 \qquad h^{(n-1)}(0) = 1 \text{ (and all other ICs zero)} \tag{4.18}$$

Note that the *highest-order* initial condition is $h^{(n-1)}(0) = 1$ and all other initial conditions are zero.

REVIEW PANEL 4.16

Impulse Response of $y^{(n)}(t) + a_1y^{(n-1)}(t) + \cdots + a_ny(t) = x(t)$

Find $h(t)$ from $h^{(n)}(t) + a_1h^{(n-1)}(t) + \cdots + a_nh(t) = 0,$ with $h^{(n-1)}(0) = 1$ (*and all other ICs zero*).

EXAMPLE 4.9 (Impulse Response for the Single-Input Case)

(a) (A First-Order System)

Find the impulse response of the system described by $y'(t) + 2y(t) = x(t)$.

The characteristic equation $s + 2 = 0$ has the root $s = -2$.

The natural response is thus $h(t) = Ke^{-2t}$.

With $h(0) = 1$, we obtain $h(0) = K = 1$, and $h(t) = e^{-2t}u(t)$.

(b) (A Second-Order System)

Find the impulse response of the system described by $y''(t) + 3y'(t) + 2y(t) = x(t)$.

The characteristic equation $s^2 + 3s + 2 = 0$ has the roots $s_1 = -1$ and $s_2 = -2$.

The natural response is $h(t) = K_1e^{-t} + K_2e^{-2t}$.

With $h(0) = 0$ and $h'(0) = 1$, we obtain

$h(0) = K_1 + K_2 = 0$ and $h'(0) = -K_1 - 2K_2 = 1$. This gives $K_1 = 1$ and $K_2 = -1$.

Thus, $h(t) = (e^{-t} - e^{-2t})u(t)$.

PGR2026-00033
Krisp Technologies EX1024 Page 97

4.5.4 Impulse Response for the General Case

To find the impulse response $h(t)$ of the general system described by

$$y^{(n)}(t) + a_1 y^{(n-1)}(t) + \cdots + a_n y(t) = b_0 x^{(m)}(t) + \cdots + b_m x(t) \tag{4.19}$$

we use linearity and superposition as follows:

1. Start with the single-input system

$$y^{(n)}(t) + a_1 y^{(n-1)}(t) + \cdots + a_n y(t) = x(t)$$

and compute its impulse response $h_0(t)$ from the homogeneous equation

$$h_0^{(n)}(t) + a_1 h_0^{(n-1)}(t) + \cdots + a_n h_0(t) = 0, \qquad h_0^{(n-1)}(0) = 1 \tag{4.20}$$

2. Then use linearity and superposition to compute $h(t)$ as

$$h(t) = b_0 h_0^{(m)}(t) + b_1 h_0^{(m-1)}(t) + \cdots + b_m h_0(t) \tag{4.21}$$

REVIEW PANEL 4.17

Impulse Response of $y^{(n)}(t) + a_1 y^{(n-1)}(t) + \cdots + a_n y(t) = b_0 x^{(m)}(t) + \cdots + b_m x(t)$

1. Find $h_0(t)$ from $h_0^{(n)}(t) + a_1 h_0^{(n-1)}(t) + \cdots + a_n h_0(t) = 0$, IC: $h_0^{(n-1)}(0) = 1$ (*all other ICs zero*).

2. Find $h(t)$ (using superposition) as $h(t) = b_0 h_0^{(m)}(t) + b_1 h_0^{(m-1)}(t) + \cdots + b_m h_0(t)$.

EXAMPLE 4.10 (Impulse Response for the General Case)

(a) Find the impulse response of the system $y'(t) + 2y(t) = 3x(t)$.

The impulse response $h_0(t)$ of the single-input system $y'(t) + 2y(t) = x(t)$ is $h_0(t) = e^{-2t}u(t)$.

The impulse response of the given system is thus $h(t) = 3h_0(t) = 3e^{-2t}u(t)$.

(b) Find the impulse response of the system $y'(t) + 2y(t) = x'(t) + 3x(t)$.

The impulse response $h_0(t)$ of the single-input system $y'(t) + 2y(t) = x(t)$ is $h_0(t) = e^{-2t}u(t)$.

The impulse response of the given system is thus

$h(t) = h_0'(t) + 3h_0(t) = \delta(t) - 2e^{-2t}u(t) + 3e^{-2t}u(t) = \delta(t) + e^{-2t}u(t)$.

(c) Find the impulse response of the system $y''(t) + 3y'(t) + 2y(t) = x''(t)$.

The impulse response $h_0(t)$ of the system $y''(t) + 3y'(t) + 2y(t) = x(t)$ is (from Example 4.9) $h_0(t) = (e^{-t} - e^{-2t})u(t)$. The required impulse response is then $h(t) = h_0''(t)$. We compute:

$$h_0'(t) = (-e^{-t} + 2e^{-2t})u(t)$$

$$h(t) = h_0''(t) = \frac{d}{dt}[h_0'(t)] = (e^{-t} - 4e^{-2t})u(t) + \delta(t)$$

PGR2026-00033
Krisp Technologies EX1024 Page 98

4.6 System Stability

Stability is important in practical design and is defined in many contexts. In the time domain, stability involves constraints on the nature of the system response. **Bounded-input**, **bounded-output** (BIBO) **stability** implies that *every* bounded input results in a bounded output. For an LTI system described by the differential equation

$$y^{(n)}(t) + a_1 y^{(n-1)}(t) + \cdots + a_n y(t) = b_0 x^{(m)}(t) + b_1 x^{(m-1)}(t) + \cdots + b_m x(t), \quad m \le n \tag{4.22}$$

the conditions for BIBO stability involve the roots of the characteristic equation. A necessary and sufficient condition for BIBO stability of an LTI system is that *every root of its characteristic equation must have a negative real part (and the highest derivative of the input must not exceed that of the output).* This criterion is based on the results of Tables 4.1 and 4.2. Roots with negative real parts ensure that the natural (and zero-input) response always decays with time (see Table 4.1), and the forced (and zero-state) response always remains bounded for *every* bounded input. Roots with zero real parts make the system unstable. Simple (non-repeated) roots with zero real parts produce a constant (or sinusoidal) natural response which is bounded, but if the input is also a constant (or a sinusoid at the same frequency), the forced response is a ramp or growing sinusoid (see Table 4.2) and hence unbounded. Repeated roots with zero real parts result in a natural response that is itself a growing sinusoid or polynomial and thus unbounded.

If the highest derivative of the input *exceeds* (not just equals) that of the output, the system is unstable. For example, if $y(t) = \frac{dx(t)}{dt}$, a step input (which is bounded) produces an impulse output (which is unbounded at $t = 0$). In the next chapter, we shall see that the stability condition described here is entirely equivalent to having an LTI system whose impulse response $h(t)$ is *absolutely integrable*. The stability of nonlinear or time-varying systems must usually be checked by other means.

REVIEW PANEL 4.18

Requirements for BIBO Stability of Analog Systems

1. The degree of the highest derivative of $x(t)$ must not exceed that of the highest derivative of $y(t)$.
2. Every root of the characteristic equation must have a *negative real part.*

EXAMPLE 4.11 (Concepts Based on Stability)

(a) The system $y''(t)+3y'(t)+2y(t) = x(t)$ is stable since the roots of its characteristic equation $s^2+3s+2 = 0$ are $s = -1, -2$ and have negative real parts.

(b) The system $y''(t) + 3y'(t) = x(t)$ is unstable. The roots of its characteristic equation $s^2 + 3s = 0$ are $s_1 = 0$, and $s_2 = -3$, and one of the roots does not have a negative real part. Although its natural response is bounded (it has the form $y_N(t) = Au(t) + Be^{-3t}u(t)$), the input $x(t) = u(t)$ produces a forced response of the form $Ctu(t)$, which becomes unbounded.

(c) The system $y'''(t) + 3y''(t) = x(t)$ is unstable. The roots of its characteristic equation $s^3 + 3s^2 = 0$ are $s_1 = s_2 = 0$, and $s_3 = -3$. They result in the natural response $y_N(t) = Au(t) + Btu(t) + Ce^{-3t}u(t)$, which becomes unbounded.

PGR2026-00033
Krisp Technologies EX1024 Page 99

4.7 Application-Oriented Examples

In this section, we analyze several circuits that are useful in understanding the concepts of signal processing and discuss some of their performance characteristics based on the results of the analysis.

4.7.1 An RC Lowpass Filter

Consider the relaxed RC lowpass filter sketched in the following review panel. Writing a node equation (based on Kirchhoff's current law) at the node labeled A, we get the result

$$y'(t) + \frac{1}{\tau}y(t) = \frac{1}{\tau}x(t) \tag{4.23}$$

where $\tau = RC$ defines the time constant of the circuit. The characteristic equation $s + \frac{1}{\tau} = 0$ yields the natural response $y_N(t) = Ke^{-t/\tau}$. To find the step response, we let $x(t) = u(t) = 1,\ t \geq 0$. So the forced response, $y_F(t) = B$, is a constant. Then, with $y_F'(t) = 0$, we obtain

$$y_F'(t) + \frac{1}{\tau}y_F(t) = \frac{1}{\tau} = 0 + B \qquad \text{or} \qquad B = 1$$

Thus, $y(t) = y_F(t) + y_N(t) = 1 + Ke^{-t/\tau}$. With $y(0) = 0$, we get $0 = 1 + K$ and

$$s(t) = y(t) = (1 - e^{-t/\tau})u(t) \qquad \text{(step response)} \tag{4.24}$$

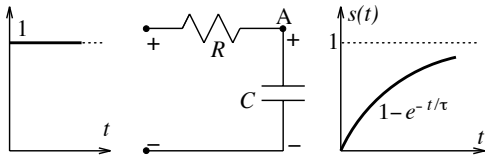
The impulse response $h(t)$ equals the derivative of the step response. Thus,

$$h(t) = s'(t) = \frac{1}{\tau}e^{-t/\tau}u(t) \qquad \text{(impulse response)} \tag{4.25}$$

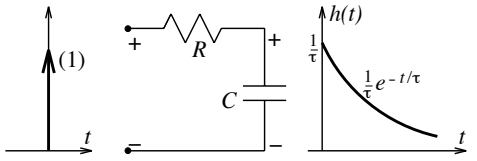
REVIEW PANEL 4.19

Unit Step Response and Unit Impulse Response of an RC Lowpass Filter

The output is the capacitor voltage. The time constant is $\tau = RC$.



Step response: $s(t) = (1 - e^{-t/\tau})u(t)$



Impulse response: $h(t) = s'(t) = \frac{1}{\tau}e^{-t/\tau}u(t)$

Performance Measures

The time-domain performance of systems is often measured in terms of their impulse response and/or step response. For an exponential signal $Ae^{-t/\tau}$, the smaller the time constant τ, the faster is the decay. For first-order systems, the time constant τ is a useful measure of the *speed* of the response, as illustrated in Figure 4.1.

The smaller the time constant, the faster the system responds, and the more the output resembles (matches) the applied input. An exponential decays to less than 1% of its peak value in about 5τ. As a result, the step response is also within 1% of its final value in about 5τ. This forms the basis for the observation that it takes about 5τ to reach steady state. For higher-order systems, the rate of decay and the

PGR2026-00033
Krisp Technologies EX1024 Page 100

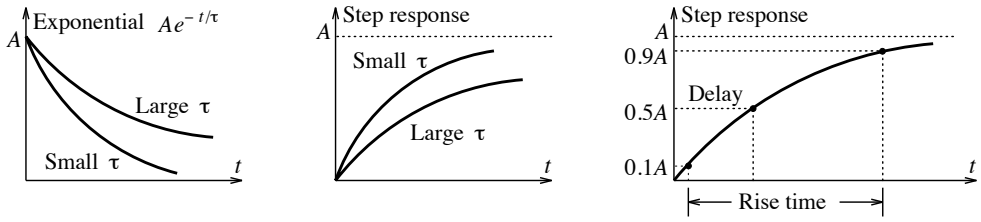


Figure 4.1 Time-domain performance measures

time to reach steady state depends on the *largest* time constant $\tau_{\max}$ (corresponding to the slowest decay) associated with the exponential terms in its impulse response. A smaller $\tau_{\max}$ implies a faster response and a shorter time to reach steady state. The speed of response is also measured by the **rise time**, which is often defined as the time it takes for the step response to rise from 10% to 90% of its final value. Another useful measure of system performance is the **delay time**, which is often defined as the time it takes for the step response to reach 50% of its final value. These measures are also illustrated in Figure 4.1. Another measure is the **settling time**, defined as the time it takes for the step response to settle to within a small fraction (typically, 5%) of its final value.

REVIEW PANEL 4.20

Time Domain Measures of System Performance

Time constant τ	The smaller the τ, the more the output resembles the input.
Speed of response	Depends on the largest time constant $\tau_{\max}$ in $h(t)$.
Steady state	Steady state is reached in about $5\tau_{\max}$.
Rise time	Time for step response to rise from 10% to 90% of its final value.
Delay time	Time for step response to reach 50% of its final value.
5% settling time	Time for step response to settle to within 5% of its final value.

EXAMPLE 4.12 (Some Analog Filters)

Find the step response and impulse response of the circuits shown in Figure E4.12.

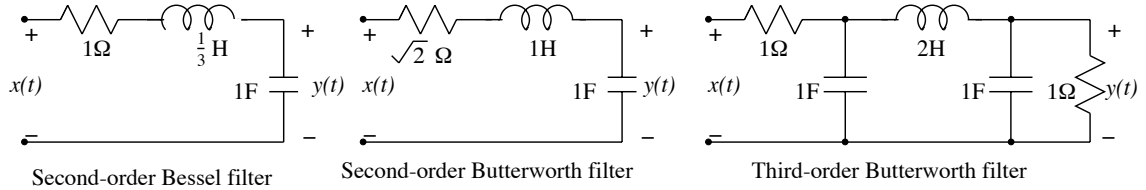


Figure E4.12. Circuits for Example 4.12.

(a) (A Second-Order Bessel Filter)

The governing differential equation is $y''(t) + 3y'(t) + 3y(t) = 3x(t)$.

The characteristic equation is $s^2 + 3s + 3 = 0$. Its roots are $s_{1,2} = -\frac{3}{2} \pm j\frac{\sqrt{3}}{2}$.

The natural response is thus $y_N(t) = e^{-3t/2}[K_1 \cos(\sqrt{3}t/2) + K_2 \sin(\sqrt{3}t/2)]$.

For the step response, $x(t) = 1,\ t \geq 0$, and we assume $y_F(t) = B$.

Then $y_F''(t) + y_F'(t) + 3y_F(t) = 0 + 0 + 3B = 3$, and $B = 1$.

PGR2026-00033
Krisp Technologies EX1024 Page 101

The total response is $y(t) = 1 + e^{-3t/2}[K_1 \cos(\sqrt{3}t/2) + K_2 \sin(\sqrt{3}t/2)]$.

Using zero initial conditions, we get $y(0) = 0 = 1 + K_1$ and $y'(0) = 0 = \frac{\sqrt{3}}{2}K_2 - \frac{3}{2}K_1$.

Thus, $K_1 = -1,\ K_2 = -\sqrt{3}$, and $y(t) = u(t) - e^{-3t/2}[\cos(\sqrt{3}t/2) + \sqrt{3}\sin(\sqrt{3}t/2)]u(t)$.

The impulse response $h(t)$ equals the derivative of the step response.

Thus, $h(t) = y'(t) = 2\sqrt{3}e^{-3t/2}\sin(\sqrt{3}t/2)u(t)$.

The impulse response decays to nearly zero in about 3 s (five time constants).

(b) (A Second-Order Butterworth Filter)

The governing differential equation is $y''(t) + \sqrt{2}y'(t) + y(t) = x(t)$.

The characteristic equation is $s^2 + \sqrt{2}s + 1 = 0$. Its roots are $s_{1,2} = -\frac{1}{\sqrt{2}} \pm j\frac{1}{\sqrt{2}}$.

The natural response is thus $y_N(t) = e^{-t/\sqrt{2}}[K_1 \cos(t/\sqrt{2}) + K_2 \sin(t/\sqrt{2})]$.

For the step response, $x(t) = 1,\ t \geq 0$, and we assume $y_F(t) = B$.

Then $y_F''(t) + \sqrt{2}y_F'(t) + y_F(t) = B = 1$.

The total response is $y(t) = 1 + e^{-t/\sqrt{2}}[K_1 \cos(t/\sqrt{2}) + K_2 \sin(t/\sqrt{2})]$.

Using zero initial conditions, we get $y(0) = 0 = 1 + K_1$ and $y'(0) = 0 = -\frac{1}{\sqrt{2}}(K_2 - K_1)$.

Thus, $K_1 = K_2 = -1$ and $y(t) = u(t) - e^{-t/\sqrt{2}}[\cos(t/\sqrt{2}) + \sin(t/\sqrt{2})]u(t)$.

The impulse response $h(t)$ equals the derivative of the step response.

Thus, $h(t) = y'(t) = \sqrt{2}e^{-t/\sqrt{2}}\sin(t/\sqrt{2})u(t)$.

The impulse response decays to nearly zero in about 7 s (five time constants).

(c) (A Third-Order Butterworth Filter)

The governing differential equation is $y'''(t) + 2y''(t) + 2y'(t) + y(t) = \frac{1}{2}x(t)$.

The characteristic equation is $s^3 + 2s^2 + 2s + 1 = 0$. Its roots are $s_{1,2} = -\frac{1}{2} \pm j\frac{\sqrt{3}}{2}$ and $s_3 = -1$.

The natural response is thus $y_N(t) = e^{-t/2}[K_1 \cos(\sqrt{3}t/2) + K_2 \sin(\sqrt{3}t/2)] + K_3 e^{-t}$.

For the step response, $x(t) = 1,\ t \geq 0$, and we assume $y_F(t) = B$.

Then $y_F'''(t) + 2y_F''(t) + 2y_F'(t) + y_F(t) = 0 + 0 + B = \frac{1}{2}$.

The total response is $y(t) = \frac{1}{2} + e^{-3t/2}[K_1 \cos(\sqrt{3}t/2) + K_2 \sin(\sqrt{3}t/2)] + K_3 e^{-t}$.

Using zero initial conditions, we get

$y(0) = 0 = \frac{1}{2} + K_1 + K_3 \qquad y'(0) = 0 = -\frac{1}{2}K_1 + \frac{\sqrt{3}}{2}K_2 - K_3 \qquad y''(0) = 0 = -\frac{1}{2}K_1 - \frac{\sqrt{3}}{2}K_2 + K_3$

Thus, $K_1 = 0,\ K_2 = -\frac{1}{\sqrt{3}},\ K_3 = \frac{1}{2}$, and $y(t) = \frac{1}{2}u(t) - \frac{1}{\sqrt{3}}e^{-t/2}\sin(\sqrt{3}t/2)u(t) - \frac{1}{2}e^{-t}u(t)$.

The impulse response $h(t)$ equals the derivative of the step response.

Thus, $h(t) = e^{-t/2}[\frac{1}{2\sqrt{3}}\sin(\sqrt{3}t/2) - \frac{1}{2}\cos(\sqrt{3}t/2)]u(t) - \frac{1}{2}e^{-t}u(t)$.

PGR2026-00033
Krisp Technologies EX1024 Page 102

CHAPTER 4 PROBLEMS

DRILL AND REINFORCEMENT

4.1 (Operators) Which of the following describe linear operators?

(a) $\mathcal{O}\{\ \} = 4\{\ \}$ **(b)** $\mathcal{O}\{\ \} = 4\{\ \} + 3$ **(c)** $y(t) = \int_{-\infty}^{t} x(t)\,dt$

(d) $\mathcal{O}\{\ \} = \sin\{\ \}$ **(e)** $y(t) = x(4t)$ **(f)** $y(t) = 4\dfrac{d\,x(t)}{dt} + 3x(t)$

4.2 (System Classification) In each of the following systems, $x(t)$ is the input and $y(t)$ is the output. Classify each system in terms of linearity, time invariance, memory, and causality.

(a) $y''(t) + 3y'(t) = 2x'(t) + x(t)$ **(b)** $y''(t) + 3y(t)y'(t) = 2x'(t) + x(t)$
(c) $y''(t) + 3tx(t)y'(t) = 2x'(t)$ **(d)** $y''(t) + 3y'(t) = 2x^2(t) + x(t+2)$
(e) $y(t) + 3 = x^2(t) + 2x(t)$ **(f)** $y(t) = 2x(t+1) + 5$
(g) $y''(t) + e^{-t}y'(t) = |x'(t-1)|$ **(h)** $y(t) = x^2(t) + 2x(t+1)$
(i) $y''(t) + \cos(2t)y'(t) = x'(t+1)$ **(j)** $y(t) + t\int_{-\infty}^{t} y(t)\,dt = 2x(t)$
(k) $y'(t) + \int_0^t y(t)\,dt = |x'(t)| - x(t)$ **(l)** $y''(t) + t\int_0^{t+1} y(t)\,dt = x'(t) + 2$

4.3 (Classification) Classify the following systems in terms of their linearity, time invariance, causality, and memory.

(a) The modulation system $y(t) = x(t)\cos(2\pi f_0 t)$.

(b) The modulation system $y(t) = [A + x(t)]\cos(2\pi f_0 t)$.

(c) The modulation system $y(t) = \cos[2\pi f_0 t x(t)]$.

(d) The modulation system $y(t) = \cos[2\pi f_0 t + x(t)]$.

(e) The sampling system $y(t) = x(t)\sum_{k=-\infty}^{\infty} \delta(t - kt_s)$.

4.4 (Forced Response) Evaluate the forced response of the following systems.

(a) $y'(t) + 2y(t) = u(t)$ **(b)** $y'(t) + 2y(t) = \cos(t)u(t)$
(c) $y'(t) + 2y(t) = e^{-t}u(t)$ **(d)** $y'(t) + 2y(t) = e^{-2t}u(t)$
(e) $y'(t) + 2y(t) = tu(t)$ **(f)** $y'(t) + 2y(t) = te^{-2t}u(t)$

4.5 (Forced Response) Evaluate the forced response of the following systems.

(a) $y''(t) + 5y'(t) + 6y(t) = 3u(t)$ **(b)** $y''(t) + 5y'(t) + 6y(t) = 6e^{-t}u(t)$
(c) $y''(t) + 5y'(t) + 6y(t) = 5\cos(t)u(t)$ **(d)** $y''(t) + 5y'(t) + 6y(t) = 2e^{-2t}u(t)$
(e) $y''(t) + 5y'(t) + 6y(t) = 2tu(t)$ **(f)** $y''(t) + 5y'(t) + 6y(t) = (6e^{-t} + 2e^{-2t})u(t)$

4.6 (Steady-State Response) The forced response of a system to sinusoidal inputs is termed the *steady-state response*. Evaluate the steady-state response of the following systems.

(a) $y'(t) + 5y(t) = 2u(t)$ **(b)** $y'(t) + y(t) = \cos(t)u(t)$
(c) $y'(t) + 3y(t) = \sin(t)u(t)$ **(d)** $y'(t) + 4y(t) = \cos(t) + \sin(2t)$
(e) $y''(t) + 5y'(t) + 6y(t) = \cos(3t)u(t)$ **(f)** $y''(t) + 4y'(t) + 4y(t) = \cos(2t)u(t)$

PGR2026-00033
Krisp Technologies EX1024 Page 103

4.7 (Zero-State Response) Evaluate the zero-state response of the following systems.

(a) $y'(t) + 2y(t) = u(t)$ **(b)** $y'(t) + y(t) = \cos(t)u(t)$
(c) $y'(t) + y(t) = r(t)$ **(d)** $y'(t) + 3y(t) = e^{-t}u(t)$
(e) $y'(t) + 2y(t) = e^{-2t}u(t)$ **(f)** $y'(t) + 2y(t) = e^{-2t}\cos(t)u(t)$

4.8 (Zero-State Response) Evaluate the zero-state response of the following systems.

(a) $y''(t) + 5y'(t) + 6y(t) = 6u(t)$ **(b)** $y''(t) + 4y'(t) + 3y(t) = 2e^{-2t}u(t)$
(c) $y''(t) + 2y'(t) + 2y(t) = 2e^{-t}u(t)$ **(d)** $y''(t) + 4y'(t) + 5y(t) = \cos(t)u(t)$
(e) $y''(t) + 4y'(t) + 3y(t) = r(t)$ **(f)** $y''(t) + 5y'(t) + 4y(t) = (2e^{-t} + 2e^{-3t})u(t)$

4.9 (System Response) Evaluate the natural, forced, zero-state, zero-input, and total response of the following systems.

(a) $y'(t) + 5y(t) = u(t)$ $y(0) = 2$
(b) $y'(t) + 3y(t) = 2e^{-2t}u(t)$ $y(0) = 1$
(c) $y'(t) + 4y(t) = 8tu(t)$ $y(0) = 2$
(d) $y'(t) + 2y(t) = 2\cos(2t)u(t)$ $y(0) = 4$
(e) $y'(t) + 2y(t) = 2e^{-2t}u(t)$ $y(0) = 6$
(f) $y'(t) + 2y(t) = 2e^{-2t}\cos(t)u(t)$ $y(0) = 8$

4.10 (System Response) Evaluate the response $y(t)$ of the following systems.

(a) $y'(t) + y(t) = 2x'(t) + x(t)$ $x(t) = 4e^{-2t}u(t)$ $y(0) = 2$
(b) $y'(t) + 3y(t) = 3x''(t)$ $x(t) = 4e^{-2t}u(t)$ $y(0) = 0$
(c) $y'(t) + 4y(t) = x'(t) - x(t)$ $x(t) = 4u(t)$ $y(0) = 6$
(d) $y'(t) + 2y(t) = x(t) + 2x(t-1)$ $x(t) = 4u(t)$ $y(0) = 0$
(e) $y'(t) + 2y(t) = x'(t) - 2x(t-1)$ $x(t) = 2e^{-t}u(t)$ $y(0) = 0$
(f) $y'(t) + 2y(t) = x''(t) - 2x'(t-1) + x(t-2)$ $x(t) = 2e^{-t}u(t)$ $y(0) = 4$

4.11 (System Response) For each of the following, evaluate the natural, forced, zero-state, zero-input, and total response. Assume $y'(0) = 1$ and all other initial conditions zero.

(a) $y''(t) + 5y'(t) + 6y(t) = 6u(t)$ $y(0) = 0$ $y'(0) = 1$
(b) $y''(t) + 5y'(t) + 6y(t) = 2e^{-t}u(t)$ $y(0) = 0$ $y'(0) = 1$
(c) $y''(t) + 4y'(t) + 3y(t) = 36tu(t)$ $y(0) = 0$ $y'(0) = 1$
(d) $y''(t) + 4y'(t) + 4y(t) = 2e^{-2t}u(t)$ $y(0) = 0$ $y'(0) = 1$
(e) $y''(t) + 4y'(t) + 4y(t) = 8\cos(2t)u(t)$ $y(0) = 0$ $y'(0) = 1$
(f) $[(s+1)^2(s+2)]y(t) = e^{-2t}u(t)$ $y(0) = 0$ $y'(0) = 1$ $y''(0) = 0$

4.12 (System Response) Evaluate the response $y(t)$ of the following systems.

(a) $y''(t) + 3y'(t) + 2y(t) = 2x'(t) + x(t)$ $x(t) = 4u(t)$ $y(0) = 2$ $y'(0) = 1$
(b) $y''(t) + 4y'(t) + 3y(t) = 3x''(t)$ $x(t) = 4e^{-2t}u(t)$ $y(0) = 0$ $y'(0) = 0$
(c) $y''(t) + 4y'(t) + 4y(t) = x'(t) - x(t)$ $x(t) = 4u(t)$ $y(0) = 6$ $y'(0) = -3$
(d) $y''(t) + 2y'(t) + 2y(t) = x(t) + 2x(t-1)$ $x(t) = 4u(t)$ $y(0) = 0$ $y'(0) = 0$
(e) $y''(t) + 5y'(t) + 6y(t) = x'(t) - 2x(t-1)$ $x(t) = 2e^{-t}u(t)$ $y(0) = 0$ $y'(0) = 0$
(f) $y''(t) + 5y'(t) + 4y(t) = x''(t) - 2x'(t-1)$ $x(t) = 3e^{-t}u(t)$ $y(0) = 4$ $y'(0) = -4$

4.13 (Impulse Response) Find the impulse response of the following systems.

(a) $y'(t) + 3y(t) = x(t)$ **(b)** $y'(t) + 4y(t) = 2x(t)$
(c) $y'(t) + 2y(t) = x'(t) - 2x(t)$ **(d)** $y'(t) + y(t) = x'(t) - x(t)$

PGR2026-00033
Krisp Technologies EX1024 Page 104

4.14 (Impulse Response) Find the impulse response of the following systems.

(a) $y''(t) + 5y'(t) + 4y(t) = x(t)$ **(b)** $y''(t) + 4y'(t) + 4y(t) = 2x(t)$
(c) $y''(t) + 4y'(t) + 3y(t) = 2x'(t) - x(t)$ **(d)** $y''(t) + 2y'(t) + y(t) = x''(t) + x'(t)$

4.15 (Stability) Which of the following systems are stable, and why?

(a) $y'(t) + 4y(t) = x(t)$ **(b)** $y'(t) - 4y(t) = 3x(t)$
(c) $y'(t) + 4y(t) = x'(t) + 3x(t)$ **(d)** $y''(t) + 5y'(t) + 4y(t) = 6x(t)$
(e) $y''(t) + 4y(t) = 2x'(t) - x(t)$ **(f)** $y''(t) + 5y'(t) + 6y(t) = x''(t)$
(g) $y''(t) - 5y'(t) + 4y(t) = x(t)$ **(h)** $y''(t) + 2y'(t) - 3y(t) = 2x'(t)$

4.16 (Impulse Response) The voltage input to a series RC circuit with a time constant τ is $\frac{1}{\alpha}e^{-t/\alpha}u(t)$.

(a) Find the analytical form for the capacitor voltage.
(b) Show that as $\alpha \to 0$, we obtain the impulse response $h(t)$ of the RC circuit.

4.17 (System Response) The step response of an LTI system is given by $s(t) = (1 - e^{-t})u(t)$.

(a) Establish its impulse response $h(t)$ and sketch both $s(t)$ and $h(t)$.
(b) Evaluate and sketch the response $y(t)$ to the input $x(t) = \text{rect}(t - 0.5)$.

REVIEW AND EXPLORATION

4.18 (System Classification) Explain why the system $y''(t) + y'(t) = t\sin(t)$ is time invariant, whereas the system $y''(t) + y'(t) = tx(t)$, where $x(t) = \sin(t)$, is time varying.

4.19 (System Classification) Investigate the linearity, time invariance, memory, causality, and stability of the following operations.

(a) $y(t) = y(0) + \int_0^t x(\lambda)\,d\lambda$ **(b)** $y(t) = \int_0^t x(\lambda)\,d\lambda,\ t > 0$
(c) $y(t) = \int_{t-1}^{t+1} x(\lambda)\,d\lambda$ **(d)** $y(t) = \int_t^{t+\alpha} x(\lambda + 2)\,d\lambda$
(e) $y(t) = \int_t^{t-\alpha} x(\lambda - 2)\,d\lambda$ **(f)** $y(t) = \int_{t-1}^{t+\alpha} x(\lambda + 1)\,d\lambda$

4.20 (Classification) Check the following for linearity, time invariance, memory, causality, and stability.

(a) The time-scaling system $y(t) = x(2t)$
(b) The folding system $y(t) = x(-t)$
(c) The time-scaling system $y(t) = x(0.5t)$
(d) The sign-inversion system $y(t) = \text{sgn}[x(t)]$
(e) The rectifying system $y(t) = |x(t)|$

4.21 (Classification) Consider the two systems (1) $y(t) = x(\alpha t)$ and (2) $y(t) = x(t + \alpha)$.

(a) For what values of α is each system linear?
(b) For what values of α is each system causal?
(c) For what values of α is each system time invariant?
(d) For what values of α is each system instantaneous?

PGR2026-00033
Krisp Technologies EX1024 Page 105

4.22 (System Response) Consider the *relaxed* system $y'(t) + y(t) = x(t)$.

(a) The input is $x(t) = u(t)$. What is the response?

(b) Use the result of part (a) (and superposition) to find the response of this system to the input $x_1(t)$ shown in Figure P4.22.

(c) The input is $x(t) = tu(t)$. What is the response?

(d) Use the result of part (c) (and superposition) to find the response of this system to the input $x_2(t)$ shown in Figure P4.22.

(e) How are the results of parts (a) and (b) related to the results of parts (c) and (d)?



Figure P4.22 Input signals for Problem 4.22

4.23 (System Response) Consider the *relaxed* system $y'(t) + \frac{1}{\tau}y(t) = x(t)$.

(a) What is the response of this system to the unit step $x(t) = u(t)$?

(b) What is the response of this system to the unit impulse $x(t) = \delta(t)$?

(c) What is the response of this system to the rectangular pulse $x(t) = u(t) - u(t - \alpha)$? Under what conditions for τ and α will the response resemble (be a good approximation to) the input?

4.24 (System Response) It is known that the response of the system $y'(t) + \alpha y(t) = x(t),\ \alpha \neq 0$, is given by $y(t) = (5 + 3e^{-2t})u(t)$.

(a) Identify the natural and forced response.

(b) Identify the values of α and $y(0)$.

(c) Identify the zero-input and zero-state response.

(d) Identify the input $x(t)$.

4.25 (System Response) It is known that the response of the system $y'(t) + y(t) = x(t)$ is given by $y(t) = (5e^{-t} + 3e^{-2t})u(t)$.

(a) Identify the zero-input and zero-state response.

(b) What is the zero-input response of the system $y'(t) + y(t) = x(t)$ if $y(0) = 10$?

(c) What is the response of the relaxed system $y'(t) + y(t) = x(t - 2)$?

(d) What is the response of the relaxed system $y'(t) + y(t) = x'(t) + 2x(t)$?

4.26 (System Response) It is known that the response of the system $y'(t) + \alpha y(t) = x(t)$ is given by $y(t) = (5 + 2t)e^{-3t}u(t)$.

(a) Identify the zero-input and zero-state response.

(b) What is the zero-input response of the system $y'(t) + \alpha y(t) = x(t)$ if $y(0) = 10$?

(c) What is the response of the relaxed system $y'(t) + \alpha y(t) = x(t - 2)$?

(d) What is the response of the relaxed system $y'(t) + \alpha y(t) = 2x(t) + x'(t)$?

(e) What is the complete response of the system $y'(t) + \alpha y(t) = x'(t) + 2x(t)$ if $y(0) = 4$?

PGR2026-00033
Krisp Technologies EX1024 Page 106

4.27 (Impulse Response) Consider the *relaxed* system $y'(t) + \frac{1}{\tau}y(t) = x(t)$.

(a) What is the impulse response of this system?

(b) What is the response of this system to the rectangular pulse $x(t) = \frac{1}{\alpha}[u(t) - u(t-\alpha)]$? Show that as $\alpha \to 0$, we obtain the system impulse response $h(t)$.

(c) What is the response of this system to the exponential input $x(t) = \frac{1}{\alpha}e^{-t/\alpha}u(t)$? Show that as $\alpha \to 0$, we obtain the system impulse response $h(t)$.

4.28 (System Response) Find the response of the following systems for $t \geq 0$.

(a) $y'(t) + 2y(t) = 2e^{-(t-1)}u(t-1)$ $\quad y(0) = 5$

(b) $y'(t) + 2y(t) = e^{-2t}u(t) + 2e^{-(t-1)}u(t-1)$ $\quad y(0) = 5$

(c) $y'(t) + 2y(t) = te^{-t} + 2e^{-(t-1)}u(t-1)$ $\quad y(0) = 5$

(d) $y'(t) + 2y(t) = \cos(2t) + 2e^{-(t-1)}u(t-1)$ $\quad y(0) = 5$

4.29 (Impulse Response) Find the step response and impulse response of each circuit in Figure P4.29.

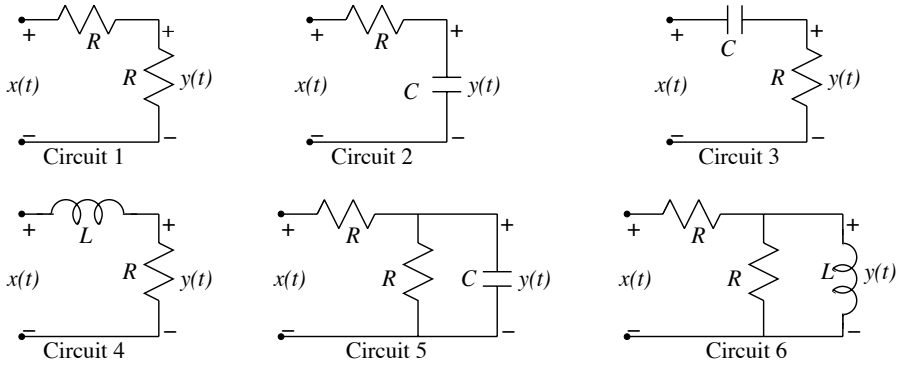


Figure P4.29 Circuits for Problem 4.29

4.30 (Impulse Response) The input-output relation for an LTI system is shown in Figure P4.30. What is the impulse response $h(t)$ of this system?

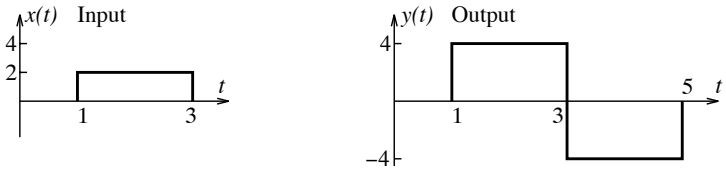


Figure P4.30 Figure for Problem 4.30

4.31 (System Response) Consider two *relaxed* RC circuits with $\tau_1 = 0.5\,\text{s}$ and $\tau_2 = 5\,\text{s}$. The input to both is the rectangular pulse $x(t) = 5\,\text{rect}(t - 0.5)$ V. The output is the capacitor voltage.

(a) Find and sketch the outputs $y_1(t)$ and $y_2(t)$ of the two circuits.

(b) At what time $t > 0$ does the output of both systems attain the same value?

4.32 (Classification and Stability) Argue for or against the following statements, assuming relaxed systems and constant element values. You may validate your arguments using simple circuits.

(a) A system with only resistors is always instantaneous and stable.

(b) A system with only inductors and/or capacitors is always stable.

(c) An RLC system with at least one resistor is always linear, causal, and stable.

PGR2026-00033
Krisp Technologies EX1024 Page 107

4.33 (Differential Equations from Impulse Response) Though there is an easy way of obtaining a system differential equation from its impulse response using transform methods, we can also obtain such a representation by working in the time domain itself. Let a system be described by $h(t) = e^{-t}u(t)$. If we compute $h'(t) = \delta(t) - e^{-t}u(t)$, we find that $h'(t) + h(t) = \delta(t)$, and the system differential equation follows as $y'(t) + y(t) = x(t)$. Using this idea, determine the system differential equation corresponding to each impulse response $h(t)$.

(a) $h(t) = e^{-\alpha t}u(t)$ **(b)** $h(t) = e^{-t}u(t) - e^{-2t}u(t)$

4.34 (Inverse Systems) If the input to a system is $x_0(t)$ and its response is $y_0(t)$, the inverse system is defined as a system that recovers $x_0(t)$ when its input is $y_0(t)$. Inverse systems are often used to undo the effects of measurement systems such as a transducers. The system equation of the inverse of many LTI systems can be found simply by switching the input and output. For example, if the system equation is $y(t) = x(t-3)$, the inverse system is $x(t) = y(t-3)$ (or $y(t) = x(t+3)$, by time invariance). Find the inverse of each system and determine whether the inverse system is stable.

(a) $y'(t) + 2y(t) = x(t)$ **(b)** $y''(t) + 2y'(t) + y(t) = x'(t) + 2x(t)$

4.35 (Inverse Systems) A requirement for a system to have an inverse is that unique inputs must produce unique outputs. Thus, the system $y(t) = |x(t)|$ does not have an inverse because of the sign ambiguity. Determine which of the following systems are invertible and, for those that are, find the inverse system.

(a) $y(t) = x^2(t)$ **(b)** $y(t) = e^{x(t)}$ **(c)** $y(t) = \cos[x(t)]$
(d) $y(t) = e^{jx(t)}$ **(e)** $y(t) = x(t-2)$ **(f)** $y'(t) + y(t) = x(t)$

COMPUTATION AND DESIGN

ctsimgui A GUI for Numerical Simulation of System Response

The graphical user interface `ctsimgui` allows you to simulate and visualize the response of analog systems, using various numerical integration methods.
You can select the input signal, the system parameters, and the simulation method.
You can display the input and simulated response for various choices of the time step.
To explore this routine, type `ctsimgui` at the MATLAB prompt.

4.36 (System Response in Symbolic Form) The ADSP routine `sysresp1` yields a symbolic result for the system response (see Chapter 21 for examples of its usage). Consider the system $y'(t) + 2y(t) = 2x(t)$. Use `sysresp1` to obtain its

(a) Step response.
(b) Impulse response.
(c) Zero-state response to $x(t) = 4e^{-3t}u(t)$.
(d) Complete response to $x(t) = 4e^{-3t}u(t)$ with $y(0) = 5$.

4.37 (System Response) Use the ADSP routine `sysresp1` to find the step response and impulse response of the following filters and plot each result over $0 \le t \le 4$. Compare the features of the step response of each filter. Compare the features of the impulse response of each filter.

(a) $y'(t) + y(t) = x(t)$ (a first-order lowpass filter)
(b) $y''(t) + \sqrt{2}y'(t) + y(t) = x(t)$ (a second-order Butterworth lowpass filter)

PGR2026-00033
Krisp Technologies EX1024 Page 108

(c) $y''(t) + y'(t) + y(t) = x'(t)$ (a bandpass filter)
(d) $y'''(t) + 2y''(t) + 2y'(t) + y(t) = x(t)$ (a third-order Butterworth lowpass filter)

4.38 (Rise Time and Settling Time) For systems whose step response rises to a nonzero final value, the rise time is commonly defined as the time it takes to rise from 10% to 90% of the final value. The settling time is another measure for such signals. The 5% settling time, for example, is defined as the time it takes for a signal to settle to within 5% of its final value. For each system, use the ADSP routine `sysresp1` to find the impulse response and step response and plot the results over $0 \le t \le 4$. For those systems whose step response rises toward a nonzero final value, use the ADSP routine `trbw` to numerically estimate the rise time and the 5% settling time.

(a) $y'(t) + y(t) = x(t)$ (a first-order lowpass filter)
(b) $y''(t) + \sqrt{2}y'(t) + y(t) = x(t)$ (a second-order Butterworth lowpass filter)
(c) $y''(t) + y'(t) + y(t) = x'(t)$ (a bandpass filter)
(d) $y'''(t) + 2y''(t) + 2y'(t) + y(t) = x(t)$ (a third-order Butterworth lowpass filter)

4.39 (System Response) Consider the system $y''(t) + 4y(t) + Cy(t) = x(t)$.

(a) Use `sysresp1` to obtain its step response and impulse response for $C = 3,\ 4,\ 5$ and plot each response over an appropriate time interval.
(b) How does the step response differ for each value of C? For what value of C would you expect the smallest rise time? For what value of C would you expect the smallest 3% settling time?
(c) Confirm your predictions in the previous part by numerically estimating the rise time and settling time, using the ADSP routine `trbw`.

4.40 (Steady-State Response in Symbolic Form) The ADSP routine `ssresp` yields a symbolic expression for the steady-state response to sinusoidal inputs (see Chapter 21 for examples of its usage). Find the steady-state response to the input $x(t) = 2\cos(3t - \frac{\pi}{3})$ for each of the following systems and plot the results over $0 \le t \le 3$, using a time step of 0.01 s.

(a) $y'(t) + \alpha y(t) = 2x(t)$, for $\alpha = 1,\ 2$
(b) $y''(t) + 4y(t) + Cy(t) = x(t)$, for $C = 3,\ 4,\ 5$

4.41 (Numerical Simulation of Analog Systems) The ADSP routine `ctsim` returns estimates of the system response using numerical integration such as Simpson's rule and Runge-Kutta methods. Consider the differential equation $y'(t) + \alpha y(t) = \alpha x(t)$. In the following, use the second-order Runge-Kutta method throughout.

(a) Let $x(t) = \text{rect}(t - 0.5)$ and $\alpha = 1$. Evaluate its response $y(t)$ analytically. Use `ctsim` to evaluate its response $y_1(t)$ over $0 \le t \le 3$, using a time step of 0.1 s. Plot both results on the same graph and compare.
(b) Let $x(t) = \sin(t),\ 0 \le t \le \pi$. Use `ctsim` to evaluate its response $y_1(t)$ over $0 \le t \le 6$, using $\alpha = 1,\ 3,\ 10$ and a time step of 0.02 s. Plot each response along with the input $x(t)$ on the same graph. Does the response begin to resemble the input as α is increased? Should it? Explain.
(c) Let $x(t) = \sin(t),\ 0 \le t \le \pi$. Use `ctsim` to evaluate its response $y_1(t)$ over $0 \le t \le 6$, using $\alpha = 100$ and a time step of 0.02 s. Plot the response along with the input $x(t)$ on the same graph. Now change the time step to 0.03 s. What is the response? To explain what is happening, find and plot the response for time steps of 0.0201 s, 0.0202 s, and 0.0203 s. Describe what happens to the computed response and why.

PGR2026-00033
Krisp Technologies EX1024 Page 109

Chapter 5

DISCRETE-TIME SYSTEMS

5.0 Scope and Objectives

The processing of discrete-time signals is accomplished by **discrete-time** (DT) systems or **digital filters**. Their description relies heavily on how they respond to arbitrary or specific signals. Many discrete-time systems may be described by difference equations relating the input and output. The class of *linear, time-invariant* systems may also be described by their impulse response, the response to an impulse input. This chapter deals with discrete-time systems and their classification, time-domain representation, and analysis based on the solution of difference equations. It also introduces the all-important concept of the impulse response, which forms a key ingredient in both system description and system analysis.

5.1 Discrete-Time Operators

In the time domain, many discrete-time systems can be modeled by **difference equations**. For example, the system described by $y[n] = y[n-1] + x[n]$ produces the present output $y[n]$ as the sum of the previous output $y[n-1]$ and the present input $x[n]$. Difference equations involve signals and their shifted versions, which can be conveniently expressed using operator notation. A *unit delay* is represented by the **backward shift operator**, denoted by z^{-1} and transforms $x[n]$ to $x[n-1]$. We express this transformation in operator notation as $z^{-1}\{x[n]\} = x[n-1]$. A delay of 2 units is denoted by z^{-2}. The **forward shift operator** z^k describes an *advance* of k units and transforms $x[n]$ to $x[n+k]$; we express this transformation in operator notation as $z^k\{x[n]\} = x[n+k]$. An operator allows us to transform one function to another. If an operator is represented by the symbol $\mathcal{O}$, the equation

$$\mathcal{O}\{x[n]\} = y[n] \tag{5.1}$$

implies that if the function $x[n]$ is treated exactly as the operator $\mathcal{O}$ tells us, we obtain the function $y[n]$. For example, the operation $\mathcal{O}\{\ \} = 4z^3\{\ \} + 6$ says that to get $y[n]$, we must shift $x[n]$ 3 units, multiply by 4, and then add 6 to the result to obtain $4z^3\{x[n]\} + 6 = 4x[n+3] + 6 = y[n]$

5.1.1 Linear Operators and Superposition

If an operation on the sum of two functions is equivalent to the sum of operations applied to each separately, the operator is said to be **additive**. In other words,

$$\mathcal{O}\{x_1[n] + x_2[n]\} = \mathcal{O}\{x_1[n]\} + \mathcal{O}\{x_2[n]\} \qquad \text{(for an additive operation)} \tag{5.2}$$

If an operation on $Kx[n]$ is equivalent to K times the linear operation on $x[n]$ where K is a scalar, the operator is said to be **homogeneous**. In other words,

$$\mathcal{O}\{Kx[n]\} = K\mathcal{O}\{x[n]\} \qquad \text{(for a homogeneous operation)} \tag{5.3}$$

PGR2026-00033
Krisp Technologies EX1024 Page 110

Together, the two describe the **principle of superposition**. An operator $\mathcal{O}$ is termed a **linear operator** if it is both additive and homogeneous:

$$\mathcal{O}\{Ax_1[n] + Bx_2[n]\} = A\mathcal{O}\{x_1[n]\} + B\mathcal{O}\{x_2[n]\} \qquad \text{(for a linear operation)} \tag{5.4}$$

Otherwise, it is nonlinear. In many instances, it suffices to test only for homogeneity (or additivity) to confirm the linearity of an operation (even though one does not imply the other). An important concept that forms the basis for the study of linear systems is that *the superposition of linear operators is also linear.*

REVIEW PANEL 5.1

A Linear Operator Obeys Superposition: $a\mathcal{O}\{x_1[n]\} + b\mathcal{O}\{x_2[n]\} = \mathcal{O}\{ax_1[n] + bx_2[n]\}$

Superposition implies both **homogeneity** and **additivity**.

Homogeneity: $\mathcal{O}\{ax[n]\} = a\mathcal{O}\{x[n]\}$ **Additivity:** $\mathcal{O}\{x_1[n]\} + \mathcal{O}\{x_2[n]\} = \mathcal{O}\{x_1[n] + x_2[n]\}$

EXAMPLE 5.1 (Testing for Linear Operators)

(a) Consider the operator $\mathcal{O}\{\ \} = C\{\ \} + D$.

Applying the homogeneity test, $\mathcal{O}\{Ax[n]\} = C(Ax[n]) + D$, but $A\mathcal{O}\{x[n]\} = A(Cx[n] + D)$.

The two differ, so the operation is nonlinear (it is linear only if $D = 0$).

(b) Consider the squaring operator $\mathcal{O}\{\ \} = \{\ \}^2$, which transforms $x[n]$ to $x^2[n]$.

We find that $A\mathcal{O}\{x[n]\} = Ax^2[n]$, but $\mathcal{O}\{Ax[n]\} = (Ax[n])^2 = A^2x^2[n]$.

The two are not equal, and the squaring operator is nonlinear.

5.2 System Classification

Discrete-time systems modeled by difference equations relate the output $y[n]$ to the input $x[n]$. The general form of an Nth-order difference equation may be written as

$$y[n] + A_1y[n-1] + \cdots + A_Ny[n-N] = B_0x[n] + B_1x[n-1] + \cdots + B_Mx[n-M] \tag{5.5}$$

The **order** N describes the output term with the largest delay. It is customary to normalize the leading coefficient to unity.

5.2.1 Linearity

A linear system is one for which superposition applies and implies that the system is relaxed (with zero initial conditions) and the system equation involves only linear operators. However, we can use superposition even for a system with nonzero initial conditions that is otherwise linear. We treat it as a multiple-input system by including the initial conditions as additional inputs. The output then equals the superposition of the outputs due to each input acting alone, and any *changes* in the input are related linearly to *changes* in the response. As a result, its response can be written as the sum of a *zero-input response* (due to the initial conditions alone) and the *zero-state response* (due to the input alone). This is the **principle of decomposition**, which allows us to analyze linear systems in the presence of nonzero initial conditions. Both the zero-input response and the zero-state response obey superposition.

PGR2026-00033
Krisp Technologies EX1024 Page 111

5.2.2 Time Invariance

Time invariance implies that the shape of the response $y[n]$ depends only on the shape of the input $x[n]$ and not on the time when it is applied. If the input is shifted to $x[n-n_0]$, the response equals $y[n-n_0]$ and is shifted by the same amount. In other words, the system does not change with time. Formally, if the operator $\mathcal{O}$ transforms the input $x[n]$ to the output $y[n]$ such that $\mathcal{O}\{x[n]\} = y[n]$, time invariance requires that

$$\mathcal{O}\{x[n-n_0]\} = y[n-n_0] \qquad \text{(for time invariance)} \tag{5.6}$$

REVIEW PANEL 5.2

Time Invariance from the Operational Relation

If $\mathcal{O}\{x[n]\} = y[n]$, then $\mathcal{O}\{x[n-n_0)\} = y[n-n_0]$ (shift input by α $\Rightarrow$ shift output by α).

EXAMPLE 5.2 (Linearity and Time Invariance of Operators)

(a) $y[n] = x[n]x[n-1]$ is nonlinear but time invariant.

The operation is $\mathcal{O}\{\ \} = (\{\ \})(z^{-1}\{\ \})$. We find that

$A\mathcal{O}\{x[n]\} = A(x[n]x[n-1])$, but $\mathcal{O}\{Ax[n]\} = (Ax[n])(Ax[n-1])$. The two are not equal.

$\mathcal{O}\{x[n-n_0]\} = x[n-n_0]x[n-n_0-1]$, and $y[n-n_0] = x[n-n_0]x[n-n_0-1]$. The two are equal.

(b) $y[n] = nx[n]$ is linear but time varying.

The operation is $\mathcal{O}\{\ \} = n\{\ \}$. We find that

$A\mathcal{O}\{x[n]\} = A(nx[n])$, and $\mathcal{O}\{Ax[n]\} = n(Ax[n])$. The two are equal.

$\mathcal{O}\{x[n-n_0]\} = n(x[n-n_0])$, but $y[n-n_0] = (n-n_0)x[n-n_0]$. The two are not equal.

(c) $y[n] = x[2n]$ is linear but time varying. The operation $n \Rightarrow 2n$ reveals that

$A\mathcal{O}\{x[n]\} = A(x[2n])$, and $\mathcal{O}\{Ax[n]\} = (Ax[2n])$. The two are equal.

$\mathcal{O}\{x[n-n_0]\} = x[2n-n_0]$, but $y[n-n_0] = x[2(n-n_0)]$. The two are not equal.

(d) $y[n] = x[n-2]$ is linear and time invariant. The operation $n \Rightarrow n-2$ reveals that

$A\mathcal{O}\{x[n]\} = A(x[n-2])$, and $\mathcal{O}\{Ax[n]\} = (Ax[n-2])$. The two are equal.

$\mathcal{O}\{x[n-n_0]\} = x[n-n_0-2]$, and $y[n-n_0] = x[n-n_0-2]$. The two are equal.

(e) $y[n] = 2^{x[n]}x[n]$ is nonlinear but time invariant.

The operation is $\mathcal{O}\{\ \} = (2)^{\{\ \}}\{\ \}$ and reveals that

$A\mathcal{O}\{x[n]\} = A(2)^{x[n]}x[n]$, but $\mathcal{O}\{Ax[n]\} = (2)^{Ax[n]}(Ax[n])$. The two are not equal.

$\mathcal{O}\{x[n-n_0]\} = (2)^{x[n-n_0]}x[n-n_0]$, and $y[n-n_0] = (2)^{x[n-n_0]}x[n-n_0]$. The two are equal.

PGR2026-00033
Krisp Technologies EX1024 Page 112

5.2.3 Linear Time-Invariant Systems

The important class of **linear time-invariant** (LTI) **systems** are described by difference equations with *constant coefficients*. To test for linearity or time-invariance, we can formally apply the linearity or time-invariance tests to each operation in the difference equation or recognize nonlinear or time-varying operations by generalizing the results of the previous example as follows:

1. *Terms* containing products of the input and/or output make a system equation nonlinear. A constant *term* also makes a system equation nonlinear.

2. *Coefficients* of the input or output that are explicit functions of n make a system equation time varying. Time-scaled inputs or outputs such as $y[2n]$ also make a system equation time varying.

REVIEW PANEL 5.3

An LTI System Is Described by a Linear Constant Coefficient Difference Equation (LCCDE)

$$y[n] + A_1 y[n-1] + \cdots + A_N y[n-N] = B_0 x[n] + B_1 x[n-1] + \cdots + B_M x[n-M]$$

All *terms* contain $x[n]$ or $y[n]$. All *coefficients* are constant (not functions of $x[n]$ or $y[n]$ or n).

REVIEW PANEL 5.4

What Makes a System Difference Equation Nonlinear or Time Varying?

It is **nonlinear** if *any term* is a constant or a *nonlinear function* of $x[n]$ or $y[n]$.
It is **time varying** if the *coefficient* of any term in $x[n]$ or $y[n]$ is an *explicit* function of n.

EXAMPLE 5.3 (Linearity and Time Invariance of Systems)
We check the following systems for linearity and time invariance.

(a) $y[n] - 2y[n-1] = 4x[n]$. This is LTI.

(b) $y[n] - 2ny[n-1] = x[n]$. This is linear but time varying.

(c) $y[n] + 2y^2[n] = 2x[n] - x[n-1]$. This is nonlinear but time invariant.

(d) $y[n] - 2y[n-1] = (2)^{x[n]}x[n]$. This is nonlinear but time invariant.

(e) $y[n] - 4y[n]y[2n] = x[n]$. This is nonlinear and time varying.

Implications of Linearity and Time Invariance

In most practical cases, if an input $x[n]$ to a relaxed LTI system undergoes a linear operation, the output $y[n]$ undergoes the same linear operation. Often, an arbitrary function may be decomposed into its simpler constituents, the response due to each analyzed separately and more effectively, and the total response found using superposition. This approach forms the basis for several methods of system analysis. Representing an arbitrary signal $x[n]$ as a weighted sum of shifted DT impulses is the basis for the discrete convolution method (Chapter 7). Representing a signal $x[n]$ as a linear combination of DT harmonics or complex exponentials is the basis for the discrete-time Fourier transform (DTFT) and the z-transform (Chapters 15 and 17).

PGR2026-00033
Krisp Technologies EX1024 Page 113

5.2.4 Causality and Memory

In a **causal system**, the present response $y[n]$ cannot depend on future values of the input, such as $x[n+2]$. Systems whose present response requires future inputs are termed **noncausal**. We can check for causality by examining the **operational transfer function** $H(z)$ derived from the difference equation in operator form. Its general form may be expressed as

$$H(z) = \frac{B_0 z^M + B_1 z^{M-1} + \cdots + B_{M-1} z + B_M}{A_0 z^N + A_1 z^{N-1} + \cdots + A_{N-1} z + A_N} = \frac{P_M(z)}{Q_N(z)} \tag{5.7}$$

If the order of the numerator polynomial exceeds the order of the denominator polynomial (i.e., $M > N$), the system is *noncausal*.

REVIEW PANEL 5.5

What Makes a System Difference Equation Noncausal?

It is **noncausal** if the *least delayed* output term is $y[n]$ and an input term like $x[n+K],\ K > 0$ is present.
Examples: $y[n] - 2y[n-2] = x[n]$ is *causal*. $y[n] - 2y[n-2] = x[n+1]$ is *noncausal*.
Examples: $y[n+1] - y[n] = x[n+1]$ is *causal*. $y[n+1] - y[n] = x[n+2]$ is *noncausal*.

Instantaneous and Dynamic Systems

If the response of a system at $n = n_0$ depends only on the input at $n = n_0$ and not at any other times (past or future), the system is called **instantaneous** or **static**. The system equation of an instantaneous system is not only *algebraic* but also the input and output display *identical arguments* such as $x[n-1]$ and $y[n-1]$. The response of a dynamic system depends on past (and/or future) inputs. Dynamic systems are usually described by (but not limited to) difference equations.

REVIEW PANEL 5.6

What Makes a System Static or Dynamic?

It is **static** if the output $y[n_0]$ depends only on the *instantaneous* value of the input $x[n_0]$.
Examples: $y[n-2] = 3x[n-2]$ is static, but $y[n] = 3x[n-2]$ is dynamic.
Most **dynamic** systems are described by *difference* equations.

EXAMPLE 5.4 (Causal and Dynamic Systems)

(a) $y[n] + A_1 y[n-1] + A_2 y[n-2] + \cdots + A_N y[n-N] = x[n+K]$ is causal for $K \le 0$. It is also dynamic (because it describes a difference equation).

(b) $y[n+N] + A_1 y[n+N-1] + A_2 y[n+N-2] + \cdots + A_N y[n] = x[n+K]$ is causal for $K \le N$ and dynamic.

(c) $y[n] = x[n+2]$ is noncausal (to find $y[0]$, we need $x[2]$) and dynamic ($y[n_0]$ does not depend on $x[n_0]$ but on $x[n_0+2]$).

(d) $y[n+4] + y[n+3] = x[n+2]$ is causal and dynamic. It is identical to $y[n] + y[n-1] = x[n-2]$.

(e) $y[n] = 2x[\alpha n]$ is causal and instantaneous for $\alpha = 1$, causal and dynamic for $\alpha < 1$, and noncausal and dynamic for $\alpha > 1$. It is also time varying if $\alpha \neq 1$.

(f) $y[n] = 2(n+1)x[n]$ is causal and instantaneous (but time varying).

PGR2026-00033
Krisp Technologies EX1024 Page 114

5.3 Digital Filters

A digital filter may be described by the difference equation

$$y[n] + A_1 y[n-1] + \cdots + A_N y[n-N] = B_0 x[n] + B_1 x[n-1] + \cdots + B_M x[n-M] \tag{5.8}$$

This describes an Nth-order **recursive filter** whose present output depends on its own past values $y[n-k]$ and on the past and present values of the input. It is also called an **infinite impulse response** (IIR) **filter** because its impulse response $h[n]$ (the response to a unit impulse input) is usually of infinite duration. Now consider the difference equation described by

$$y[n] = B_0 x[n] + B_1 x[n-1] + \cdots + B_M x[n-M] \qquad \text{(FIR filter)} \tag{5.9}$$

Its present response depends only on the input terms and shows no dependence (recursion) on past values of the response. It is called a **nonrecursive filter**, or a **moving average filter**, because its response is just a weighted sum (moving average) of the input terms. It is also called a **finite impulse response** (FIR) **filter** (because its impulse response is of finite duration).

REVIEW PANEL 5.7

The Terminology of Digital Filters

Recursive (IIR): $y[n] + A_1 y[n-1] + \cdots + A_N y[n-N] = B_0 x[n] + B_1 x[n-1] + \cdots + B_M x[n-M]$

Nonrecursive (FIR): $y[n] = B_0 x[n] + B_1 x[n-1] + \cdots + B_M x[n-M]$

5.3.1 Digital Filter Realization

Digital filters described by difference equations can be *realized* by using elements corresponding to the operations of *scaling* (or multiplication), *shift* (or delay), and *summing* (or addition) that naturally occur in such equations. These elements describe the **gain** (scalar multiplier), **delay**, and **summer** (or **adder**), represented symbolically in Figure 5.1.

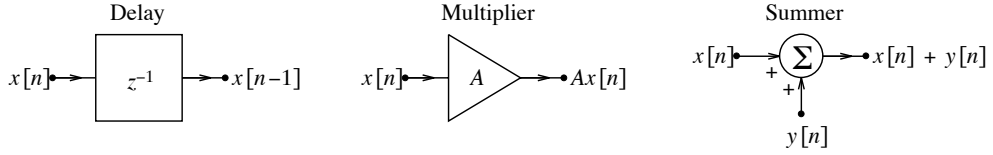


Figure 5.1 The building blocks for digital filter realization

Delay elements in cascade result in an output delayed by the sum of the individual delays. The operational notation for a delay of k units is z^{-k}. A nonrecursive filter described by

$$y[n] = B_0 x[n] + B_1 x[n-1] + \cdots + B_N x[n-N] \tag{5.10}$$

can be realized using a **feed-forward** structure with N delay elements, and a recursive filter of the form

$$y[n] = -A_1 y[n-1] - \cdots - A_N y[n-N] + x[n] \tag{5.11}$$

requires a **feedback** structure (because the output depends on its own past values). Each realization is shown in Figure 5.2 and requires N delay elements. The general form described by

$$y[n] = -A_1 y[n-1] - \cdots - A_N y[n-N] + B_0 x[n] + B_1 x[n-1] + \cdots + B_N x[n-N] \tag{5.12}$$

PGR2026-00033
Krisp Technologies EX1024 Page 115

requires both feed-forward and feedback and $2N$ delay elements, as shown in Figure 5.3. However, since LTI systems may be cascaded in any order (as we shall learn in the next chapter), we can switch the two subsystems to obtain a **canonical realization** with only N delays, as also shown in Figure 5.3.

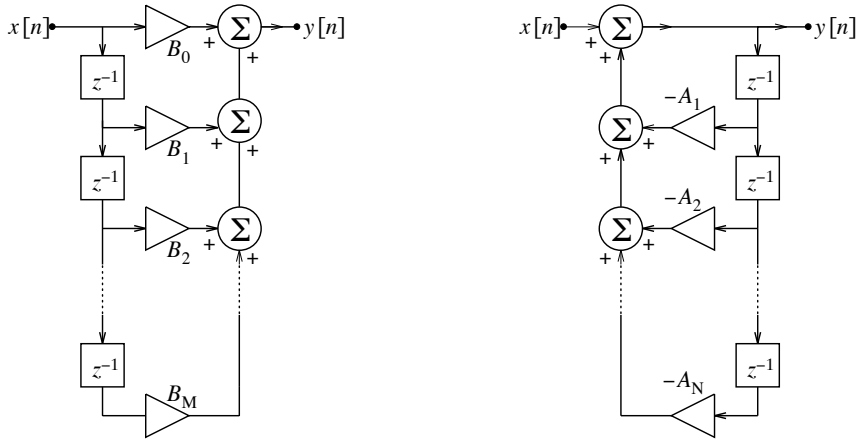


Figure 5.2 Realization of a nonrecursive (left) and recursive (right) digital filter

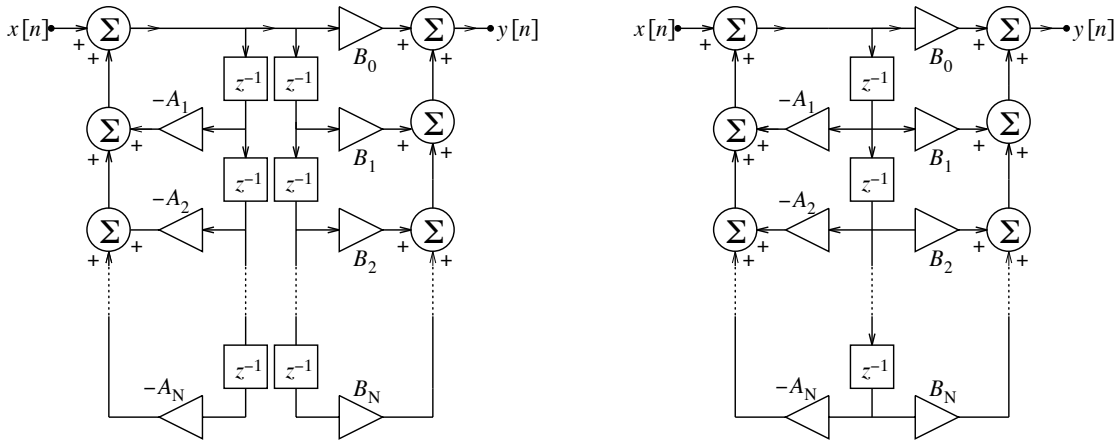


Figure 5.3 Direct (left) and canonical (right) realization of a digital filter

5.3.2 Analysis of Digital Filters

Digital filters or DT systems can be analyzed in the time domain using any of the following models:

The difference equation representation applies to linear, nonlinear, and time-varying systems. For LTI systems, it allows computation of the response using superposition even if initial conditions are present.

The impulse response representation describes a *relaxed* LTI system by its impulse response $h[n]$. The output $y[n]$ appears explicitly in the governing relation called the *convolution sum*. It also allows us to relate time-domain and transformed-domain methods of system analysis. We discuss the details of this method in Chapter 7.

PGR2026-00033
Krisp Technologies EX1024 Page 116

The state variable representation describes an nth-order system by n simultaneous first-order difference equations called **state equations** in terms of n state variables. It is useful for complex or nonlinear systems and those with multiple inputs and outputs. For LTI systems, state equations can be solved using matrix methods. The state variable form is also readily amenable to numerical solution. We do not pursue this method in this book.

5.4 Digital Filters Described by Difference Equations

We now examine how to find the response $y[n]$ of a digital filter excited by an input $x[n]$. For nonrecursive filters, whose output depends only on the input, the response is simply the weighted sum of the input terms exactly as described by its system equation.

For recursive systems, whose difference equation incorporates dependence on past values of the input, we can use recursion and successively compute the values of the output as far as desired. This approach is simple, but it is not always easy to discern a closed-form solution for the output. We can also use an analytical approach that exactly parallels the solution of differential equations. Typically, we specify $y[-1]$) for a first-order system, $y[-1]$ and $y[-2]$ for a second-order system, and $y[-1], y[-2], \ldots, y[-N]$ for an Nth-order system. In other words, an Nth-order difference equation requires N consecutive initial conditions for its complete solution. The following example illustrates this approach.

5.4.1 Solution by Recursion

The solution of a difference equation by recursion is straightforward. Given an Nth-order difference equation subject to the initial conditions $y[-1], y[-2], \ldots, y[-N]$, we successively generate values of $y[0], y[1], \ldots$, as far as desired.

EXAMPLE 5.5 (System Response Using Recursion)

(a) Consider a system described by $y[n] = a_1 y[n-1] + b_0 u[n]$. Let the initial condition be $y[-1] = 0$. We then successively compute

$$\begin{aligned} y[0] &= a_1 y[-1] + b_0 u[0] = b_0 \\ y[1] &= a_1 y[0] + b_0 u[1] = a_1 b_0 + b_0 = b_0[1 + a_1] \\ y[2] &= a_1 y[1] + b_0 u[2] = a_1[a_1 b_0 + b_0] + b_0 = b_0[1 + a_1 + a_1^2] \end{aligned}$$

The form of $y[n]$ may be discerned as

$$y[n] = b_0[1 + a_1 + a_1^2 + \cdots + a_1^{n-1} + a_1^n]$$

Using the closed form for the geometric sequence results in

$$y[n] = \frac{b_0(1 - a_1^{n+1})}{1 - a_1}$$

(b) Consider a system described by $y[n] = a_1 y[n-1] + b_0 n u[n]$. Let the initial condition be $y[-1] = 0$. We then successively compute

$$\begin{aligned} y[0] &= a_1 y[-1] = 0 \\ y[1] &= a_1 y[0] + b_0 u[1] = b_0 \\ y[2] &= a_1 y[1] + 2b_0 u[2] = a_1 b_0 + 2b_0 \\ y[3] &= a_1 y[2] + 3b_0 u[3] = a_1[a_1 b_0 + 2b_0] + 3b_0 = a_1^2 + 2a_1 b_0 + 3b_0 \end{aligned}$$

PGR2026-00033
Krisp Technologies EX1024 Page 117

The general form is thus $y[n] = a_1^{n-1} + 2b_0a_1^{n-2} + 3b_0a_1^{n-3} + (n-1)b_0a_1 + nb_0$.

We can find a more compact form for this, but not without some effort. By adding and subtracting $b_0a_1^{n-1}$ and factoring out a_1^n, we obtain

$$y[n] = a_1^n - b_0a_1^{n-1} + b_0a_1^n[a_1^{-1} + 2a_1^{-2} + 3a_1^{-3} + \cdots + na_1^{-n}]$$

Using the closed form for the sum Σkx^k from $k = 1$ to $k = N$ (with $x = a^{-1}$), we get

$$y[n] = a_1^n - b_0a_1^{n-1} + b_0a_1^n\frac{a^{-1}[1-(n+1)a^{-n} + na^{-(n+1)}]}{(1-a^{-1})^2}$$

What a chore! More elegant ways of solving difference equations are described later in this chapter.

(c) Consider the recursive system $y[n] = y[n-1] + x[n] - x[n-3]$. If $x[n]$ equals $\delta[n]$ and $y[-1] = 0$, we successively obtain

$$\begin{aligned} y[0] &= y[-1] + \delta[0] - \delta[-3] = 1 & \qquad y[3] &= y[2] + \delta[3] - \delta[0] = 1 - 1 = 0 \\ y[1] &= y[0] + \delta[1] - \delta[-2] = 1 & \qquad y[4] &= y[3] + \delta[4] - \delta[1] = 0 \\ y[2] &= y[1] + \delta[2] - \delta[-1] = 1 & \qquad y[5] &= y[4] + \delta[5] - \delta[2] = 0 \end{aligned}$$

The impulse response of this "recursive" filter is zero after the first three values and has a finite length. It is actually a nonrecursive (FIR) filter in disguise!

5.4.2 Formal Solution of Difference Equations

In analogy with differential equations, we start with an Nth-order difference equation with a single input:

$$y[n] + A_1y[n-1] + A_2y[n-2] + \cdots + A_Ny[n-N] = x[n] \tag{5.13}$$

Its solution, subject to the given initial conditions $y[-1], y[-2], y[-3], \ldots, y[-N]$, may be expressed as a superposition of the **natural response** and the **forced response**. The natural response contains only exponentials whose exponents correspond to the roots (real or complex) of the *characteristic equation* defined by

$$1 + A_1z^{-1} + A_2z^{-2} + \cdots + A_Nz^{-N} = z^N + A_1z^{N-1} + \cdots + A_N = 0 \tag{5.14}$$

This equation has N roots, $z_1, z_2, \ldots, z_N$. The natural response is a linear combination of N discrete-time exponentials of the form

$$y_N[n] = K_1z_1^n + K_2z_2^n + \cdots + K_Nz_N^n \tag{5.15}$$

This form must be modified for multiple roots. Table 5.1 summarizes the preferred forms for multiple or complex roots. The N constants $K_1,\ K_2, \ldots, K_N$ are evaluated by using the specified initial conditions but *only after setting up the total response*.

The forced response satisfies the given difference equation and has the same form as the input. Table 5.2 summarizes these forms for various types of input signals. The constants in the forced response can be found uniquely and independently of the natural response or initial conditions by satisfying the given difference equation.

The **total response** is found by *first* adding the forced and natural response and then evaluating the undetermined constants (in the natural component), using the prescribed initial conditions.

For stable systems, the natural response is also called the **transient response** since it decays to zero with time. For a system with harmonic or switched harmonic inputs, the forced response is also a harmonic at the input frequency and is called the **steady-state response**.

PGR2026-00033
Krisp Technologies EX1024 Page 118

Table 5.1 Form of the Natural Response for Discrete LTI Systems

Entry	**Root of Characteristic Equation**	**Form of Natural Response**
1	Real and distinct: r	Kr^n
2	Complex conjugate: $re^{j\Omega}$	$r^n[K_1\cos(n\Omega) + K_2\sin(n\Omega)]$
3	Real, repeated: r^{p+1}	$r^n(K_0 + K_1 n + K_2 n^2 + \cdots + K_p n^p)$
4	Complex, repeated: $\left(re^{j\Omega}\right)^{p+1}$	$r^n\cos(n\Omega)(A_0 + A_1 n + A_2 n^2 + \cdots + A_p n^p)$ $+ r^n\sin(n\Omega)(B_0 + B_1 n + B_2 n^2 + \cdots + B_p n^p)$

Table 5.2 Form of the Forced Response for Discrete LTI Systems

Note: If the right-hand side (RHS) is α^n, where α is also a root of the characteristic equation repeated p times, the forced response form must be multiplied by n^p.		
Entry	**Forcing Function (RHS)**	**Form of Forced Response**
1	C_0 (constant)	C_1 (another constant)
2	α^n (see note above)	$C\alpha^n$
3	$\cos(n\Omega + \beta)$	$C_1\cos(n\Omega) + C_2\sin(n\Omega)$ or $C\cos(n\Omega + \phi)$
4	$\alpha^n\cos(n\Omega + \beta)$ (see note above)	$\alpha^n[C_1\cos(n\Omega) + C_2\sin(n\Omega)]$
5	n	$C_0 + C_1 n$
6	n^p	$C_0 + C_1 n + C_2 n^2 + \cdots + C_p n^p$
7	$n\alpha^n$ (see note above)	$\alpha^n(C_0 + C_1 n)$
8	$n^p\alpha^n$ (see note above)	$\alpha^n(C_0 + C_1 n + C_2 n^2 + \cdots + C_p n^p)$
9	$n\cos(n\Omega + \beta)$	$(C_1 + C_2 n)\cos(n\Omega) + (C_3 + C_4 n)\sin(n\Omega)$

REVIEW PANEL 5.8

Response of LTI Systems Described by Difference Equations

Total Response = Natural Response + Forced Response

The roots of the characteristic equation determine *only the form* of the natural response.
The input terms (RHS) of the difference equation *completely determine* the forced response.
Initial conditions *satisfy the total response* to yield the constants in the natural response.

PGR2026-00033
Krisp Technologies EX1024 Page 119

EXAMPLE 5.6 (Forced and Natural Response)

(a) Consider the system shown in Figure E5.6A.
Find its response if $x[n] = (0.4)^n, \quad n \geq 0$ and the initial condition is $y[-1] = 10$.

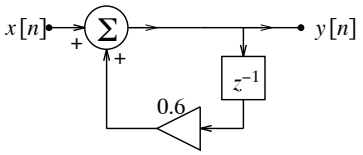


Figure E5.6A The system for Example 5.6(a)

The difference equation describing this system is $y[n] - 0.6y[n-1] = x[n] = (0.4)^n, \quad n \geq 0$.

Its characteristic equation is $1 - 0.6z^{-1} = 0$ or $z - 0.6 = 0$.

Its root $z = 0.6$ gives the form of the natural response $y_N[n] = K(0.6)^n$.

Since $x[n] = (0.4)^n$, the forced response is $y_F[n] = C(0.4)^n$.

We find C by substituting for $y_F[n]$ into the difference equation

$y_F[n] - 0.6y_F[n-1] = (0.4)^n = C(0.4)^n - 0.6C(0.4)^{n-1}$.

Cancel out $(0.4)^n$ from both sides and solve for C to get

$C - 1.5C = 1$ or $C = -2$.

Thus, $y_F[n] = -2(0.4)^n$. The total response is $y[n] = y_N[n] + y_F[n] = -2(0.4)^n + K(0.6)^n$.

We use the initial condition $y[-1] = 10$ on the *total response* to find K:

$y[-1] = 10 = -5 + \frac{K}{0.6}$ and $K = 9$.

Thus, $y[n] = -2(0.4)^n + 9(0.6)^n, \ n \geq 0$.

(b) Consider the difference equation $y[n] - 0.5y[n-1] = 5\cos(0.5n\pi), \quad n \geq 0$ with $y[-1] = 4$.

Its characteristic equation is $1 - 0.5z^{-1} = 0$ or $z - 0.5 = 0$.

Its root $z = 0.5$ gives the form of the natural response $y_N[n] = K(0.5)^n$.

Since $x[n] = 5\cos(0.5n\pi)$, the forced response is $y_F[n] = A\cos(0.5n\pi) + B\sin(0.5n\pi)$.

We find $y_F[n-1] = A\cos[0.5(n-1)\pi] + B\sin[0.5(n-1)\pi] = -A\sin(0.5n\pi) - B\cos(0.5n\pi)$. Then

$$y_F[n] - 0.5y_F[n-1] = (A + 0.5B)\cos(0.5n\pi) - (0.5A - B)\sin(0.5n\pi) = 5\cos(0.5n\pi)$$

Equate the coefficients of the cosine and sine terms to get

$(A + 0.5B) = 5, \quad (0.5A - B) = 0$ or $A = 4, \ B = 2$, and $y_F[n] = 4\cos(0.5n\pi) + 2\sin(0.5n\pi)$.

The total response is $y[n] = K(0.5)^n + 4\cos(0.5n\pi) + 2\sin(0.5n\pi)$. With $y[-1] = 4$, we find

$y[-1] = 4 = 2K - 2$ or $K = 3$, and thus $y[n] = 3(0.5)^n + 4\cos(0.5n\pi) + 2\sin(0.5n\pi), \ n \geq 0$.

The steady-state response is $4\cos(0.5n\pi) + 2\sin(0.5n\pi)$, and the transient response is $3(0.5)^n$.

PGR2026-00033
Krisp Technologies EX1024 Page 120

(c) Consider the difference equation $y[n] - 0.5y[n-1] = 3(0.5)^n, \quad n \geq 0$ with $y[-1] = 2$.
Its characteristic equation is $1 - 0.5z^{-1} = 0$ or $z - 0.5 = 0$.
Its root, $z = 0.5$, gives the form of the natural response $y_N[n] = K(0.5)^n$.
Since $x[n] = (0.5)^n$ has the same form as the natural response, the forced response is $y_F[n] = Cn(0.5)^n$.
We find C by substituting for $y_F[n]$ into the difference equation:
$y_F[n] - 0.5y_F[n-1] = 3(0.5)^n = Cn(0.5)^n - 0.5C(n-1)(0.5)^{n-1}$.
Cancel out $(0.5)^n$ from both sides and solve for C to get $Cn - C(n-1) = 3$, or $C = 3$.
Thus, $y_F[n] = 3n(0.5)^n$. The total response is $y[n] = y_N[n] + y_F[n] = K(0.5)^n + 3n(0.5)^n$.
We use the initial condition $y[-1] = 2$ on the *total response* to find K:
$y[-1] = 2 = 2K - 6$, and $K = 4$.
Thus, $y[n] = 4(0.5)^n + 3n(0.5)^n = (4 + 3n)(0.5)^n, \ n \geq 0$.

(d) (A Second-Order System) Consider the system shown in Figure E5.6D.
Find the forced and natural response of this system if $x[n] = u[n]$ and $y[-1] = 0, \ y[-2] = 12$.

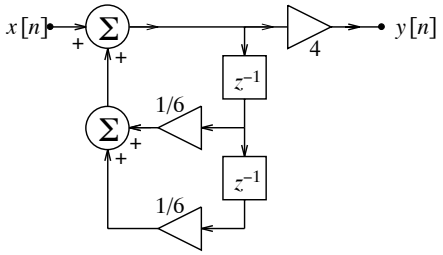


Figure E5.6D The system for Example 5.6(d)

Comparison with the generic realization of Figure 5.2 reveals that the system difference equation is

$$y[n] - \tfrac{1}{6}y[n-1] - \tfrac{1}{6}y[n-2] = 4x[n] = 4u[n]$$

Its characteristic equation is $1 - \frac{1}{6}z^{-1} - \frac{1}{6}z^{-2} = 0$ or $z^2 - \frac{1}{6}z - \frac{1}{6} = 0$.
Its roots are $z_1 = \frac{1}{2}$ and $z_2 = -\frac{1}{3}$.
The natural response is thus $y_N[n] = K_1(z_1)^n + K_2(z_2)^n = K_1(\frac{1}{2})^n + K_2(-\frac{1}{3})^n$.
Since the forcing function is $4u[n]$ (a constant for $n \geq 0$), the forced response $y_F[n]$ is constant.
Let $y_F[n] = C$. Then $y_F[n-1] = C, y_F[n-2] = C$, and
$y_F[n] - \frac{1}{6}y_F[n-1] - \frac{1}{6}y_F[n-2] = C - \frac{1}{6}C - \frac{1}{6}C = 4$. This yields $C = 6$.
Thus, $y_F[n] = 6$. The total response $y[n]$ is $y[n] = y_N[n] + y_F[n] = K_1(\frac{1}{2})^n + K_2(-\frac{1}{3})^n + 6$.
To find the constants K_1 and K_2, we use the initial conditions on the *total response* to obtain
$y[-1] = 0 = 2K_1 - 3K_2 + 6$, and $y[-2] = 12 = 4K_1 + 9K_2 + 6$. We find $K_1 = -1.2$ and $K_2 = 1.2$.
Thus, $y[n] = -1.2(\frac{1}{2})^n + 1.2(-\frac{1}{3})^n + 6, \ n \geq 0$.
Its transient response is $-1.2(\frac{1}{2})^n + 1.2(-\frac{1}{3})^n$. Its steady-state response is a constant that equals 6.

PGR2026-00033
Krisp Technologies EX1024 Page 121

5.4.3 The Zero-Input Response and Zero-State Response

It is often more convenient to describe the response $y(t)$ of an LTI system as the sum of its zero-state response (ZSR) $y_{zs}[n]$ (assuming zero initial conditions) and zero-input response (ZIR) $y_{zi}[n]$ (assuming zero input). Each component is found using the method of undetermined coefficients. Note that the natural and forced components $y_N[n]$ and $y_F[n]$ do not, in general, correspond to the zero-input and zero-state response, respectively, even though each pair adds up to the total response. Remember also that the zero-state response obeys superposition (as does the zero-input response).

REVIEW PANEL 5.9

The Zero-Input Response (ZIR) and Zero-State Response (ZSR) of LTI Systems

Total Response = ZIR (assume zero input, use given IC) + **ZSR** (assume zero IC).
Each obeys superposition. Each is found from its own *natural* and *forced* components.

REVIEW PANEL 5.10

Solving $y[n] + A_1 y[n-1] + \cdots + A_N y[n-N] = x[n]$

1. Find ZSR from $y_{zs}[n] + A_1 y_{zs}[n-1] + \cdots + A_N y_{zs}[n-N] = x[n]$, **with zero initial conditions**.
2. Find ZIR from $y_{zi}[n] + A_1 y_{zi}[n-1] + \cdots + A_N y_{zi}[n-N] = 0$, **using given initial conditions**.
3. Find complete response as $y[n] = y_{zs}[n] + y_{zi}[n]$.

EXAMPLE 5.7 (Zero-Input and Zero-State Response for the Single-Input Case)

(a) Consider the difference equation $y[n] - 0.6y[n-1] = (0.4)^n,\ n \geq 0$, with $y[-1] = 10$.

The forced response and the form of the natural response were found in Example 5.6(a) as:

$$y_F[n] = -2(0.4)^n \qquad y_N[n] = K(0.6)^n$$

1. Its ZSR is found from the form of the total response is $y_{zs}[n] = -2(0.4)^n + K(0.6)^n$, with zero initial conditions:

$$y_{zs}[-1] = 0 = -5 + \tfrac{K}{0.6} \qquad K = 3 \qquad y_{zs}[n] = -2(0.4)^n + 3(0.6)^n,\ n \geq 0$$

2. Its ZIR is found from the natural response $y_{zi}[n] = K(0.6)^n$, with given initial conditions:

$$y_{zi}[-1] = 10 = \tfrac{K}{0.6} \qquad K = 6 \qquad y_{zi}[n] = 6(0.6)^n,\ n \geq 0$$

3. The total response is $y[n] = y_{zi}[n] + y_{zs}[n] = -2(0.4)^n + 9(0.6)^n,\ n \geq 0$.
This matches the results of Example 5.6(a).

(b) Let $y[n] - \frac{1}{6}y[n-1] - \frac{1}{6}y[n-2] = 4,\ n \geq 0$, with $y[-1] = 0$ and $y[-2] = 12$.

1. The ZIR has the form of the natural response $y_{zi}[n] = K_1(\frac{1}{2})^n + K_2(-\frac{1}{3})^n$ (see Example 5.6(d)). To find the constants, we use the given initial conditions $y[-1] = 0$ and $y[-2] = 12$:

$$0 = K_1(\tfrac{1}{2})^{-1} + K_2(-\tfrac{1}{3})^{-1} = 2K_1 - 3K_2 \qquad 12 = K_1(\tfrac{1}{2})^{-2} + K_2(-\tfrac{1}{3})^{-2} = 4K_1 + 9K_2$$

Thus, $K_1 = 1.2,\ K_2 = 0.8$, and

$$y_{zi}[n] = 1.2(\tfrac{1}{2})^n + 0.8(-\tfrac{1}{3})^n, \quad n \geq 0$$

PGR2026-00033
Krisp Technologies EX1024 Page 122

2. The ZSR has the same form as the total response. Since the forced response (found in Example 5.6(d)) is $y_F[n] = 6$, we have

$$y_{zs}[n] = K_1(\tfrac{1}{2})^n + K_2(-\tfrac{1}{3})^n + 6$$

To find the constants, we assume zero initial conditions, $y[-1] = 0$ and $y[-2] = 0$, to get

$$y[-1] = 0 = 2K_1 - 3K_2 + 6 \qquad y[-2] = 0 = 4K_1 + 9K_2 + 6$$

We find $K_1 = -2.4$ and $K = 0.4$, and thus

$$y_{zs}[n] = -2.4(\tfrac{1}{2})^n + 0.4(-\tfrac{1}{3})^n + 6, \quad n \geq 0$$

3. The total response is $y[n] = y_{zi}[n] + y_{zs}[n] = -1.2(\frac{1}{2})^n + 1.2(-\frac{1}{3})^n + 6,\ n \geq 0$.
This matches the result of Example 5.6(d).

(c) (Linearity of the ZSR and ZIR) An IIR filter is described by $y[n] - y[n-1] - 2y[n-2] = x[n]$, with $x[n] = 6u[n]$ and initial conditions $y[-1] = -1,\ y[-2] = 4$.

1. Find the zero-input response, zero-state response, and total response.
2. How does the total response change if $y[-1] = -1,\ y[-2] = 4$ as given, but $x[n] = 12u[n]$?
3. How does the total response change if $x[n] = 6u[n]$ as given, but $y[-1] = -2,\ y[-2] = 8$?

1. We find the characteristic equation as $(1 - z^{-1} - 2z^{-2}) = 0$ or $(z^2 - z - 2) = 0$.
The roots of the characteristic equation are $z_1 = -1$ and $z_2 = 2$.
The form of the *natural* response is $y_N[n] = A(-1)^n + B(2)^n$.
Since the input $x[n]$ is constant for $n \geq 0$, the form of the *forced response* is also constant.
So, choose $y_F[n] = C$ in the system equation and evaluate C:

$$y_F[n] - y_F[n-1] - 2y_F[n-2] = C - C - 2C = 6 \qquad C = -3 \qquad y_F[n] = -3$$

For the ZSR, we use the *form* of the total response and zero initial conditions:

$$y_{zs}[n] = y_F[n] + y_N[n] = -0.5 + A(-1)^n + B(2)^n, \quad y[-1] = y[-2] = 0$$

We obtain $y_{zs}[-1] = 0 = -3 - A + 0.5B$ and $y_{zs}[-2] = 0 = -3 + A + 0.25B$
Thus, $A = 1$, $B = 8$, and $y_{zs}[n] = -3 + (-1)^n + 8(2)^n,\ n \geq 0$.
For the ZIR, we use the *form* of the *natural* response and the *given* initial conditions:

$$y_{zi}[n] = y_N[n] = A(-1)^n + B(2)^n \qquad y[-1] = -1 \qquad y[-2] = 4$$

This gives $y_{zi}[-1] = -1 = -A + 0.5B$, and $y_{zi}[-2] = 4 = A + 0.25B$.
Thus, $A = 3$, $B = 4$, and $y_{zi}[n] = 3(-1)^n + 4(2)^n,\ n \geq 0$.
The total response is the sum of the zero-input and zero-state response:

$$y[n] = y_{zi}[n] + y_{zs}[n] = -3 + 4(-1)^n + 12(2)^n, \quad n \geq 0$$

2. If $x[n] = 12u[n]$, the zero-state response doubles to $y_{zs}[n] = -6 + 2(-1)^n + 16(2)^n$.
3. If $y[-1] = -2$ and $y[-2] = 8$, the zero-input response doubles to $y_{zi}[n] = 6(-1)^n + 8(2)^n$.

PGR2026-00033
Krisp Technologies EX1024 Page 123

5.4.4 Solution of the General Difference Equation

The solution of the general difference equation described by

$$y[n] + A_1 y[n-1] + \cdots + A_N y[n-N] = B_0 x[n] + B_1 x[n-1] + \cdots + B_M x[n-M] \tag{5.16}$$

is found by invoking linearity and superposition as follows:

1. Compute the zero-state response $y_0[n]$ of the single-input system

$$y_0[n] + A_1 y_0[n-1] + A_2 y_0[n-2] + \cdots + A_N y_0[n-N] = x[n] \tag{5.17}$$

2. Use linearity and superposition to find $y_{zs}[n]$ as

$$y_{zs}[n] = B_0 y_0[n] + B_1 y_0[n-1] + \cdots + B_M y_0[n-M] \tag{5.18}$$

3. Find the zero-input response $y_{zi}[n]$ using initial conditions.
4. Find the total response as $y[n] = y_{zs}[n] + y_{zi}[n]$.

Note that the zero-input response is computed and included just once.

REVIEW PANEL 5.11

Solving $y[n] + A_1 y[n-1] + \cdots + A_N y[n-N] = B_0 x[n] + B_1 x[n-1] + \cdots + B_M x[n-M]$

1. Find $y_0[n]$ from $y_0[n] + A_1 y_0[n-1] + \cdots + A_N y_0[n-N] = x[n]$, **using zero initial conditions**.
2. Find ZSR (using superposition) as $y_{zs}[n] = B_0 y_0[n] + B_1 y_0[n-1] + \cdots + B_M y_0[n-M]$.
3. Find ZIR from $y_{zi}[n] + A_1 y_{zi}[n-1] + \cdots + A_N y_{zi}[n-N] = 0$, **using given initial conditions**.
4. Find complete response as $y[n] = y_{zs}[n] + y_{zi}[n]$.

EXAMPLE 5.8 (Response of a General System)

Consider the recursive digital filter whose realization is shown in Figure E5.8.
What is the response of this system if $x[n] = 6u[n]$ and $y[-1] = -1,\ y[-2] = 4$?

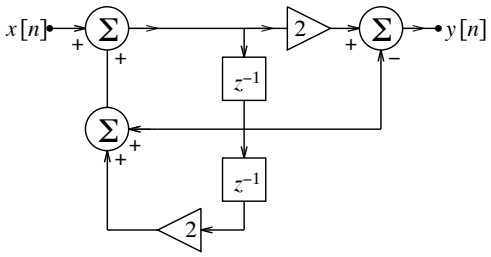


Figure E5.8 The digital filter for Example 5.8

Comparison with the generic realization of Figure 5.3 reveals that the system difference equation is

$$y[n] - y[n-1] - 2y[n-2] = 2x[n] - x[n-1]$$

From Example 5.7(c), the ZSR of $y[n] - y[n-1] - 2y[n-2] = x[n]$ is $y_0[n] = [-3 + (-1)^n + 8(2)^n]u[n]$. The ZSR for the input $2x[n] - x[n-1]$ is thus

$$y_{zs}[n] = 2y_0[n] - y_0[n-1] = [-6 + 2(-1)^n + 16(2)^n]u[n] - [3 + (-1)^{n-1} + 8(2)^{n-1}]u[n-1]$$

From Example 5.7(c), the ZIR of $y[n] - y[n-1] - 2y[n-2] = x[n]$ is $y_{zi}[n] = [3(-1)^n + 4(2)^n]u[n]$. The total response is $y[n] = y_{zi}[n] + y_{zs}[n]$:

$$y[n] = [3(-1)^n + 4(2)^n]u[n] + [-6 + 2(-1)^n + 16(2)^n]u[n] - [3 + (-1)^{n-1} + 8(2)^{n-1}]u[n-1]$$

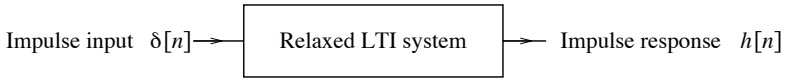
PGR2026-00033
Krisp Technologies EX1024 Page 124

5.5 Impulse Response of Digital Filters

The impulse response $h[n]$ of a *relaxed* LTI system is simply the response to a unit impulse input $\delta[n]$. The impulse response provides us with a powerful method for finding the zero-state response of LTI systems to arbitrary inputs using superposition (as described in the next chapter). The impulse response and the step response are often used to assess the time-domain performance of digital filters.

REVIEW PANEL 5.12

The Impulse Response and Step Response Is Defined Only for Relaxed LTI Systems



Impulse response $h[n]$: The output of a **relaxed** LTI system if the input is a unit impulse $\delta[n]$
Step response $s[n]$: The output of a **relaxed** LTI system if the input is a unit step $u[n]$

5.5.1 Impulse Response of Nonrecursive Filters

For a nonrecursive (FIR) filter of length $M+1$ described by

$$y[n] = B_0x[n] + B_1x[n-1] + \cdots + B_Mx[n-M] \tag{5.19}$$

the impulse response $h[n]$ (with $x[n] = \delta[n]$) is an $M+1$ term sequence of the input terms, which may be written as

$$h[n] = B_0\delta[n] + B_1\delta[n-1] + \cdots + B_M\delta[n-M] \quad \text{or} \quad h[n] = \{\overset{\Downarrow}{B_0}, B_1, \ldots, B_M\} \tag{5.20}$$

The sequence $h[n]$ represents the **FIR filter coefficients**.

5.5.2 Impulse Response of Recursive Filters

Recursion provides a simple means of obtaining as many terms of the impulse response $h[n]$ of a *relaxed* recursive filter as we please, though we may not always be able to discern a closed form from the results. Remember that we must use zero initial conditions.

EXAMPLE 5.9 (Impulse Response by Recursion)
Find the impulse response of the system described by $y[n] - \alpha y[n-1] = x[n]$.

We find $h[n]$ as the solution to $h[n] = \alpha h[n-1] + \delta[n]$ subject to the initial condition $y[-1] = 0$. By recursion, we obtain

$$h[0] = \alpha h[-1] + \delta[0] = 1 \qquad h[2] = \alpha h[1] = \alpha^2$$
$$h[1] = \alpha h[0] = \alpha \qquad h[3] = \alpha h[2] = \alpha^3$$

The general form of $h[n]$ is easy to discern as $h[n] = \alpha^n u[n]$.

5.5.3 Impulse Response for the Single-Input Case

Consider the Nth-order difference equation with a single input:

$$y[n] + A_1y[n-1] + A_2y[n-2] + \cdots + A_Ny[n-N] = x[n] \tag{5.21}$$

PGR2026-00033
Krisp Technologies EX1024 Page 125

To find its impulse response, we solve the difference equation

$$h[n] + A_1h[n-1] + A_2h[n-2] + \cdots + A_Nh[n-N] = \delta[n] \qquad \text{(zero initial conditions)} \tag{5.22}$$

Since the input $\delta[n]$ is zero for $n > 0$, we must apparently assume a forced response that is zero and thus solve for the natural response using initial conditions (leading to a trivial result). The trick is to use at least one nonzero initial condition, which we must find by recursion. By recursion, we find $h[0] = 1$. Since $\delta[n] = 0,\ n > 0$, the impulse response is found as the *natural response* of the homogeneous equation

$$h[n] + A_1h[n-1] + A_2h[n-2] + \cdots + A_Nh[n-N] = 0, \quad h[0] = 1 \tag{5.23}$$

subject to the nonzero initial condition $h[0] = 1$. All the other initial conditions are assumed to be zero ($h[-1] = 0$ for a second-order system, $h[-1] = h[-2] = 0$ for a third-order system, and so on).

REVIEW PANEL 5.13

The Impulse Response of $y[n] + A_1y[n-1] + \cdots + A_Ny[n-N] = x[n]$

Find $h[n]$ from $h[n] + A_1h[n-1] + \cdots + A_Nh[n-N] = 0$, with IC $h[0] = 1$ (*all other ICs zero*).

EXAMPLE 5.10 (Impulse Response Computation for the Single-Input Case)

(a) Consider the difference equation $y[n] - 0.6y[n-1] = x[n]$.

Its impulse response is found by solving $h[n] - 0.6h[n-1] = 0,\ h[0] = 1$.

Its natural response is $h[n] = K(0.6)^n$.

With $h[0] = 1$, we find $K = 1$, and thus $h[n] = (0.6)^nu[n]$.

(b) Let $y[n] - \frac{1}{6}y[n-1] - \frac{1}{6}y[n-2] = x[n]$.

Its impulse response is found by solving $h[n] - \frac{1}{6}h[n-1] - \frac{1}{6}h[n-2] = 0,\ h[0] = 1,\ h[-1] = 0$.

Its characteristic equation is $1 - \frac{1}{6}z^{-1} - \frac{1}{6}z^{-2} = 0$ or $z^2 - \frac{1}{6}z - \frac{1}{6} = 0$.

Its roots, $z_1 = \frac{1}{2}$ and $z_2 = -\frac{1}{3}$, give the natural response $h[n] = K_1(\frac{1}{2})^n + K_2(-\frac{1}{3})^n$.

With $h[0] = 1$ and $h[-1] = 0$, we find $1 = K_1 + K_2$ and $0 = 2K_1 - 3K_2$.

Solving for the constants, we obtain $K_1 = 0.6$ and $K_2 = 0.4$.

Thus, $h[n] = [0.6(\frac{1}{2})^n + 0.4(-\frac{1}{3})^n]u[n]$.

5.5.4 Impulse Response for the General Case

To find the impulse response of the general system described by

$$y[n] + A_1y[n-1] + A_2y[n-2] + \cdots + A_Ny[n-N] = B_0x[n] + B_1x[n-1] + \cdots + B_Mx[n-M] \tag{5.24}$$

we use linearity and superposition as follows:

1. Find the impulse response $h_0[n]$ of the single-input system

$$y_0[n] + A_1y_0[n-1] + A_2y_0[n-2] + \cdots + A_Ny_0[n-N] = x[n] \tag{5.25}$$

PGR2026-00033
Krisp Technologies EX1024 Page 126

by solving the homogeneous equation

$$h_0[n] + A_1 h_0[n-1] + \cdots + A_N h_0[n-N] = 0, \quad y[0] = 1 \text{ (all other ICs zero)} \tag{5.26}$$

2. Then, invoke superposition to find the actual impulse response $h[n]$ as

$$h[n] = B_0 h_0[n] + B_1 h_0[n-1] + \cdots + B_M h_0[n-M] \tag{5.27}$$

REVIEW PANEL 5.14

Impulse Response of $y[n] + A_1 y[n-1] + \cdots + A_N y[n-N] = B_0 x[n] + B_1 x[n-1] + \cdots + B_M x[n-M]$

1. Find $h_0[n]$ from $h_0[n] + A_1 h_0[n-1] + \cdots + A_N h_0[n-N] = 0$, IC: $h_0[0] = 1$ (*all others zero*).

2. Find $h[n]$ (using superposition) as $h[n] = B_0 h_0[n] + B_1 h_0[n-1] + \cdots + B_M h_0[n-M]$.

EXAMPLE 5.11 (Impulse Response for the General Case)

(a) Find the impulse response of $y[n] - 0.6y[n-1] = 4x[n]$ and $y[n] - 0.6y[n-1] = 3x[n+1] - x[n]$.

We start with the single-input system $y_0[n] - 0.6y_0[n-1] = x[n]$.

Its impulse response $h_0[n]$ was found in Example 5.10(a) as $h_0[n] = (0.6)^n u[n]$.

Then, for the first system, $h_[n] = 4h_0[n] = 4(0.6)^n u[n]$.

For the second system, $h[n] = 3h_0[n+1] - h_0[n] = 3(0.6)^{n+1} u[n+1] - (0.6)^n u[n]$.

This may also be expressed as $h[n] = 3\delta[n+1] + 0.8(0.6)^n u[n]$.

Comment: The general approach can be used for causal or noncausal systems.

(b) Let $y[n] - \frac{1}{6}y[n-1] - \frac{1}{6}y[n-2] = 2x[n] - 6x[n-1]$.

To find $h[n]$, start with the single-input system $y[n] - \frac{1}{6}y[n-1] - \frac{1}{6}y[n-2] = x[n]$.

Its impulse response $h_0[n]$ was found in Example 5.10(b) as

$$h_0[n] = [0.6(\tfrac{1}{2})^n + 0.4(-\tfrac{1}{3})^n]u[n]$$

The impulse response of the given system is $h[n] = 4h_0[n] - h_0[n-1]$. We find

$$h[n] = [1.2(\tfrac{1}{2})^n + 0.8(-\tfrac{1}{3})^n]u[n] - [3.6(\tfrac{1}{2})^{n-1} + 2.4(-\tfrac{1}{3})^{n-1}]u[n-1]$$

Comment: This may be simplified to $h[n] = [-6(\tfrac{1}{2})^n + 8(-\tfrac{1}{3})^n]u[n]$.

5.5.5 Recursive Form for Nonrecursive Digital Filters

The terms *FIR* and *nonrecursive* are synonymous. A nonrecursive filter always has a finite impulse response. The terms *IIR* and *recursive* are often, but not always, synonymous. Not all recursive filters have an infinite impulse response. In fact, nonrecursive filters can often be implemented in recursive form. A recursive filter may also be described by a nonrecursive filter of the form $y[n] = B_0 x[n] + B_1 x[n-1] + \cdots + B_M x[n-M]$ if we know all the past inputs. In general, this implies $M \to \infty$.

PGR2026-00033
Krisp Technologies EX1024 Page 127

EXAMPLE 5.12 (Recursive Form for Nonrecursive Filters)
Consider the nonrecursive filter $y[n] = x[n] + x[n-1] + x[n-2]$.

Its impulse response is $h[n] = \delta[n] + \delta[n-1] + \delta[n-2]$.

To cast this filter in recursive form, we compute $y[n-1] = x[n-1] + x[n-2] + x[n-3]$.

Upon subtraction from the original equation, we obtain the recursive form

$$y[n] - y[n-1] = x[n] - x[n-3]$$

Comment: Remember that the impulse response of this recursive system is of finite length.

5.5.6 The Response of Anti-Causal Systems

So far, we have focused on the response of systems described by difference equations to causal inputs. However, by specifying an anti-causal input and appropriate initial conditions, the same difference equation can be solved backward in time for $n < 0$ to generate an anti-causal response. For example, to find the causal impulse response $h[n]$, we assume that $h[n] = 0,\ n < 0$; but to find the anti-causal impulse response $h_A[n]$ of the same system, we would assume that $h[n] = 0,\ n > 0$. This means that the same system can be described by two different impulse response functions. How we distinguish between them is easily handled using the z-transform, which we study in Chapter 17.

EXAMPLE 5.13 (Causal and Anti-Causal Impulse Response)

(a) Find the causal impulse response of the first-order system $y[n] - 0.4y[n-1] = x[n]$.

For the causal impulse response, we assume $h[n] = 0,\ n < 0$, and solve for $h[n],\ n > 0$, by recursion from $h[n] = 0.4h[n-1] + \delta[n]$. With $h[0] = 0.4h[-1] + \delta[0] = 1$ and $\delta[n] = 0,\ n \neq 0$, we find

$$h[1] = 0.4h[0] = 0.4 \qquad h[2] = 0.4h[1] = (0.4)^2 \quad h[3] = 0.4h[2] = (0.4)^3 \quad \text{etc.}$$

The general form is easily discerned as $h[n] = (0.4)^n$ and is valid for $n \geq 0$.

Comment: The causal impulse response of $y[n] - \alpha y[n-1] = x[n]$ is $h[n] = \alpha^n u[n]$.

(b) Find the anti-causal impulse response of the first-order system $y[n] - 0.4y[n-1] = x[n]$.

For the anti-causal impulse response, we assume $h[n] = 0,\ n \geq 0$, and solve for $h[n],\ n < 0$, by recursion from $h[n-1] = 2.5(h[n] - \delta[n])$. With $h[-1] = 2.5(h[0] - \delta[0]) = -2.5$, and $\delta[n] = 0,\ n \neq 0$, we find

$$h[-2] = 2.5h[-1] = -(2.5)^2 \qquad h[-3] = 2.5h[-2] = -(2.5)^3 \quad h[-4] = 2.5h[-3] = -(2.5)^4 \quad \text{etc.}$$

The general form is easily discerned as $h[n] = -(2.5)^{-n} = -(0.4)^n$ and is valid for $n \leq -1$.

Comment: The anti-causal impulse response of $y[n] - \alpha y[n-1] = x[n]$ is $h[n] = -\alpha^n u[-n-1]$.

PGR2026-00033
Krisp Technologies EX1024 Page 128

5.6 Stability of Discrete-Time LTI Systems

Stability is important in practical design and involves constraints on the nature of the system response. **Bounded-input**, **bounded-output** (BIBO) **stability** implies that *every* bounded input must result in a bounded output. For an LTI system described by the difference equation

$$y[n] + A_1 y[n-1] + \cdots + A_N y[n-N] = B_0 x[n] + B_1 x[n-1] + \cdots + B_M x[n-M] \tag{5.28}$$

the conditions for BIBO stability involve the roots of the characteristic equation. A necessary and sufficient condition for BIBO stability of such an LTI system is that *every root of its characteristic equation must have a magnitude less than unity*. This criterion is based on the results of Tables 5.1 and 5.2. Root magnitudes less than unity ensure that the natural (and zero-input) response always decays with time (see Table 5.1), and the forced (and zero-state) response always remains bounded for *every* bounded input. Roots with magnitudes that equal unity make the system unstable. Simple (non-repeated) roots with unit magnitude produce a constant (or sinusoidal) natural response that is bounded; but if the input is also a constant (or sinusoid at the same frequency), the forced response is a ramp or growing sinusoid (see Table 5.2) and hence unbounded. Repeated roots with unit magnitude result in a natural response that is itself a growing sinusoid or polynomial and thus unbounded. In the next chapter, we shall see that the stability condition is equivalent to having an LTI system whose impulse response $h[n]$ is *absolutely summable*. The stability of nonlinear or time-varying systems usually must be checked by other means.

The Stability of FIR Filters: The system equation of an FIR filter describes the output as a weighted sum of shifted inputs. If the input remains bounded, the weighted sum of the inputs is also bounded. In other words, FIR *filters are always stable*.

REVIEW PANEL 5.15

Requirement for BIBO Stability of Discrete-Time LTI Systems

Every root of the characteristic equation must have magnitude (absolute value) *less than* unity.
Note: FIR *filters are always stable*.

EXAMPLE 5.14 (Concepts Based on Stability)

(a) The system $y[n] - \frac{1}{6}y[n-1] - \frac{1}{6}y[n-2] = x[n]$ is stable since the roots of its characteristic equation $z^2 - \frac{1}{6}z - \frac{1}{6} = 0$ are $z_1 = \frac{1}{2}$ and $z_2 = -\frac{1}{3}$ and their magnitudes are less than 1.

(b) The system $y[n] - y[n-1] = x[n]$ is unstable. The root of its characteristic equation $z - 1 = 0$ is $z = 1$ gives the natural response $y_N = Ku[n]$, which is actually bounded. However, for an input $x[n] = u[n]$, the forced response will have the form $Cnu[n]$, which becomes unbounded.

(c) The system $y[n] - 2y[n-1] + y[n-2] = x[n]$ is unstable. The roots of its characteristic equation $z^2 - 2z + 1 = 0$ are equal and produce the unbounded natural response $y_N[n] = Au[n] + Bnu[n]$.

(d) The system $y[n] - \frac{1}{2}y[n-1] = nx[n]$ is linear, time varying and unstable. The (bounded) step input $x[n] = u[n]$ results in a response that includes the ramp $nu[n]$, which becomes unbounded.

(e) The system $y[n] = x[n] - 2x[n-1]$ is stable because it describes an FIR filter.

PGR2026-00033
Krisp Technologies EX1024 Page 129

5.7 Connections: System Representation in Various Forms

An LTI system may be described by a difference equation, impulse response, or input-output data. All three are related and, given one form, we should be able to access the others. We have already studied how to obtain the impulse response from a difference equation. Here we shall describe how to obtain the system difference equation from its impulse response or from input-output data.

5.7.1 Difference Equations from the Impulse Response

In the time domain, the process of finding the difference equation from its impulse response is tedious. It is much easier implemented by other methods (such as the z-transform, which we explore in Chapter 17). The central idea is that the terms in the impulse response are an indication of the natural response (and the roots of the characteristic equation) from which the difference equation may be reconstructed if we can describe the combination of the impulse response and its delayed versions by a sum of impulses. The process is best illustrated by some examples.

EXAMPLE 5.15 (Difference Equations from the Impulse Response)

(a) Let $h[n] = u[n]$. Then $h[n-1] = u[n-1]$, and $h[n] - h[n-1] = u[n] - u[n-1] = \delta[n]$.

The difference equation corresponding to $h[n] - h[n-1] = \delta[n]$ is simply $y[n] - y[n-1] = x[n]$.

(b) Let $h[n] = 3(0.6)^n u[n]$. This suggests a difference equation whose left-hand side is $y[n] - 0.6y[n-1]$.

We then set up $h[n] - 0.6h[n-1] = 3(0.6)^n u[n] - 1.8(0.6)^{n-1}u[n-1]$. This simplifies to

$$h[n] - 0.6h[n-1] = 3(0.6)^n u[n] - 3(0.6)^n u[n-1] = 3(0.6)^n(u[n] - u[n-1]) = 3(0.6)^n\delta[n] = 3\delta[n]$$

The difference equation corresponding to $h[n] - 0.6h[n-1] = 3\delta[n]$ is $y[n] - 0.6y[n-1] = 3x[n]$.

(c) Let $h[n] = 2(-0.5)^n u[n] + (0.5)^n u[n]$. This suggests a characteristic equation $(z - 0.5)(z + 0.5)$.

The left-hand side of the difference equation is thus $y[n] - 0.25y[n-2]$. We now compute

$$h[n] - 0.25h[n-2] = 2(-0.5)^n u[n] + (0.5)^n u[n] - 0.25(2(-0.5)^{n-1}u[n-1] + (0.5)^{n-1}u[n-1])$$

This simplifies to

$$h[n] - 0.25h[n-2] = [2(-0.5)^n + (0.5)^n](u[n] - u[n-2]) = [2(-0.5)^n + (0.5)^n](\delta[n] + \delta[n-2])$$

This simplifies further to $h[n] - 0.25h[n-2] = 3\delta[n] - 0.5\delta[n-1]$.

Finally, the difference equation is $y[n] - 0.25y[n-2] = 3x[n] - 0.5x[n-1]$.

5.7.2 Difference Equations from Input-Output Data

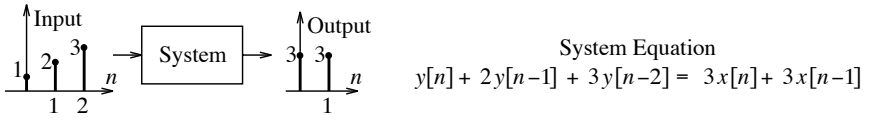
The difference equation of LTI systems may also be obtained from input-output data. The response of the system described by $y[n] = 3x[n] + 2x[n-1]$ to $x[n] = \delta[n]$ is $y[n] = 3\delta[n] + 2\delta[n-1]$. Turning things around, the input $\delta[n]$ and output $3\delta[n] + 2\delta[n-1]$ then corresponds to the difference equation $y[n] = 3x[n] + 2x[n-1]$. Note how the coefficients of the input match the output data (and vice versa).

PGR2026-00033
Krisp Technologies EX1024 Page 130

REVIEW PANEL 5.16

Difference Equation of an LTI System from Input-Output Information

Example: If $x[n] = \{1,\ 2,\ 3\}$ and $y[n] = \{3,\ 3\}$, then $y[n] + 2y[n-1] + 3y[n-2] = 3x[n] + 3x[n-1]$.



5.8 Application-Oriented Examples

5.8.1 Inverse Systems

Inverse systems are quite important in practical applications. For example, a measurement system (such as a transducer) invariably affects (distorts) the signal being measured. To undo the effects of the distortion requires a system that acts as the inverse of the measurement system. If an input $x[n]$ to a system results in an output $y[n]$, then its inverse is a system that recovers the signal $x[n]$ in response to the input $y[n]$, as illustrated in Figure 5.4.

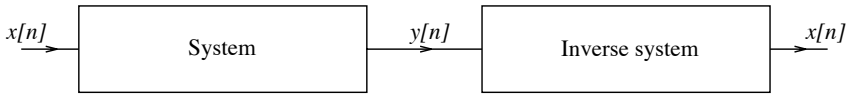


Figure 5.4 A system and its inverse

Not all systems have an inverse. For a system to have an inverse, or be **invertible**, distinct inputs must lead to distinct outputs. If a system produces an identical output for two different inputs, it does not have an inverse. For invertible LTI systems described by difference equations, finding the inverse system is as easy as switching the input and output variables.

REVIEW PANEL 5.17

Finding the Inverse of an LTI System? Try Switching the Input and Output

System: $y[n] + 2y[n-1] = 3x[n] + 4x[n-1]$ **Inverse system:** $3y[n] + 4y[n-1] = x[n] + 2x[n-1]$

How to find the inverse from the impulse response $h[n]$? Find the difference equation first.

EXAMPLE 5.16 (Inverse Systems)

(a) Refer to the interconnected system shown in Figure E5.16A(1). Find the difference equation of the inverse system, sketch a realization of each system, and find the output of each system.

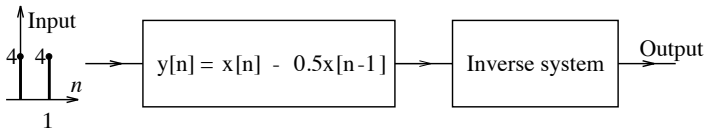


Figure E5.16A(1) The interconnected system for Example 5.16(a)

The original system is described by $y[n] = x[n] - 0.5x[n-1]$. By switching the input and output, the inverse system is described by $y[n] - 0.5y[n-1] = x[n]$. The realization of each system is shown in

PGR2026-00033
Krisp Technologies EX1024 Page 131

Figure E5.16A(2). Are they related? Yes. If you flip the realization of the echo system end-on-end *and change the sign of the feedback signal*, you get the inverse realization.

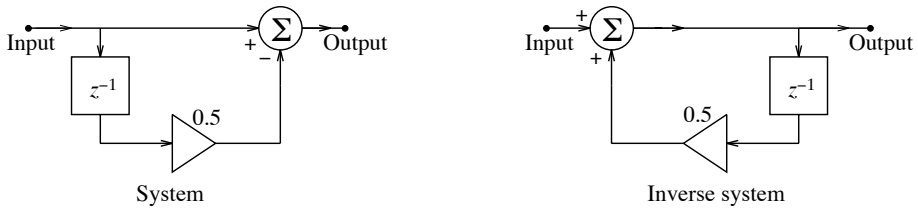


Figure E5.16A(2) Realization of the system and its inverse for Example 5.16(a)

The response $g[n]$ of the first system is simply

$$g[n] = (4\delta[n] + 4\delta[n-1]) - (2\delta[n-1] + 2\delta[n-2]) = 4\delta[n] + 2\delta[n-1]) - 2\delta[n-2])$$

If we let the output of the second system be $y_0[n]$, we have

$$y_0[n] = 0.5y_0[n-1] + 4\delta[n] + 2\delta[n-1]) - 2\delta[n-2])$$

Recursive solution gives

$$y_0[0] = 0.5y_0[-1] + 4\delta[0] = 4 \quad y_0[1] = 0.5y_0[0] + 2\delta[0] = 4 \quad y_0[2] = 0.5y_0[1] - 2\delta[0] = 0$$

All subsequent values of $y_0[n]$ are zero since the input terms are zero for $n > 2$. The output is thus $y_0[n] = \{\overset{\Downarrow}{4},\ 4\}$, the same as the input to the overall system.

(b) The nonlinear system $y[n] = x^2[n]$ does not have an inverse. Two inputs, differing in sign, yield the same output. If we try to recover $x[n]$ as $\sqrt{y[n]}$, we run into a sign ambiguity.

(c) The linear (but time-varying) decimating system $y[n] = x[2n]$ does not have an inverse. Two inputs, which differ in the samples discarded (for example, the signals $\{1, 2, 4, 5\}$ and $\{1, 3, 4, 8\}$ yield the same output $\{1,\ 4\}$). If we try to recover the original signal by interpolation, we cannot uniquely identify the original signal.

(d) The linear (but time-varying) interpolating system $y[n] = x[n/2]$ does have an inverse. Its inverse is a decimating system that discards the very samples inserted during interpolation and thus recovers the original signal.

(e) The LTI system $y[n] = x[n] + 2x[n-1]$ also has an inverse. Its inverse is found by switching the input and output as $y[n] + 2y[n-1] = x[n]$. This example also shows that the inverse of an FIR filter results in an IIR filter.

PGR2026-00033
Krisp Technologies EX1024 Page 132

5.8.2 Echo and Reverb

The digital processing of audio signals often involves digital filters to create various special effects such as echo and reverb, which are typical of modern-day studio sounds. An echo filter has the form

$$y[n] = x[n] + \alpha x[n - D] \qquad \text{(an echo filter)} \tag{5.29}$$

This describes an FIR filter whose output $y[n]$ equals the input $x[n]$ and its delayed (by D samples) and attenuated (by α) replica of $x[n]$ (the echo term). Its realization is sketched in Figure 5.5. The D-sample delay is implemented by a cascade of D delay elements and represented by the block marked z^{-D}. This filter is also called a **comb filter** (for reasons to be explained in later chapters).

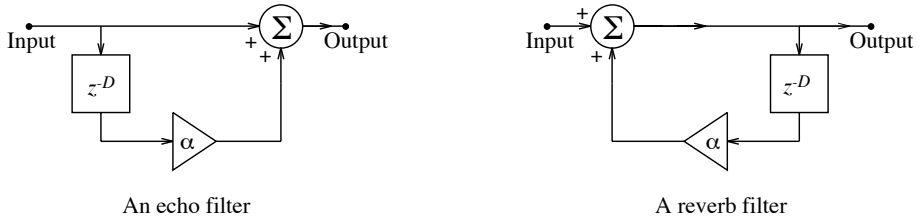


Figure 5.5 Echo and reverb filters

Reverberations are due to multiple echoes (from the walls and other structures in a concert hall, for example). For simplicity, if we assume that the signal suffers the same delay D and the same attenuation α in each round-trip to the source, we may describe the action of reverb by

$$y[n] = x[n] + \alpha x[n - D] + \alpha^2 x[n - 2D] + \alpha^3 x[n - 3D] + \cdots \tag{5.30}$$

If we delay both sides by D units, we get

$$\alpha y[n - D] = \alpha x[n - D] + \alpha^2 x[n - 2D] + \alpha^3 x[n - 3D] + \alpha^4 x[n - 4D] + \cdots \tag{5.31}$$

Subtracting the second equation from the first, we obtain a compact form for a reverb filter:

$$y[n] - \alpha y[n - D] = x[n] \qquad \text{(a reverb filter)} \tag{5.32}$$

This is an IIR filter whose realization is also sketched in Figure 5.5. Its form is reminiscent of the inverse of the echo system $y[n] + \alpha y[n - D] = x[n]$, but with α replaced by $-\alpha$.

In concept, it should be easy to tailor the simple reverb filter to simulate realistic effects by including more terms with different delays and attenuation. In practice, however, this is no easy task, and the filter designs used by commercial vendors in their applications are often proprietary.

5.8.3 Periodic Sequences and Wave-Table Synthesis

Electronic music synthesizers typically possess *tone banks* of stored tone sequences that are replicated and used in their original or delayed forms or combined to synthesize new sounds and generate various musical effects. The periodic versions of such sequences can be generated by filters that have a recursive form developed from a nonrecursive filter whose impulse response corresponds to one period of the periodic signal. The difference equation of such a recursive filter is given by

$$y[n] - y[n - N] = x_1[n] \qquad (N\text{-sample periodic signal generator}) \tag{5.33}$$

PGR2026-00033
Krisp Technologies EX1024 Page 133

where $x_1[n]$ corresponds to one period (N samples) of the signal $x[n]$. This form actually describes a reverb system with no attenuation whose delay equals the period N. Hardware implementation often uses a circular buffer or *wave-table* (in which one period of the signal is stored), and cycling over it generates the periodic signal. The same wave-table can also be used to change the frequency (or period) of the signal (to double the frequency for example, we would cycle over alternate samples) or for storing a new signal.

EXAMPLE 5.17 (Generating Periodic Signals Using Recursive Filters)
Suppose we wish to generate the periodic signal $x[n]$ described by

$$x[n] = \{2,\ 3,\ 1,\ 6,\ 2,\ 3,\ 1,\ 6,\ 2,\ 3,\ 1,\ 6,\ 2,\ 3,\ 1,\ 6, \ldots\}$$

The impulse response sequence of a nonrecursive filter that generates this signal is simply

$$h[n] = \{2,\ 3,\ 1,\ 6,\ 2,\ 3,\ 1,\ 6,\ 2,\ 3,\ 1,\ 6,\ 2,\ 3,\ 1,\ 6, \ldots\}$$

If we delay this sequence by one period (four samples), we obtain

$$h[n-4] = \{0,\ 0,\ 0,\ 0,\ 2,\ 3,\ 1,\ 6,\ 2,\ 3,\ 1,\ 6,\ 2,\ 3,\ 1,\ 6,\ 2,\ 3,\ 1,\ 6, \ldots\}$$

Subtracting the delayed sequence from the original gives us $h[n] - h[n-4] = \{2,\ 3,\ 1,\ 6\}$.

Its recursive form is $y[n] - y[n-4] = x_1[n]$, where $x_1[n] = \{2,\ 3,\ 1,\ 6\}$ describes one period of $x[n]$.

5.8.4 How Difference Equations Arise

We conclude with some examples of difference equations, which arise in many ways in various fields ranging from mathematics and engineering to economics and biology.

1. $y[n] = y[n-1] + n,\ y[-1] = 1$

   This difference equation describes the number of regions $y[n]$ into which n lines divide a plane if no two lines are parallel and no three lines intersect.

2. $y[n+1] = (n+1)(y[n]+1),\ y[0] = 0$

   This difference equation describes the number of multiplications $y[n]$ required to compute the determinant of an $n \times n$ matrix using cofactors.

3. $y[n+2] = y[n+1] + y[n],\ y[0] = 0,\ y[1] = 1$

   This difference equation generates the Fibonacci sequence $\{y[n] = 0,\ 1,\ 1,\ 2,\ 3,\ 5, \ldots\}$, where each number is the sum of the previous two.

4. $y[n+2] - 2xy[n+1] + y[n] = 0,\ y[0] = 1,\ y[1] = x$

   This difference equation generates the **Chebyshev polynomials** $T_n(x) = y[n]$ *of the first kind.* We find that $T_2(x) = y[2] = 2x^2 - 1,\ T_3(x) = y[3] = 4x^3 - 3x$, etc. Similar difference equations called *recurrence relations* form the basis for generating other polynomial sets.

5. $y[n+1] = \alpha y[n](1 - y[n])$

   In biology, this difference equation, called a *logistic equation*, is used to model the growth of populations that reproduce at discrete intervals.

6. $y[n+1] = (1+\alpha)y[n] + d[n]$

   This difference equation describes the bank balance $y[n]$ at the beginning of the nth-compounding period (day, month, etc.) if the percent interest rate is α per compounding period and $d[n]$ is the amount deposited in that period.

PGR2026-00033
Krisp Technologies EX1024 Page 134

CHAPTER 5 PROBLEMS

DRILL AND REINFORCEMENT

5.1 (Operators) Which of the following describe linear operators?

(a) $\mathcal{O}\{\ \} = 4\{\ \}$ **(b)** $\mathcal{O}\{\ \} = 4\{\ \} + 3$ **(c)** $\mathcal{O}\{\ \} = \alpha^{\{\ \}}$

5.2 (System Classification) In each of the systems below, $x[n]$ is the input and $y[n]$ is the output. Check each system for linearity, shift invariance, memory, and causality.

(a) $y[n] - y[n-1] = x[n]$
(b) $y[n] + y[n+1] = nx[n]$
(c) $y[n] - y[n+1] = x[n+2]$
(d) $y[n+2] - y[n+1] = x[n]$
(e) $y[n+1] - x[n]y[n] = nx[n+2]$
(f) $y[n] + y[n-3] = x^2[n] + x[n+6]$
(g) $y[n] - 2^n y[n] = x[n]$
(h) $y[n] = x[n] + x[n-1] + x[n-2]$

5.3 (Response by Recursion) Find the response of the following systems by recursion to $n = 4$ and try to discern the general form for $y[n]$.

(a) $y[n] - ay[n-1] = \delta[n]$ $y[-1] = 0$
(b) $y[n] - ay[n-1] = u[n]$ $y[-1] = 1$
(c) $y[n] - ay[n-1] = nu[n]$ $y[-1] = 0$
(d) $y[n] + 4y[n-1] + 3y[n-2] = u[n-2]$ $y[-1] = 0$ $y[-2] = 1$

5.4 (Forced Response) Find the forced response of the following systems.

(a) $y[n] - 0.4y[n-1] = u[n]$
(b) $y[n] - 0.4y[n-1] = (0.5)^n$
(c) $y[n] + 0.4y[n-1] = (0.5)^n$
(d) $y[n] - 0.5y[n-1] = \cos(n\pi/2)$

5.5 (Forced Response) Find the forced response of the following systems.

(a) $y[n] - 1.1y[n-1] + 0.3y[n-2] = 2u[n]$
(b) $y[n] - 0.9y[n-1] + 0.2y[n-2] = (0.5)^n$
(c) $y[n] + 0.7y[n-1] + 0.1y[n-2] = (0.5)^n$
(d) $y[n] - 0.25y[n-2] = \cos(n\pi/2)$

5.6 (Zero-State Response) Find the zero-state response of the following systems.

(a) $y[n] - 0.5y[n-1] = 2u[n]$
(b) $y[n] - 0.4y[n-1] = (0.5)^n$
(c) $y[n] - 0.4y[n-1] = (0.4)^n$
(d) $y[n] - 0.5y[n-1] = \cos(n\pi/2)$

5.7 (Zero-State Response) Find the zero-state response of the following systems.

(a) $y[n] - 1.1y[n-1] + 0.3y[n-2] = 2u[n]$
(b) $y[n] - 0.9y[n-1] + 0.2y[n-2] = (0.5)^n$
(c) $y[n] + 0.7y[n-1] + 0.1y[n-2] = (0.5)^n$
(d) $y[n] - 0.25y[n-2] = \cos(n\pi/2)$

5.8 (System Response) Let $y[n] - 0.5y[n-1] = x[n]$, with $y[-1] = -1$. Find the response of this system for the following inputs.

(a) $x[n] = 2u[n]$
(b) $x[n] = (0.25)^n u[n]$
(c) $x[n] = n(0.25)^n u[n]$
(d) $x[n] = (0.5)^n u[n]$
(e) $x[n] = n(0.5)^n$
(f) $x[n] = (0.5)^n \cos(0.5n\pi)$

PGR2026-00033
Krisp Technologies EX1024 Page 135

5.9 (System Response) Find the response of the following systems.

(a) $y[n] + 0.1y[n-1] - 0.3y[n-2] = 2u[n]$ $\quad y[-1] = 0 \quad y[-2] = 0$
(b) $y[n] - 0.9y[n-1] + 0.2y[n-2] = (0.5)^n$ $\quad y[-1] = 1 \quad y[-2] = -4$
(c) $y[n] + 0.7y[n-1] + 0.1y[n-2] = (0.5)^n$ $\quad y[-1] = 0 \quad y[-2] = 3$
(d) $y[n] - 0.25y[n-2] = (0.4)^n$ $\quad y[-1] = 0 \quad y[-2] = 3$
(e) $y[n] - 0.25y[n-2] = (0.5)^n$ $\quad y[-1] = 0 \quad y[-2] = 0$

5.10 (System Response) Sketch a realization for each system, assuming zero initial conditions. Then evaluate the complete response from the information given. Check your answer by computing the first few values by recursion.

(a) $y[n] - 0.4y[n-1] = x[n]$ $\quad x[n] = (0.5)^n u[n] \quad y[-1] = 0$
(b) $y[n] - 0.4y[n-1] = 2x[n] + x[n-1]$ $\quad x[n] = (0.5)^n u[n] \quad y[-1] = 0$
(c) $y[n] - 0.4y[n-1] = 2x[n] + x[n-1]$ $\quad x[n] = (0.5)^n u[n] \quad y[-1] = 5$
(d) $y[n] + 0.5y[n-1] = x[n] - x[n-1]$ $\quad x[n] = (0.5)^n u[n] \quad y[-1] = 2$
(e) $y[n] + 0.5y[n-1] = x[n] - x[n-1]$ $\quad x[n] = (-0.5)^n u[n] \quad y[-1] = 0$

5.11 (System Response) For each system, evaluate the natural, forced, and total response. Assume that $y[-1] = 0, y[-2] = 1$. Check your answer for the total response by computing its first few values by recursion.

(a) $y[n] + 4y[n-1] + 3y[n-2] = u[n]$
(b) $y[n] + 4y[n-1] + 4y[n-2] = 2^n u[n]$
(c) $y[n] + 4y[n-1] + 8y[n-2] = \cos(n\pi)u[n]$
(d) $\{(1 + 2z^{-1})^2\}y[n] = n(2)^n u[n]$
(e) $\{1 + \frac{3}{4}z^{-1} + \frac{1}{8}z^{-2}\}y[n] = (\frac{1}{3})^n u[n]$
(f) $\{1 + 0.5z^{-1} + 0.25z^{-2}\}y[n] = \cos(0.5n\pi)u[n]$
(g) $\{z^2 + 4z + 4\}y[n] = 2^n u[n]$
(h) $\{1 - 0.5z^{-1}\}y[n] = (0.5)^n \cos(0.5n\pi)u[n]$

5.12 (System Response) For each system, set up a difference equation and compute the zero-state, zero-input, and total response, assuming $x[n] = u[n]$ and $y[-1] = y[-2] = 1$.

(a) $\{1 - z^{-1} - 2z^{-2}\}y[n] = x[n]$
(b) $\{z^2 - z - 2\}y[n] = x[n]$
(c) $\{1 - \frac{3}{4}z^{-1} + \frac{1}{8}z^{-2}\}y[n] = \{z^{-1}\}x[n]$
(d) $\{1 - \frac{3}{4}z^{-1} + \frac{1}{8}z^{-2}\}y[n] = \{1 + z^{-1}\}x[n]$
(e) $\{1 - 0.25z^{-2}\}y[n] = x[n]$
(f) $\{z^2 - 0.25\}y[n] = \{2z^2 + 1\}x[n]$

5.13 (Impulse Response by Recursion) Find the impulse response $h[n]$ by recursion up to $n = 4$ for each of the following systems.

(a) $y[n] - y[n-1] = 2x[n]$
(b) $y[n] - 3y[n-1] + 6y[n-2] = x[n-1]$
(c) $y[n] - 2y[n-3] = x[n-1]$
(d) $y[n] - y[n-1] + 6y[n-2] = nx[n-1] + 2x[n-3]$

5.14 (Analytical Form for Impulse Response) Classify each filter as recursive or FIR (nonrecursive), and causal or noncausal, and find an expression for its impulse response $h[n]$.

(a) $y[n] = x[n] + x[n-1] + x[n-2]$
(b) $y[n] = x[n+1] + x[n] + x[n-1]$
(c) $y[n] + 2y[n-1] = x[n]$
(d) $y[n] + 2y[n-1] = x[n-1]$
(e) $y[n] + 2y[n-1] = 2x[n] + 6x[n-1]$
(f) $y[n] + 2y[n-1] = x[n+1] + 4x[n] + 6x[n-1]$
(g) $\{1 + 4z^{-1} + 3z^{-2}\}y[n] = \{z^{-2}\}x[n]$
(h) $\{z^2 + 4z + 4\}y[n] = \{z + 3\}x[n]$
(i) $\{z^2 + 4z + 8\}y[n] = x[n]$
(j) $y[n] + 4y[n-1] + 4y[n-2] = x[n] - x[n+2]$

5.15 (Stability) Investigate the causality and stability of the following systems.

(a) $y[n] = x[n-1] + x[n] + x[n+1]$
(b) $y[n] = x[n] + x[n-1] + x[n-2]$
(c) $y[n] - 2y[n-1] = x[n]$
(d) $y[n] - 0.2y[n-1] = x[n] - 2x[n+2]$
(e) $y[n] + y[n-1] + 0.5y[n-2] = x[n]$
(f) $y[n] - y[n-1] + y[n-2] = x[n] - x[n+1]$
(g) $y[n] - 2y[n-1] + y[n-2] = x[n] - x[n-3]$
(h) $y[n] - 3y[n-1] + 2y[n-2] = 2x[n+3]$

PGR2026-00033
Krisp Technologies EX1024 Page 136

REVIEW AND EXPLORATION

5.16 (System Classification) Classify the following systems in terms of their linearity, time invariance, memory, causality, and stability.

(a) $y[n] = 3^n x[n]$ **(b)** $y[n] = e^{jn\pi} x[n]$
(c) $y[n] = \cos(0.5n\pi)x[n]$ **(d)** $y[n] = [1 + \cos(0.5n\pi)]x[n]$
(e) $y[n] = e^{x[n]}$ **(f)** $y[n] = x[n] + \cos[0.5(n+1)\pi]$

5.17 (System Classification) Classify the following systems in terms of their linearity, time invariance, memory, causality, and stability.

(a) $y[n] = x[n/3]$ (zero interpolation)
(b) $y[n] = \cos(n\pi)x[n]$ (modulation)
(c) $y[n] = [1 + \cos(n\pi)]x[n]$ (modulation)
(d) $y[n] = \cos(n\pi x[n])$ (frequency modulation)
(e) $y[n] = \cos(n\pi + x[n])$ (phase modulation)
(f) $y[n] = x[n] - x[n-1]$ (differencing operation)
(g) $y[n] = 0.5x[n] + 0.5x[n-1]$ (averaging operation)
(h) $y[n] = \frac{1}{N} \sum_{k=0}^{N-1} x[n-k]$ (moving average)
(i) $y[n] - \alpha y[n-1] = \alpha x[n], \quad 0 < \alpha < 1$ (exponential averaging)
(j) $y[n] = 0.4(y[n-1] + 2) + x[n]$

5.18 (Classification) Classify each system in terms of its linearity, time invariance, memory, causality, and stability.

(a) The folding system $y[n] = x[-n]$.
(b) The decimating system $y[n] = x[2n]$.
(c) The zero-interpolating system $y[n] = x[n/2]$.
(d) The sign-inversion system $y[n] = \text{sgn}\{x[n]\}$.
(e) The rectifying system $y[n] = |x[n]|$.

5.19 (Classification) Classify each system in terms of its linearity, time invariance, causality, and stability.

(a) $y[n] = \text{round}\{x[n]\}$ **(b)** $y[n] = \text{median}\{x[n+1], x[n], x[n-1]\}$
(c) $y[n] = x[n]\,\text{sgn}(n)$ **(d)** $y[n] = x[n]\,\text{sgn}\{x[n]\}$

5.20 (Inverse Systems) Are the following systems invertible? If not, explain why; if invertible, find the inverse system.

(a) $y[n] = x[n] - x[n-1]$ (differencing operation)
(b) $y[n] = \frac{1}{3}(x[n] + x[n-1] + x[n-2])$ (moving average operation)
(c) $y[n] = 0.5x[n] + x[n-1] + 0.5x[n-2]$ (weighted moving average operation)
(d) $y[n] - \alpha y[n-1] = (1-\alpha)x[n], \quad 0 < \alpha < 1$ (exponential averaging operation)
(e) $y[n] = \cos(n\pi)x[n]$ (modulation)
(f) $y[n] = \cos(x[n])$
(g) $y[n] = e^{x[n]}$

PGR2026-00033
Krisp Technologies EX1024 Page 137

5.21 (An Echo System and Its Inverse) An echo system is described by $y[n] = x[n] + 0.5x[n-N]$. Assume that the echo arrives after 1 ms and the sampling rate is 2 kHz.

(a) What is the value of N? Sketch a realization of this echo system.
(b) What is the impulse response and step response of this echo system?
(c) Find the difference equation of the inverse system. Then, sketch its realization and find its impulse response and step response.

5.22 (Reverb) A reverb filter is described by $y[n] = x[n] + 0.25y[n-N]$. Assume that the echoes arrive every millisecond and the sampling rate is 2 kHz.

(a) What is the value of N? Sketch a realization of this reverb filter.
(b) What is the impulse response and step response of this reverb filter?
(c) Find the difference equation of the inverse system. Then, sketch its realization and find its impulse response and step response.

5.23 (System Response) Consider the system $y[n] - 0.5y[n-1] = x[n]$. Find its zero-state response to the following inputs.

(a) $x[n] = u[n]$ **(b)** $x[n] = (0.5)^n u[n]$ **(c)** $x[n] = \cos(0.5n\pi)u[n]$
(d) $x[n] = (-1)^n u[n]$ **(e)** $x[n] = j^n u[n]$ **(f)** $x[n] = (\sqrt{j})^n u[n] + (\sqrt{j})^{-n} u[n]$

5.24 (System Response) For the system realization shown in Figure P5.24, find the response to the following inputs and initial conditions.

(a) $x[n] = u[n]$ $y[-1] = 0$ **(b)** $x[n] = u[n]$ $y[-1] = 4$
(c) $x[n] = (0.5)^n u[n]$ $y[-1] = 0$ **(d)** $x[n] = (0.5)^n u[n]$ $y[-1] = 6$
(e) $x[n] = (-0.5)^n u[n]$ $y[-1] = 0$ **(f)** $x[n] = (-0.5)^n u[n]$ $y[-1] = -2$

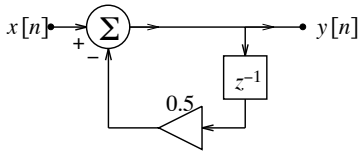


Figure P5.24 System realization for Problem 5.24

5.25 (System Response) Find the response of the following systems.

(a) $y[n] - 0.4y[n-1] = 2(0.5)^{n-1}u[n-1]$ $y[-1] = 0$
(b) $y[n] - 0.4y[n-1] = (0.4)^n u[n] + 2(0.5)^{n-1}u[n-1]$ $y[-1] = 2.5$
(c) $y[n] - 0.4y[n-1] = n(0.5)^n u[n] + 2(0.5)^{n-1}u[n-1]$ $y[-1] = 2.5$

5.26 (System Response) Find the impulse response of the following filters.

(a) $y[n] = x[n] - x[n-1]$ (differencing operation)
(b) $y[n] = 0.5x[n] + 0.5x[n-1]$ (averaging operation)
(c) $y[n] = \frac{1}{N}\sum_{k=0}^{N-1} x[n-k],\ N = 3$ (moving average)
(d) $y[n] = \frac{2}{N(N+1)}\sum_{k=0}^{N-1} (N-k)x[n-k],\ N = 3$ (weighted moving average)
(e) $y[n] - \alpha y[n-1] = (1-\alpha)x[n],\ N = 3,\ \alpha = \frac{N-1}{N+1}$ (exponential averaging)

PGR2026-00033
Krisp Technologies EX1024 Page 138

5.27 (System Response) It is known that the response of the system $y[n] + \alpha y[n-1] = x[n],\ \alpha \neq 0$, is given by $y[n] = [5 + 3(0.5)^n]u[n]$.

(a) Identify the natural response and forced response.
(b) Identify the values of α and $y[-1]$.
(c) Identify the zero-input response and zero-state response.
(d) Identify the input $x[n]$.

5.28 (System Response) It is known that the response of the system $y[n]+0.5y[n-1] = x[n]$ is described by $y[n] = [5(0.5)^n + 3(-0.5)^n)]u[n]$.

(a) Identify the zero-input response and zero-state response.
(b) What is the zero-input response of the system $y[n] + 0.5y[n-1] = x[n]$ if $y[-1] = 10$?
(c) What is the response of the relaxed system $y[n] + 0.5y[n-1] = x[n-2]$?
(d) What is the response of the relaxed system $y[n] + 0.5y[n-1] = x[n-1] + 2x[n]$?

5.29 (System Response) It is known that the response of the system $y[n] + \alpha y[n-1] = x[n]$ is described by $y[n] = (5 + 2n)(0.5)^n u[n]$.

(a) Identify the zero-input response and zero-state response.
(b) What is the zero-input response of the system $y[n] + \alpha y[n-1] = x[n]$ if $y[-1] = 10$?
(c) What is the response of the relaxed system $y[n] + \alpha y[n-1] = x[n-1])$?
(d) What is the response of the relaxed system $y[n] + \alpha y[n-1] = 2x[n-1] + x[n]$?
(e) What is the complete response of the $y[n] + \alpha y[n-1] = x[n] + 2x[n-1]$ if $y[-1] = 4$?

5.30 (System Interconnections) Two systems are said to be in cascade if the output of the first system acts as the input to the second. Find the response of the following cascaded systems if the input is a unit step and the systems are described as follows. In which instances does the response differ when the order of cascading is reversed? Can you use this result to justify that the order in which the systems are cascaded does not matter in finding the overall response if both systems are LTI?

(a) System 1: $y[n] = x[n] - x[n-1]$ System 2: $y[n] = 0.5y[n-1] + x[n]$
(b) System 1: $y[n] = 0.5y[n-1] + x[n]$ System 2: $y[n] = x[n] - x[n-1]$
(c) System 1: $y[n] = x^2[n]$ System 2: $y[n] = 0.5y[n-1] + x[n]$
(d) System 1: $y[n] = 0.5y[n-1] + x[n]$ System 2: $y[n] = x^2[n]$

5.31 (Systems in Cascade and Parallel) Consider the realization of Figure P5.31.

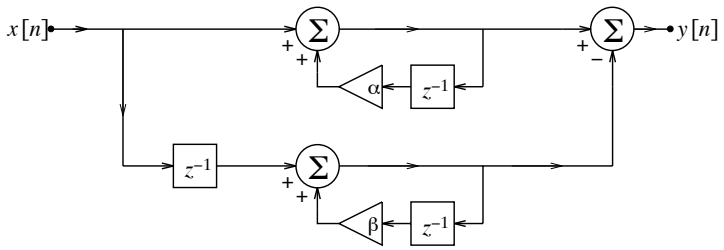


Figure P5.31 System realization for Problem 5.31

(a) Find its impulse response if $\alpha \neq \beta$. Is the overall system FIR or IIR?
(b) Find its difference equation and impulse response if $\alpha = \beta$. Is the overall system FIR or IIR?
(c) Find its difference equation and impulse response if $\alpha = \beta = 1$. What is the function of the overall system?

PGR2026-00033
Krisp Technologies EX1024 Page 139

5.32 (Difference Equations from Impulse Response) Find the difference equations describing the following systems.

(a) $h[n] = \delta[n] + 2\delta[n-1]$ **(b)** $h[n] = \{2, \underset{\Downarrow}{3}, -1\}$
(c) $h[n] = (0.3)^n u[n]$ **(d)** $h[n] = (0.5)^n u[n] - (-0.5)^n u[n]$

5.33 (Difference Equations from Impulse Response) A system is described by the impulse response $h[n] = (-1)^n u[n]$. Find the difference equation of this system. Then find the difference equation of the inverse system. Does the inverse system describe an FIR filter or IIR filter? What function does it perform?

5.34 (Difference Equations from Differential Equations) Consider an analog system described by the differential equation $y''(t) + 3y'(t) + 2y(t) = 2u(t)$.

(a) Confirm that this describes a stable analog system.
(b) Convert this to a difference equation using the backward Euler algorithm and check the stability of the resulting digital filter.
(c) Convert this to a difference equation using the forward Euler algorithm and check the stability of the resulting digital filter.
(d) Which algorithm is better in terms of preserving stability? Can the results be generalized to any arbitrary analog system?

5.35 (Difference Equations) For the filter realization shown in Figure P5.35, find the difference equation relating $y[n]$ and $x[n]$ if the impulse response of the filter is given by

(a) $h[n] = \delta[n] - \delta[n-1]$ **(b)** $h[n] = 0.5\delta[n] + 0.5\delta[n-1]$

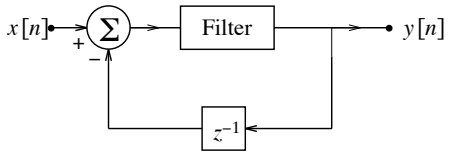


Figure P5.35 Filter realization for Problem 5.35

5.36 (Periodic Signal Generators) Find the difference equation of a filter whose impulse response is a periodic sequence with first period $x[n] = \{\underset{\Downarrow}{1}, 2, 3, 4, 6, 7, 8\}$. Sketch a realization for this filter.

5.37 (Recursive and IIR Filters) The terms *recursive* and *IIR* are not always synonymous. A recursive filter could in fact have a finite impulse response. Use recursion to find the the impulse response $h[n]$ for each of the following recursive filters. Which filters (if any) describe IIR filters?

(a) $y[n] - y[n-1] = x[n] - x[n-2]$
(b) $y[n] - y[n-1] = x[n] - x[n-1] - 2x[n-2] + 2x[n-3]$

5.38 (Recursive Forms of FIR Filters) An FIR filter may always be recast in recursive form by the simple expedient of including identical factors on the left-hand and right-hand side of its difference equation in operational form. For example, the filter $y[n] = (1 - z^{-1})x[n]$ is FIR, but the identical filter $(1+z^{-1})y[n] = (1+z^{-1})(1-z^{-1})x[n]$ has the difference equation $y[n]+y[n-1] = x[n]-x[n-2]$ and can be implemented recursively. Find two different recursive difference equations (with different orders) for each of the following filters.

(a) $y[n] = x[n] - x[n-2]$ **(b)** $h[n] = \{1, \underset{\Downarrow}{2}, 1\}$

PGR2026-00033
Krisp Technologies EX1024 Page 140

5.39 (Nonrecursive Forms of IIR Filters) An FIR filter may always be exactly represented in recursive form, but we can only approximately represent an IIR filter by an FIR filter by truncating its impulse response to N terms. The larger the truncation index N, the better is the approximation. Consider the IIR filter described by $y[n]-0.8y[n-1]=x[n]$. Find its impulse response $h[n]$ and truncate it to three terms to obtain $h_3[n]$, the impulse response of the approximate FIR equivalent. Would you expect the greatest mismatch in the response of the two filters to identical inputs to occur for lower or higher values of n? Compare the step response of the two filters up to $n=6$ to justify your expectations.

5.40 (Nonlinear Systems) One way to solve nonlinear difference equations is by recursion. Consider the nonlinear difference equation $y[n]y[n-1]-0.5y^2[n-1]=0.5Au[n]$.

(a) What makes this system nonlinear?
(b) Using $y[-1]=2$, recursively obtain $y[0]$, $y[1]$, and $y[2]$.
(c) Use $A=2$, $A=4$, and $A=9$ in the results of part (b) to confirm that this system finds the square root of A.
(d) Repeat parts (b) and (c) with $y[-1]=1$ to check whether the choice of the initial condition affects system operation.

5.41 (LTI Concepts and Stability) Argue that neither of the following describes an LTI system. Then, explain how you might check for their stability and determine which of the systems are stable.

(a) $y[n]+2y[n-1]=x[n]+x^2[n]$ **(b)** $y[n]-0.5y[n-1]=nx[n]+x^2[n]$

5.42 (Response of Causal and Noncausal Systems) A difference equation may describe a causal or noncausal system depending on how the initial conditions are prescribed. Consider a first-order system governed by $y[n]+\alpha y[n-1]=x[n]$.

(a) With $y[n]=0, n<0$, this describes a causal system. Assume $y[-1]=0$ and find the first few terms $y[0], y[1], \ldots$ of the impulse response and step response, using recursion, and establish the general form for $y[n]$.
(b) With $y[n]=0, n>0$, we have a noncausal system. Assume $y[0]=0$ and rewrite the difference equation as $y[n-1]=\{-y[n]+x[n]\}/\alpha$ to find the first few terms $y[0]$, $y[-1]$, $y[-2], \ldots$ of the impulse response and step response, using recursion, and establish the general form for $y[n]$.

COMPUTATION AND DESIGN

dtsimgui A GUI for System Response Simulation

The graphical user interface `dtsimgui` allows you to simulate and visualize the response of discrete-time systems.
You can select the input signal, the system parameters, and initial conditions.
To explore this routine, type `dtsimgui` at the MATLAB prompt.

5.43 (Numerical Integration Algorithms) Numerical integration algorithms approximate the area $y[n]$ from $y[n-1]$ or $y[n-2]$ (one or more time steps away). Consider the following integration algorithms.

(a) $y[n]=y[n-1]+t_s x[n]$ (rectangular rule)
(b) $y[n]=y[n-1]+\frac{t_s}{2}(x[n]+x[n-1])$ (trapezoidal rule)

PGR2026-00033
Krisp Technologies EX1024 Page 141

(c) $y[n] = y[n-1] + \frac{t_s}{12}(5x[n] + 8x[n-1] - x[n-2])$ (Adams-Moulton rule)
(d) $y[n] = y[n-2] + \frac{t_s}{3}(x[n] + 4x[n-1] + x[n-2])$ (Simpson's rule)
(e) $y[n] = y[n-3] + \frac{3t_s}{8}(x[n] + 3x[n-1] + 3x[n-2] + x[n-3])$ (Simpson's three-eighths rule)

Use each of the rules to approximate the area of $x(t) = \text{sinc}(t),\ 0 \le t \le 3$, with $t_s = 0.1\,\text{s}$ and $t_s = 0.3\,\text{s}$, and compare with the expected result of 0.53309323761827. How does the choice of the time step t_s affect the results? Which algorithm yields the most accurate results?

5.44 (System Response) Use the MATLAB routine `filter` to obtain and plot the response of the filter described by $y[n] = 0.25(x[n] + x[n-1] + x[n-2] + x[n-3])$ to the following inputs and comment on your results.

(a) $x[n] = 1,\ 0 \le n \le 60$
(b) $x[n] = 0.1n,\ 0 \le n \le 60$
(c) $x[n] = \sin(0.1n\pi),\ 0 \le n \le 60$
(d) $x[n] = 0.1n + \sin(0.5n\pi),\ 0 \le n \le 60$
(e) $x[n] = \sum_{k=-\infty}^{\infty} \delta[n-5k],\ 0 \le n \le 60$
(f) $x[n] = \sum_{k=-\infty}^{\infty} \delta[n-4k],\ 0 \le n \le 60$

5.45 (System Response) Use the MATLAB routine `filter` to obtain and plot the response of the filter described by $y[n] - y[n-4] = 0.25(x[n] + x[n-1] + x[n-2] + x[n-3])$ to the following inputs and comment on your results.

(a) $x[n] = 1,\ 0 \le n \le 60$
(b) $x[n] = 0.1n,\ 0 \le n \le 60$
(c) $x[n] = \sin(0.1n\pi),\ 0 \le n \le 60$
(d) $x[n] = 0.1n + \sin(0.5n\pi),\ 0 \le n \le 60$
(e) $x[n] = \sum_{k=-\infty}^{\infty} \delta[n-5k],\ 0 \le n \le 60$
(f) $x[n] = \sum_{k=-\infty}^{\infty} \delta[n-4k],\ 0 \le n \le 60$

5.46 (System Response) Use MATLAB to obtain and plot the response of the following systems over the range $0 \le n \le 199$.

(a) $y[n] = x[n/3],\ x[n] = (0.9)^n u[n]$ (assume zero interpolation)
(b) $y[n] = \cos(0.2n\pi)x[n],\ x[n] = \cos(0.04n\pi)$ (modulation)
(c) $y[n] = [1 + \cos(0.2n\pi)]x[n],\ x[n] = \cos(0.04n\pi)$ (modulation)

5.47 (System Response) Use MATLAB to obtain and plot the response of the following filters, using direct commands (where possible) and also using the routine `filter`, and compare your results. Assume that the input is given by $x[n] = 0.1n + \sin(0.1n\pi),\ 0 \le n \le 60$. Comment on your results.

(a) $y[n] = \frac{1}{N}\sum_{k=0}^{N-1} x[n-k],\ N = 4$ (moving average)

PGR2026-00033
Krisp Technologies EX1024 Page 142

(b) $y[n] = \frac{2}{N(N+1)} \sum_{k=0}^{N-1} (N-k)x[n-k],\ N = 4$ (weighted moving average)

(c) $y[n] - \alpha y[n-1] = (1-\alpha)x[n],\ N = 4,\ \alpha = \frac{N-1}{N+1}$ (exponential average)

5.48 (System Response) Use MATLAB to obtain and plot the response of the following filters, using direct commands and using the routine `filter`, and compare your results. Use an input that consists of the sum of the signal $x[n] = 0.1n + \sin(0.1n\pi),\ 0 \le n \le 60$ and uniformly distributed random noise with a mean of 0. Comment on your results.

(a) $y[n] = \frac{1}{N} \sum_{k=0}^{N-1} x[n-k],\ N = 4$ (moving average)

(b) $y[n] = \frac{2}{N(N+1)} \sum_{k=0}^{N-1} (N-k)x[n-k],\ N = 4$ (weighted moving average)

(c) $y[n] - \alpha y[n-1] = (1-\alpha)x[n],\ N = 4,\ \alpha = \frac{N-1}{N+1}$ (exponential averaging)

5.49 (System Response) Use the MATLAB routine `filter` to obtain and plot the response of the following FIR filters. Assume that $x[n] = \sin(n\pi/8),\ 0 \le n \le 60$. Comment on your results. From the results, can you describe the the function of these filters?

(a) $y[n] = x[n] - x[n-1]$ (first difference)

(b) $y[n] = x[n] - 2x[n-1] + x[n-2]$ (second difference)

(c) $y[n] = \frac{1}{3}(x[n] + x[n-1] + x[n-2])$ (moving average)

(d) $y[n] = 0.5x[n] + x[n-1] + 0.5x[n-2]$ (weighted average)

5.50 (System Response in Symbolic Form) The ADSP routine `sysresp1` returns the system response in symbolic form. See Chapter 21 for examples of its usage. Obtain the response of the following filters and plot the response for $0 \le n \le 30$.

(a) The step response of $y[n] - 0.5y[n] = x[n]$

(b) The impulse response of $y[n] - 0.5y[n] = x[n]$

(c) The zero-state response of $y[n] - 0.5y[n] = (0.5)^n u[n]$

(d) The complete response of $y[n] - 0.5y[n] = (0.5)^n u[n],\ y[-1] = -4$

(e) The complete response of $y[n] + y[n-1] + 0.5y[n-2] = (0.5)^n u[n],\ y[-1] = -4,\ y[-2] = 3$

5.51 (Inverse Systems and Echo Cancellation) A signal $x(t)$ is passed through the echo-generating system $y(t) = x(t) + 0.9x(t-\tau) + 0.8x(t-2\tau)$, with $\tau = 93.75$ ms. The resulting echo signal $y(t)$ is sampled at $S = 8192$ Hz to obtain the sampled signal $y[n]$.

(a) The difference equation of a digital filter that generates the output $y[n]$ from $x[n]$ may be written as $y[n] = x[n] + 0.9x[n-N] + 0.8x[n-2N]$. What is the value of the index N?

(b) What is the difference equation of an echo-canceling filter (inverse filter) that could be used to recover the input signal $x[n]$?

(c) The echo signal is supplied as `echosig.mat`. Load this signal into MATLAB (using the command `load echosig`). Listen to this signal using the MATLAB command `sound`. Can you hear the echoes? Can you make out what is being said?

(d) Filter the echo signal using your inverse filter and listen to the filtered signal. Have you removed the echoes? Can you make out what is being said? Do you agree with what is being said? If so, please thank Prof. Tim Schulz (`http://www.ee.mtu.edu/faculty/schulz`) for this problem.

PGR2026-00033
Krisp Technologies EX1024 Page 143

Chapter 6

CONTINUOUS CONVOLUTION

6.0 Scope and Objectives

In the time domain, convolution may be regarded as a method of finding the zero-state response of a relaxed LTI system. This chapter describes the convolution operation and its properties, and establishes the key connections between time-domain and transformed-domain methods for signal and system analysis based on convolution. Frequency-domain viewpoints of convolution will be presented in later chapters.

6.1 Introduction

The convolution method for finding the *zero-state* response $y(t)$ of a system to an input $x(t)$ applies to linear time-invariant (LTI) systems. The system is assumed to be described by its impulse response $h(t)$. An informal way to establish a mathematical form for $y(t)$ is illustrated in Figure 6.1.

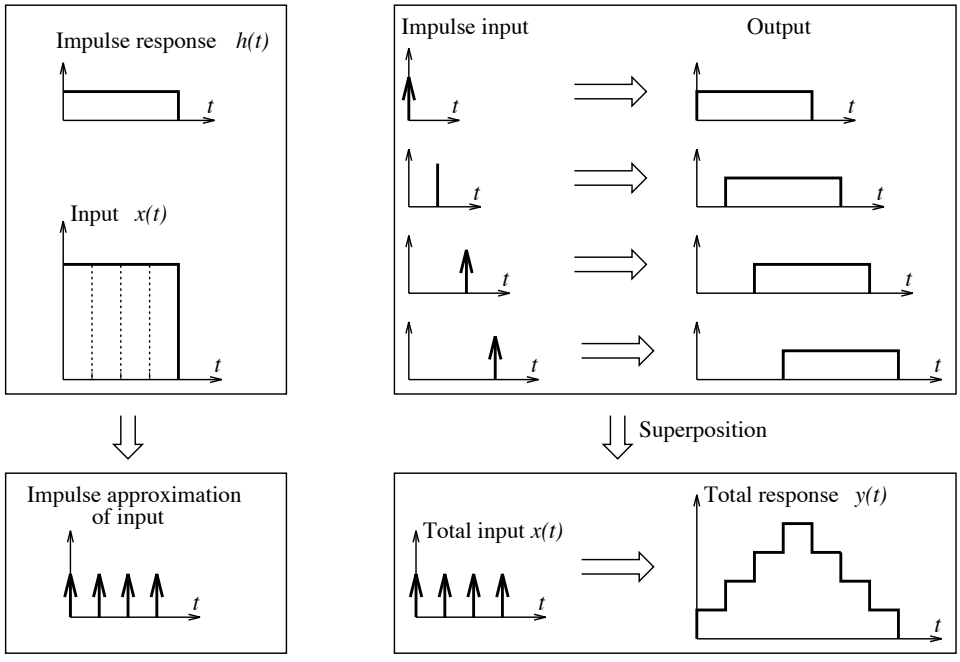


Figure 6.1 The process of convolution

PGR2026-00033
Krisp Technologies EX1024 Page 144

We divide $x(t)$ into narrow rectangular strips of width t_s at kt_s, $k = 0, \pm 1, \pm 2, \ldots$ and replace each strip by an impulse whose strength $t_s x(kt_s)$ equals the area under each strip:

$$x(t) \approx \sum_{k=-\infty}^{\infty} t_s x(kt_s)\delta(t - kt_s) \qquad \text{(sum of shifted impulses)} \tag{6.1}$$

Since $x(t)$ is a sum of weighted shifted impulses, the response $y(t)$, by superposition, is a sum of the weighted shifted impulse responses:

$$y(t) = \sum_{k=-\infty}^{\infty} t_s x(kt_s)h(t - kt_s) \qquad \text{(sum of shifted impulse responses)} \tag{6.2}$$

In the limit as $t_s \to d\lambda \to 0, kt_s$ describes a continuous variable λ, and both $x(t)$ and $y(t)$ may be represented by integral forms to give

$$x(t) = \int_{-\infty}^{\infty} x(\lambda)\delta(t - \lambda)\, d\lambda \qquad y(t) = \int_{-\infty}^{\infty} x(\lambda)h(t - \lambda)\, d\lambda \tag{6.3}$$

Note that the result for $x(t)$ is a direct consequence of the sifting property of impulses. The result

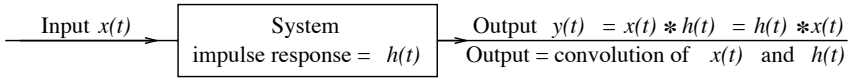
$$y(t) = x(t) \star h(t) = \int_{-\infty}^{\infty} x(\lambda)h(t - \lambda)\, d\lambda \tag{6.4}$$

describes the **convolution integral** for finding the zero-state response of a system. In this book, we use the shorthand notation $x(t) \star h(t)$ to describe the convolution of the signals $x(t)$ and $h(t)$.

Notation: We use $x(t) \star h(t)$ (or $x(t) * h(t)$ in figures) as a shorthand notation for $\displaystyle\int_{-\infty}^{\infty} x(\lambda)h(t - \lambda)\, d\lambda$

REVIEW PANEL 6.1

Convolution Yields the Zero-State Response of an LTI System



6.1.1 The Convolution Process

In the convolution integral, the time t determines the relative location of $h(t - \lambda)$ with respect to $x(\lambda)$. The convolution will yield a nonzero result only for those values of t over which $h(t - \lambda)$ and $x(\lambda)$ overlap. The response $y(t)$ for all time requires the convolution for every value of t. We must evaluate the area of the product $x(\lambda)h(t - \lambda)$ as t varies. Varying t amounts to sliding the folded function $h(-\lambda)$ past $x(\lambda)$ by the chosen values of t, as shown in Figure 6.2.

The impulse response $h(t)$ of many physical systems decays to zero for large t. At any instant t, the response $y(t)$ may be regarded as the *cumulative interaction* of the input $x(\lambda)$ with the folded impulse response $h(t-\lambda)$. Due to the folding, later values of the input are multiplied or weighted by correspondingly earlier (and larger) values of the impulse response and summed to yield $y(t)$. The response depends on both the present and past values of the input and impulse response. Inputs in the more recent past are weighted more heavily by the (folded) impulse response than less recent inputs, but future values of the input (or impulse response) have absolutely no influence on the present or past response.

PGR2026-00033
Krisp Technologies EX1024 Page 145

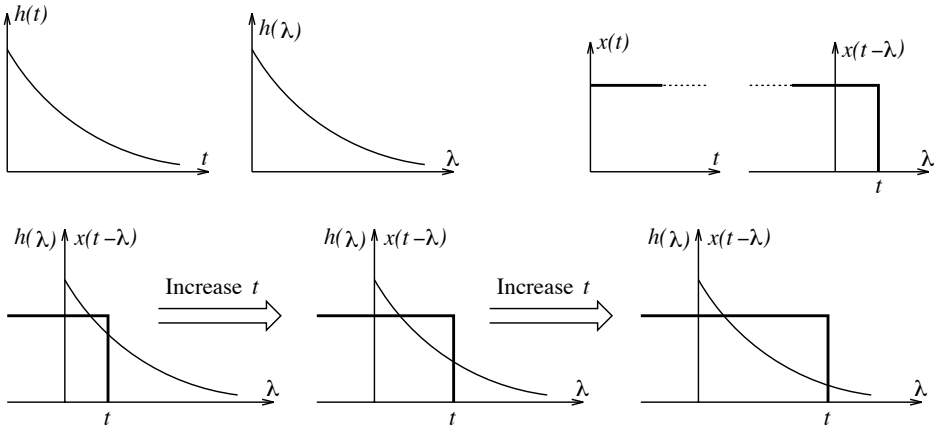


Figure 6.2 Convolution as a process of sliding a folded signal past another

Apart from its physical significance, the convolution integral is just another mathematical operation. It takes only a change of variable $\xi = t - \lambda$ to show that

$$x(t) \star h(t) = \int_{-\infty}^{\infty} x(\lambda)h(t-\lambda)\,d\lambda = -\int_{\infty}^{-\infty} x(t-\xi)h(\xi)\,d\xi = \int_{-\infty}^{\infty} x(t-\xi)h(\xi)\,d\xi = h(t) \star x(t) \qquad (6.5)$$

This is the *commutative property*, one where the order is unimportant. It says that, at least mathematically, we can switch the roles of the input and the impulse response for any system.

REVIEW PANEL 6.2

Convolution May Be Performed in Any Order

$$y(t) = x(t) \star h(t) = h(t) \star x(t) \qquad y(t) = \int_{-\infty}^{\infty} x(\lambda)h(t-\lambda)\,d\lambda = \int_{-\infty}^{\infty} h(\lambda)x(t-\lambda)\,d\lambda$$

For two causal signals $x(t)u(t)$ and $h(t)u(t)$, the product $x(\lambda)u(\lambda)h(t-\lambda)u(t-\lambda)$ is nonzero only over the range $0 \le \lambda \le t$ (because $u(\lambda)$ is zero for $\lambda < 0$ and $u(t-\lambda)$ is a left-sided step, which is zero for $\lambda > t$). Since both $u(\lambda)$ and $u(t-\lambda)$ are unity in this range, the convolution integral simplifies to

$$y(t) = \int_0^t x(\lambda)h(t-\lambda)\,d\lambda, \qquad x(t) \text{ and } h(t) \text{ zero for } t < 0 \qquad (6.6)$$

This result generalizes to the fact that the convolution of two right-sided signals is also right-sided and the convolution of two left-sided signals is also left-sided.

REVIEW PANEL 6.3

Sketching $u(\lambda)$ and $u(t-\lambda)$ Versus λ



PGR2026-00033
Krisp Technologies EX1024 Page 146

6.2 Convolution of Some Common Signals

The convolution of any signal $h(t)$ with an impulse reproduces the signal $h(t)$. The sifting property gives

$$\delta(t) \star h(t) = h(t) \tag{6.7}$$

This is simply another way of describing $h(t)$ as the impulse response of a system. With $h(t) = \delta(t)$, we have the less obvious result $\delta(t) \star \delta(t) = \delta(t)$. These two results are illustrated in Figure 6.3.

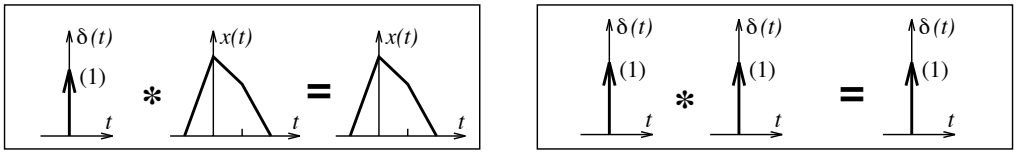


Figure 6.3 Convolution of a signal with an impulse replicates the signal

Convolution is a linear operation and obeys superposition. It is also a time-invariant operation and implies that shifting the input (or the impulse response) by α shifts the output (the convolution) by α.

REVIEW PANEL 6.4

Basic Concepts in Convolution

From time invariance: Shift $x(t)$ *or* $h(t)$ by α and you shift $y(t)$ by α.

From linearity: Superposition applies. $[x_1(t) + x_2(t)] \star h(t) = x_1(t) \star h(t) + x_2(t) \star h(t)$

Convolution with an impulse replicates the signal: $x(t) \star \delta(t) = x(t)$

EXAMPLE 6.1 (Convolution with Impulses)

Let $x(t) = 4\,\mathrm{tri}(t/2)$ and $h(t) = \delta(t+1) + \delta(t-1)$. Sketch their convolution $y(t) = x(t) \star h(t)$ and their product $p(t) = x(t)h(t)$ and find the area of each result.

The two signals and their convolution and product are sketched in Figure E6.1. The convolution is found by shifting the triangular pulse 1 unit to the left and right and using superposition. The area of the convolution $y(t)$ equals 16. To sketch the product $p(t)$, we use $\int x(t)\delta(t-\alpha) = x(\alpha)\delta(t-\alpha)$ (the product property of impulses). The area of $p(t)$ equals 4.

Comment: Note that convolution of a signal with an impulse replicates the signal, whereas the product of a signal with an impulse replicates the impulse.

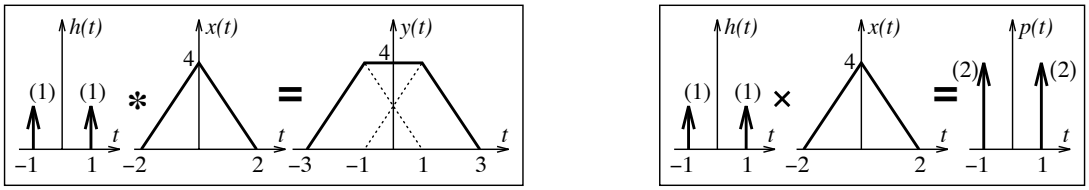


Figure E6.1 The signals of Example 6.1 and their convolution and product

PGR2026-00033
Krisp Technologies EX1024 Page 147

6.2.1 Analytical Evaluation of Convolution

Convolution is an integral operation that may be evaluated analytically, graphically, or numerically. The result depends on the nature of the signals being convolved. For example, we should expect to see a linear form if both signals are piecewise constant or a quadratic form if one is linear and the other is piecewise constant. In analytical convolution, we describe $x(t)$ and $h(t)$ by expressions using step functions, set up the forms for $x(\lambda)$ and $h(t-\lambda)$, and integrate their product.

While evaluating the integral, keep in mind that $x(\lambda)$ and $h(t-\lambda)$ are functions of λ (and not t), but t is a constant with respect to λ. Simplification of the integration limits due to the step functions in the convolution kernel occurs frequently in problem solving. For piecewise signals, it is best to break up and solve each integral separately because the limits on each integral will determine the range of of each result.

REVIEW PANEL 6.5

Concepts in Analytical Convolution: Express $x(t)$ and $h(t)$ Using Step Functions

The step function in $x(\lambda)$ changes the *lower* limit; the step in $h(t-\lambda)$ changes the *upper* limit.
Example: $u(\lambda-1)$ changes *lower* limit to 1 and $u(t-\lambda+3)$ changes *upper* limit to $t+3$.

Then: $\displaystyle\int_{-\infty}^{\infty}(.)u(\lambda-1)u(t-\lambda+3)\,d\lambda = \int_{1}^{t+3}(.)\,d\lambda = (\text{result of integration})u(t+2)$

A function of t is *constant* with respect to λ and can be pulled out of the convolution integral.

EXAMPLE 6.2 (Analytical Convolution)

(a) Let $x(t)=e^{-2t}u(t)$ and $h(t)=e^{-t}u(t)$. Then $x(\lambda)=e^{-2\lambda}u(\lambda)$ and $h(t-\lambda)=e^{-(t-\lambda)}u(t-\lambda)$. Since $u(\lambda)=0,\ \lambda<0$, and $u(t-\lambda)=0,\ \lambda>t$, we may write

$$y(t)=x(t)\star h(t)=\int_{-\infty}^{\infty}e^{-2\lambda}e^{-(t-\lambda)}u(\lambda)u(t-\lambda)\,d\lambda=e^{-t}\int_{0}^{t}e^{-\lambda}\,d\lambda=e^{-t}(1-e^{-t})=e^{-t}-e^{-2t},\quad t\geq 0$$

Comment: Note how we pulled e^{-t} out of the first integral because it was not a function of λ.

(b) Let $x(t)=e^{-\alpha t}u(t+3)$ and $h(t)=e^{-\alpha t}u(t-1)$. Then $h(\lambda)=e^{-\alpha\lambda}u(\lambda+3)$ and $x(t-\lambda)=e^{-\alpha(t-\lambda)}u(t-\lambda-1)$. Since $u(\lambda+3)=0,\ \lambda<-3$, and $u(t-\lambda-1)=0,\ \lambda>t-1$, we obtain

$$y(t)=\int_{-\infty}^{\infty}e^{-\alpha\lambda}u(\lambda+1)e^{-\alpha(t-\lambda)}u(t-\lambda-1)\,d\lambda=\int_{-3}^{t-1}e^{-\alpha\lambda}e^{-\alpha(t-\lambda)}\,d\lambda$$

Since $e^{-\alpha t}$ is not a function of λ, we can pull it out of the integral to get

$$y(t)=e^{-\alpha t}\int_{-3}^{t-1}d\lambda=(t+2)e^{-\alpha t},\quad t-1\geq -3 \qquad \text{or} \qquad y(t)=(t+2)e^{-\alpha t}u(t+2)$$

(c) Consider the convolution of $x(t)=u(t+1)-u(t-1)$ with itself. Changing the arguments to $x(\lambda)$ and $x(t-\lambda)$ results in the convolution

$$y(t)=\int_{-\infty}^{\infty}[u(\lambda+1)-u(\lambda-1)][u(t-\lambda+1)-u(t-\lambda-1)]\,d\lambda$$

Since $u(t-\lambda+1)=0,\ \lambda<t+1$, and $u(t-\lambda-1)=0,\ \lambda>t-1$, the integration limits for the four integrals can be simplified and result in

$$y(t)=\int_{-1}^{t+1}d\lambda-\int_{-1}^{t-1}d\lambda-\int_{1}^{t+1}d\lambda+\int_{1}^{t-1}d\lambda$$

PGR2026-00033
Krisp Technologies EX1024 Page 148

Evaluating each integral separately, we obtain

$$\int_{-1}^{t+1} d\lambda = t+2,\ t \geq -2 \qquad \int_{-1}^{t-1} d\lambda = t,\ t \geq 0 \qquad \int_{1}^{t+1} d\lambda = t,\ t \geq 0 \qquad \int_{1}^{t-1} d\lambda = t-2,\ t \geq 2$$

Based on each result and its range, we can express the convolution $y(t)$ as

$$y(t) = (t+2)u(t+2) - tu(t) - tu(t) + (t-2)u(t-2) = r(t+2) - 2r(t) + r(t-2)$$

We can also express $y(t)$ by ranges, if required.

Caution: Had we simply added the results without regard to their range, we would have obtained $y(t) = (t+2) - t - t + (t-2) = 0$. This result holds only for $t \geq 2$ when all terms contribute to $y(t)$.

6.3 Some Properties of Convolution

Example 6.2 reveals the following important results and properties that serve as consistency checks while finding the convolution $y(t)$ of any arbitrary signals $x(t)$ and $h(t)$.

Duration and Area The starting time of $y(t)$ equals the sum of the starting times of $x(t)$ and $h(t)$. The ending time of $y(t)$ equals the sum of the ending times of $x(t)$ and $h(t)$. The duration of $y(t)$ equals the sum of the durations of $x(t)$ and $h(t)$. The area of $y(t)$ equals the product of the areas of $x(t)$ and $h(t)$. A formal proof involves interchanging the order of integration and noting that areas are shift invariant. We have

$$\begin{aligned}\int_{-\infty}^{\infty} y(t)\,dt &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x(\lambda)h(t-\lambda)\,d\lambda\,dt \\ &= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} h(t-\lambda)\,dt\right] x(\lambda)\,d\lambda = \int_{-\infty}^{\infty} h(t)\,dt \int_{-\infty}^{\infty} x(t)\,dt\end{aligned}$$

Properties Based on Linearity A linear operation on the input to a system results in a similar operation on the response. Thus, the input $x'(t)$ results in the response $y'(t)$, and we have $x'(t) \star h(t) = y'(t)$. In fact, the derivative of any one of the convolved signals results in the derivative of the convolution. Repeated derivatives of either $x(t)$ or $h(t)$ lead to the general result

$$x^{(m)}(t) \star h(t) = x(t) \star h^{(m)}(t) = y^{(m)}(t) \qquad x^{(m)}(t) \star h^{(n)}(t) = y^{(m+n)}(t) \tag{6.8}$$

Integration of the input to a system results in integration of the response. The step response thus equals the running integral of the impulse response. More generally, the convolution $x(t) \star u(t)$ equals the *running integral* of $x(t)$ because

$$x(t) \star u(t) = \int_{-\infty}^{\infty} x(\lambda)u(t-\lambda)\,d\lambda = \int_{-\infty}^{t} x(\lambda)\,d\lambda \tag{6.9}$$

Properties Based on Time Invariance If the input to a system is shifted by α, so too is the response. In other words, $x(t-\alpha) \star h(t) = y(t-\alpha)$. In fact, shifting any one of the convolved signals by α shifts the convolution by α. If both $x(t)$ and $h(t)$ are shifted, we can use this property in succession to obtain

$$x(t-\alpha) \star h(t-\beta) = y(t-\alpha-\beta) \tag{6.10}$$

The concepts of linearity and shift invariance lie at the heart of many other properties of convolution.

PGR2026-00033
Krisp Technologies EX1024 Page 149

Time Scaling If both $x(t)$ and $h(t)$ are scaled by α to $x(\alpha t)$ and $h(\alpha t)$, the duration property suggests that the convolution $y(t)$ is also scaled by α. In fact, $x(\alpha t) \star h(\alpha t) = |\frac{1}{\alpha}| y(\alpha t)$, where the scale factor $|\frac{1}{\alpha}|$ is required to satisfy the area property. The time-scaling property is valid only when *both functions are scaled by the same factor.*

Symmetry If both signals are folded ($\alpha = -1$), so is their convolution. As a consequence of this, the convolution of an odd symmetric and an even symmetric signal is odd symmetric, whereas the convolution of two even symmetric (or two odd symmetric) signals is even symmetric. Interestingly, the convolution of $x(t)$ with its folded version $x(-t)$ is also even symmetric, with a maximum at $t = 0$. The convolution $x(t) \star x(-t)$ is called the **autocorrelation** of $x(t)$ and is discussed later in this chapter.

REVIEW PANEL 6.6

Useful Properties of Convolution

Shift: $x(t-\alpha) \star h(t) = y(t-\alpha)$ **Derivative:** $x'(t) \star h(t) = y'(t)$ **Scale:** $x(\alpha t) \star h(\alpha t) = \frac{1}{|\alpha|} y(\alpha t)$

Convolution with impulses: $x(t) \star \delta(t) = x(t)$ **and steps:** $x(t) \star u(t) = \int_0^t x(\lambda)\, d\lambda$

Convolution is generally a smoothing and stretching operation (unless the convolved signals contain impulses or their derivatives).

6.3.1 Some Useful Convolution Results

Several useful convolution results are sketched in Figure 6.4. Here is how some of these results come about.

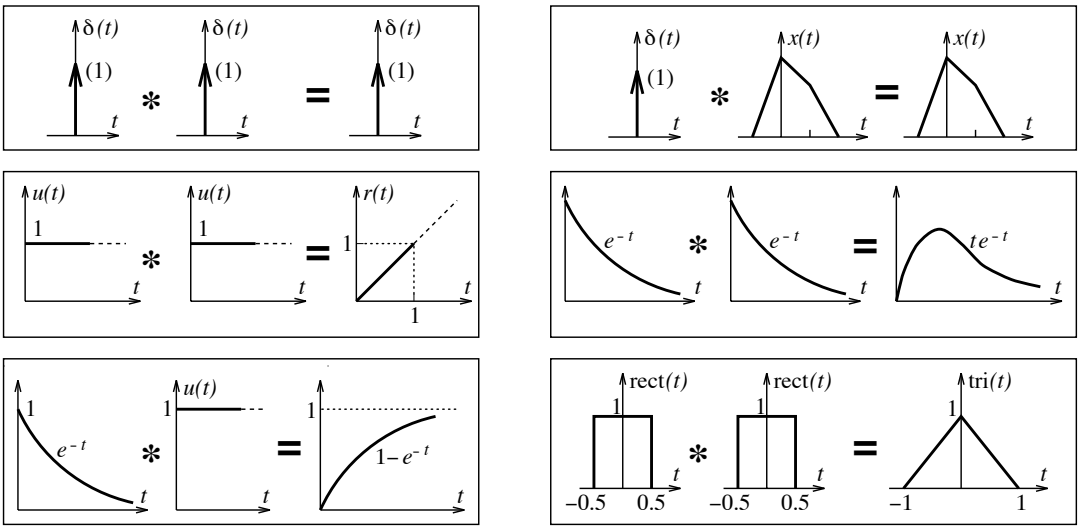


Figure 6.4 The convolution of some useful signals

1. $u(t) \star u(t) = \int_{-\infty}^{\infty} u(\lambda)u(t-\lambda)\, d\lambda = \int_0^t d\lambda = tu(t) = r(t)$

2. $e^{-t}u(t) \star e^{-t}u(t) = \int_{-\infty}^{\infty} e^{-\lambda}e^{-(t-\lambda)}u(\lambda)u(t-\lambda)\, d\lambda = e^{-t}\int_0^t d\lambda = te^{-t}u(t)$

PGR2026-00033
Krisp Technologies EX1024 Page 150

3. $u(t) \star e^{-t}u(t) = \displaystyle\int_{-\infty}^{\infty} u(t-\lambda)e^{-\lambda}u(\lambda)\,d\lambda = \int_0^t e^{-\lambda}\,d\lambda = (1-e^{-t})u(t)$

4. $\text{rect}(t) \star \text{rect}(t) = [u(t+0.5) - u(t-0.5)] \star [u(t+0.5) - u(t-0.5)] = r(t+1) - 2r(t) + r(t-1) = \text{tri}(t)$

REVIEW PANEL 6.7

Three Useful Convolution Results

$u(t) \star u(t) = r(t)$ $\qquad$ $\text{rect}(t) \star \text{rect}(t) = \text{tri}(t)$ $\qquad$ $e^{-\alpha t}u(t) \star e^{-\alpha t}u(t) = te^{-\alpha t}u(t)$

EXAMPLE 6.3 (Convolution Properties)

(a) The impulse response of an RC lowpass filter is $h(t) = e^{-t}u(t)$. Find its step response.

The response $s(t)$ to the step input $u(t)$ is the running integral of $h(t)$, and thus

$$s(t) = \int_{-\infty}^{t} e^{-\lambda}u(\lambda)\,d\lambda = (1-e^{-t})u(t)$$

(b) Using linearity, the convolution $y_r(t) = r(t) \star e^{-t}u(t) = u(t) \star u(t) \star e^{-t}u(t)$ is the running integral of the step response $s(t) = u(t) \star e^{-t}u(t)$ and equals

$$y_r(t) = \int_0^t s(t)\,dt = \int_0^t 1 - e^{-t}\,dt = r(t) - (1-e^{-t})u(t)$$

(c) Using shifting and superposition, the response $y_1(t)$ to the input $x(t) = u(t) - u(t-2)$ equals

$$y_1(t) = [1-e^{-t}]u(t) - [1-e^{-(t-2)}]u(t-2)$$

(d) Using the area property, the area of $y_1(t)$ equals $[\int e^{-t}\,dt][\int x(t)\,dt] = 2$.

Comment: Try integrating $y_1(t)$ directly at your own risk to arrive at the same answer!

(e) Starting with $e^{-t}u(t) \star e^{-t}u(t) = te^{-t}u(t)$, and using the scaling property and $u(\alpha t) = u(t)$,

$$e^{-\alpha t}u(\alpha t) \star e^{-\alpha t}u(\alpha t) = \frac{1}{\alpha}(\alpha t)e^{-\alpha t}u(t) = te^{-\alpha t}u(t)$$

(f) Starting with $u(t) \star e^{-t}u(t) = (1-e^{-t})u(t)$, and using the scaling property and $u(\alpha t) = u(t)$,

$$u(\alpha t) \star e^{-\alpha t}u(\alpha t) = \frac{1}{\alpha}(1-e^{-\alpha t})u(t)$$

(g) With $\alpha = -1$ in the scaling property, we obtain $u(-t) \star e^{t}u(-t) = (1-e^{t})u(-t)$.
This confirms that the convolution of left-sided signals is also left-sided.

PGR2026-00033
Krisp Technologies EX1024 Page 151

(h) Let $x(t) = u(t+3) - u(t-1)$ and $h(t) = u(t+1) - u(t-1)$.
Using superposition, the convolution $y(t) = x(t) \star x(t)$ may be described as

$$y(t) = u(t+3) \star u(t+1) - u(t+3) \star u(t-1) - u(t-1) \star u(t+1) + u(t-1) \star u(t-1)$$

Since $u(t) \star u(t) = r(t)$, we invoke time invariance for each term to get

$$y(t) = r(t+4) - r(t+2) - r(t) + r(t+2)$$

The signals and their convolution are shown in Figure E6.3H. The convolution $y(t)$ is a trapezoid extending from $t = -4$ to $t = 2$ whose duration is 6 units, whose starting time equals the sum of the starting times of $x(t)$ and $h(t)$, and whose area equals the product of the areas of $x(t)$ and $h(t)$.

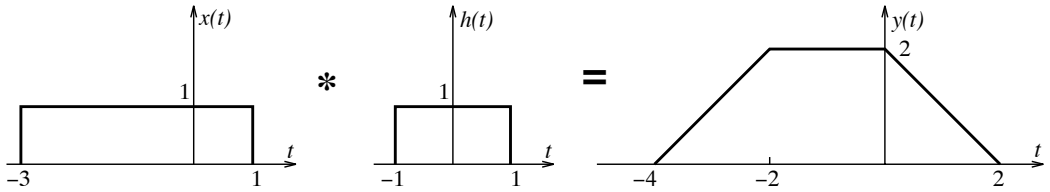


Figure E6.3H The signals for Example 6.3(h) and their convolution

6.4 Convolution by Ranges (Graphical Convolution)

Analytical convolution is tedious to express and sketch range by range. It is often easier to find such range-by-range results directly. In some cases, the results may even be obtained by graphical evaluation of areas without solving integrals, which leads to the name **graphical convolution**. In this approach, we describe $h(\lambda)$ and $x(t-\lambda)$ by intervals. As we increase t (slide $x(t-\lambda)$ to the right) from a position of no overlap with $x(t)$, the overlap region of $x(\lambda)$ and $h(t-\lambda)$ changes, as do the expressions for their product and the convolution (the area of their product).

The Pairwise Sum for Finding Convolution Ranges The mechanics of convolution by ranges may be viewed as a bookkeeping operation. As we fold and slide, we must keep track of changes in the ranges, integration limits, and so on. The choice of which function to fold and slide is arbitrary. But choosing the one with the simpler representation leads to simpler integrals. The key, however, is to be able to establish the correct ranges. Each range represents the largest duration over which the convolution is described by the same expression. A new range begins each time a range end point of the folded function slides past a range end point of the other. This is the basis for the following foolproof **pairwise sum rule** for obtaining the convolution ranges.

1. Set up two sequences containing the range end points of $x(t)$ and $h(t)$.
2. Form their pairwise sum (by summing each value from one sequence with the other).
3. Arrange the pairwise sum in increasing order and discard duplications.

The resulting sequence yields the end points for the convolution ranges. An example follows.

EXAMPLE 6.4 (The Pairwise Sum Rule)
The pairwise sum of the sequences $\{0, 1, 3\}$ and $\{-2, 0, 2\}$ gives $\{-2, 0, 2, -1, 1, 3, 1, 3, 5\}$.
The ordered sequence becomes $\{-2, -1, 0, 1, 1, 2, 3, 3, 5\}$.
Discarding duplications, we get $\{-2, -1, 0, 1, 2, 3, 5\}$. The ranges for nonzero convolution are then

$$-2 \le t \le -1 \qquad -1 \le t \le 0 \qquad 0 \le t \le 1 \qquad 1 \le t \le 2 \qquad 2 \le t \le 3 \qquad 3 \le t \le 5$$

PGR2026-00033
Krisp Technologies EX1024 Page 152

The Recipe for Convolution by Ranges is summarized in the following review panel. To sketch $x(\lambda)$ versus λ, simply relabel the axes. To sketch $x(t-\lambda)$ versus λ, fold $x(\lambda)$ and delay by t. For example, if the end points of $x(\lambda)$ are $(-4,\ 3)$, the end points of (the folded) $x(t-\lambda)$ will be $(t-3,\ t+4)$.

REVIEW PANEL 6.8

Recipe for Convolution by Ranges

1. Express $x(t)$ and $h(t)$ *by intervals* and find the convolution ranges using the **pairwise sum**.
2. For each range, locate $x(t-\lambda)$ relative to $h(\lambda)$. Keep the end points of $x(t-\lambda)$ in terms of t.
3. For each range, integrate $x(t-\lambda)h(\lambda)$ over their overlap duration to find the convolution.

EXAMPLE 6.5 (Convolution by Ranges)
Find the convolution of $x(t)$ and $h(t)$ sketched in Figure E6.5.

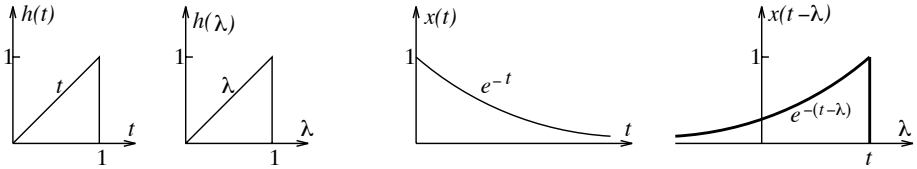
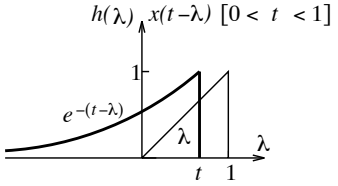
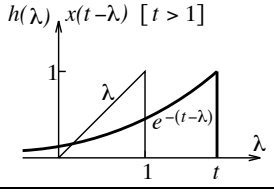


Figure E6.5 The signals for Example 6.5

The pairwise sum gives the end points of the convolution ranges as $[0,\ 1,\ \infty]$. For each range, we superpose $x(t-\lambda) = e^{-(t-\lambda)}u(t-\lambda)$ and $h(\lambda) = \lambda$ to obtain the following results:

Overlapping Signals	Range and Convolution $y(t)$	End Values
	Range: $0 \le t \le 1$ $\int_0^t \lambda e^{-(t-\lambda)}\,d\lambda = -1 + t + e^{-t}$	$\nearrow\ y(0) = 0$ $\searrow\ y(1) = e^{-1}$
	Range: $t \ge 1$ $\int_0^1 \lambda e^{-(t-\lambda)}\,d\lambda = e^{-t}\int_0^1 \lambda e^{\lambda}\,d\lambda = e^{-t}$	$\nearrow\ y(1) = e^{-1}$ $\searrow\ y(\infty) = 0$

We used the indefinite integral $\int \lambda e^{\lambda}\,d\lambda = (\lambda-1)e^{\lambda}$ to simplify the results. The convolution results match at the range end points. The convolution is plotted in Figure E6.5A.

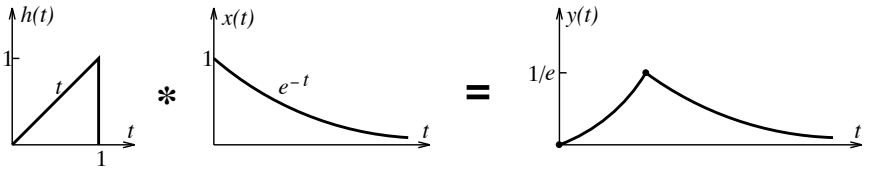


Figure E6.5A The convolution of the signals for Example 6.5

PGR2026-00033
Krisp Technologies EX1024 Page 153

EXAMPLE 6.6 (Convolution by Ranges)
Find the convolution of $x(t)$ and $h(t)$ sketched in Figure E6.6.

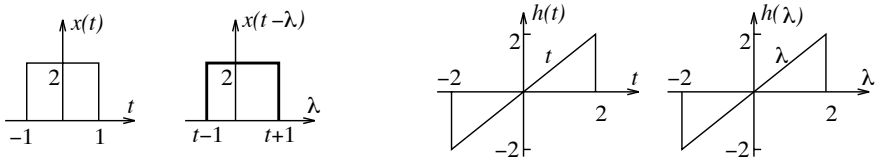
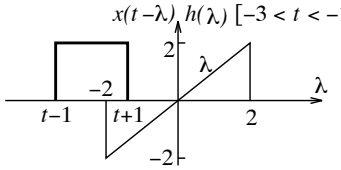
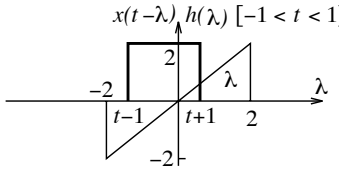
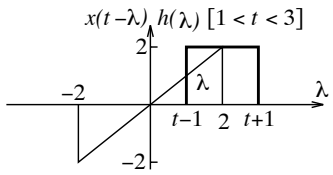


Figure E6.6 The signals for Example 6.6

The pairwise sum gives the end points of the convolution ranges as $[-3,\ -1,\ -1,\ 3]$. For each range, we superpose $x(t-\lambda) = 2,\ t-1 \le \lambda \le t+1$, and $h(\lambda) = \lambda,\ -2 \le \lambda \le 2$ to obtain the following results:

Overlapping Signals	**Range and Convolution $y(t)$**	**End Values**
	Range: $-3 \le t \le -1$ $\int_{-2}^{t+1} 2\lambda\, d\lambda = t^2 + 2t - 3$	$y(-3) = 0$ $y(-1) = -4$
	Range: $-1 \le t \le 1$ $\int_{t-1}^{t+1} 2\lambda\, d\lambda = 4t$	$y(-1) = -4$ $y(1) = 4$
	Range: $1 \le t \le 3$ $\int_{t-1}^{2} 2\lambda\, d\lambda = -t^2 + 2t + 3$	$y(1) = 4$ $y(3) = 0$

The convolution results match at the range end points and are plotted in Figure E6.6A.

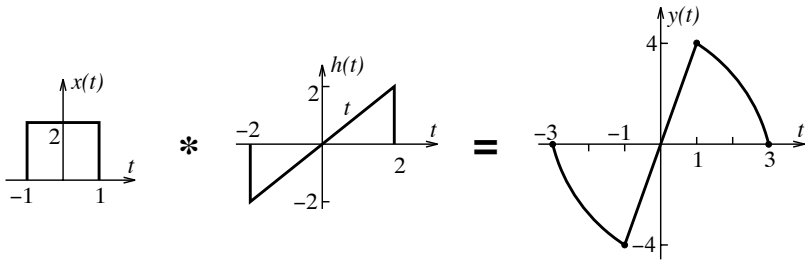


Figure E6.6A The convolution of the signals for Example 6.6

As a consistency check, note how the convolution results match at the end points of each range. Note that one of the convolved signals has even symmetry, the other has odd symmetry, and the convolution result has odd symmetry.

PGR2026-00033
Krisp Technologies EX1024 Page 154

EXAMPLE 6.7 (Convolution by Ranges)
Find the convolution of the signals $x(t)$ and $h(t)$ sketched in Figure E6.7.

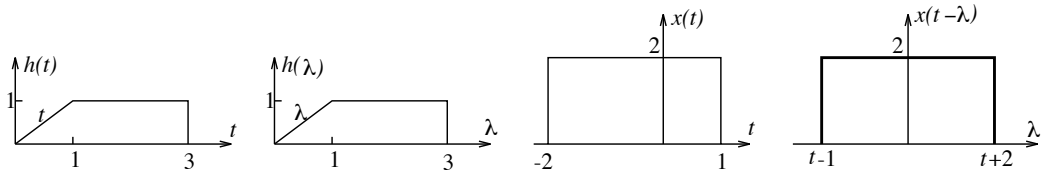
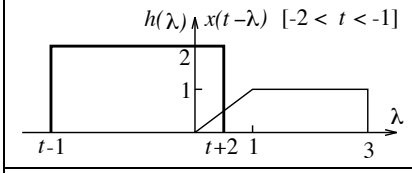
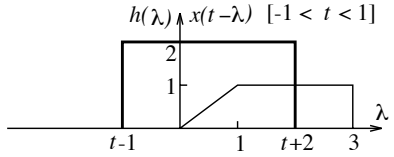
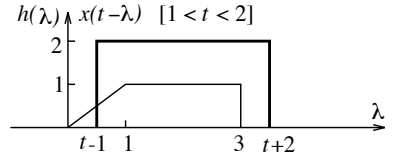
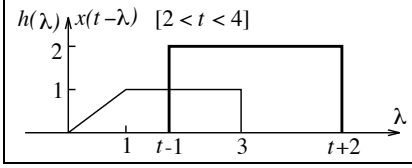


Figure E6.7 The signals for Example 6.7

The end points of $x(t)$ and $h(t)$ are $\{-2, 1\}$ and $\{0, 1, 3\}$. Their pairwise sum gives $\{-2, -1, 1, 1, 2, 4\}$. Discarding duplications, we get the range end points as $\{-2, -1, 1, 2, 4\}$. For each range, we superpose $x(t-\lambda) = 1,\ t-1 \le \lambda \le t+2$, and $h(\lambda) =$ to obtain the following results:

Overlapping Signals	**Range and Convolution $y(t)$**	**End Values**
	Range: $-2 \le t \le -1$ $\int_0^{t+2} 2\lambda\, d\lambda = (t+2)^2$	$y(-2) = 0$ $y(-1) = 1$
	Range: $-1 \le t \le 1$ $\int_0^1 2\lambda\, d\lambda + \int_1^{t+2} 2\, d\lambda = 2t+3$	$y(-1) = 1$ $y(1) = 5$
	Range: $1 \le t \le 2$ $\int_{t-1}^1 2\lambda\, d\lambda + \int_1^3 2\, d\lambda = -t^2 + 2t + 4$	$y(1) = 5$ $y(2) = 4$
	Range: $2 \le t \le 4$ $\int_{t-1}^3 2\lambda\, d\lambda = -2t + 8$	$y(2) = 4$ $y(4) = 0$

The convolution is plotted in Figure E6.7A. The convolution results match at the range end points. Since $x(t)$ is constant while $h(t)$ is piecewise linear, their convolution must yield only linear or quadratic forms. Our results also confirm this.

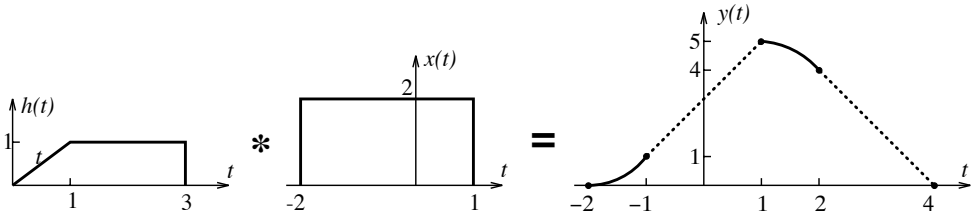


Figure E6.7A The convolution of the signals for Example 6.7

PGR2026-00033
Krisp Technologies EX1024 Page 155

REVIEW PANEL 6.9

Consistency Checks for the Convolution $y(t) = x(t) \star h(t)$

Start: sum of start times of $x(t)$ and $h(t)$. **Duration:** sum of durations of $x(t)$ and $h(t)$.
End: sum of end times of $x(t)$ and $h(t)$. **Area:** *product* of areas of $x(t)$ and $h(t)$.

6.4.1 Graphical Convolution of Piecewise Constant Signals

The convolution of *piecewise constant* signals is linear over every range. All we then need is to find the convolution at the end points of each range and connect the dots, so to speak. The convolution values can easily be calculated graphically. In particular, the convolution of two rectangular pulses of heights A and B and widths $W_{\min}$ and $W_{\max}$ is a trapezoid. You can mentally visualize this result by folding one of the pulses and sliding it past the other. The convolution increases from zero to a maximum in $W_{\min}$ units (partial increasing overlap), stays constant for $W_{\max} - W_{\min}$ units (total overlap), and decreases to zero in $W_{\min}$ units (partial decreasing overlap). The maximum convolution equals $ABW_{\min}$ (the area of the product during total overlap). The starting time of the convolution can be found from the sum of the starting times of the two pulses. Note that if the pulse widths are equal, the convolution is a triangular function.

REVIEW PANEL 6.10

Sketching the Convolution $y(t)$ **of Piecewise Constant Signals?**

1. Find the convolution values at the *end points* of each range. **2.** Connect by straight lines.

Rectangular pulses: $y(t)$ is a *trapezoid* for *unequal* pulse widths, a *triangle* for equal pulse widths.

EXAMPLE 6.8 (Convolution of Piecewise Constant Signals)
Consider the piecewise linear signals $x(t)$ and $h(t)$ and their convolution, shown in Figure E6.8.

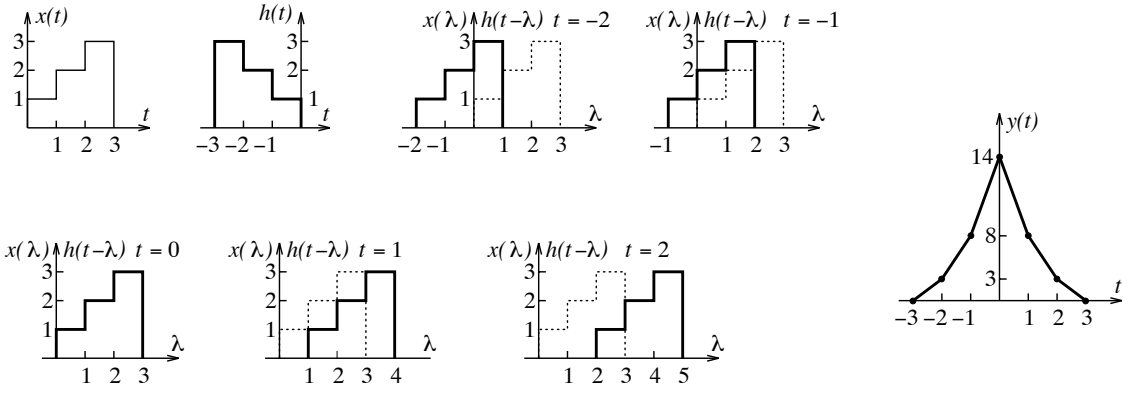


Figure E6.8 The signals for Example 6.8 and their convolution

The convolution starts at $t = -3$. The convolution ranges cover unit intervals up to $t = 3$. The area of $x(\lambda)h(t - \lambda)$ with t chosen for each end point yields the following results:

$$y(-3) = 0 \qquad y(-2) = 3 \qquad y(-1) = 8 \qquad y(0) = 14 \qquad y(1) = 8 \qquad y(2) = 3 \qquad y(3) = 0$$

Note that $h(t) = x(-t)$. The convolution $x(t) \star x(-t)$ is called the *autocorrelation* of $x(t)$ and is always even symmetric, with a maximum at the origin.

PGR2026-00033
Krisp Technologies EX1024 Page 156

6.4.2 Impulse Response of LTI Systems in Cascade and Parallel

Consider the ideal cascade (with no loading effects) of two LTI systems, as shown in Figure 6.5.

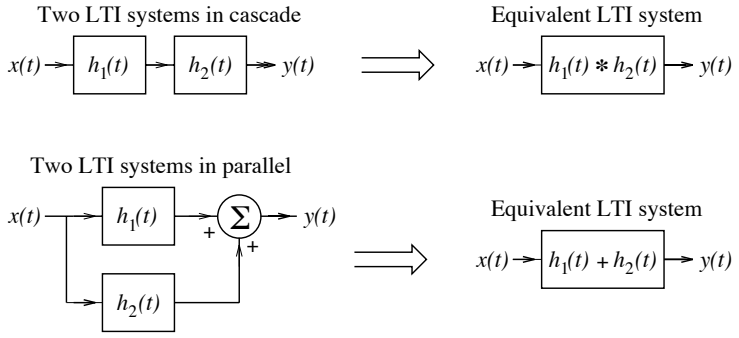


Figure 6.5 Cascaded and parallel systems and their equivalents

The response of the first system is $y_1(t) = x(t) \star h_1(t)$. The response $y(t)$ of the second system is

$$y(t) = y_1(t) \star h_2(t) = [x(t) \star h_1(t)] \star h_2(t) = x(t) \star [h_1(t) \star h_2(t)] \tag{6.11}$$

If we wish to replace the cascaded system by an equivalent LTI system with impulse response $h(t)$ such that $y(t) = x(t) \star h(t)$, it follows that $h(t) = h_1(t) \star h_2(t)$. Generalizing this result, the impulse response $h(t)$ of N ideally cascaded LTI systems is simply the convolution of the N individual impulse responses:

$$h(t) = h_1(t) \star h_2(t) \star \cdots \star h_N(t) \qquad \text{(for a cascade combination)} \tag{6.12}$$

If the $h_k(t)$ are energy signals, the order of cascading is unimportant. The overall impulse response of systems in parallel equals the sum of the individual impulse responses, as shown in Figure 6.5.

$$h_P(t) = h_1(t) + h_2(t) + \cdots + h_N(t) \qquad \text{(for a parallel combination)} \tag{6.13}$$

REVIEW PANEL 6.11

The Impulse Response of N Interconnected LTI Systems

In cascade: *Convolve* the impulse responses: $h_C(t) = h_1(t) \star h_2(t) \star \cdots \star h_N(t)$
In parallel: *Add* the impulse responses: $h_P(t) = h_1(t) + h_2(t) + \cdots + h_N(t)$

EXAMPLE 6.9 (Interconnected Systems)

(a) Refer to the system shown in Figure E6.9A. Find the impulse response of each subsystem, the overall impulse response, and the outputs $f(t)$ and $g(t)$.

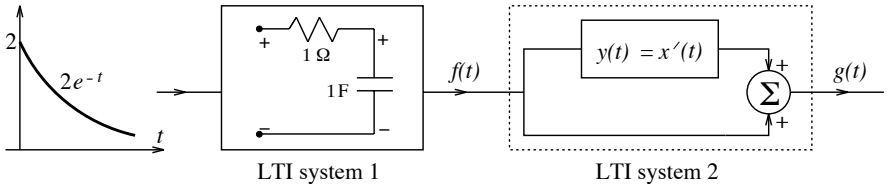


Figure E6.9A The interconnected system for Example 6.9(a)

PGR2026-00033
Krisp Technologies EX1024 Page 157

The time constant of the RC circuit is $\tau = 1$. Its impulse response is thus $h(t) = e^{-t}u(t)$. The input-output relation for the second system has the form $y_0(t) = x_0(t) + x_0'(t)$. Its impulse response is thus $h_2(t) = \delta(t) + \delta'(t)$.

The overall impulse response $h(t)$ is given by their convolution:

$$h(t) = e^{-t}u(t) \star [\delta(t) + \delta'(t)] = e^{-t}u(t) + \delta(t) - e^{-t}u(t) = \delta(t)$$

This means that the overall system output equals the applied input and the second system acts as the inverse of the first.

The output $g(t)$ is thus $g(t) = 2e^{-t}u(t)$.

The output $f(t)$ is given by the convolution $f(t) = 2e^{-t}u(t) \star e^{-t}u(t) = 2te^{-t}u(t)$.

(b) Refer to the cascaded system shown in Figure E6.9B. Will the outputs $g(t)$ and $w(t)$ be equal? Explain.

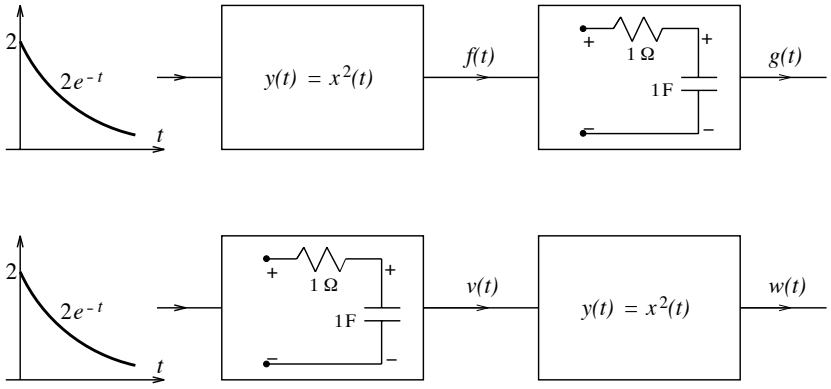


Figure E6.9B The cascaded system for Example 6.9(b)

The impulse response of the RC circuit is $h(t) = e^{-t}u(t)$. For the first system, the output $f(t)$ is $f(t) = 4e^{-2t}u(t)$. Using convolution, the output $g(t)$ is given by

$$g(t) = e^{-t}u(t) \star 4e^{-2t}u(t) = 4(e^{-t} - e^{-2t})u(t)$$

For the second system, the outputs $v(t)$ and $w(t)$ are

$$v(t) = e^{-t}u(t) \star 2e^{-t}u(t) = 2te^{-t}u(t) \qquad w(t) = v^2(t) = 4t^2e^{-2t}u(t)$$

Clearly, $w(t)$ and $g(t)$ are not equal. The reason is that the order of cascading is unimportant *only for* LTI *systems* and the squaring block is nonlinear.

6.5 Stability and Causality

In Chapter 4 we examined the stability of LTI systems described by differential equations. Bounded-input, bounded-output (BIBO) stability of such systems required every root of the characteristic equation to have negative real parts. For systems described by their impulse response, this actually translates to an equivalent condition that involves constraints on the impulse response $h(t)$.

PGR2026-00033
Krisp Technologies EX1024 Page 158

If $x(t)$ is bounded such that $|x(t)| < M$, then its folded, shifted version $x(t-\lambda)$ is also bounded. Using the *fundamental theorem of calculus* (the absolute value of any integral cannot exceed the integral of the absolute value of its integrand), the convolution integral yields the following inequality:

$$|y(t)| < \int_{-\infty}^{\infty} |h(\tau)||x(t-\tau)|\,d\tau < M\int_{-\infty}^{\infty} |h(\tau)|\,d\tau \tag{6.14}$$

It follows immediately that for $y(t)$ to be bounded, we require

$$\int_{-\infty}^{\infty} |h(\tau)|\,d\tau < \infty \qquad \text{(for a stable LTI system)} \tag{6.15}$$

For BIBO stability, therefore, $h(t)$ must be *absolutely integrable*. This is both a necessary and sufficient condition. If satisfied, we are guaranteed a stable LTI system. In particular, if $h(t)$ is an energy signal, we have a stable system.

6.5.1 The Impulse Response of Causal Systems

For a causal system, we require an impulse response of the form $h(t)u(t)$, which equals zero for $t < 0$. This ensures that an input $x(t)u(t-t_0)$ that starts at $t = t_0$ results in a response $y(t)$ that also starts at $t = t_0$. We see this by setting up the convolution integral as

$$y(t) = \int_{-\infty}^{\infty} x(\lambda)u(\lambda - t_0)h(t-\lambda)u(t-\lambda)\,d\lambda = \int_{t_0}^{t} x(\lambda)h(t-\lambda)\,d\lambda \tag{6.16}$$

Causal systems are also called **physically realizable**. Causality actually imposes a powerful constraint on $h(t)$. The even and odd parts of a causal $h(t)$ cannot be independent, and $h(t)$ can in fact be found from its even symmetric (or odd symmetric) part alone.

REVIEW PANEL 6.12

Stability and Causality of CT Systems

For stability: $h(t)$ must be *absolutely integrable* with $\displaystyle\int_{-\infty}^{\infty} |h(\xi)|\,d\xi < \infty$.

For causality: $h(t)$ must be zero for negative time with $h(t) = 0,\ t < 0$.

EXAMPLE 6.10 (Stability of LTI Systems)

(a) (Stability of a Differentiating System)

The system $y(t) = x'(t)$ has the impulse response $h(t) = \delta'(t)$. It is not BIBO stable since $\int |h(t)|\,dt$ is not finite. The (bounded) input $u(t)$, for example, results in the response $\delta(t)$, which is clearly unbounded (at $t = 0$). Extending this concept, for a stable system we require the highest derivative of the input not to exceed the highest derivative of the output. This implies a proper operational transfer function (the degree of whose numerator cannot exceed that of the denominator).

(b) (Stability of Ideal Filters)

An ideal lowpass filter is described by an impulse response of the form $h(t) = \text{sinc}(Bt)$. Since $h(t)$ is not zero for $t < 0$, this filter is physically unrealizable. Since the sinc function is not absolutely integrable, an ideal filter is also unstable.

PGR2026-00033
Krisp Technologies EX1024 Page 159

6.6 The Response to Periodic Inputs

Consider a system with impulse response $h(t) = e^{-t}u(t)$ excited by a harmonic input $x(t) = e^{j\omega_0 t}$. The response $y(t)$ then equals

$$y(t) = \int_{-\infty}^{\infty} e^{-(t-\lambda)}u(t-\lambda)e^{j\omega_0\lambda}\,d\lambda = \int_{-\infty}^{t} e^{-t}e^{\lambda(1+j\omega_0)}\,d\lambda \tag{6.17}$$

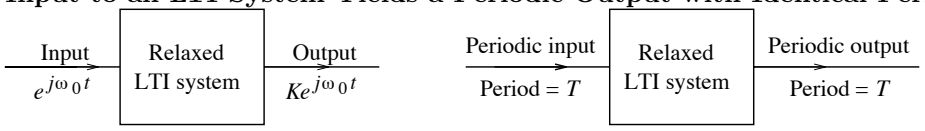
Upon simplification, we obtain

$$y(t) = \frac{e^{-t}}{1+j\omega_0}\left[e^{\lambda(1+j\omega_0)}\right]\Big|_{-\infty}^{t} = \frac{e^{-t}}{1+j\omega_0}e^{t(1+j\omega_0)} = \frac{1}{1+j\omega_0}e^{j\omega_0 t} = Kx(t) \tag{6.18}$$

where $K = 1/(1+j\omega_0)$ is a (complex) constant. The response $y(t) = Kx(t)$ is also a harmonic at the input frequency ω_0. More generally, the response of LTI systems to any periodic input is also periodic with the same period as the input. In the parlance of convolution, *the convolution of two signals, one of which is periodic, is also periodic and has the same period as the input.*

REVIEW PANEL 6.13

A Periodic Input to an LTI System Yields a Periodic Output with Identical Period



6.6.1 Periodic Extension and Wraparound

Conceptually, if we know the response to one period of the input, we can find the entire response using superposition. It is important to realize that although the response to one period is not periodic, superposition of its infinitely many shifted versions yields a periodic signal. More formally, the sum of an absolutely integrable signal (or energy signal) $x(t)$ and *all* its replicas shifted by multiples of T yields a periodic signal with period T called its **periodic extension** $x_{\text{pe}}(t)$. Thus,

$$x_{\text{pe}}(t) = \sum_{k=-\infty}^{\infty} x(t+kT) \tag{6.19}$$

The following review panel lists the periodic extensions of two useful signals. The area of one period of the periodic extension $x_{\text{pe}}(t)$ equals the total area of $x(t)$. In fact, adding $y(t)$ and its infinitely many shifted versions to obtain the periodic extension is equivalent to wrapping $y(t)$ around in one-period segments and adding them up instead. The **wraparound** method can thus be used to find the periodic output as the periodic extension of the response to one period.

REVIEW PANEL 6.14

Periodic Extension with Period T of Two Signals Over $(0,\ T)$

$$x(t) = e^{-\alpha t}u(t) \Rightarrow x_{\text{pe}}(t) = \frac{e^{-\alpha t}}{1-e^{-\alpha T}} \qquad x(t) = te^{-\alpha t}u(t) \Rightarrow x_{\text{pe}}(t) = \frac{te^{-\alpha t}}{1-e^{-\alpha T}} + \frac{Te^{-\alpha(t+T)}}{(1-e^{-\alpha T})^2}$$

PGR2026-00033
Krisp Technologies EX1024 Page 160

EXAMPLE 6.11 (Periodic Extension)

(a) If $x(t)$ is a pulse of duration $\alpha \leq T$, its periodic extension $x_{\text{pe}}(t)$ is just $x(t)$ with period T, as shown in Figure E6.11A. If $\alpha > T$, the periodic extension $x_{\text{pe}}(t)$ is found by wrapping around one-period segments of $x(t)$ past T units and adding them to $x(t)$.

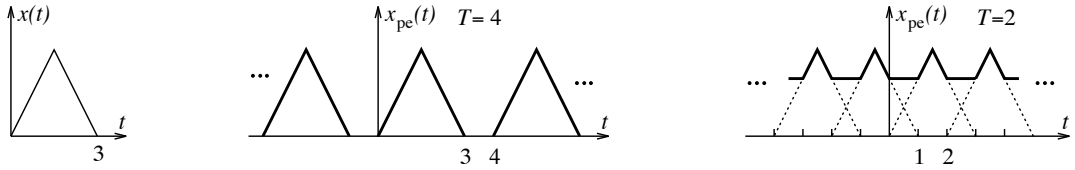


Figure E6.11A The pulse signal for Example 6.11(a) and its periodic extension

(b) The periodic extension of $x(t) = e^{-\alpha t}u(t)$ with period T may be expressed, using wraparound, as

$$x_{\text{pe}}(t) = e^{-\alpha t} + e^{-\alpha(t+T)} + e^{-\alpha(t+2T)} + \cdots = e^{-\alpha t}\sum_{k=0}^{\infty} e^{-\alpha T} = \frac{e^{-\alpha t}}{1 - e^{-\alpha T}} \tag{6.20}$$

The signal and its periodic extension is shown in Figure E6.11B.

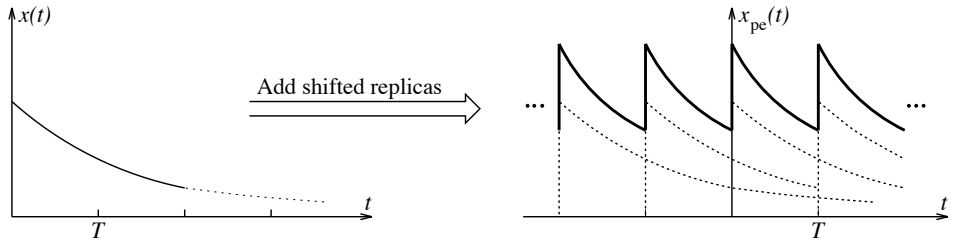


Figure E6.11B The exponential signal for Example 6.11(b) and its periodic extension

6.6.2 Finding the Response to Periodic Inputs

Rather than wrap around the response to one period, we may also find the response $y_p(t)$ to a periodic input $x_p(t)$ by first creating the periodic extension $h_{\text{pe}}(t)$ of $h(t)$ and then finding the convolution of one period of $h_{\text{pe}}(t)$ and one period of $x_p(t)$. However, we must use wraparound to generate $y_p(t)$ because the regular convolution of one period of $h_{\text{pe}}(t)$ and one period of $x_p(t)$ extends for *two periods*.

REVIEW PANEL 6.15

System Response to a Periodic Input $x_p(t)$ with Period T

1. Find the periodic extension $h_{\text{pe}}(t)$ of $h(t)$ with period T.
2. Find the regular convolution $y(t)$ of one period of $x_p(t)$ and $h_{\text{pe}}(t)$ (this will last $2T$ units).
3. Wrap around $y(t)$, $t \geq T$, using $y_W(t) = y(t+T)$. Add $y_W(t)$ to the unwrapped portion of $y(t)$.

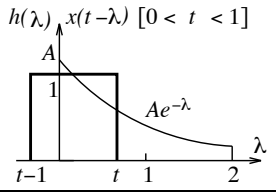
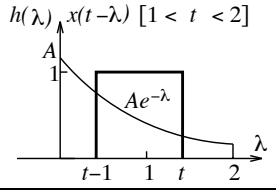
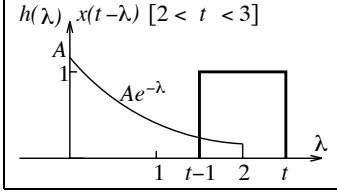
PGR2026-00033
Krisp Technologies EX1024 Page 161

The Cyclic Approach

Yet another way to find the system response is to slide the *entire* periodic signal $x_p(t-\lambda)$ past only one period of the periodic extension $h_{\rm pe}(\lambda)$. Since $x_p(t-\lambda)$ is periodic, as portions of $x_p(t-\lambda)$ slide out to the right of the one period window of $h_{\rm pe}(\lambda)$, identical portions come into view (are wrapped around) from the left. This is the *cyclic* method with wraparound built in. After $x_p(t-\lambda)$ slides one full period, the convolution replicates.

EXAMPLE 6.12 (The Response to Periodic Inputs)
Find the response of a lowpass filter with impulse response $h_0(t) = e^{-t}u(t)$ excited by a periodic rectangular pulse train $x(t)$ with unit height, unit pulse width, and period $T = 2$.

(a) (Periodic Extension) One period of the periodic extension of $h_0(t)$ is given by $h(t) = Ae^{-t}$, where $A = 1/(1-e^{-2})$. We first find the regular convolution of one period of $x(t)$ with $h(t)$. The pairwise sum gives the end points of the convolution ranges as [0, 1, 2, 3]. For each range, we superpose $x(t-\lambda) = 1,\ t-1 \le \lambda \le t$, and $h(\lambda) = Ae^{-\lambda},\ 0 \le \lambda \le 2$, to obtain the following results:

Overlapping Signals	**Range and Convolution $y(t)$**
	Range: $0 \le t \le 1$ $\int_0^t Ae^{-\lambda}\,d\lambda = A(1-e^{-t})$
	Range: $1 \le t \le 2$ $\int_{t-1}^t Ae^{-\lambda}\,d\lambda = A\left[e^{-(t-1)} - e^{-t}\right]$
	Range: $2 \le t \le 3$ $\int_{t-1}^2 Ae^{-\lambda}\,d\lambda = A\left[e^{-(t-1)} - e^{-2}\right]$

We wrap around the last 1-unit range past $t = 2$ (replacing t by $t+2$), and add to the first term, to get one period of the periodic output $y_p(t)$ as

$$y_p(t) = \begin{cases} A(1-e^{-t}) + A\left[e^{-(t+1)} - e^{-2}\right] &= 1 - \dfrac{e^{-(t-1)}}{1+e}, \qquad 0 \le t \le 1 \\ A\left[e^{-(t-1)} - e^{-t}\right] &= \dfrac{e^{-(t-2)}}{1+e}, \qquad 1 \le t \le 2 \end{cases}$$

The regular convolution and periodic convolution is plotted in Figure E6.12A.

PGR2026-00033
Krisp Technologies EX1024 Page 162

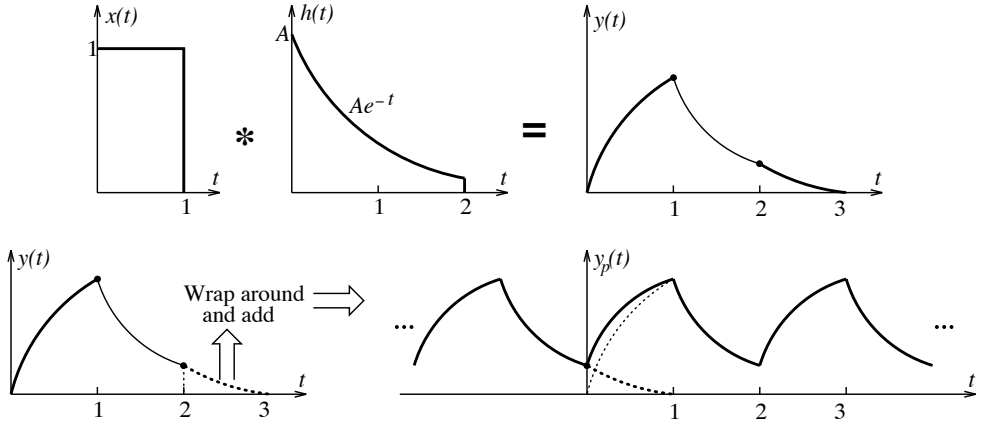


Figure E6.12A The regular and periodic convolution of the signals for Example 6.12(a)

(b) (The Cyclic Method) The output for one period may be computed using the cyclic approach by first creating $x(t-\lambda)$ and a one-period segment of $h(\lambda)$, as shown in Figure E6.12B.

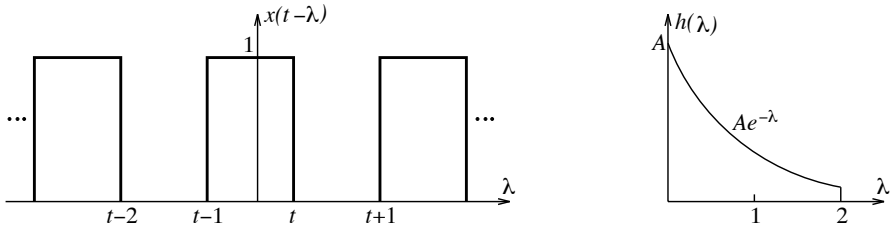
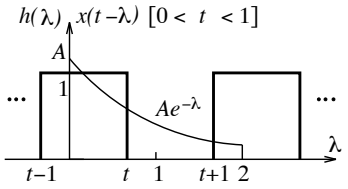
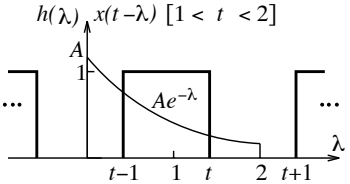


Figure E6.12B The signals $x(t-\lambda)$ and one period of $h(\lambda)$ for Example 6.12(b)

We then slide the folded signal $x(t-\lambda)$ past $h(\lambda)$ for a one-period duration (2 units), and find the periodic convolution as follows:

Overlapping Signals	**Periodic Convolution** $y_p(t)$
	Range: $0 \le t \le 1$ $$\int_0^t Ae^{-\lambda}\,d\lambda + \int_{t+1}^1 Ae^{-\lambda}\,d\lambda = 1 - \frac{e^{-(t-1)}}{1+e}$$
	Range: $1 \le t \le 2$ $$\int_{t-1}^t Ae^{-\lambda}\,d\lambda = \frac{e^{-(t-2)}}{1+e}$$

As $x(t-\lambda)$ slides right over one period of $h(\lambda)$, we see portions of two pulses in partial view for $0 \le t \le 1$, and one pulse in full view for $1 \le t \le 2$. As expected, both methods yield identical results.

PGR2026-00033
Krisp Technologies EX1024 Page 163

6.6.3 Existence of Convolution

Signals for which the convolution does not exist are not hard to find. The convolution of $u(t)$ and $u(-t)$ is one such example. Unfortunately, the question of existence can be answered only in terms of sufficient conditions that do not preclude the convolution of functions for which such conditions are violated. These conditions are related primarily to signal energy, area, or one-sidedness. The convolution of two energy signals, two absolutely integrable signals, or two left-sided signals, or two right-sided signals always exists. The regular convolution of two periodic signals does not exist because it is either identically zero (if the periods are non-commensurate and one of them has a zero average value) or infinite. This is why we must turn to averages and resort to the concept of periodic convolution.

6.7 Periodic Convolution

Convolution involves finding the area under the product of one signal and a folded shifted version of another. Periodic signals are of infinite extent. For such signals, this product will be either *infinite* or *zero*. Clearly, *the convolution of two periodic signals does not exist in the usual sense*. This is where the concept of **periodic convolution** comes in. The distinction between regular and periodic convolution is rather like the distinction between energy and power. For signals with infinite energy, a much more useful measure is the signal power, defined as the energy averaged over all time. Similarly, for signals whose convolution is infinite, a much more useful measure is the average convolution. This average is called **periodic convolution** and denoted $x_p(t) \circledast h_p(t)$.

For two periodic signals $x_p(t)$ and $h_p(t)$ with identical or common period T, the periodic convolution $y_p(t)$ for one period T may be found by convolving one period T of each signal and wrapping around the result past T.

$$y_p(t) = x_p(t) \circledast h_p(t) = \frac{1}{T}\int_0^T x_{1p}(\lambda)h_p(t-\lambda)\,d\lambda, \qquad 0 \le t \le T \tag{6.21}$$

The process is exactly like finding the system response to periodic inputs, except that no periodic extension is required. The "periodic" convolution of other power signals with non-commensurate periods must be found from a limiting form, by averaging the convolution of one periodic signal with a *finite stretch* T_0 of the other, as $T_0 \to \infty$.

$$y_p(t) = x(t) \circledast h(t) = \lim_{T_0 \to \infty} \frac{1}{T_0}\int_{T_0} x_T(\lambda)h(t-\lambda)\,d\lambda \qquad \text{(for } \textit{nonperiodic} \text{ power signals)} \tag{6.22}$$

REVIEW PANEL 6.16

Periodic Convolution of Periodic Signals with Identical Periods T

1. Find the regular convolution $y(t)$ of one-period segments of each (this will last $2T$ units).
2. Wrap around $y(t)$, $t \ge T$, using $y_W(t) = y(t+T)$. Add $y_W(t)$ to the unwrapped portion of $y(t)$.
3. Divide the result by the period T.

EXAMPLE 6.13 (Periodic Convolution)

Consider the periodic signals $x(t)$ and $y(t)$, shown at the left in Figure E6.13, for three cases. To find their periodic convolution, we first find the regular convolution of their one-period segments; create its periodic extension with period T, using wraparound where necessary; and normalize (divide) the result by T.

PGR2026-00033
Krisp Technologies EX1024 Page 164

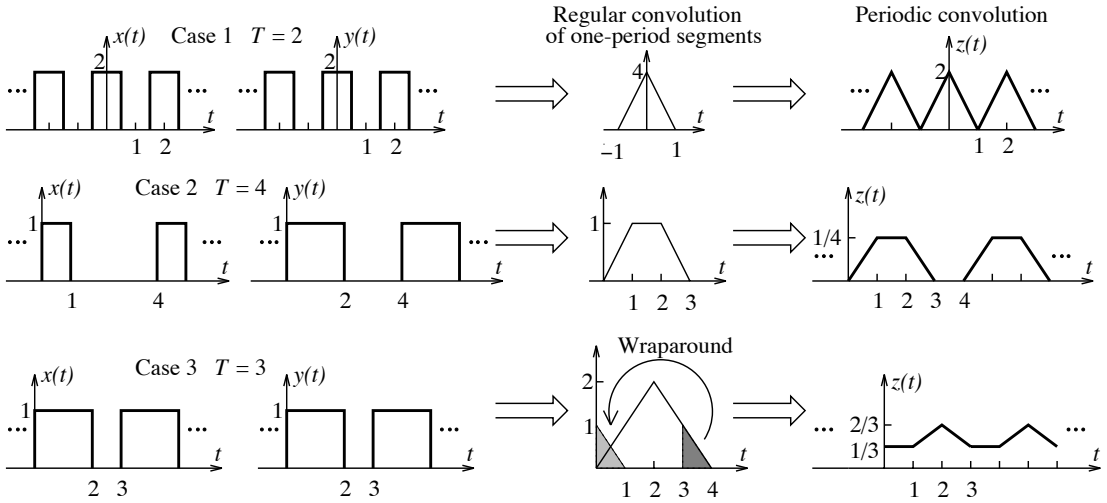


Figure E6.13 The periodic signals for Example 6.13 and their periodic convolution

The periodic convolution $z(t)$ (shown on the far right in Figure E6.13) requires wraparound only in the third case where the regular convolution extends beyond $T = 3$.

6.8 Connections: Convolution and Transform Methods

Not only is convolution an important method of system analysis in the time domain, but it also leads to concepts that form the basis of every transform method described in this text. Its role in linking the time domain and the transformed domain is intimately tied to the concept of eigensignals and eigenvalues.

6.8.1 Eigensignals and Eigenvalues

The analysis of linear systems using convolution is based on expressing an input $x(t)$ as a weighted sum of shifted impulses, finding the response to each such impulse, and using superposition to obtain the total response. The signal $x(t)$ can also be described by a weighted combination of other useful signals besides impulses. Let $\phi_k(t)$ form a class of functions that result in the response $\psi_k(t)$. If $x(t)$ can be expressed as a weighted sum of these $\phi_k(t)$, the response $y(t)$ is a corresponding weighted sum of $\psi_k(t)$:

$$x(t) = \sum_k \alpha_k \phi_k \qquad y(t) = \sum_k \beta_k \psi_k \tag{6.23}$$

What better choice for $\phi_k(t)$ than one that yields a response that is just a scaled version of itself such that $\psi_k(t) = A_k \phi_k(t)$? Then

$$y(t) = \sum_k \beta_k \psi_k = \sum_k \beta_k A_k \phi_k \tag{6.24}$$

Finding the output thus reduces to finding just the scale factors β_k, which may be real or complex. Signals ϕ_k that are preserved in form by a system except for a scale factor β_k are called **eigensignals**, **eigenfunctions**, or **characteristic functions** because they are intrinsic (in German, *eigen*) to the system. The factor β_k by which the eigensignal is scaled is called the **eigenvalue** of the system or the **system function**.

REVIEW PANEL 6.17

System Response to an Eigensignal = (Eigensignal) × (System Function)

PGR2026-00033
Krisp Technologies EX1024 Page 165

6.8.2 Harmonics as Eigensignals of Linear Systems

Everlasting harmonic signals form the most important class of eigensignals of linear systems. The system response to the everlasting harmonic signal $e^{j\omega t}$ is another everlasting harmonic signal $Ke^{j\omega t}$ at the same frequency ω but scaled in magnitude and phase by the (possibly complex) constant K.

More generally, everlasting complex exponentials of the form e^{st}, where $s = \sigma + j\omega = s + j2\pi f$ is a complex quantity, are also eigensignals of every linear system.

REVIEW PANEL 6.18

The Everlasting Exponential e^{st} Is an Eigensignal of Linear Systems

In this everlasting exponential e^{st}, the quantity s has the general form $s = \sigma + j\omega$.
The response of a linear system to e^{st} is Ce^{st} where C is a (possibly complex) constant.

6.8.3 Transformed Domain Relations

Consider the eigensignal $x(t) = e^{st}$ as the input to a linear system with impulse response $h(t)$. The response $y(t)$ equals the convolution:

$$y(t) = x(t) \star h(t) = \int_{-\infty}^{\infty} e^{s(t-\lambda)} h(\lambda)\, d\lambda = e^{st} \int_{-\infty}^{\infty} h(\lambda) e^{-s\lambda}\, d\lambda \tag{6.25}$$

The response equals the product of the eigensignal e^{st} and the system function (which is a function only of the variable s). If we denote this system function by $H(s)$, we have

$$H(s) = \int_{-\infty}^{\infty} h(\lambda) e^{-s\lambda}\, d\lambda \tag{6.26}$$

This is also called the **transfer function**. It is actually a description of $h(t)$ by a weighted sum of complex exponentials and is, in general, also complex. Now, the signal $x(t)$ also yields a similar description, called the **two-sided Laplace transform**:

$$X(s) = \int_{-\infty}^{\infty} x(\lambda) e^{-s\lambda}\, d\lambda \qquad \text{(two-sided Laplace transform)} \tag{6.27}$$

For a *causal* signal of the form $x(t)u(t)$, we obtain the **one-sided Laplace transform**:

$$X(s) = \int_{0}^{\infty} x(\lambda) e^{-s\lambda}\, d\lambda \qquad \text{(one-sided Laplace transform)} \tag{6.28}$$

The transform of the response $y(t) = x(t)H(s)$ may also be written as

$$Y(s) = \int_{-\infty}^{\infty} y(t) e^{-st}\, dt = \int_{-\infty}^{\infty} H(s)x(t) e^{-st}\, dt = H(s)X(s) \tag{6.29}$$

Since $y(t)$ also equals $x(t) \star h(t)$, *convolution in the time domain is equivalent to multiplication in the transformed domain*. This is one of the most fundamental results, and one that we shall use repeatedly in subsequent chapters.

REVIEW PANEL 6.19

Convolution in One Domain Corresponds to Multiplication in the Other

PGR2026-00033
Krisp Technologies EX1024 Page 166

With only slight modifications, we can describe several other transformed-domain relations as follows:

1. With $s = j2\pi f$, we use $e^{j2\pi f}$ as the eigensignals and transform $h(t)$ to the frequency domain in terms of its **steady-state transfer function** $H(f)$ and the signal $x(t)$ to its **Fourier transform** $X(f)$:

$$X(f) = \int_{-\infty}^{\infty} x(\lambda) e^{-j2\pi f\lambda}\, d\lambda \qquad \text{(Fourier transform)} \tag{6.30}$$

2. For a single harmonic $x(t) = e^{j2\pi f_0 t}$, the impulse response $h(t)$ transforms to a complex constant $H(f_0) = Ke^{j\theta}$. This produces the response $Ke^{j(2\pi f_0 t+\theta)}$ and describes the method of **phasor analysis**.

3. For a periodic signal $x_p(t)$ described by a combination of harmonics $e^{jk2\pi f_0 t}$ at the discrete frequencies kf_0, we use superposition to obtain a frequency-domain description of $x_p(t)$ over one period in terms of its **Fourier series** coefficients $X[k]$.

REVIEW PANEL 6.20

The Type of Convolution Depends on the Transform Method

Transform Method	Time-Domain Convolution	Frequency-Domain Convolution
Laplace Transform	$x(t) \star h(t)$ (regular)	$X(s) \star H(s)$ (regular, complex)
Fourier Transform	$x(t) \star h(t)$ (regular)	$X(f) \star H(f)$ (regular, complex)
Fourier Series	$x_p(t) \circledast h_p(t)$ (periodic)	$X[k] \star H[k]$ (regular, discrete)

6.9 Convolution Properties Based on Moments

The nth moment of a signal $x(t)$ is defined as $m_n = \int t^n x(t)\, dt$. Thus, $m_0 = \int x(t)\, dt$ is just the area under $x(t)$. The area property suggests the following obvious result for the zeroth moments

$$\int_{-\infty}^{\infty} y(t)\, dt = \left[\int_{-\infty}^{\infty} x(t)\, dt\right]\left[\int_{-\infty}^{\infty} h(t)\, dt\right] \qquad \text{or} \qquad m_0(y) = m_0(x)m_0(h) \tag{6.31}$$

The quantity $D_x = m_x = \frac{m_1}{m_0} = \int tx(t)\, dt \Big/ \int x(t)\, dt$ corresponds to the **effective delay** or mean for an energy signal. The delays are also related by

$$D_y = D_x + D_h \qquad \text{or} \qquad \frac{m_1(y)}{m_0(y)} = \frac{m_1(h)}{m_0(h)} + \frac{m_1(x)}{m_0(x)} \tag{6.32}$$

Central moments (about the mean) are defined as $\mu_n = \int (t - m_x)^n x(t)\, dt$. The variance of $x(t)$ is defined by $\mu_2 = \int (t - m_x)^2 x(t)\, dt = [m_2(x)/m_0(x)] - m_x^2 = T_x^2$, and is a measure of the **effective duration** for an energy signal. The durations are related by

$$T_y = \sqrt{T_x^2 + T_h^2} \qquad \text{or} \qquad \mu_2(y) = \mu_2(x) + \mu_2(h) \tag{6.33}$$

6.9.1 Moments of Cascaded Systems

The impulse response of a cascade of linear, time-invariant systems with impulse responses $h_1(t), h_2(t), \ldots$ is described by the repeated convolution

$$h(t) = h_1(t) \star h_2(t) \star h_3(t) \star \cdots \tag{6.34}$$

The moment properties suggest that the overall system delay D equals the sum of delays of the individual systems and the overall duration T equals the square root of the sum of squares of the individual durations.

PGR2026-00033
Krisp Technologies EX1024 Page 167

EXAMPLE 6.14 (Moment Properties and Cascaded Systems)
Consider a lowpass filter with time constant $\tau = 1$ and $h(t) = e^{-t}u(t)$. The moments of $h(t)$ are

$$m_0(h) = \int_0^\infty h(t)\,dt = 1 \qquad m_1(h) = \int_0^\infty th(t)\,dt = 1 \qquad m_2(h) = \int_0^\infty t^2 h(t)\,dt = 2 \qquad D_h = 1$$

(a) If $x(t) = h(t)$, the response is $y(t) = x(t) \star h(t) = te^{-t}u(t)$. The moments of $y(t)$ are

$$m_0(y) = \int_0^\infty y(t)\,dt = 1 \qquad m_1(y) = \int_0^\infty ty(t)\,dt = 2 \qquad D_y = 2$$

We see that $m_0(y) = m_0(x)m_0(h) = 1$ and $D_y = D_x + D_h = 2$.

(b) We compute the effective filter duration T_h as

$$T_h^2 = \int_0^\infty (t - m_h)^2 h(t)\,dt = \frac{m_2(h)}{m_0(h)} - m_h^2 = 2 - 1 = 1$$

Comment: For a cascade of N identical lowpass filters (with $\tau = 1$), the overall effective delay D_e is $D_e = ND_h = N$, and the overall effective duration T_e is $T_e = \sqrt{NT_h^2} = T_h\sqrt{N}$.

6.9.2 Repeated Convolution and the Central Limit Theorem

Repeated convolution is associative as long as the intermediate convolutions exist. In particular, the repeated convolution of *energy* signals is always associative, and order is unimportant. With three signals, for example,

$$x_1 \star (x_2 \star x_3) = (x_1 \star x_2) \star x_3 = (x_1 \star x_3) \star x_2 = (x_3 \star x_2) \star x_1 = \cdots \tag{6.35}$$

For energy signals, we also observe that the convolution is smoother than any of the signals convolved, and repeated convolution begins to take on the bell-shaped Gaussian form as the number n of functions convolved becomes large and may be expressed as

$$y_n(t) = x_1(t) \star x_2(t) \star \cdots \star x_n(t) \approx \frac{K}{\sqrt{2\pi\sigma_n^2}} \exp\left[-\frac{(t - m_n)^2}{2\sigma_n^2}\right] \tag{6.36}$$

Here, m_n is the sum of the individual means (delays), σ_n^2 is the sum of the individual variances, and the constant K equals the product of the areas under each of the convolved functions.

$$K = \prod_{k=1}^{n} \int x_k(t)\,dt \tag{6.37}$$

This result is one manifestation of the **central limit theorem**. It allows us to assert that the response of a complex system composed of many subsystems is Gaussian, since its response is based on repeated convolution. The individual responses need not be Gaussian and need not even be known.

The central limit theorem fails if any function has zero area, making $K = 0$. Sufficient conditions for it to hold require finite values of the average, the variance, and the absolute third moment. All time-limited functions and many others satisfy these rather weak conditions. The system function $H(f)$ of a large number of cascaded systems is also a Gaussian because convolution in the time domain is equivalent to multiplication in the frequency domain. In probability theory, the central limit theorem asserts that the sum of n statistically independent random variables approaches a Gaussian for large n, regardless of the nature of their distributions.

PGR2026-00033
Krisp Technologies EX1024 Page 168

REVIEW PANEL 6.21

The Convolution of n Energy Signals Approaches a Gaussian Form for Large n

If m_n = sum of the means, σ_n^2 = sum of the variances, and K = *product* of the areas, then

$$x_1(t) \star x_2(t) \star \cdots \star x_n(t) \quad \approx \quad \frac{K}{\sqrt{2\pi\sigma_n^2}} \exp\left[-\frac{(t-m_n^2)^2}{2\sigma_n^2}\right]$$

EXAMPLE 6.15 (The Central Limit Theorem)

(a) Consider a cascade of identical RC lowpass filters with $h(t) = e^{-t}u(t)$. The impulse response $y_n(t)$ of the cascaded system is simply the repeated convolution of $h(t)$ with itself. The first few convolutions yield

$$y_1(t) = te^{-t}u(t) \qquad y_2(t) = \tfrac{1}{2}t^2e^{-t}u(t) \qquad y_3(t) = \tfrac{1}{3}t^3e^{-t}u(t)$$

If the convolution is repeated n times, the result generalizes to

$$y_n(t) = e^{-t}u(t) \star e^{-t}u(t) \star \cdots \star \epsilon^{-t}u(t) = \frac{t^n}{n!}e^{-t}u(t) \tag{6.38}$$

The peak shifts to the right by 1 unit for each successive convolution. We expect to see a progression toward the bell-shaped Gaussian form with increasing n. The result for $n = 40$ is illustrated in Figure E6.15(a) along with its Gaussian approximation.

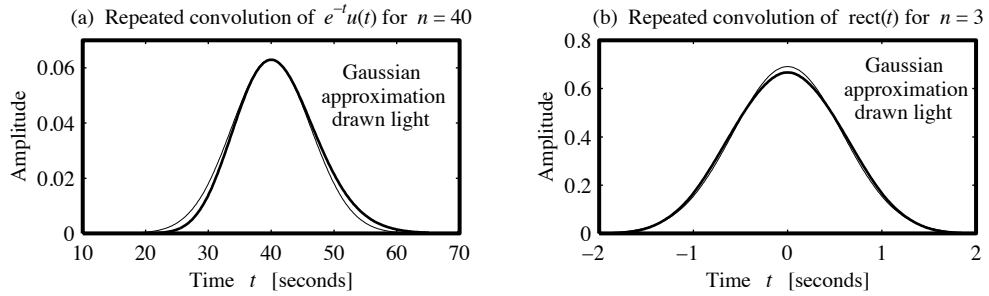


Figure E6.15 The repeated convolution and its Gaussian approximation for the signals of Example 6.15

To find the Gaussian form as $n \to \infty$, we start with the mean m_h, variance σ^2, and area A for $e^{-t}u(t)$. We find

$$A = m_0 = \int_0^\infty h(t)\,dt = 1 \qquad m_h = \frac{m_1}{m_0} = 1 \qquad \sigma^2 = 1$$

For n cascaded systems, we have $m_N = nm_h = n, \sigma_N^2 = n\sigma^2 = n$, and $K = A^n = 1$. These values lead to the Gaussian approximation for $y_n(t)$ as

$$y_n(t) \approx \frac{1}{\sqrt{2\pi n}} \exp\left[-\frac{(t-n)^2}{2n}\right]$$

(b) An even more striking example is provided by the convolution of even symmetric rectangular pulses, shown in Figure E6.15(b) for $n = 3$. Notice how the result begins to take on a Gaussian look after only a few repeated convolutions.

PGR2026-00033
Krisp Technologies EX1024 Page 169

6.10 Correlation

Correlation is an operation similar to convolution. It involves sliding one function past the other and finding the area under the resulting product. Unlike convolution, however, no folding is performed. The correlation $r_{xx}(t)$ of two identical functions $x(t)$ is called **autocorrelation**. For two different functions $x(t)$ and $y(t)$, the correlation $r_{xy}(t)$ or $r_{yx}(t)$ is referred to as **cross-correlation**.

Using the symbol $\star\star$ to denote correlation, we define the two operations as

$$r_{xx}(t) = x(t) \star\star x(t) = \int_{-\infty}^{\infty} x(\lambda)x(\lambda - t)\, d\lambda \tag{6.39}$$

$$r_{xy}(t) = x(t) \star\star y(t) = \int_{-\infty}^{\infty} x(\lambda)y(\lambda - t)\, d\lambda \tag{6.40}$$

$$r_{yx}(t) = y(t) \star\star x(t) = \int_{-\infty}^{\infty} y(\lambda)x(\lambda - t)\, d\lambda \tag{6.41}$$

The variable t is often referred to as the **lag**. The definitions of cross-correlation are not standard, and some authors prefer to switch the definitions of $r_{xy}(t)$ and $r_{yx}(t)$.

6.10.1 Some Properties of Correlation

Cross-correlation is a measure of similarity between different functions. Since the order matters in cross-correlation, we have two cross-correlation functions given by $r_{xy}(t)$ and $r_{yx}(t)$ as

$$r_{xy}(t) = \int_{-\infty}^{\infty} x(\lambda)y(\lambda - t)\, d\lambda \qquad\qquad r_{yx}(t) = \int_{-\infty}^{\infty} y(\lambda)x(\lambda - t)\, d\lambda \tag{6.42}$$

At $t = 0$, we have

$$r_{xy}(0) = \int_{-\infty}^{\infty} x(\lambda)y(\lambda)\, d\lambda = r_{yx}(0) \tag{6.43}$$

Thus, $r_{xy}(0) = r_{yx}(0)$. The cross-correlation also satisfies the inequality

$$|r_{xy}(t)| \le \sqrt{r_{xx}(0)r_{yy}(0)} = \sqrt{E_x E_y} \tag{6.44}$$

where E_x and E_y represent the signal energy in $x(t)$ and $y(t)$, respectively.

Correlation as Convolution

The absence of folding actually implies that the correlation of $x(t)$ and $y(t)$ is equivalent to the convolution of $x(t)$ with the folded version $y(-t)$, and we have $r_{xy}(t) = x(t) \star\star y(t) = x(t) \star y(-t)$.

Area and Duration

Since folding does not affect the area or duration, the area and duration properties for convolution also apply to correlation. The starting and ending time for the cross-correlation $r_{xy}(t)$ may be found by using the starting and ending times of $x(t)$ and the folded signal $y(-t)$.

Commutation

The absence of folding means that the correlation depends on which function is shifted and, in general, $x(t) \star\star y(t) \neq y(t) \star\star x(t)$. Since shifting one function to the right is actually equivalent to shifting the other function to the left by an equal amount, the correlation $r_{xy}(t)$ is related to $r_{yx}(t)$ by $r_{xy}(t) = r_{yx}(-t)$.

PGR2026-00033
Krisp Technologies EX1024 Page 170

REVIEW PANEL 6.22

Correlation Is the Convolution of One Signal with a Folded Version of the Other

$r_{xh}(t) = x(t) \star\star h(t) = x(t) \star h(-t)$ $\qquad$ $r_{hx}(t) = h(t) \star\star x(t) = h(t) \star x(-t)$

Duration = *sum* of durations of $x(t)$ and $h(t)$. $\qquad$ **Area** = *product* of areas of $x(t)$ and $h(t)$.

Periodic Correlation

The correlation of two periodic signals or power signals is defined in the same sense as periodic convolution:

$$r_{xy}(t) = \frac{1}{T}\int_T x(\lambda)y(\lambda - t)\,d\lambda \qquad r_{xy}(t) = \lim_{T_0\to\infty}\frac{1}{T_0}\int_{T_0} x(\lambda)y(\lambda - t)\,d\lambda \tag{6.45}$$

The first form defines the correlation of periodic signals with identical periods T, which is also periodic with the same period T. The second form is reserved for nonperiodic power signals or random signals.

6.10.2 Autocorrelation

The autocorrelation operation involves identical functions. It can thus be performed in any order and represents a commutative operation.

What Autocorrelation Means

Autocorrelation may be viewed as a measure of similarity, or *coherence*, between a function $x(t)$ and its shifted version. Clearly, under no shift, the two functions "match" and result in a maximum for the autocorrelation. But with increasing shift, it would be natural to expect the similarity and hence the correlation between $x(t)$ and its shifted version to decrease. As the shift approaches infinity, all traces of similarity vanish, and the autocorrelation decays to zero.

Symmetry

Since $r_{xy}(t) = r_{yx}(-t)$, we have $r_{xx}(t) = r_{xx}(-t)$. This means that *the autocorrelation of a real function is even.* The autocorrelation of an even function $x(t)$ also equals the convolution of $x(t)$ with itself, because the folding operation leaves an even function unchanged.

Maximum Value

It turns out that the autocorrelation function is symmetric about the origin where it attains its maximum value. It thus satisfies

$$r_{xx}(t) \le r_{xx}(0) \tag{6.46}$$

It follows that *the autocorrelation $r_{xx}(t)$ is finite and nonnegative for all t.*

Periodic Autocorrelation

For periodic signals, we define periodic autocorrelation in much the same way as periodic convolution. If we shift a periodic signal with period T past itself, the two line up after every period, and the periodic autocorrelation also has period T.

PGR2026-00033
Krisp Technologies EX1024 Page 171

Signal Energy and Power

The value of the autocorrelation (or periodic autocorrelation) at the origin is related to the signal energy (or signal power for periodic signals) by

$$E_x = r_{xx}(0) = \int_{-\infty}^{\infty} x(\lambda)x(\lambda - 0)\,d\lambda = \int_{-\infty}^{\infty} x^2(\lambda)\,d\lambda \qquad P_x = r_{xx}(0) = \frac{1}{T}\int_T x_p^2(\lambda)\,d\lambda \tag{6.47}$$

REVIEW PANEL 6.23

The Autocorrelation Is Always Even Symmetric with a Maximum at the Origin

$r_{xx}(t) = x(t) \star\star x(t) = x(t) \star x(-t)$ $\qquad$ $r_{xx}(t) = r_{xx}(-t)$ $\qquad$ $r_{xx}(t) \le r_{xx}(0)$

EXAMPLE 6.16 (Autocorrelation and Cross-Correlation)

(a) The autocorrelation of the rectangular pulse $x(t) = \text{rect}(t - 0.5)$ equals the convolution of $x(t)$ and $x(-t)$ and is illustrated in Figure E6.16A.

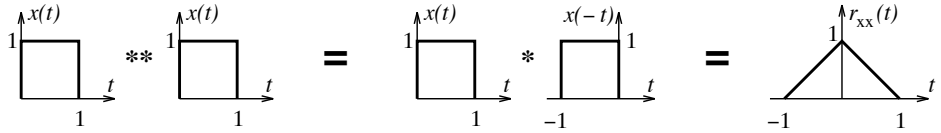
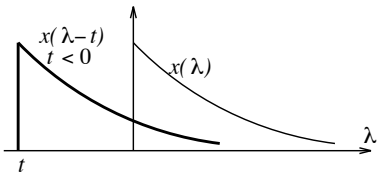
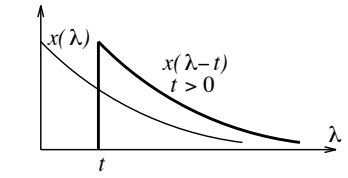


Figure E6.16A The signal for Example 6.16(a) and its autocorrelation

---

(b) Consider the autocorrelation of $x(t) = e^{-t}u(t)$. As we shift $x(\lambda - t) = e^{t-\lambda}u(\lambda - t)$ past $x(\lambda) = e^{-\lambda}u(\lambda)$, we obtain two ranges ($t < 0$ and $t > 0$) over which the autocorrelation results are described as follows:

Overlapping Signals	Autocorrelation $r_{xx}(t)$	End Values
	Range: $t < 0$ $\int_0^{\infty} e^{-\lambda}e^{t-\lambda}\,d\lambda = 0.5e^{t}$	↗ $y(-\infty) = 0$ ↘ $y(0) = 0.5$
	Range: $t > 0$ $\int_t^{\infty} e^{-\lambda}e^{t-\lambda}\,d\lambda = 0.5e^{-t}$	↗ $y(0) = 0.5$ ↘ $y(\infty) = 0$

The result may be expressed as $r_{xx}(t) = 0.5e^{-|t|}$.

(c) The cross-correlation of the signals $x(t)$ and $h(t)$, shown in Figure E6.16C, may be found using the convolution of one signal and the folded version of the other. Observe that $r_{xh}(t) = r_{hx}(-t)$.

PGR2026-00033
Krisp Technologies EX1024 Page 172

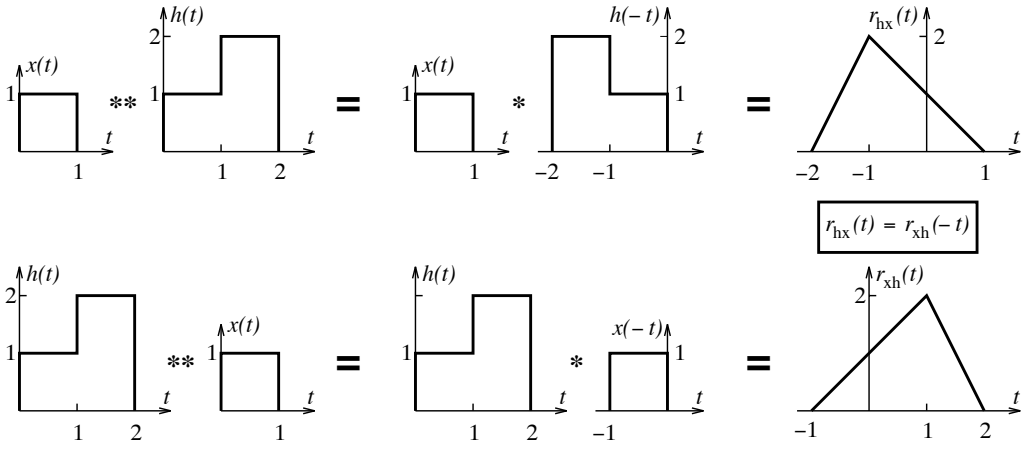


Figure E6.16C The signals of Example 6.16(c) and their cross-correlations

6.10.3 Matched Filters

Correlation forms the basis for many methods of signal detection and delay estimation (usually in the presence of noise). An example is target ranging by radar, illustrated in Figure 6.6, where the objective is to estimate the target distance (or range) R.

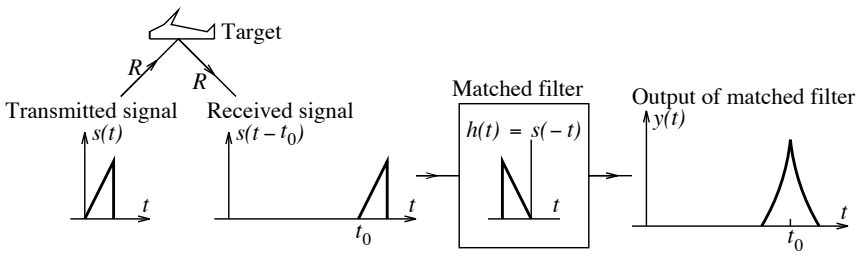


Figure 6.6 Illustrating the concept of matched filtering

A transmitter sends out an interrogating signal $s(t)$, and the reflected and delayed signal (the echo) $s(t - t_0)$ is processed by a correlation receiver, or **matched filter**, whose impulse response is *matched* to the signal to obtain the target range. In fact, its impulse response is chosen as $h(t) = s(-t)$, a folded version of the transmitted signal, in order to maximize the signal-to-noise ratio. The response $y(t)$ of the matched filter is the convolution of the received echo and the folded signal $h(t) = s(-t)$ or the correlation of $s(t - t_0)$ (the echo) and $s(t)$ (the signal). This response attains a maximum at $t = t_0$, which represents the time taken to cover the round-trip distance $2R$. The target range R is then given by $R = 0.5ct_0$, where c is the velocity of signal propagation.

Why not use the received signal directly to estimate the delay? The reason is that we may not be able to detect the presence (let alone the exact onset) of the received signal because it is usually much weaker than the transmitted signal and also contaminated by additive noise. However, if the noise is uncorrelated with the original signal (as it usually is), their cross-correlation is very small (ideally zero), and the cross-correlation of the original signal with the noisy echo yields a peak (at $t = t_0$) that stands out and is much easier to detect. Ideally, of course, we would like to transmit narrow pulses (approximating impulses) whose autocorrelation attains a sharp peak.

PGR2026-00033
Krisp Technologies EX1024 Page 173

CHAPTER 6 PROBLEMS

DRILL AND REINFORCEMENT

6.1 (Convolution Kernel) For each signal $x(t)$, sketch $x(\lambda)$ vs. λ and $x(t-\lambda)$ vs. λ, and identify significant points along each axis.

(a) $x(t) = r(t)$ **(b)** $x(t) = u(t-2)$ **(c)** $x(t) = 2\text{tri}[0.5(t-1)]$ **(d)** $x(t) = e^{-|t|}$

6.2 (Convolution Concepts) Using the defining relation, compute $y(t) = x(t) \star h(t)$ at $t = 0$.

(a) $x(t) = u(t-1)$ $h(t) = u(t+2)$
(b) $x(t) = u(t)$ $h(t) = tu(t-1)$
(c) $x(t) = tu(t+1)$ $h(t) = (t+1)u(t)$
(d) $x(t) = u(t)$ $h(t) = \cos(0.5\pi t)\text{rect}(0.5t)$

6.3 (Analytical Convolution) Evaluate each convolution $y(t) = x(t) \star h(t)$ and sketch $y(t)$.

(a) $x(t) = e^{-t}u(t)$ $h(t) = r(t)$
(b) $x(t) = te^{-\alpha t}u(t)$ $h(t) = u(t)$
(c) $x(t) = e^{-t}u(t)$ $h(t) = \cos(t)u(t)$
(d) $x(t) = e^{-t}u(t)$ $h(t) = \cos(t)$
(e) $x(t) = 2t[u(t+2) - u(t-2)]$ $h(t) = u(t) - u(t-4)$
(f) $x(t) = 2tu(t)$ $h(t) = \text{rect}(t/2)$
(g) $x(t) = r(t)$ $h(t) = \dfrac{1}{t}u(t-1)$

6.4 (Convolution with Impulses) Sketch the convolution $y(t) = x(t) \star h(t)$ for each pair of signals shown in Figure P6.4.

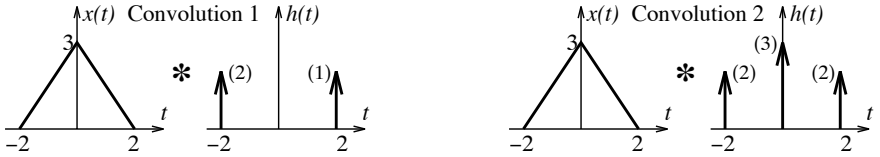


Figure P6.4 The signals $x(t)$ and $h(t)$ for Problem 6.4

6.5 (Convolution by Ranges) For each pair of signals $x(t)$ and $h(t)$ shown in Figure P6.5, establish the convolution ranges. Then sketch $x(\lambda)h(t-\lambda)$ for each range, evaluate the convolution over each range, and sketch the convolution result $y(t)$.

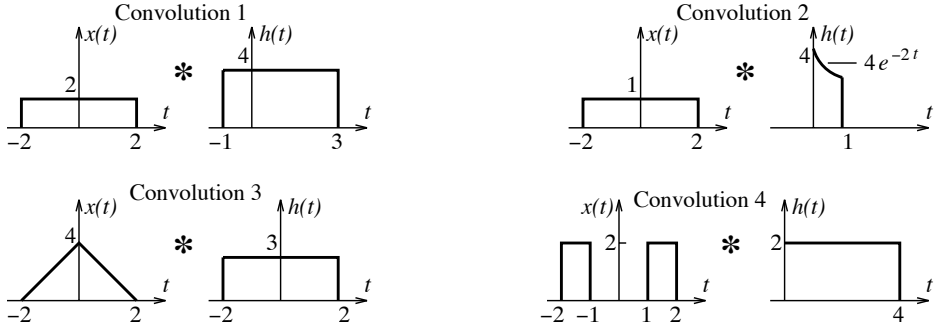


Figure P6.5 The signals $x(t)$ and $h(t)$ for Problem 6.5

PGR2026-00033
Krisp Technologies EX1024 Page 174

6.6 (Convolution and Its Properties)

(a) Use the result $u(t) \star u(t) = r(t)$ to find and sketch $y(t) = \text{rect}(t/2) \star \text{rect}(t/4)$.

(b) Use the result $\text{rect}(t) \star \text{rect}(t) = \text{tri}(t)$ to find and sketch $y(t) = [\text{rect}(t+\frac{1}{2}) - \text{rect}(t-\frac{1}{2})] \star \text{rect}(t-\frac{1}{2})$.

6.7 (Properties) The step response of a system is $s(t) = e^{-t}u(t)$. What is the system impulse response $h(t)$? Compute the response of this system to the following inputs.

(a) $x(t) = r(t)$ **(b)** $x(t) = \text{rect}(t/2)$ **(c)** $x(t) = \text{tri}[(t-2)/2]$ **(d)** $x(t) = \delta(t+1) - \delta(t-1)$

6.8 (Properties) The step response of each system is $s(t)$ and the input is $x(t)$. Compute the response of each system.

(a) $s(t) = r(t) - r(t-1)$ $\quad x(t) = \sin(2\pi t)u(t)$

(b) $s(t) = e^{-t}u(t)$ $\quad x(t) = e^{-t}u(t)$

6.9 (System Analysis) Consider the following system:

$$x(t) \longrightarrow \boxed{h(t) = 2e^{-t}u(t) - \delta(t)} \longrightarrow y(t)$$

(a) Let $x(t) = e^{-t}u(t)$. Find the output $y(t)$.

(b) Let $x(t) = u(t)$. Find the output $y(t)$.

(c) Let $x(t) = \cos(t)$. Find the output $y(t)$.

6.10 (Cascaded Systems) Find the response $y(t)$ of the following cascaded systems.

(a) $x(t) = u(t) \longrightarrow \boxed{h_1(t) = e^{-t}u(t)} \longrightarrow \boxed{h_2(t) = e^{-t}u(t)} \longrightarrow y(t)$

(b) $x(t) = e^{-t}u(t) \longrightarrow \boxed{h_1(t) = e^{-t}u(t)} \longrightarrow \boxed{h_2(t) = e^{-t}u(t)} \longrightarrow y(t)$

(c) $x(t) = e^{-t}u(t) \longrightarrow \boxed{h_1(t) = e^{-t}u(t)} \longrightarrow \boxed{h_2(t) = \delta(t) - e^{-t}u(t)} \longrightarrow y(t)$

6.11 (Stability) Investigate the stability and causality of the following systems.

(a) $h(t) = e^{t+1}u(t-1)$ **(b)** $h(t) = e^{-t}u(t+1)$ **(c)** $h(t) = \delta(t)$

(d) $h(t) = (1 - e^{-t})u(t)$ **(e)** $h(t) = \delta(t) - e^{-t}u(t)$ **(f)** $h(t) = \text{sinc}(t-1)$

6.12 (Causality) Argue that the impulse response $h(t)$ of a causal system must be zero for $t < 0$. Based on this result, if the input to a causal system starts at $t = t_0$, at what time does the response start?

6.13 (Signal-Averaging Filter) Consider a a signal-averaging filter whose impulse response is described by $h(t) = \frac{1}{T}\,\text{rect}\left(\frac{t-0.5T}{T}\right)$.

(a) What is the response of this filter to the unit step input $x(t) = u(t)$?

(b) What is the response of this filter to a periodic sawtooth signal $x(t)$ with peak value A, duty ratio D, and period T?

6.14 (Periodic Extension) For each signal shown in Figure P6.14, sketch the periodic extension with period $T = 6$ and $T = 4$.

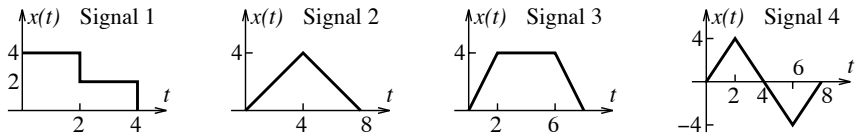


Figure P6.14 The signals for Problem 6.14

PGR2026-00033
Krisp Technologies EX1024 Page 175

6.15 (Convolution and Periodic Inputs) The voltage input to a series RC circuit with time constant $\tau = 1$ is a rectangular pulse train starting at $t = 0$. The pulses are of unit width and unit height and repeat every 2 seconds. The output is the capacitor voltage.

(a) Use convolution to compute the output at $t = 1$ s and $t = 2$ s.

(b) Assume that the input has been applied for a long time. What is the steady-state output?

6.16 (Periodic Convolution) Find and sketch the periodic convolution $y_p(t) = x(t) \circledast h(t)$ of each pair of periodic signals shown in Figure P6.16.

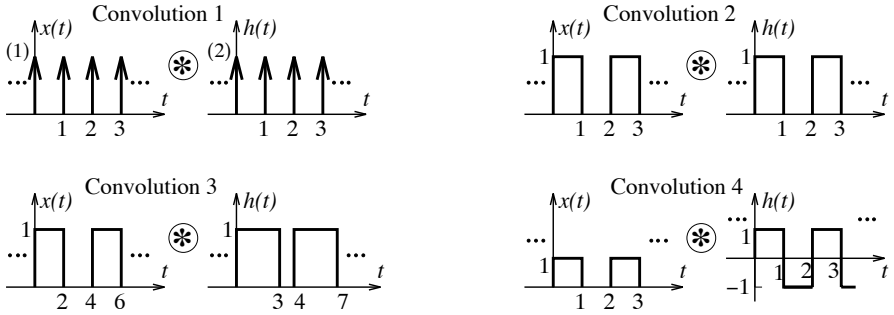


Figure P6.16 The periodic signals for Problem 6.16

6.17 (Inverse Systems) Given a system whose impulse response is $h(t) = e^{-t}u(t)$, we wish to find the impulse response $h_I(t)$ of an inverse system such that $h(t) \star h_I(t) = \delta(t)$. The form that we require for the inverse system $h_I(t)$ is $h_I(t) = K_1\delta(t) + K_2\delta'(t)$.

(a) For what values of K_1 and K_2 will $h(t) \star h_I(t) = \delta(t)$?

(b) Is the inverse system stable? Is it causal?

(c) What is the impulse response $h_I(t)$ of the inverse system if $h(t) = 2e^{-3t}u(t)$.

6.18 (Correlation) Let $x(t) = \text{rect}(t + 0.5)$ and $h(t) = t\,\text{rect}(t - 0.5)$.

(a) Find the autocorrelation $r_{xx}(t)$.

(b) Find the autocorrelation $r_{hh}(t)$.

(c) Find the cross-correlation $r_{hx}(t)$.

(d) Find the cross-correlation $r_{xh}(t)$.

(e) How are the results of parts (c) and (d) related?

REVIEW AND EXPLORATION

6.19 (The Convolution Concept) Let $x(t) = 2[u(t) - u(t - 5)]$ and $h(t) = u(t) - u(t - 5)$.

(a) Approximate $x(t)$ by five impulses located at $t = 0, 1, 2, 3, 4$. Find the convolution of $h(t)$ with each impulse, and use superposition to find the approximate form of $y(t) = x(t) \star h(t)$.

(b) How long does the convolution last, and what is its maximum?

(c) If $x(t)$ were approximated by ten impulses, how long would the convolution last, and what would be its maximum value?

(d) What would you expect the exact form of $y(t)$ to approach?

PGR2026-00033
Krisp Technologies EX1024 Page 176

6.20 (Operations on the Impulse) Explain the difference between each of the following operations on the impulse $\delta(t-1)$. Use sketches to plot results if appropriate.

(a) $[e^{-t}u(t)]\delta(t-1)$ **(b)** $\displaystyle\int_{-\infty}^{\infty} e^{-t}\delta(t-1)\,dt$ **(c)** $e^{-t}u(t) \star \delta(t-1)$

6.21 (Convolution) Compute and sketch the convolution of the following pairs of signals.

(a) $x(t) = \displaystyle\sum_{k=-\infty}^{\infty} \delta(t-k)$ $h(t) = \text{rect}(t)$

(b) $x(t) = \displaystyle\sum_{k=-\infty}^{\infty} \delta(t-3k)$ $h(t) = \text{tri}(t)$

(c) $x(t) = \displaystyle\sum_{k=-\infty}^{\infty} \text{rect}(t-2k)$ $h(t) = \text{rect}(t)$

6.22 (Impulse Response and Step Response) Find the step response $s(t)$ of each system whose impulse response $h(t)$ is given.

(a) $h(t) = \text{rect}(t-0.5)$ **(b)** $h(t) = \sin(2\pi t)u(t)$
(c) $h(t) = \sin(2\pi t)\text{rect}(t-0.5)$ **(d)** $h(t) = e^{-|t|}$

6.23 (Convolution and System Response) Consider a system described by the differential equation $y'(t) + 2y(t) = x(t)$.

(a) What is the impulse response $h(t)$ of this system?
(b) Find its output if $x(t) = e^{-2t}u(t)$ by convolution.
(c) Find its output if $x(t) = e^{-2t}u(t)$ and $y(0) = 0$ by solving the differential equation.
(d) Find its output if $x(t) = e^{-2t}u(t)$ and $y(0) = 1$ by solving the differential equation.
(e) Are any of the outputs identical? Should they be? Explain.

6.24 (System Response) Consider the two inputs and two circuits shown in Figure P6.24.

(a) Find the impulse response of each circuit.
(b) Use convolution to find the response of circuit 1 to input 1. Assume $R = 1\,\Omega$, $C = 1$ F.
(c) Use convolution to find the response of circuit 2 to input 1. Assume $R = 1\,\Omega$, $C = 1$ F.
(d) Use convolution to find the response of circuit 1 to input 2. Assume $R = 1\,\Omega$, $C = 1$ F.
(e) Use convolution to find the response of circuit 2 to input 2. Assume $R = 1\,\Omega$, $C = 1$ F.
(f) Use convolution to find the response of circuit 1 to input 1. Assume $R = 2\,\Omega$, $C = 1$ F.
(g) Use convolution to find the response of circuit 1 to input 2. Assume $R = 2\,\Omega$, $C = 1$ F.

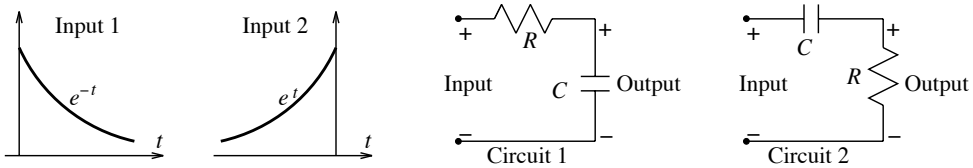


Figure P6.24 The circuits for Problem 6.24

6.25 (Impulse Response and Step Response) The step response of a system is $s(t) = \delta(t)$. The input to the system is a periodic square wave described for one period by $x(t) = \text{sgn}(t)$, $-1 \le t \le 1$. Sketch the system output.

PGR2026-00033
Krisp Technologies EX1024 Page 177

6.26 (Impulse Response and Step Response) The input to a system is a periodic square wave with period $T = 2$ s described for one period by $x_p(t) = -\text{sgn}(t),\ -1 \le t \le 1$. The output is a periodic triangular wave described by $y_p(t) = \text{tri}(t) - 0.5,\ -1 \le t \le 1$. What is the impulse response of the system? What is the response of this system to the single pulse $x(t) = \text{rect}(t - 0.5)$?

6.27 (Convolution) An RC lowpass filter has the impulse response $h(t) = \frac{1}{\tau}e^{-t/\tau}u(t)$, where τ is the time constant. Find its response to the following inputs for $\tau = 0.5$ and $\tau = 1$.

(a) $x(t) = e^{-2t}u(t)$ **(b)** $x(t) = e^{2t}u(-t)$ **(c)** $x(t) = e^{-2|t|}$

6.28 (Convolution) Find the convolution of each pair of signals.

(a) $x(t) = e^{-|t|}$ $h(t) = e^{-|t|}$
(b) $x(t) = e^{t}u(-t) - e^{-t}u(t)$ $h(t) = x(t)$
(c) $x(t) = e^{t}u(-t) - e^{-t}u(t)$ $h(t) = x(-t)$

6.29 (Convolution by Ranges) Consider a series RC lowpass filter with $\tau = 1$. Use convolution by ranges to find the capacitor voltage, its maximum value, and time of maximum for each input $x(t)$.

(a) $x(t) = \text{rect}(t - 0.5)$ **(b)** $x(t) = t\,\text{rect}(t - 0.5)$ **(c)** $x(t) = (1 - t)\text{rect}(t - 0.5)$

6.30 (Cascading) The impulse response of two cascaded systems equals the convolution of their impulse responses. Does the step response $s_C(t)$ of two cascaded systems equal $s_1(t) \star s_2(t)$, the convolution of their step responses? If not, how is $s_C(t)$ related to $s_1(t)$ and $s_2(t)$?

6.31 (Cascading) System 1 compresses a signal by a factor of 2, and system 2 is an RC lowpass filter with $\tau = 1$. Find the output of each cascaded combination. Will their output be identical? Should it be? Explain.

(a) $x(t) = 2e^{-t}u(t) \longrightarrow \boxed{\text{system 1}} \longrightarrow \boxed{\text{system 2}} \longrightarrow y(t)$

(b) $x(t) = 2e^{-t}u(t) \longrightarrow \boxed{\text{system 2}} \longrightarrow \boxed{\text{system 1}} \longrightarrow y(t)$

6.32 (Cascading) System 1 is a squaring circuit, and system 2 is an RC lowpass filter with $\tau = 1$. Find the output of each cascaded combination. Will their output be identical? Should it be? Explain.

(a) $x(t) = 2e^{-t}u(t) \longrightarrow \boxed{\text{system 1}} \longrightarrow \boxed{\text{system 2}} \longrightarrow y(t)$

(b) $x(t) = 2e^{-t}u(t) \longrightarrow \boxed{\text{system 2}} \longrightarrow \boxed{\text{system 1}} \longrightarrow y(t)$

6.33 (Cascading) System 1 is a highpass RC circuit with $h(t) = \delta(t) - e^{-t}u(t)$, and system 2 is an RC lowpass filter with $\tau = 1$. Find the output of each cascaded combination. Will their output be identical? Should it be? Explain.

(a) $x(t) = 2e^{-t}u(t) \longrightarrow \boxed{\text{system 1}} \longrightarrow \boxed{\text{system 2}} \longrightarrow y(t)$

(b) $x(t) = 2e^{-t}u(t) \longrightarrow \boxed{\text{system 2}} \longrightarrow \boxed{\text{system 1}} \longrightarrow y(t)$

6.34 (Cascading) System 1 is a highpass RC circuit with $h(t) = \delta(t) - e^{-t}u(t)$, and system 2 is an RC lowpass filter with $\tau = 1$.

(a) Find the impulse response $h_P(t)$ of their parallel connection.
(b) Find the impulse response $h_{12}(t)$ of the cascade of system 1 and system 2.
(c) Find the impulse response $h_{21}(t)$ of the cascade of system 2 and system 1.
(d) Are $h_{12}(t)$ and $h_{21}(t)$ identical? Should they be? Explain.
(e) Find the impulse response $h_I(t)$ of a system whose parallel connection with $h_{12}(t)$ yields $h_P(t)$.

PGR2026-00033
Krisp Technologies EX1024 Page 178

6.35 (Cascading) System 1 is described by $y(t) = x'(t) + x(t)$, and system 2 is an RC lowpass filter with $\tau = 1$.

(a) What is the output of the cascaded system to the input $x(t) = 2e^{-t}u(t)$?

(b) What is the output of the cascaded system to the input $x(t) = \delta(t)$?

(c) How are system 1 and system 2 related? Should they be? Explain.

6.36 (Cascading) Consider the three circuits shown in Figure P6.36.

(a) Find their impulse responses $h_1(t)$, $h_2(t)$, and $h_3(t)$.

(b) What is the impulse response $h_{12}(t)$ of the ideal cascade of the first two circuits? Does this equal the impulse response $h_3(t)$? Explain.

(c) What value of R will ensure an impulse response $h_3(t)$ that differs from $h_{12}(t)$ by no more than 1% at $t = 0$?

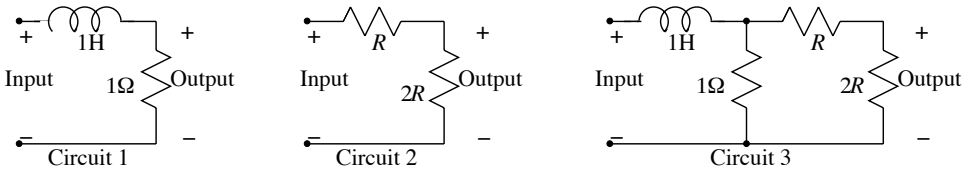


Figure P6.36 The circuits for Problem 6.36

6.37 (Stability and Causality) Check for the causality and stability of each of the following systems.

(a) $h(t) = e^{-(t+1)}u(t)$ **(b)** $h(t) = e^{-t-1}u(t+1)$

(c) $h(t) = \delta(t) - e^{-t}u(t)$ **(d)** $h(t) = \delta(t) - e^{t}u(-1-t)$

6.38 (Stability and Causality) Check for the causality and stability of the parallel connection and cascade connection of each pair of systems.

(a)	$h_1(t) = e^{-t}u(t)$	$h_2(t) = \delta(t)$
(b)	$h_1(t) = e^{-t+3}u(t-3)$	$h_2(t) = \delta(t+2)$
(c)	$h_1(t) = e^{-t}u(t)$	$h_2(t) = e^{-t+2}u(t-1)$
(d)	$h_1(t) = e^{-t}u(t)$	$h_2(t) = e^{t}u(-t)$
(e)	$h_1(t) = e^{-\lvert t\rvert}$	$h_2(t) = e^{-\lvert t-1\rvert}$
(f)	$h_1(t) = e^{\lvert t\rvert}$	$h_2(t) = e^{-\lvert t-1\rvert}$
(g)	$h_1(t) = e^{\lvert t-1\rvert}$	$h_2(t) = -e^{-\lvert t-1\rvert}$

6.39 (Stability) Investigate the stability of the following systems.

(a) $y'(t) = x(t)$ **(b)** $y'(t) + \alpha y(t) = x(t)$ **(c)** $y^{(n)}(t) = x(t)$ **(d)** $y(t) = x^{(n)}(t)$

6.40 (Convolution and System Classification) The impulse response of three systems is

$$h_1(t) = 2\delta(t) \qquad h_2(t) = \delta(t) + \delta(t-3) \qquad h_3(t) = e^{-t}u(t)$$

(a) Find the response of each to the input $x(t) = u(t) - u(t-1)$.

(b) For system 1, the input is zero at $t = 2$ s, and so is the response. Does the statement "zero output if zero input" apply to dynamic or instantaneous systems or both? Explain.

(c) Argue that system 1 is instantaneous. What about the other two?

(d) What must be the form of $h(t)$ for an instantaneous system?

PGR2026-00033
Krisp Technologies EX1024 Page 179

6.41 (Convolution and Smoothing) Convolution is usually a smoothing operation unless one signal is an impulse or its derivative, but exceptions occur even for smooth signals. Evaluate and comment on the duration and smoothing effects of the following convolutions.

(a) $y(t) = \text{rect}(t) \star \text{tri}(t)$ **(b)** $y(t) = \text{rect}(t) \star \delta(t)$
(c) $y(t) = \text{rect}(t) \star \delta'(t)$ **(d)** $y(t) = \text{sinc}(t) \star \text{sinc}(t)$
(e) $y(t) = e^{-\pi t^2} \star e^{-\pi t^2}$ **(f)** $y(t) = \sin(2\pi t) \star \text{rect}(t)$

6.42 (Eigensignals) The input $x(t)$ and response $y(t)$ of two systems are given. Which of the systems are linear, and why?

(a) $x(t) = \cos(t), \quad y(t) = 0.5\sin(t - 0.25\pi)$
(b) $x(t) = \cos(t), \quad y(t) = \cos(2t)$

6.43 (Eigensignals) If the input to a system is its eigensignal, the response has the same form as the eigensignal. Justify the following statements by computing the system response by convolution (if the impulse response is given) or by solving the given differential equation. You may pick convenient numerical values for the parameters.

(a) Every signal is an eigensignal of the system described by $h(t) = A\delta(t)$.
(b) The signal $x(t) = e^{j\alpha t}$ is an eigensignal of any LTI system such as that described by the impulse response $h(t) = e^{-\beta t}u(t)$.
(c) The signal $x(t) = \cos(\alpha t)$ is an eigensignal of any LTI system described by a differential equation such as $y'(t) + \beta y(t) = x(t)$.
(d) The signal $x(t) = \text{sinc}(\alpha t)$ is an eigensignal of ideal filters described by $h(t) = \text{sinc}(\beta t), \ \beta \geq \alpha$.

6.44 (Eigensignals) Which of the following can be the eigensignal of an LTI system?

(a) $x(t) = e^{-2t}u(t)$ **(b)** $x(t) = e^{-j2t}$ **(c)** $x(t) = \cos(2t)$ **(d)** $x(t) = e^{jt} + e^{j2t}$

6.45 (Stability) Investigate the causality and stability of the following systems.

(a) $h(t) = u(t)$ **(b)** $h(t) = e^{-2t}u(t)$ **(c)** $h(t) = \delta(t-1)$
(d) $h(t) = \text{rect}(t)$ **(e)** $h(t) = \text{sinc}(t)$ **(f)** $h(t) = \text{sinc}^2(t)$

6.46 (Invertibility) Determine which of the following systems are invertible and, for those that are, find the impulse response of the inverse system.

(a) $h(t) = e^{-2t}u(t)$ **(b)** $h(t) = \delta(t-1)$ **(c)** $h(t) = \text{sinc}(t)$

6.47 (Periodic Extension) The periodic extension $x(t)$ with period T has the same form and area as $x(t)$. Use this concept to find the constants in the following assumed form for $x_{\text{pe}}(t)$ of each signal.

(a) Signal: $x(t) = e^{-t/\tau}u(t)$ Periodic extension for $0 \leq t \leq T$: $x_{\text{pe}}(t) = Ke^{-t/\tau}$
(b) Signal: $x(t) = te^{-t/\tau}u(t)$ Periodic extension for $0 \leq t \leq T$: $x_{\text{pe}}(t) = (A + Bt)e^{-t/\tau}$

6.48 (The Duration Property) The convolution duration usually equals the sum of the durations of the convolved signals. But consider the following convolutions.

$$y_1(t) = u(t) \star \sin(\pi t)[u(t) - u(t-2)] \qquad y_2(t) = \text{rect}(t) \star \cos(2\pi t)u(t)$$

(a) Evaluate each convolution and find its duration. Is the duration infinite? If not, what causes it to be finite?
(b) In the first convolution, replace the sine pulse by an arbitrary signal $x(t)$ of zero area and finite duration T_d and argue that the convolution is nonzero only for a duration T_d.

PGR2026-00033
Krisp Technologies EX1024 Page 180

(c) In the second convolution, replace the cosine by an arbitrary periodic signal $x_p(t)$ with zero average value and period $T = 1$. Argue that the convolution is nonzero for only for 1 unit.

6.49 (Convolutions that Replicate) Signals that replicate under self-convolution include the impulse, sinc, Gaussian, and Lorentzian. For each of the following known results, determine the constant A using the area property of convolution.

(a) $\delta(\alpha t) \star \delta(\alpha t) = A\delta(\alpha t)$

(b) $\text{sinc}(\alpha t) \star \text{sinc}(\alpha t) = A\,\text{sinc}(\alpha t)$

(c) $e^{-\pi t^2} \star e^{-\pi t^2} = Ae^{-\pi t^2/2}$

(d) $\dfrac{1}{1+t^2} \star \dfrac{1}{1+t^2} = \dfrac{A}{1+0.25t^2}$

6.50 (Convolution and Moments) For each of the following signal pairs, find the moments m_0, m_1, and m_2 and verify each of the convolution properties based on moments (as discussed in the text).

(a) $x(t) = h(t)$ $\quad h(t) = \text{rect}(t)$

(b) $x(t) = e^{-t}u(t)$ $\quad h(t) = e^{-2t}u(t)$

6.51 (Central Limit Theorem) Show that the n-fold repeated convolution of the signal $h(t) = e^{-t}u(t)$ with itself has the form

$$h_n(t) = \frac{t^n e^{-t}}{n!}u(t).$$

(a) Show that $h_n(t)$ has a maximum at $t = n$.

(b) Assume a Gaussian approximation $h_n(t) \approx g_n(t) = K\exp[-\alpha(t-n)^2]$. Equating $h_n(t)$ and $g_n(t)$ at $t = n$, show that $K = (e^{-n}n^n/n!)$.

(c) Use the **Stirling limit** to show that $K = \sqrt{1/(2\pi n)}$. The Stirling limit is defined by

$$\lim_{n\to\infty}\left[\sqrt{n}\frac{n^n}{n!}e^{-n}\right] = \frac{1}{\sqrt{2\pi}}$$

(d) Show that $\alpha = 1/(2n)$ by equating the areas of $h_n(t)$ and $g_n(t)$.

6.52 (Matched Filters) A folded, shifted version of a signal $s(t)$ defines the impulse response of a **matched filter** corresponding to $s(t)$.

(a) Find and sketch the impulse response of a matched filter for the signal $s(t) = u(t) - u(t-1)$.

(b) Find the response $y(t)$ of this matched filter to the signal $x(t) = s(t-D)$, where $D = 2$ s corresponds to the signal delay.

(c) At what time t_m does the response $y(t)$ attain its maximum value, and how is t_m related to the signal delay D?

6.53 (Autocorrelation Functions) A signal $x(t)$ can qualify as an autocorrelation function only if it satisfies certain properties. Which of the following qualify as valid autocorrelation functions, and why?

(a) $r_{xx}(t) = e^{-\alpha t}u(t)$ **(b)** $r_{xx}(t) = e^{-\alpha|t|}$ **(c)** $r_{xx}(t) = te^{-\alpha t}u(t)$

(d) $r_{xx}(t) = |t|e^{-\alpha|t|}$ **(e)** $r_{xx}(t) = \text{sinc}^2(\alpha t)$ **(f)** $r_{xx}(t) = \dfrac{1}{1+t^2}$

(g) $r_{xx}(t) = \dfrac{t}{1+t^2}$ **(h)** $r_{xx}(t) = \dfrac{1+t^2}{4+t^2}$ **(i)** $r_{xx}(t) = \dfrac{t^2-1}{t^2+4}$

PGR2026-00033
Krisp Technologies EX1024 Page 181

6.54 (Correlation) Find the cross-correlations $r_{xh}(t)$ and $r_{hx}(t)$ for each pair of signals.

(a) $x(t) = e^{-t}u(t)$ $\quad h(t) = e^{-t}u(t)$
(b) $x(t) = e^{-t}u(t)$ $\quad h(t) = e^{t}u(-t)$
(c) $x(t) = e^{-|t|}$ $\quad h(t) = e^{-|t|}$
(d) $x(t) = e^{-(t-1)}u(t-1)$ $\quad h(t) = e^{-t}u(t)$

COMPUTATION AND DESIGN

ctcongui A GUI for Visualization of the Convolution Process

The graphical user interface `ctcongui` allows you to visualize a movie of the convolution or correlation of two analog signals.
You can select the signals and the operation (convolution or correlation).
The interface displays a movie of how the product function and the convolution or correlation result changes as one function slides past the other.
Use the mouse to manually drag one function past the other and see the results change.
To explore this routine, type `ctcongui` at the MATLAB prompt.

6.55 (Animation of Convolution) Use `ctcongui` to animate the convolution of each of the following pairs of signals and determine whether it is possible to visually identify the convolution ranges.

(a) $x(t) = \text{rect}(t)$ $\quad h(t) = \text{rect}(t)$
(b) $x(t) = \text{rect}(t)$ $\quad h(t) = e^{-t}[u(t) - u(t-5)]$
(c) $x(t) = \text{rect}(t)$ $\quad h(t) = \text{tri}(t)$
(d) $x(t) = e^{-t}u(t)$ $\quad h(t) = (1-t)[u(t) - u(t-1)]$
(e) $x(t) = e^{-t}u(t)$ $\quad h(t) = t[u(t) - u(t-1)]$

6.56 (Convolution of Sinc Functions) It is claimed that the convolution of the two identical signals $x(t) = h(t) = \text{sinc}(t)$ is $y(t) = x(t) \star h(t) = \text{sinc}(t)$. Let both $x(t)$ and $h(t)$ be described over the symmetric limits $-\alpha \le t \le \alpha$. Use `ctcongui` to animate the convolution of $x(t)$ and $h(t)$ for $\alpha = 1$, $\alpha = 5$, and $\alpha = 10$ (you may want to choose a smaller time step in `ctcongui` for larger values of α). Does the convolution begin to approach the required result as α increases?

PGR2026-00033
Krisp Technologies EX1024 Page 182

Chapter 7

DISCRETE CONVOLUTION

7.0 Scope and Objectives

Convolution is a central concept in relating the time and frequency domains. In the time domain, convolution may be regarded as a method of finding the zero-state response of a relaxed linear time-invariant system. This chapter describes the operation and properties of discrete convolution and establishes the key connections between time-domain and transformed-domain methods for signal and system analysis based on convolution. Frequency-domain viewpoints of discrete convolution are presented later, both in this chapter and others.

7.1 Discrete Convolution

Discrete-time convolution is a method of finding the *zero-state response of relaxed* linear time-invariant (LTI) systems. It is based on the concepts of linearity and time invariance and assumes that the system information is known in terms of its impulse response $h[n]$. In other words, if the input is $\delta[n]$, a unit sample at the origin $n = 0$, the system response is $h[n]$. Now, if the input is $x[0]\delta[n]$, a scaled impulse at the origin, the response is $x[0]h[n]$ (by linearity). Similarly, if the input is the shifted impulse $x[1]\delta[n-1]$ at $n = 1$, the response is $x[1]h[n-1]$ (by time invariance). The response to the shifted impulse $x[k]\delta[n-k]$ at $n = k$ is $x[k]h[n-k]$. Since an arbitrary input $x[n]$ is simply a sequence of samples, it can be described by a sum of scaled and shifted impulses:

$$x[n] = \sum_{k=-\infty}^{\infty} x[k]\delta[n-k] \tag{7.1}$$

By superposition, the response to $x[n]$ is the sum of scaled and shifted versions of the impulse response:

$$y[n] = \sum_{k=-\infty}^{\infty} x[k]h[n-k] = x[n] \star h[n] \tag{7.2}$$

This is the defining relation for the convolution operation, which we call **linear convolution**, and denote by $y[n] = x[n] \star h[n]$ (or by $x[n] * h[n]$ in the figures) in this book. The expression for computing $y[n]$ is called the **convolution sum**. As with continuous-time convolution, the order in which we perform the operation does not matter, and we can interchange the arguments of x and h without affecting the result. Thus,

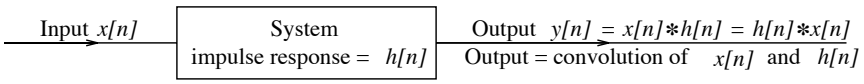
$$y[n] = \sum_{k=-\infty}^{\infty} x[n-k]h[k] = h[n] \star x[n] \tag{7.3}$$

Notation: We use $x[n] \star h[n]$ to denote $\displaystyle\sum_{k=-\infty}^{\infty} x[k]h[n-k]$

PGR2026-00033
Krisp Technologies EX1024 Page 183

REVIEW PANEL 7.1

Convolution Yields the Zero-State Response of an LTI System



7.1.1 Analytical Evaluation of Discrete Convolution

The procedure for analytical convolution mimics the continuous case and may be implemented quite readily if $x[n]$ and $h[n]$ are described by simple enough analytical expressions. Often, we must resort to a table of closed-form solutions for finite or infinite series. When evaluating the convolution sum, keep in mind that $x[k]$ and $h[n-k]$ are functions of the summation variable k. The summations frequently involve step functions of the form $u[k]$ and $u[n-k]$. Since $u[k]=0,\ k<0$ and $u[n-k]=0,\ k>n$, these can be used to simplify the lower and upper summation limits to $k=0$ and $k=n$, respectively.

EXAMPLE 7.1 (Analytical Evaluation of Discrete Convolution)

(a) Let $x[n]=h[n]=u[n]$. Then $x[k]=u[k]$ and $h[n-k]=u[n-k]$. The lower limit on the convolution sum simplifies to $k=0$ (because $u[k]=0,\ k<0$), the upper limit to $k=n$ (because $u[n-k]=0,\ k>n$), and we get

$$y[n]=\sum_{k=-\infty}^{\infty}u[k]u[n-k]=\sum_{k=0}^{n}1=(n+1)u[n]=r[n+1]$$

Note that $(n+1)u[n]$ also equals $r[n+1]$, and thus $u[n]\star u[n]=r[n+1]$.

(b) Let $x[n]=h[n]=a^nu[n]$, $a<1$. Then $x[k]=a^ku[k]$ and $h[n-k]=a^{n-k}u[n-k]$. The lower limit on the convolution sum simplifies to $k=0$ (because $u[k]=0,\ k<0$), the upper limit to $k=n$ (because $u[n-k]=0,\ k>n$), and we get

$$y[n]=\sum_{k=-\infty}^{\infty}a^ka^{n-k}u[k]u[n-k]=\sum_{k=0}^{n}a^ka^{n-k}=a^n\sum_{k=0}^{n}1=(n+1)a^nu[n]$$

(c) Let $x[n]=u[n]$ and $h[n]=\alpha^nu[n],\ \alpha<1$. Then

$$u[n]\star\alpha^nu[n]=\sum_{k=-\infty}^{\infty}\alpha^ku[k]u[n-k]=\sum_{k=0}^{n}\alpha^k=\frac{1-\alpha^{n+1}}{1-\alpha},\quad|\alpha|<1$$

(d) Let $x[n]=(0.8)^nu[n]$ and $h[n]=(0.4)^nu[n]$. Then

$$y[n]=\sum_{k=-\infty}^{\infty}(0.8)^ku[k](0.4)^{n-k}u[n-k]=\sum_{k=0}^{n}(0.8)^k(0.4)^{n-k}=(0.4)^n\sum_{k=0}^{n}2^k$$

Using the closed-form result for the sum, we get $y[n]=(0.4)^n\dfrac{1-2^{n+1}}{1-2}=(0.4)^n(2^{n+1}-1)u[n]$.

PGR2026-00033
Krisp Technologies EX1024 Page 184

(e) Let $x[n] = nu[n+1]$ and $h[n] = a^{-n}u[n],\ a < 1$. With $h[n-k] = a^{-(n-k)}u[n-k]$ and $x[k] = ku[k+1]$, the lower and upper limits on the convolution sum become $k = -1$ and $k = n$. Then

$$\begin{aligned} y[n] &= \sum_{k=-1}^{n} ka^{-(n-k)} = -a^{-n-1} + a^{-n}\sum_{k=0}^{n} ka^{k} \\ &= -a^{-n-1} + \frac{a^{-n+1}}{(1-a)^2}[1-(n+1)a^n + na^{n+1}] \end{aligned}$$

The results of analytical discrete convolution often involve finite or infinite summations. In the last part, we used known results to generate the closed-form solution. This may not always be feasible, however.

7.2 Convolution Properties

Many of the properties of discrete convolution are based on linearity and time invariance. For example, if $x[n]$ *or* $h[n]$ is shifted by n_0, so is $y[n]$. Thus, if $y[n] = x[n] \star h[n]$, then

$$x[n-n_0] \star h[n] = x[n] \star h[n-n_0] = y[n-n_0] \tag{7.4}$$

The sum of the samples in $x[n],\ h[n]$, and $y[n]$ are related by

$$\sum_{n=-\infty}^{\infty} y[n] = \left(\sum_{n=-\infty}^{\infty} x[n]\right)\left(\sum_{n=-\infty}^{\infty} h[n]\right) \tag{7.5}$$

For causal systems ($h[n] = 0,\ n < 0$) and causal signals ($x[n] = 0,\ n < 0$), $y[n]$ is also causal. Thus,

$$y[n] = x[n] \star h[n] = h[n] \star x[n] = \sum_{k=0}^{n} x[k]h[n-k] = \sum_{k=0}^{n} h[k]x[n-k] \tag{7.6}$$

An extension of this result is that the convolution of two left-sided signals is also left-sided and the convolution of two right-sided signals is also right-sided.

EXAMPLE 7.2 (Properties of Convolution)

(a) Here are two useful convolution results that are readily found from the defining relation:

$$\delta[n] \star x[n] = x[n] \qquad \delta[n] \star \delta[n] = \delta[n]$$

(b) We find $y[n] = u[n] \star x[n]$. Since the step response is the running sum of the impulse response, the convolution of a signal $x[n]$ with a unit step is the running sum of the signal $x[n]$:

$$x[n] \star u[n] = \sum_{k=-\infty}^{n} x[k]$$

(c) We find $y[n] = \text{rect}(n/2N) \star \text{rect}(n/2N)$ where $\text{rect}(n/2N) = u[n+N] - u[n-N-1]$.

The convolution contains four terms:

$$y[n] = u[n+N] \star u[n+N] - u[n+N] \star u[n-N-1] - u[n-N-1] \star u[n+N] + u[n-N-1] \star u[n-N-1]$$

PGR2026-00033
Krisp Technologies EX1024 Page 185

Using the result $u[n] \star u[n] = r[n+1]$ and the shifting property, we obtain

$$y[n] = r[n+2N+1] - 2r[n] + r[n-2N-1] = (2N+1)\text{tri}\left(\frac{n}{2N+1}\right)$$

The convolution of two rect functions (with identical arguments) is thus a tri function.

7.3 Convolution of Finite Sequences

In practice, we often deal with sequences of finite length, and their convolution may be found by several methods. The convolution $y[n]$ of two finite-length sequences $x[n]$ and $h[n]$ is also of finite length and is subject to the following rules, which serve as useful consistency checks:

1. The starting index of $y[n]$ equals the sum of the starting indices of $x[n]$ and $h[n]$.
2. The ending index of $y[n]$ equals the sum of the ending indices of $x[n]$ and $h[n]$.
3. The length L_y of $y[n]$ is related to the lengths L_x and L_h of $x[n]$ and $h[n]$ by $L_y = L_x + L_h - 1$.

7.3.1 The Sum-by-Column Method

This method is based on the idea that the convolution $y[n]$ equals the sum of the (shifted) impulse responses due to each of the impulses that make up the input $x[n]$. To find the convolution, we set up a row of index values beginning with the starting index of the convolution and $h[n]$ and $x[n]$ below it. We regard $x[n]$ as a sequence of weighted shifted impulses. Each element (impulse) of $x[n]$ generates a shifted impulse response (product with $h[n]$) starting at its index (to indicate the shift). Summing the response (by columns) gives the discrete convolution. *Note that none of the sequences is folded.* It is better (if only to save paper) to let $x[n]$ be the shorter sequence. The starting index (and the marker location corresponding to $n = 0$) for the convolution $y[n]$ is found from the starting indices of $x[n]$ and $h[n]$.

REVIEW PANEL 7.2

Discrete Convolution Using the Sum-by-Column Method

1. Line up the sequence $x[n]$ below the sequence $h[n]$.
2. Line up with each sample of $x[n]$, the product of the entire array $h[n]$ with that sample of $x[n]$.
3. Sum the columns of the (successively shifted) arrays to generate the convolution sequence.

EXAMPLE 7.3 (Convolution of Finite-Length Signals)

(a) An FIR (finite impulse response) filter has an impulse response given by $h[n] = \{\overset{\Downarrow}{1},\ 2,\ 2,\ 3\}$. Find its response $y[n]$ to the input $x[n] = \{\overset{\Downarrow}{2},\ -1,\ 3\}$. Assume that both $x[n]$ and $h[n]$ start at $n = 0$.

The paper-and-pencil method expresses the input as $x[n] = 2\delta[n] - \delta[n-1] + 3\delta[n-2]$ and tabulates the response to each impulse and the total response as follows:

$h[n]$		=	1	2	2	3		
$x[n]$		=	2	−1	3			
Input	**Response**							
$2\delta[n]$	$2h[n]$	=	2	4	4	6		
$-\delta[n-1]$	$-h[n-1]$	=		−1	−2	−2	−3	
$3\delta[n-2]$	$3h[n-2]$	=			3	6	6	9
Sum $= x[n]$	Sum $= y[n]$	=	2	3	5	10	3	9

PGR2026-00033
Krisp Technologies EX1024 Page 186

So, $y[n] = \{\overset{\Downarrow}{2},\ 3,\ 5,\ 10,\ 3,\ 9\} = 2\delta[n] + 3\delta[n-1] + 5\delta[n-2] + 10\delta[n-3] + 3\delta[n-4] + 9\delta[n-5]$.

The convolution process is illustrated graphically in Figure E7.3A.

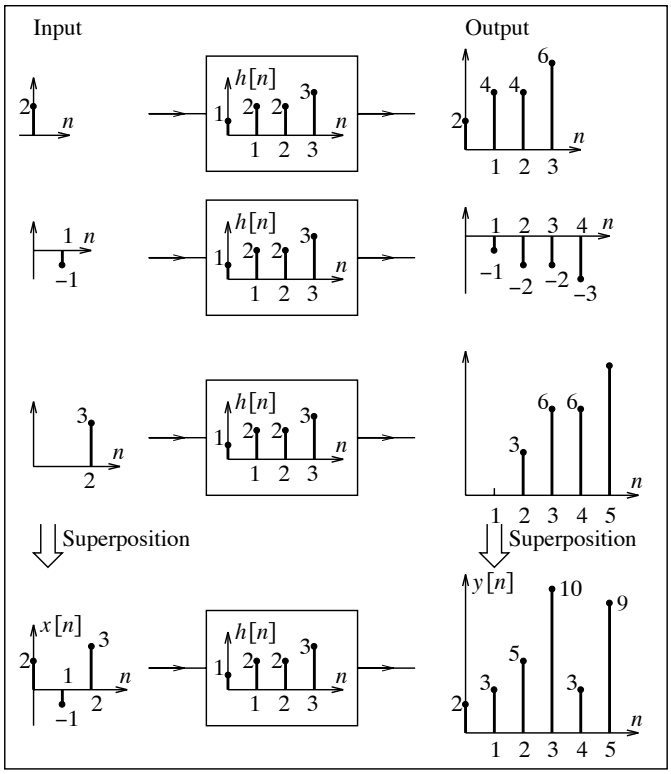


Figure E7.3A The discrete convolution for Example 7.3(a)

(b) Let $h[n] = \{2, 5, \overset{\Downarrow}{0}, 4\}$ and $x[n] = \{4, \overset{\Downarrow}{1}, 3\}$.

We note that the convolution starts at $n = -3$ and use this to set up the index array and generate the convolution as follows:

n	-3	-2	-1	0	1	2
$h[n]$	2	5	0	4		
$x[n]$	4	1	3			
	8	20	0	16		
		2	5	0	4	
			6	15	0	12
$y[n]$	8	22	11	$\overset{\Downarrow}{31}$	4	12

The marker is placed by noting that the convolution starts at $n = -3$, and we get

$$y[n] = \{8,\ 22,\ 11,\ \overset{\Downarrow}{31},\ 4,\ 12\}$$

PGR2026-00033
Krisp Technologies EX1024 Page 187

(c) (Response of an Averaging Filter) Let $x[n] = \{\overset{\Downarrow}{2}, 4, 6, 8, 10, 12, \ldots\}$.

What system will result in the response $y[n] = \{\overset{\Downarrow}{1}, 3, 5, 7, 9, 11, \ldots\}$?

At each instant, the response is the average of the input and its previous value. This system describes an averaging or moving average filter. Its difference equation is simply $y[n] = \frac{1}{2}(x[n] + x[n-1])$.

Its impulse response is thus $h[n] = 0.5\{\delta[n] + \delta[n-1]\}$, or $h[n] = \{\overset{\Downarrow}{0.5},\ 0.5\}$.

Using discrete convolution, we find the response as follows:

x:	2	4	6	8	10	12	...	
h:	$\frac{1}{2}$	$\frac{1}{2}$						
	1	2	3	4	5	6	...	
		1	2	3	4	5	6	...
y:	1	3	5	7	9	11	...	

This result is indeed the averaging operation we expected.

7.3.2 The Fold, Shift, Multiply, and Sum Concept

The convolution sum may also be interpreted as follows. We fold $x[n]$ and shift $x[-n]$ to line up its last element with the first element of $h[n]$. We then successively shift $x[-n]$ (to the right) past $h[n]$, one index at a time, and find the convolution at each index as the sum of the pointwise products of the overlapping samples. One method of computing $y[n]$ is to list the values of the folded function on a strip of paper and slide it along the stationary function, to better visualize the process. This technique has prompted the name **sliding strip method**. We simulate this method by showing the successive positions of the stationary and folded sequence along with the resulting products, the convolution sum, and the actual convolution.

EXAMPLE 7.4 (Convolution by the Sliding Strip Method)

Find the convolution of $h[n] = \{\overset{\Downarrow}{2}, 5, 0, 4\}$ and $x[n] = \{\overset{\Downarrow}{4}, 1, 3\}$.

Since both sequences start at $n = 0$, the folded sequence is $x[-k] = \{3, 1, \overset{\Downarrow}{4}\}$.
We line up the folded sequence below $h[n]$ to begin overlap and shift it successively, summing the product sequence as we go, to obtain the discrete convolution. The results are computed in Figure E7.4.

The discrete convolution is $y[n] = \{\overset{\Downarrow}{8}, 22, 11, 31, 4, 12\}$.

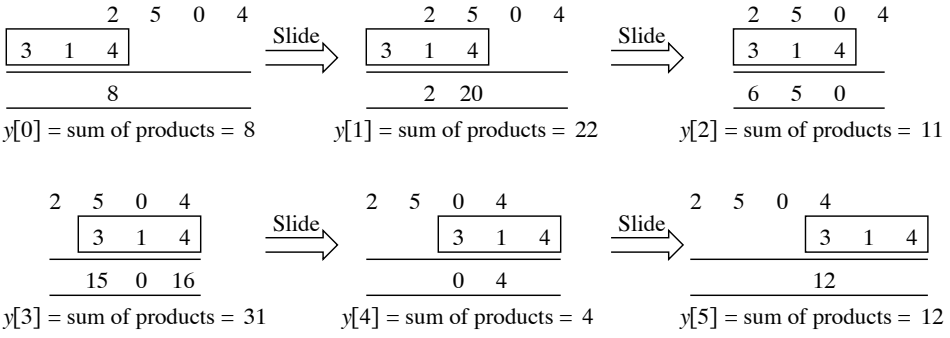


Figure E7.4 The discrete signals for Example 7.4 and their convolution

PGR2026-00033
Krisp Technologies EX1024 Page 188

7.3.3 Discrete Convolution, Multiplication, and Zero Insertion

The discrete convolution of two *finite-length* sequences $x[n]$ and $h[n]$ is entirely equivalent to multiplication of two polynomials whose coefficients are described by the arrays $x[n]$ and $h[n]$ (in ascending or descending order). The convolution sequence corresponds to the coefficients of the product polynomial. Based on this result, if we insert zeros between adjacent samples of each signal to be convolved, their convolution corresponds to the original convolution sequence with zeros inserted between its adjacent samples.

EXAMPLE 7.5 (Polynomial Multiplication and Zero Insertion)

(a) Let $h[n] = \{\overset{\Downarrow}{2}, 5, 0, 4\}$ and $x[n] = \{\overset{\Downarrow}{4}, 1, 3\}$. To find their convolution, we set up the polynomials

$$h(z) = 2z^3 + 5z^2 + 0z + 4 \qquad x(z) = 4z^2 + 1z + 3$$

Their product is $y(z) = 8z^5 + 22z^4 + 11z^3 + 31z^2 + 4z + 12$.

The convolution is thus $y[n] = \{\overset{\Downarrow}{8}, 22, 11, 31, 4, 12\}$.

(b) Zero insertion of each convolved sequence gives

$$h_1[n] = \{\overset{\Downarrow}{2}, 0, 5, 0, 0, 0, 4\} \qquad x_1[n] = \{\overset{\Downarrow}{4}, 0, 1, 0, 3\}$$

To find their convolution, we set up the polynomials

$$h_1(z) = 2z^6 + 5z^4 + 0z^2 + 4 \qquad x_1(z) = 4z^4 + 1z^2 + 3$$

Their product is $y_1(z) = 8z^{10} + 22z^8 + 11z^6 + 31z^4 + 4z^2 + 12$.

The convolution is then $y_1[n] = \{\overset{\Downarrow}{8}, 0, 22, 0, 11, 0, 31, 0, 4, 0, 12\}$.

This result is just $y[n]$ with zeros inserted between adjacent samples.

REVIEW PANEL 7.3

Convolution of Finite-Length Signals Corresponds to Polynomial Multiplication

Example: $\{\overset{\Downarrow}{1},\ 1,\ 3\} \star \{\overset{\Downarrow}{1},\ 0,\ 2\} = \{\overset{\Downarrow}{1},\ 1,\ 5,\ 2,\ 6\}$ $\qquad (x^2 + x + 3)(x^2 + 2) = x^4 + x^3 + 5x^2 + 2x + 6$

If we append zeros to one of the convolved sequences, the convolution result will not change but will show as many appended zeros. In fact, the zeros may be appended to the beginning or end of a sequence and will will appear at the corresponding location in the convolution result.

REVIEW PANEL 7.4

Zero-Padding: If One Sequence Is Zero-Padded, So Is the Convolution

Example: If $x[n] \star h[n] = y[n]$ then $\{0,\ 0, x[n],\ 0,\ 0\} \star \{h[n],\ 0\} = \{0,\ 0,\ y[n],\ 0,\ 0,\ 0\}$.

PGR2026-00033
Krisp Technologies EX1024 Page 189

7.3.4 Impulse Response of LTI Systems in Cascade and Parallel

Consider the ideal cascade of two LTI systems shown in Figure 7.1. The response of the first system is $y_1[n] = x[n] \star h_1[n]$. The response $y[n]$ of the second system is

$$y[n] = y_1[n] \star h_2[n] = (x[n] \star h_1[n]) \star h_2[n] = x[n] \star (h_1[n] \star h_2[n]) \tag{7.7}$$

If we wish to replace the cascaded system by an equivalent LTI system with impulse response $h[n]$ such that $y[n] = x[n] \star h[n]$, it follows that $h[n] = h_1[n] \star h_2[n]$. Generalizing this result, the impulse response $h[n]$ of N ideally cascaded LTI systems is simply the convolution of the N individual impulse responses

$$h[n] = h_1[n] \star h_2[n] \star \cdots \star h_N[n] \qquad \text{(for a cascade combination)} \tag{7.8}$$

If the $h_k[n]$ are energy signals, the order of cascading is unimportant.

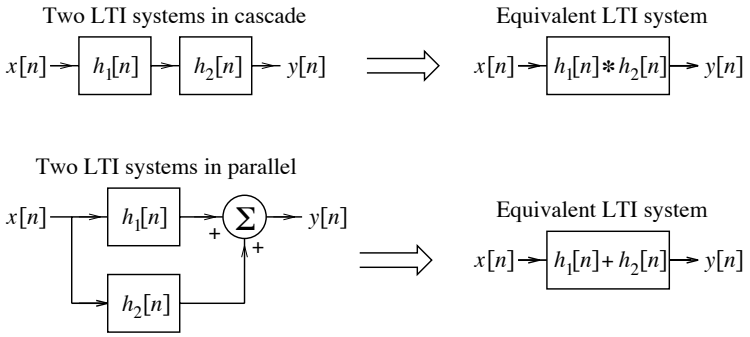


Figure 7.1 Cascaded and parallel systems and their equivalents

The overall impulse response of systems in parallel equals the sum of the individual impulse responses, as shown in Figure 7.1:

$$h_P[n] = h_1[n] + h_2[n] + \cdots + h_N[n] \qquad \text{(for a parallel combination)} \tag{7.9}$$

REVIEW PANEL 7.5

Impulse Response of N Interconnected Discrete LTI Systems

In cascade: *Convolve* the impulse responses: $h_C[n] = h_1[n] \star h_2[n] \star \cdots \star h_N[n]$

In parallel: *Add* the impulse responses: $h_P[n] = h_1[n] + h_2[n] + \cdots + h_N[n]$

EXAMPLE 7.6 (Interconnected Systems)

Consider the interconnected system of Figure E7.6. Find its overall impulse response and the output. Comment on the results.

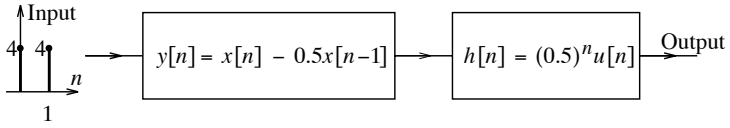


Figure E7.6 The interconnected system of Example 7.6

PGR2026-00033
Krisp Technologies EX1024 Page 190

The impulse response of the first system is $h_1[n] = \delta[n] - 0.5\delta[n-1]$. The overall impulse response $h_C[n]$ is given by the convolution

$$h_C[n] = (\delta[n] - 0.5\delta[n-1]) \star (0.5)^n u[n] = (0.5)^n u[n] - 0.5(0.5)^{n-1} u[n-1]$$

This simplifies to

$$h_C[n] = (0.5)^n (u[n] - u[n-1]) = (0.5)^n \delta[n] = \delta[n]$$

What this means is that the overall system output equals the applied input. The second system thus acts as the inverse of the first.

7.4 Stability and Causality of LTI Systems

In Chapter 5, we found that the BIBO (bounded-input, bounded-output) stability of LTI systems described by difference equations requires every root of the characteristic equation to have a magnitude less than unity. For systems described by their impulse response, this is entirely equivalent to requiring the impulse response $h[n]$ to be absolutely summable. Here is why. If $x[n]$ is bounded such that $|x[n]| < M$, so too is its shifted version $x[n-k]$. The convolution sum then yields the following inequality:

$$|y[n]| < \sum_{k=-\infty}^{\infty} |h[k]||x[n-k]| < M \sum_{k=-\infty}^{\infty} |h[k]| \tag{7.10}$$

If the output is to remain bounded ($|y[n]| < \infty$), then

$$\sum_{k=-\infty}^{\infty} |h[k]| < \infty \qquad \text{(for a stable LTI system)} \tag{7.11}$$

In other words, $h[n]$ must be absolutely summable. This is both a necessary and sufficient condition, and if met we are assured a stable system. We always have a stable system if $h[n]$ is an energy signal. Since the impulse response is defined only for linear systems, the stability of nonlinear systems must be investigated by other means.

7.4.1 Causality

In analogy with analog systems, causality of discrete-time systems implies a non-anticipating system with an impulse response $h[n] = 0, n < 0$. This ensures that an input $x[n]u[n-n_0]$ starting at $n = n_0$ results in a response $y[n]$ also starting at $n = n_0$ (and not earlier). This follows from the convolution sum:

$$y[n] = \sum_{k-\infty}^{\infty} x[k]u(k-n_0)h[n-k]u[n-k] = \sum_{n_0}^{n} x[k]h[n-k] \tag{7.12}$$

REVIEW PANEL 7.6

Stability and Causality of Discrete LTI Systems

For stability: $h[n]$ must be *absolutely summable* with $\sum_{k=-\infty}^{\infty} |h[k]| < \infty$.

For causality: $h[n]$ must be zero for negative indices with $h[n] = 0,\ n < 0$.

PGR2026-00033
Krisp Technologies EX1024 Page 191

EXAMPLE 7.7 (The Concept of Stability)

(a) The FIR filter described by $y[n] = x[n+1] - x[n]$ has the impulse response $h[n] = \{1, -\overset{\Downarrow}{1}\}$. It is a stable system, since $\sum |h[n]| = |1| + |-1| = 2$. It is also noncausal because $h[n] = \delta[n+1] - \delta[n]$ is not zero for $n < 0$. We emphasize that FIR filters are always stable because $\sum |h[n]|$ is the absolute sum of a finite sequence and is thus always finite.

(b) A filter described by $h[n] = (-0.5)^n u[n]$ is causal. It describes a system with the difference equation $y[n] = x[n] + ay[n-1]$. It is also stable because $\sum |h[n]|$ is finite. In fact, we find that

$$\sum_{n=-\infty}^{\infty} |h[n]| = \sum_{n=0}^{\infty} (-0.5)^n = \frac{1}{1-0.5} = 2$$

(c) A filter described by the difference equation $y[n] - 0.5y[n-1] = nx[n]$ is causal but time varying. It is also unstable. If we apply a step input $u[n]$ (bounded input), then $y[n] = nu[n] + 0.5y[n-1]$. The term $nu[n]$ grows without bound and makes this system unstable. We caution you that this approach is not a formal way of checking for the stability of time-varying systems.

7.5 System Response to Periodic Inputs

In analogy with analog systems, the response of a discrete-time system to a periodic input with period N is also periodic with the same period N. A simple example demonstrates this concept.

EXAMPLE 7.8 (Response to Periodic Inputs)

(a) Let $x[n] = \{\overset{\Downarrow}{1},\ 2,\ -3,\ 1,\ 2,\ -3,\ 1,\ 2,\ -3, \ldots\}$ and $h[n] = \{\overset{\Downarrow}{1},\ 1\}$.

The convolution $y[n] = x[n] \star h[n]$, using the sum-by-column method, is

Index n	0	1	2	3	4	5	6	7	8	9	10
$x[n]$	1	2	−3	1	2	−3	1	2	−3	1	...
$h[n]$	1	1									
	1	2	−3	1	2	−3	1	2	−3	1	...
		1	2	−3	1	2	−3	1	2	−3	...
$y[n]$	1	3	−1	−2	3	−1	−2	3	−1	−2	...

The convolution $y[n]$ is periodic with period $N = 3$, except for start-up effects (which last for one period). One period of the convolution is $y[n] = \{\overset{\Downarrow}{-2},\ 3,\ -1\}$.

(b) Let $x[n] = \{\overset{\Downarrow}{1},\ 2,\ -3,\ 1,\ 2,\ -3,\ 1,\ 2,\ -3, \ldots\}$ and $h[n] = \{\overset{\Downarrow}{1},\ 1,\ 1\}$.

The convolution $y[n] = x[n] \star h[n]$, using the sum-by-column method, is found as follows:

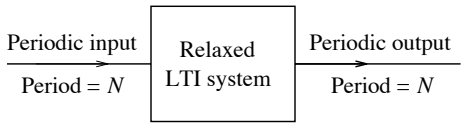
PGR2026-00033
Krisp Technologies EX1024 Page 192

Index n	0	1	2	3	4	5	6	7	8	9	10
$x[n]$	1	2	-3	1	2	-3	1	2	-3	1	...
$h[n]$	1	1	1								
	1	2	-3	1	2	-3	1	2	-3	1	...
		1	2	-3	1	2	-3	1	2	-3	...
			1	2	-3	1	2	-3	1	2	...
$y[n]$	1	3	0	0	0	0	0	0	0	0	...

Except for start-up effects, the convolution is *zero*. The system $h[n] = \{\overset{\Downarrow}{1},\ 1,\ 1\}$ is a moving average filter. It extracts the 3-point running sum, which is always zero for the given periodic signal $x[n]$.

REVIEW PANEL 7.7

The Response of LTI Systems to Periodic Signals Is Also Periodic with Identical Period



One way to find the system response to periodic inputs is to find the response to one period of the input and then use superposition. In analogy with analog signals, if we add an absolutely summable signal (or energy signal) $x[n]$ and its infinitely many replicas shifted by multiples of N, we obtain a *periodic signal* with period N, which is called the periodic extension of $x[n]$:

$$x_{\text{pe}}[n] = \sum_{k=-\infty}^{\infty} x[n + kN] \tag{7.13}$$

An equivalent way of finding one period of the periodic extension is to wrap around N-sample sections of $x[n]$ and add them all up. If $x[n]$ is shorter than N, we obtain one period of its periodic extension simply by padding $x[n]$ with zeros (to increase its length to N).

The methods for finding the response of a discrete-time system to periodic inputs rely on the concepts of periodic extension and wraparound. In analogy with the continuous case, we find the regular convolution due to one period of the input and use wraparound to generate one period of the periodic output.

EXAMPLE 7.9 (Convolution if Only One Sequence Is Periodic)

(a) Let $x[n] = \{\overset{\Downarrow}{1},\ 2,\ -3\}$ describe one period of a periodic input with period $N = 3$ to a system whose impulse response is $h[n] = \{\overset{\Downarrow}{1},\ 1\}$. The response $y[n]$ is also periodic with period $N = 3$. To find $y[n]$ for one period, we find the regular convolution $y_1[n]$ of $h[n]$ and one period of $x[n]$ to give

$$y_1[n] = \{\overset{\Downarrow}{1},\ 1\} \star \{\overset{\Downarrow}{1}, 2, -3\} = \{\overset{\Downarrow}{1},\ 3,\ -1,\ -3\}$$

We then wrap around $y_1[n]$ past three samples to obtain one period of $y[n]$ as

$$\{\overset{\Downarrow}{1},\ 3,\ -1,\ -3\} \implies \text{wrap around} \implies \left\{ \begin{array}{ccc} \overset{\Downarrow}{1} & 3 & -1 \\ -3 & & \end{array} \right\} \implies \text{sum} \implies \{\overset{\Downarrow}{-2},\ 3,\ -1\}$$

This is identical to the result obtained in the previous example.

PGR2026-00033
Krisp Technologies EX1024 Page 193

(b) We find the response $y_p[n]$ of a moving average filter described by $h[n] = \{\overset{\Downarrow}{2}, 1, 1, 3, 1\}$ to a periodic signal whose one period is $x_p[n] = \{\overset{\Downarrow}{2}, 1, 3\}$, with $N = 3$, using two methods.

1. We find the regular convolution $y[n] = x_p[n] \star h[n]$ to obtain

$$y[n] = \{\overset{\Downarrow}{2}, 1, 3\} \star \{\overset{\Downarrow}{2}, 1, 1, 3, 1\} = \{\overset{\Downarrow}{4}, 4, 9, 10, 8, 10, 3\}$$

To find $y_p[n]$, values past $N = 3$ are wrapped around and summed to give

$$\{\overset{\Downarrow}{4}, 4, 9, 10, 8, 10, 3\} \implies \text{wrap around} \implies \left\{ \begin{array}{ccc} \overset{\Downarrow}{4} & 4 & 9 \\ 10 & 8 & 10 \\ 3 & & \end{array} \right\} \implies \text{sum} \implies \{\overset{\Downarrow}{17},\ 12,\ 19\}$$

2. In analogy with continuous convolution, we could also create the periodic extension $h_p[n]$, with $N = 3$, and use wraparound to get $h_p[n] = \{\overset{\Downarrow}{5},\ 2,\ 1\}$. The regular convolution of one period of each sequence gives $y[n] = \{\overset{\Downarrow}{2},\ 1\ , 3\} \star \{\overset{\Downarrow}{5},\ 2\ , 1\} = \{\overset{\Downarrow}{1}\ 0,\ 9,\ 19,\ 7,\ 3\}$. This result is then wrapped around past these samples and gives $y_p[n] = \{\overset{\Downarrow}{1}\ 7,\ 12,\ 19\}$, as before.

7.6 Periodic Convolution

The regular convolution of two signals, both of which are periodic, does not exist. For this reason, we resort to **periodic convolution** by using averages. If *both* $x_p[n]$ and $h_p[n]$ are periodic with identical period N, their periodic convolution generates a convolution result $y_p[n]$ that is also periodic with the same period N. The periodic convolution or **circular convolution** $y_p[n]$ of $x_p[n]$ and $h_p[n]$ is denoted $y_p[n] = x_p[n] \circledast h_p[n]$ and, over one period ($n = 0, 1, \ldots, N-1$), it is defined by

$$y_p[n] = x_p[n] \circledast h_p[n] = h_p[n] \circledast x_p[n] = \sum_{k=0}^{N-1} x_p[k] h_p[n-k] = \sum_{k=0}^{N-1} h_p[k] x_p[n-k] \tag{7.14}$$

An averaging factor of $1/N$ is sometimes included with the summation. Periodic convolution can be implemented using wraparound. We find the *linear* convolution of one period of $x_p[n]$ and $h_p[n]$, which will have $(2N-1)$ samples. We then extend its length to $2N$ (by appending a zero), slice it in two halves (of length N each), line up the second half with the first, and add the two halves to get the periodic convolution.

REVIEW PANEL 7.8

Periodic Convolution of Periodic Discrete-Time Signals with Identical Period N

1. Find the regular convolution of one-period segments of each (this will contain $2N-1$ samples).
2. Append a zero. Wrap around the last N samples and add to the first N samples.

PGR2026-00033
Krisp Technologies EX1024 Page 194

EXAMPLE 7.10 (Periodic Convolution)

(a) Find the periodic convolution of $x_p[n] = \{\overset{\Downarrow}{1},\ 0,\ 1,\ 1\}$ and $h_p[n] = \{\overset{\Downarrow}{1},\ 2,\ 3,\ 1\}$.

The period is $N = 4$. First, we find the linear convolution $y[n]$.

Index n	0	1	2	3	4	5	6
$h_p[n]$	1	2	3	1			
$x_p[n]$	1	0	1	1			
	1	2	3	1			
		0	0	0	0		
			1	2	3	1	
				1	2	3	1
$y[n]$	1	2	4	4	5	4	1

Then, we append a zero, wrap around the last four samples, and add.

Index n	0	1	2	3
First half of $y[n]$	1	2	4	4
Wrapped around half of $y[n]$	5	4	1	0
Periodic convolution $y_p[n]$	6	6	5	4

(b) Find the periodic convolution of $x_p[n] = \{\overset{\Downarrow}{1}, 2, 3\}$ and $h_p[n] = \{\overset{\Downarrow}{1}, 0, 2\}$, with period $N = 3$.

The regular convolution is easily found to be $y_R[n] = \{\overset{\Downarrow}{1}, 2, 5, 4, 6\}$.

Appending a zero and wrapping around the last three samples gives $y_p[n] = \{\overset{\Downarrow}{5}, 8, 5\}$.

7.6.1 Periodic Convolution By the Cyclic Method

To find the periodic convolution, we shift the folded signal $x_p[-n]$ past $h_p[n]$, one index at a time, and find the convolution at each index as the sum of the pointwise product of their samples but only over a one-period window $(0, N-1)$. Values of $x_p[n]$ and $h_p[n]$ outside the range $(0, N-1)$ are generated by **periodic extension**. One way to visualize the process is to line up $x[k]$ clockwise around a circle and $h[k]$ counterclockwise (folded), on a concentric circle positioned to start the convolution, as shown in Figure 7.2.

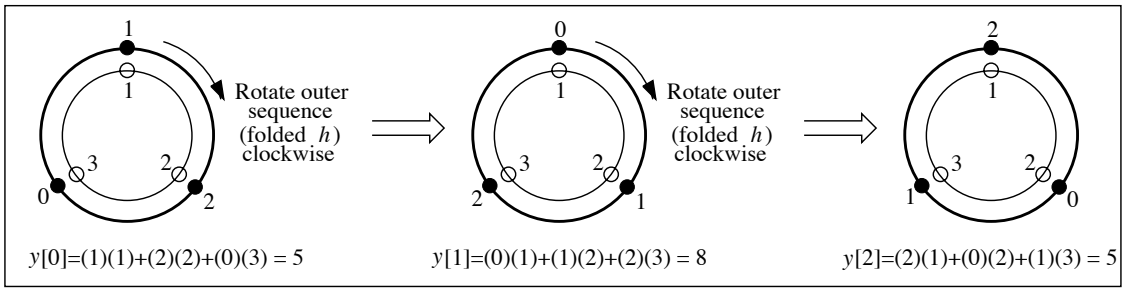


Figure 7.2 The cyclic method of circular (periodic) convolution

Shifting the folded sequence turns it clockwise. At each turn, the convolution equals the sum of the pairwise products. This approach clearly brings out the cyclic nature of periodic convolution.

PGR2026-00033
Krisp Technologies EX1024 Page 195

7.6.2 Periodic Convolution By the Circulant Matrix

Periodic convolution may also be expressed as a matrix multiplication. We set up an $N \times N$ matrix whose columns equal $x[n]$ and its cyclically shifted versions (or whose rows equal successively shifted versions of the *first* period of the folded signal $x[-n]$). This is called the **circulant matrix** or **convolution matrix**. An $N \times N$ circulant matrix $\mathbf{C}_x$ for $x[n]$ has the general form

$$\mathbf{C}_x = \begin{bmatrix} x[0] & x[N-1] & \dots & x[2] & x[1] \\ x[1] & x[0] \dots & & & x[2] \\ x[2] & x[1] & \dots & & x[3] \\ \vdots & \vdots & & & \vdots \\ x[N-2] & & \dots & x[0] & x[N-1] \\ x[N-1] & x[N-2] & \dots & x[1] & x[0] \end{bmatrix} \tag{7.15}$$

Note that each diagonal of the circulant matrix has equal values. Such a *constant diagonal matrix* is also called a **Toeplitz matrix**. Its matrix product with an $N \times 1$ column matrix $\mathbf{h}$ describing $h[n]$ yields the periodic convolution $\mathbf{y = Ch}$ as an $N \times 1$ column matrix.

EXAMPLE 7.11 (Periodic Convolution By the Circulant Matrix)

Consider $x[n] = \{\overset{\Downarrow}{1}, 0, 2\}$ and $h[n] = \{\overset{\Downarrow}{1}, 2, 3\}$, described over one period ($N = 3$).

(a) The circulant matrix $\mathbf{C}_x$ and periodic convolution $y_1[n]$ are given by

$$\mathbf{C}_x = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 2 \\ 2 & 0 & 1 \end{bmatrix} \qquad \mathbf{h} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \qquad y_1[n] = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 2 \\ 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 5 \\ 8 \\ 5 \end{bmatrix}$$

Comment: Though not required, normalization by $N = 3$ gives $y_{p1}[n] = \dfrac{y_1[n]}{3} = \{\overset{\Downarrow}{\frac{5}{3}}, \frac{8}{3}, \frac{5}{3}\}$.

(b) The periodic convolution $y_2[n]$ of $x[n]$ and $h[n]$ over a two-period window yields

$$\mathbf{C}_2 = \begin{bmatrix} 1 & 2 & 0 & 1 & 2 & 0 \\ 0 & 1 & 2 & 0 & 1 & 2 \\ 2 & 0 & 1 & 2 & 0 & 1 \\ 1 & 2 & 0 & 1 & 2 & 0 \\ 0 & 1 & 2 & 0 & 1 & 2 \\ 2 & 0 & 1 & 2 & 0 & 1 \end{bmatrix} \qquad \mathbf{h}_2 = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 1 \\ 2 \\ 3 \end{bmatrix} \qquad y_2[n] = \begin{bmatrix} 10 \\ 16 \\ 10 \\ 10 \\ 16 \\ 10 \end{bmatrix}$$

We see that $y_2[n]$ has double the length (and values) of $y_1[n]$, but it is still periodic with $N = 3$.

Comment: Normalization by a two-period window width (6 samples) gives

$$y_{p2}[n] = \frac{y_2[n]}{6} = \{\overset{\Downarrow}{\tfrac{5}{3}}, \tfrac{8}{3}, \tfrac{5}{3}, \tfrac{5}{3}, \tfrac{8}{3}, \tfrac{5}{3}\}$$

Note that one period ($N = 3$) of $y_{p2}[n]$ is identical to the normalized result $y_{p1}[n]$ of part (a).

PGR2026-00033
Krisp Technologies EX1024 Page 196

7.6.3 Regular Convolution from Periodic Convolution

The linear convolution of $x[n]$ (with length N_x) and $h[n]$ (with length N_h) may also be found using the *periodic* convolution of two zero-padded signals $x_z[n]$ and $h_z[n]$ (each of length $N_y = N_x + N_h - 1$). The *regular* convolution of the original, unpadded sequences equals the *periodic* convolution of the zero-padded sequences.

REVIEW PANEL 7.9

Regular Convolution from Periodic Convolution by Zero-Padding

$x[n] \star h[n]$ = Periodic convolution of their zero-padded (*each* to length $N_x + N_h - 1$) versions.

EXAMPLE 7.12 (Regular Convolution by the Circulant Matrix)

Let $x[n] = \{\overset{\Downarrow}{2}, 5, 0, 4\}$ and $h[n] = \{\overset{\Downarrow}{4}, 1, 3\}$. Their regular convolution has $S = M + N - 1 = 6$ samples. Using trailing zeros, we create the padded sequences

$$x_{\text{zp}}[n] = \{\overset{\Downarrow}{2}, 5, 0, 4, 0, 0\} \qquad h_{\text{zp}}[n] = \{\overset{\Downarrow}{4}, 1, 3, 0, 0, 0\}$$

The periodic convolution $x_{\text{zp}}[n] \circledast h_{\text{zp}}[n]$, using the circulant matrix, equals

$$\mathbf{C}_{\text{xzp}} = \begin{bmatrix} 2 & 0 & 0 & 4 & 0 & 5 \\ 5 & 2 & 0 & 0 & 4 & 0 \\ 0 & 5 & 2 & 0 & 0 & 4 \\ 4 & 0 & 5 & 2 & 0 & 0 \\ 0 & 4 & 0 & 5 & 2 & 0 \\ 0 & 0 & 4 & 0 & 5 & 2 \end{bmatrix} \quad \mathbf{h}_{\text{zp}} = \begin{bmatrix} 4 \\ 1 \\ 3 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad y_p[n] = \begin{bmatrix} 8 \\ 22 \\ 11 \\ 31 \\ 4 \\ 12 \end{bmatrix}$$

This is identical to the regular convolution $y[n] = x[n] \star h[n]$ obtained previously by several other methods in previous examples.

7.7 Connections: Discrete Convolution and Transform Methods

Discrete-time convolution provides a connection between the time-domain and frequency-domain methods of system analysis for discrete-time signals. It forms the basis for every transform method described in this text, and its role in linking the time domain and the transformed domain is intimately tied to the concept of discrete eigensignals and eigenvalues.

7.7.1 Discrete-Time Harmonics, Eigensignals, and Discrete Transforms

Just as the everlasting analog harmonic $e^{j2\pi ft}$ is an eigensignal of analog systems, the everlasting discrete-time harmonic $e^{j2\pi nF}$ is an eigensignal of discrete-time systems. The response of a linear system to such a harmonic is also a harmonic at the same frequency but with changed magnitude and phase. This key concept leads to idea of discrete transforms analogous to Fourier and Laplace transforms.

REVIEW PANEL 7.10

The Everlasting Exponential z^n Is an Eigensignal of Discrete-Time Linear Systems

In this complex exponential z^n, the quantity z has the general form $z = re^{j2\pi F}$.
The response of a linear system to z^n is Cz^n, where C is a (possibly complex) constant.

PGR2026-00033
Krisp Technologies EX1024 Page 197

7.7.2 The Discrete-Time Fourier Transform

For the harmonic input $x[n] = e^{j2\pi nF}$, the response $y[n]$ equals

$$y[n] = \sum_{k=-\infty}^{\infty} e^{j2\pi(n-k)F} h[k] = e^{j2\pi nF} \sum_{k=-\infty}^{\infty} h[k] e^{-j2\pi kF} = x[n] H_p(F) \tag{7.16}$$

This is just the input modified by the system function $H_p(F)$, where

$$H_p(F) = \sum_{k=-\infty}^{\infty} h[k] e^{-j2\pi kF} \tag{7.17}$$

The quantity $H_p(F)$ describes the **discrete-time Fourier transform** (DTFT) or **discrete-time frequency response** of $h[n]$. Unlike the analog system function $H(f)$, the DT system function $H_p(F)$ is periodic with a period of unity because $e^{-j2\pi kF} = e^{-j2\pi k(F+1)}$. This periodicity is also a direct consequence of the discrete nature of $h[n]$.

As with the impulse response $h[n]$, which is just a signal, any signal $x[n]$ may similarly be described by its DTFT $X_p(F)$. The response $y[n] = x[n]H_p[F]$ may then be transformed to its DTFT $Y_p[n]$ to give

$$Y_p(F) = \sum_{k=-\infty}^{\infty} y[k] e^{-j2\pi Fk} = \sum_{k=-\infty}^{\infty} x[k] H_p(F) e^{-j2\pi Fk} = H_p(F) X_p(F) \tag{7.18}$$

Once again, convolution in the time domain corresponds to multiplication in the frequency domain.

7.7.3 The *z*-Transform

For an input $x[n] = r^n e^{j2\pi nF} = \left(re^{j2\pi F}\right)^n = z^n$, where z is complex, with magnitude $|z| = r$, the response may be written as

$$y[n] = x[n] \star h[n] = \sum_{k=-\infty}^{\infty} z^{n-k} h[k] = z^n \sum_{k=-\infty}^{\infty} h[k] z^{-k} = x[n] H(z) \tag{7.19}$$

The response equals the input (eigensignal) modified by the system function $H(z)$, where

$$H(z) = \sum_{k=-\infty}^{\infty} h[k] z^{-k} \qquad \text{(two-sided } z\text{-transform)} \tag{7.20}$$

The complex quantity $H(z)$ describes the ***z*-transform** of $h[n]$ and is not, in general, periodic in z. Denoting the z-transform of $x[n]$ and $y[n]$ by $X(z)$ and $Y(z)$, we write

$$Y(z) = \sum_{k=-\infty}^{\infty} y[k] z^{-k} = \sum_{k=-\infty}^{\infty} x[k] H(z) z^{-k} = H(z) X(z) \tag{7.21}$$

Convolution in the time domain thus corresponds to multiplication in the z-domain.

If $r = 1$, $x[n] = \left(e^{j2\pi F}\right)^n = e^{j2\pi nF}$, and the z-transform reduces to the DTFT. We can thus obtain the DTFT of $x[n]$ from its z-transform $X(z)$ by letting $z = e^{j2\pi F}$ or $|z| = 1$ to give

$$H(F) = H(z)|_{z=\exp(j2\pi F)} = H(z)|_{|z|=1} \tag{7.22}$$

The DTFT is thus the z-transform evaluated on the unit circle $|z| = 1$.

PGR2026-00033
Krisp Technologies EX1024 Page 198

7.7.4 Sampling and Periodic Extension

Another key concept is that *sampling in the time domain leads to a periodic extension in the frequency domain and vice versa*. This implies that the spectrum (DTFT) of a discrete-time signal is periodic, whereas a discrete spectrum (the Fourier series) corresponds to a periodic time signal. If we sample the periodic DTFT, we obtain the **discrete Fourier transform** (DFT), which corresponds to a *periodic*, *discrete-time* signal. If we sample a periodic time signal, we obtain the **discrete Fourier series** (DFS), which is also discrete and periodic.

REVIEW PANEL 7.11

Sampling in One Domain Leads to Periodic Extension in the Other Domain

7.7.5 What to Keep in Mind

Convolution plays a central role in signal processing and system analysis. The main reason for the popularity of transformed domain methods of system analysis is that *convolution in one domain corresponds to multiplication in the other*. The type of convolution required depends on the type of signals under consideration. For example, we perform *discrete* convolution when dealing with discrete-time signals and periodic convolution when dealing with periodic signals. The frequency-domain representation of such signals gives rise to periodic spectra and requires *periodic* convolution in the frequency domain.

REVIEW PANEL 7.12

The Type of Convolution Depends on the Transform Method

Transform Method	Time-Domain Convolution	Frequency-Domain Convolution
DTFT	$x[n] \star h[n]$ (regular, discrete)	$X_p(F) \circledast H_p(F)$ (periodic, complex)
DFT	$x[n] \circledast h[n]$ (periodic, discrete)	$X_T[k] \circledast H_T[k]$ (periodic, discrete)
***z*-Transform**	$x[n] \star h[n]$ (regular, discrete)	$X(z) \star H(z)$ (regular, complex)

7.8 Deconvolution

Given the system impulse response $h[n]$, the response $y[n]$ of the system to an input $x[n]$ is simply the convolution of $x[n]$ and $h[n]$. Given $x[n]$ and $y[n]$ instead, how do we find $h[n]$? This situation arises very often in practice and is referred to as **deconvolution** or **system identification**.

For discrete-time systems, we have a partial solution to this problem. Since discrete convolution may be thought of as polynomial multiplication, *discrete* deconvolution may be regarded as polynomial division. One approach to discrete deconvolution is to use the idea of *long division*, a familiar process, illustrated in the following example.

REVIEW PANEL 7.13

Deconvolution May Be Regarded as Polynomial Division or Matrix Inversion

EXAMPLE 7.13 (Deconvolution by Polynomial Division)

Consider $x[n] = \{\overset{\Downarrow}{2}, 5, 0, 4\}$ and $y[n] = \{\overset{\Downarrow}{8}, 22, 11, 31, 4, 12\}$. We regard these as being the coefficients, in descending order, of the polynomials

$$x(w) = 2w^3 + 5w^2 + 0w + 4 \qquad y(w) = 8w^5 + 22w^4 + 11w^3 + 31w^2 + 4w + 12$$

PGR2026-00033
Krisp Technologies EX1024 Page 199

The polynomial $h(w)$ may be deconvolved out of $x(w)$ and $y(w)$ by performing the division $y(w)/x(w)$:

$$
\begin{array}{r}
\multicolumn{1}{r}{4w^2 + w + 3} \\
2w^3 + 5w^2 + 0w + 4 \;\overline{)\;8w^5 + 22w^4 + 11w^3 + 31w^2 + 4w + 12} \\
\multicolumn{1}{r}{\underline{8w^5 + 20w^4 + \;\; 0w^3 + 16w^2}} \\
\multicolumn{1}{r}{2w^4 + 11w^3 + 15w^2 + 4w + 12} \\
\multicolumn{1}{r}{\underline{2w^4 + \;\; 5w^3 + \;\; 0w^2 + 4w}} \\
\multicolumn{1}{r}{6w^3 + 15w^2 + 0w + 12} \\
\multicolumn{1}{r}{\underline{6w^3 + 15w^2 + 0w + 12}} \\
\multicolumn{1}{r}{0}
\end{array}
$$

The coefficients of the quotient polynomial describe the sequence $h[n] = \{\overset{\Downarrow}{4},\ 1,\ 3\}$.

7.8.1 Deconvolution By Recursion

Deconvolution may also be recast as a recursive algorithm. The convolution

$$y[n] = x[n] \star h[n] = \sum_{k=0}^{n} h[k]x[n-k] \tag{7.23}$$

when evaluated at $n = 0$, provides the seed value $h[0]$ as

$$y[0] = x[0]h[0] \qquad\qquad h[0] = y[0]/x[0] \tag{7.24}$$

We now separate the term containing $h[n]$ in the convolution relation

$$y[n] = \sum_{k=0}^{n} h[k]x[n-k] = h[n]x[0] + \sum_{k=0}^{n-1} h[k]x[n-k] \tag{7.25}$$

and evaluate $h[n]$ for successive values of $n > 0$ from

$$h[n] = \frac{1}{x[0]}\left[y[n] - \sum_{k=0}^{n-1} h[k]x[n-k]\right] \tag{7.26}$$

If all goes well, we need to evaluate $h[n]$ only at $M - N + 1$ points, where M and N are the lengths of $y[n]$ and $x[n]$, respectively.

Naturally, problems arise if a remainder is involved. This may well happen in the presence of noise, which could modify the values in the output sequence even slightly. In other words, the approach is quite susceptible to noise or roundoff error and not very practical.

EXAMPLE 7.14 (Deconvolution by Recursion)
Let $x[n] = \{2,\ 5,\ 0,\ 4\}$ and $y[n] = \{8,\ 22,\ 11,\ 31,\ 4,\ 12\}$. We need only $6 - 4 + 1$ or 3 evaluations of $h[n]$. If both $x[n]$ and $h[n]$ start at $n = 0$, we compute the seed value as $h[0] = y[0]/x[0] = 4$. Then

$$
\begin{aligned}
h[1] &= \frac{1}{x[0]}\left[y[1] - \sum_{k=0}^{0} h[k]x[1-k]\right] = \frac{y[1] - h[0]x[1]}{x[0]} = 1 \\
h[2] &= \frac{1}{x[0]}\left[y[2] - \sum_{k=0}^{1} h[k]x[2-k]\right] = \frac{y[2] - h[0]x[2] - h[1]x[1]}{x[0]} = 3
\end{aligned}
$$

As before, $h[n] = \{4,\ 1,\ 3\}$.

PGR2026-00033
Krisp Technologies EX1024 Page 200

7.9 Discrete Correlation

Correlation is a measure of similarity between two signals and is found using a process similar to convolution. The discrete **cross-correlation** (denoted $\star\star$) of $x[n]$ and $h[n]$ is defined by

$$r_{xh}[n] = x[n] \star\star h[n] = \sum_{k=-\infty}^{\infty} x[k]h[k-n] = \sum_{k=-\infty}^{\infty} x[k+n]h[k] \tag{7.27}$$

$$r_{hx}[n] = h[n] \star\star x[n] = \sum_{k=-\infty}^{\infty} h[k]x[k-n] = \sum_{k=-\infty}^{\infty} h[k+n]x[k] \tag{7.28}$$

Some authors prefer to switch the definitions of $r_{xh}[n]$ and $r_{hx}[n]$.

To find $r_{xh}[n]$, we line up the last element of $h[n]$ with the first element of $x[n]$ and start shifting $h[n]$ past $x[n]$, one index at a time. We sum the pointwise product of the overlapping values to generate the correlation at each index. This is equivalent to performing the *convolution* of $x[n]$ and the *folded* signal $h[-n]$. The starting index of the correlation equals the sum of the starting indices of $x[n]$ and $h[-n]$. Similarly, $r_{hx}[n]$ equals the convolution of $x[-n]$ and $h[n]$, and its starting index equals the sum of the starting indices of $x[-n]$ and $h[n]$. However, $r_{xh}[n]$ does not equal $r_{hx}[n]$. The two are folded versions of each other and related by $r_{xh}[n] = r_{hx}[-n]$.

REVIEW PANEL 7.14

Correlation Is the Convolution of One Signal with a Folded Version of the Other

$r_{xh}[n] = x[n] \star\star h[n] = x[n] \star h[-n]$ $\qquad$ $r_{hx}[n] = h[n] \star\star x[n] = h[n] \star x[-n]$

Correlation length: $N_x + N_h - 1$ $\qquad$ **Correlation sum:** $\sum r[n] = (\sum x[n])(\sum h[n])$

7.9.1 Autocorrelation

The correlation $r_{xx}[n]$ of a signal $x[n]$ with itself is called the *autocorrelation*. It is an even symmetric function ($r_{xx}[n] = r_{xx}[-n]$) with a maximum at $n = 0$ and satisfies the inequality $|r_{xx}[n]| \le r_{xx}[0]$.

Correlation is an effective method of detecting signals buried in noise. Noise is essentially uncorrelated with the signal. This means that if we correlate a noisy signal with itself, the correlation will be due only to the signal (if present) and will exhibit a sharp peak at $n = 0$.

REVIEW PANEL 7.15

The Autocorrelation Is Always Even Symmetric with a Maximum at the Origin

$r_{xx}[n] = x[n] \star\star x[n] = x[n] \star x[-n]$ $\qquad$ $r_{xx}[n] = r_{xx}[-n]$ $\qquad$ $r_{xx}[n] \le r_{xx}[0]$

EXAMPLE 7.15 (Discrete Autocorrelation and Cross-Correlation)

(a) Let $x[n] = a^n u[n]$, $|a| < 1$. Since $x[k-n] = a^{k-n}u[k-n]$ starts at $k = n$, we obtain the autocorrelation $r_{xx}[n]$ for $n \ge 0$ as

$$r_{xx}[n] = \sum_{k=-\infty}^{\infty} x[k]x[k-n] = \sum_{k=n}^{\infty} a^k a^{k-n} = \sum_{m=0}^{\infty} a^{m+n}a^m = a^n \sum_{m=0}^{\infty} a^{2m} = a^n \frac{1}{1-a^2}$$

Since autocorrelation is an even symmetric function, we have $r_{xx}[n] = \dfrac{a^{|n|}}{1-a^2}$.

PGR2026-00033
Krisp Technologies EX1024 Page 201

(b) Let $x[n] = a^n u[n],\ |a| < 1$, and $y[n] = \text{rect}(n/2N)$. To find $r_{xy}[n]$, we shift $y[k]$ and sum the products over different ranges. Since $y[k-n]$ shifts the pulse to the right over the limits $(-N+n,\ N+n)$, the correlation $r_{xy}[n]$ equals zero until $n = -N$. We then obtain

$$-N \le n \le N-1 \text{ (partial overlap):} \quad r_{xy}[n] = \sum_{k=-\infty}^{\infty} x[k]y[k-n] = \sum_{k=0}^{N+1} a^k = \frac{1-a^{N+n+1}}{1-a}$$

$$n \ge N \text{ (total overlap):} \quad r_{xy}[n] = \sum_{k=-N+1}^{N+1} a^k = \sum_{m=0}^{2N} a^{m-N+1} = a^{-N+1}\frac{1-a^{2N+1}}{1-a}$$

7.9.2 Periodic Discrete Correlation

For periodic sequences with identical period N, the periodic discrete correlation is defined in analogy with periodic continuous correlation as

$$r_{xh}[n] = x[n] \circledast\circledast h[n] = \sum_{k=0}^{N-1} x[k]h[k-n] \qquad r_{hx}[n] = h[n] \circledast\circledast x[n] = \sum_{k=0}^{N-1} h[k]x[k-n] \tag{7.29}$$

As with discrete periodic convolution, an averaging factor of $1/N$ is sometimes included in the summation. We can also find the periodic correlation $r_{xh}[n]$ using convolution and wraparound, provided we use one period of the *folded*, *periodic extension* of the sequence $h[n]$.

7.9.3 Applications of Correlation

Correlation finds widespread use in applications such as target ranging and estimation of periodic signals buried in noise. For target ranging, a sampled interrogating signal $x[n]$ is transmitted toward the target. The signal reflected from the target is $s[n] = \alpha x[n-D] + p[n]$, a delayed (by D) and attenuated (by α) version of $x[n]$, contaminated by noise $p[n]$. If the noise is uncorrelated with the signal $x[n]$, its correlation with $x[n]$ is essentially zero. The cross-correlation of $x[n]$ and its delayed version $\alpha x[n-D]$ yield a result that attains a peak at $n = D$. It is thus quite easy to identify the index D from the correlation (rather than from the reflected signal directly), even in the presence of noise. The (round-trip) delay index D may then be related to the target distance d by $d = 0.5vD/S$, where v is the propagation velocity and S is the rate at which the signal is sampled.

Correlation methods may also be used to identify the presence of a periodic signal $x[n]$ buried in the noisy signal $s[n] = x[n] + p[n]$, where $p[n]$ is the noise component (presumably uncorrelated with the signal), and to extract the signal itself. The idea is to first identify the period of the signal from the *periodic* autocorrelation of the noisy signal. If the noisy signal contains a periodic component, the autocorrelation will show peaks at multiples of the period N. Once the period N is established, we can recover $x[n]$ as the *periodic* cross-correlation of an impulse train $i[n] = \delta(n-kN)$, with period N, and the noisy signal $s[n]$. Since $i[n]$ is uncorrelated with the noise, the periodic cross-correlation of $i[n]$ and $s[n]$ yields (an amplitude scaled version of) the periodic signal $x[n]$.

REVIEW PANEL 7.16

How to Identify a Periodic Signal $x[n]$ Buried in a Noisy Signal $s[n]$

Find the period N from the *periodic* autocorrelation of the noisy signal $s[n]$.
Find the signal $x[n]$ as the *periodic* cross-correlation of $s[n]$ and an *impulse train* with period N.

PGR2026-00033
Krisp Technologies EX1024 Page 202

CHAPTER 7 PROBLEMS

DRILL AND REINFORCEMENT

7.1 (Folding) For each signal $x[n]$, sketch $g[k] = x[3-k]$ vs. k and $h[k] = x[2+k]$ vs. k.

(a) $x[n] = \{\overset{\Downarrow}{1},\ 2,\ 3,\ 4\}$ **(b)** $x[n] = \{3,\ 3,\ \overset{\Downarrow}{3},\ 2,\ 2,\ 2\}$

7.2 (Closed-Form Convolution) Find the convolution $y[n] = x[n] \star h[n]$ for the following:

(a) $x[n] = u[n]$ $h[n] = u[n]$
(b) $x[n] = (0.8)^n u[n]$ $h[n] = (0.4)^n u[n]$
(c) $x[n] = (0.5)^n u[n]$ $h[n] = (0.5)^n\{u[n+3] - u[n-4]\}$
(d) $x[n] = \alpha^n u[n]$ $h[n] = \alpha^n u[n]$
(e) $x[n] = \alpha^n u[n]$ $h[n] = \beta^n u[n]$
(f) $x[n] = \alpha^n u[n]$ $h[n] = \text{rect}(n/2N)$

7.3 (Convolution of Finite Sequences) Find the convolution $y[n] = x[n] \star h[n]$ for each of the following signal pairs. Use a marker to indicate the origin $n = 0$.

(a) $x[n] = \{\overset{\Downarrow}{1}, 2, 0, 1\}$ $h[n] = \{\overset{\Downarrow}{2}, 2, 3\}$
(b) $x[n] = \{\overset{\Downarrow}{0}, 2, 4, 6\}$ $h[n] = \{\overset{\Downarrow}{6}, 4, 2, 0\}$
(c) $x[n] = \{-3, -2, \overset{\Downarrow}{-1}, 0, 1\}$ $h[n] = \{\overset{\Downarrow}{4}, 3, 2\}$
(d) $x[n] = \{3, 2, \overset{\Downarrow}{1}, 1, 2\}$ $h[n] = \{4, \overset{\Downarrow}{2}, 3, 2\}$
(e) $x[n] = \{3, 0, 2, 0, \overset{\Downarrow}{1}, 0, 1, 0, 2\}$ $h[n] = \{4, 0, \overset{\Downarrow}{2}, 0, 3, 0, 2\}$
(f) $x[n] = \{\overset{\Downarrow}{0}, 0, 0, 3, 1, 2\}$ $h[n] = \{4, \overset{\Downarrow}{2}, 3, 2\}$

7.4 (Properties) Let $x[n] = h[n] = \{\overset{\Downarrow}{3}, 4, 2, 1\}$. Compute the following:

(a) $y[n] = x[n] \star h[n]$ **(b)** $g[n] = x[-n] \star h[-n]$
(c) $p[n] = x[n] \star h[-n]$ **(d)** $f[n] = x[-n] \star h[n]$
(e) $r[n] = x[n-1] \star h[n+1]$ **(f)** $s[n] = x[n-1] \star h[n+4]$

7.5 (Properties) Let $x[n] = h[n] = \{\overset{\Downarrow}{2}, 6, 0, 4\}$. Compute the following:

(a) $y[n] = x[2n] \star h[2n]$
(b) Find $g[n] = x[n/2] \star h[n/2]$, assuming zero interpolation.
(c) Find $p[n] = x[n/2] \star h[n]$, assuming step interpolation where necessary.
(d) Find $r[n] = x[n] \star h[n/2]$, assuming linear interpolation where necessary.

7.6 (Application) Consider a 2-point averaging filter whose present output equals the average of the present and previous input.

(a) Set up a difference equation for this system.
(b) What is the impulse response of this system?
(c) What is the response of this system to the sequence $\{\overset{\Downarrow}{1}, 2, 3, 4, 5\}$?
(d) Use convolution to show that the system performs the required averaging operation.

PGR2026-00033
Krisp Technologies EX1024 Page 203

7.7 (Stability) Investigate the causality and stability of the following systems.

(a) $h[n] = (2)^n u[n-1]$ **(b)** $y[n] = 2x[n+1] + 3x[n] - x[n-1]$

(c) $h[n] = (-0.5)^n u[n]$ **(d)** $h[n] = \{3, 2, \overset{\Downarrow}{1}, 1, 2\}$

(e) $h[n] = (0.5)^{-n} u[-n]$ **(f)** $h[n] = (0.5)^{|n|}$

7.8 (Periodic Convolution) Find the regular convolution $y[n] = x[n] \star h[n]$ of one period of each pair of periodic signals. Then, use wraparound to compute the periodic convolution $y_p[n] = x[n] \circledast h[n]$. In each case, specify the *minimum* number of padding zeros we must use if we wish to find the regular convolution from the periodic convolution of the zero-padded signals.

(a) $x[n] = \{\overset{\Downarrow}{1}, 2, 0, 1\}$ $h[n] = \{\overset{\Downarrow}{2}, 2, 3, 0\}$

(b) $x[n] = \{\overset{\Downarrow}{0}, 2, 4, 6\}$ $h[n] = \{\overset{\Downarrow}{6}, 4, 2, 0\}$

(c) $x[n] = \{-3, -2, \overset{\Downarrow}{-1}, 0, 1\}$ $h[n] = \{\overset{\Downarrow}{4}, 3, 2, 0, 0\}$

(d) $x[n] = \{3, 2, \overset{\Downarrow}{1}, 1, 2\}$ $h[n] = \{4, \overset{\Downarrow}{2}, 3, 2, 0\}$

7.9 (Periodic Convolution) Find the periodic convolution $y_p[n] = x[n] \circledast h[n]$ for each pair of signals using the circulant matrix for $x[n]$.

(a) $x[n] = \{\overset{\Downarrow}{1}, 2, 0, 1\}$ $h[n] = \{\overset{\Downarrow}{2}, 2, 3, 0\}$

(b) $x[n] = \{\overset{\Downarrow}{0}, 2, 4, 6\}$ $h[n] = \{\overset{\Downarrow}{6}, 4, 2, 0\}$

7.10 (Convolution and Interpolation) Consider the following system with $x[n] = \{\overset{\Downarrow}{0}, 3, 9, 12, 15, 18\}$.

$$x[n] \longrightarrow \boxed{\text{zero interpolate by } N} \longrightarrow \boxed{\text{filter}} \longrightarrow y[n]$$

(a) Find the response $y[n]$ if $N = 2$ and the filter impulse response is $h[n] = \{\overset{\Downarrow}{1},\ 1\}$. Show that, except for end effects, the output describes a step interpolation between the samples of $x[n]$.

(b) Find the response $y[n]$ if $N = 3$ and the filter impulse response is $h[n] = \{\overset{\Downarrow}{1},\ 1,\ 1\}$. Does the output describe a step interpolation between the samples of $x[n]$?

(c) Pick N and $h[n]$ if the system is to perform step interpolation by 4.

7.11 (Convolution and Interpolation) Consider the following system with $x[n] = \{\overset{\Downarrow}{0}, 3, 9, 12, 15, 18\}$.

$$x[n] \longrightarrow \boxed{\text{zero interpolate by } N} \longrightarrow \boxed{\text{filter}} \longrightarrow y[n]$$

(a) Find the response $y[n]$ if $N = 2$ and the filter impulse response is $h[n] = \text{tri}(n/2)$. Show that, except for end effects, the output describes a linear interpolation between the samples of $x[n]$.

(b) Find the response $y[n]$ if $N = 3$ and the filter impulse response is $h[n] = \text{tri}(n/3)$. Does the output describe a linear interpolation between the samples of $x[n]$?

(c) Pick N and $h[n]$ if the system is to perform linear interpolation by 4.

7.12 (Correlation) For each pair of signals, compute the autocorrelation $r_{xx}[n]$, the autocorrelation $r_{hh}[n]$, the cross-correlation $r_{xh}[n]$, and the cross-correlation $r_{hx}[n]$. For each result, indicate the location of the origin $n = 0$ by a marker.

PGR2026-00033
Krisp Technologies EX1024 Page 204

(a) $x[n] = \{\overset{\Downarrow}{1}, 2, 0, 1\}$ $\quad h[n] = \{\overset{\Downarrow}{2}, 2, 3\}$
(b) $x[n] = \{\overset{\Downarrow}{0}, 2, 4, 6\}$ $\quad h[n] = \{\overset{\Downarrow}{6}, 4, 2\}$
(c) $x[n] = \{-3, -2, \overset{\Downarrow}{-1}, 2\}$ $\quad h[n] = \{\overset{\Downarrow}{4}, 3, 2\}$
(d) $x[n] = \{3, 2, \overset{\Downarrow}{1}, 1, 2\}$ $\quad h[n] = \{4, \overset{\Downarrow}{2}, 3, 2\}$

7.13 (Correlation) Let $x[n] = \text{rect}[(n-4)/2]$ and $h[n] = \text{rect}[n/4]$.

(a) Find the autocorrelation $r_{xx}[n]$.
(b) Find the autocorrelation $r_{hh}[n]$.
(c) Find the cross-correlation $r_{xh}[n]$.
(d) Find the cross-correlation $r_{hx}[n]$.
(e) How are the results of parts (c) and (d) related?

7.14 (Periodic Correlation) For each pair of periodic signals described for one period, compute the periodic autocorrelations $r_{xx}[n]$ and $r_{hh}[n]$, and the periodic cross-correlations $r_{xh}[n]$ and $r_{hx}[n]$. For each result, indicate the location of the origin $n = 0$ by a marker.

(a) $x[n] = \{\overset{\Downarrow}{1}, 2, 0, 1\}$ $\quad h[n] = \{\overset{\Downarrow}{2}, 2, 3, 0\}$
(b) $x[n] = \{\overset{\Downarrow}{0}, 2, 4, 6\}$ $\quad h[n] = \{\overset{\Downarrow}{6}, 4, 2, 0\}$
(c) $x[n] = \{-3, -2, \overset{\Downarrow}{-1}, 2\}$ $\quad h[n] = \{0, \overset{\Downarrow}{4}, 3, 2\}$
(d) $x[n] = \{3, 2, \overset{\Downarrow}{1}, 1, 2\}$ $\quad h[n] = \{4, \overset{\Downarrow}{2}, 3, 2, 0\}$

REVIEW AND EXPLORATION

7.15 (Convolution with Impulses) Find the convolution $y[n] = x[n] \star h[n]$ of the following signals.

(a) $x[n] = \delta[n-1]$ $\quad h[n] = \delta[n-1]$
(b) $x[n] = \cos(0.25n\pi)$ $\quad h[n] = \delta[n] - \delta[n-1]$
(c) $x[n] = \cos(0.25n\pi)$ $\quad h[n] = \delta[n] - 2\delta[n-1] + \delta[n-2]$
(d) $x[n] = (-1)^n$ $\quad h[n] = \delta[n] + \delta[n-1]$

7.16 (Convolution) Find the convolution $y[n] = x[n] \star h[n]$ for each pair of signals.

(a) $x[n] = (0.4)^{-n}u[n]$ $\quad h[n] = (0.5)^{-n}u[n]$
(b) $x[n] = \alpha^{-n}u[n]$ $\quad h[n] = \beta^{-n}u[n]$
(c) $x[n] = \alpha^{n}u[-n]$ $\quad h[n] = \beta^{n}u[-n]$
(d) $x[n] = \alpha^{-n}u[-n]$ $\quad h[n] = \beta^{-n}u[-n]$

7.17 (Step Response) Given the impulse response $h[n]$, find the step response $s[n]$ of each system.

(a) $h[n] = (0.5)^n u[n]$ $\quad$ **(b)** $h[n] = (0.5)^n \cos(n\pi)u[n]$
(c) $h[n] = (0.5)^n \cos(n\pi + 0.5\pi)u[n]$ $\quad$ **(d)** $h[n] = (0.5)^n \cos(n\pi + 0.25\pi)u[n]$
(e) $h[n] = n(0.5)^n u[n]$ $\quad$ **(f)** $h[n] = n(0.5)^n \cos(n\pi)u[n]$

7.18 (Convolution and System Response) Consider the system $y[n] - 0.5y[n] = x[n]$.

(a) What is the impulse response $h[n]$ of this system?
(b) Find its output if $x[n] = (0.5)^n u[n]$ by convolution.

PGR2026-00033
Krisp Technologies EX1024 Page 205

(c) Find its output if $x[n] = (0.5)^n u[n]$ and $y[-1] = 0$ by solving the difference equation.
(d) Find its output if $x[n] = (0.5)^n u[n]$ and $y[-1] = 2$ by solving the difference equation.
(e) Are any of the outputs identical? Should they be? Explain.

7.19 (Convolution of Symmetric Sequences) The convolution of sequences that are symmetric about their midpoint is also endowed with symmetry (about its midpoint). Compute $y[n] = x[n] \star h[n]$ for each pair of signals and use the results to establish the type of symmetry (about the midpoint) in the convolution if the convolved signals are both even symmetric (about their midpoint), both odd symmetric (about their midpoint), or one of each type.

(a) $x[n] = \{2,\ 1,\ 2\}$ $\quad h[n] = \{1,\ 0,\ 1\}$
(b) $x[n] = \{2,\ 1,\ 2\}$ $\quad h[n] = \{1,\ 1\}$
(c) $x[n] = \{2,\ 2\}$ $\quad h[n] = \{1,\ 1\}$
(d) $x[n] = \{2,\ 0,\ -2\}$ $\quad h[n] = \{1,\ 0,\ -1\}$
(e) $x[n] = \{2,\ 0,\ -2\}$ $\quad h[n] = \{1,\ -1\}$
(f) $x[n] = \{2,\ -2\}$ $\quad h[n] = \{1,\ -1\}$
(g) $x[n] = \{2,\ 1,\ 2\}$ $\quad h[n] = \{1,\ 0,\ -1\}$
(h) $x[n] = \{2,\ 1,\ 2\}$ $\quad h[n] = \{1,\ -1\}$
(i) $x[n] = \{2,\ 2\}$ $\quad h[n] = \{1,\ -1\}$

7.20 (Convolution and Interpolation) Let $x[n] = \{\overset{\Downarrow}{2}, 4, 6, 8\}$.
(a) Find the convolution $y[n] = x[n] \star x[n]$.
(b) Find the convolution $y_1[n] = x[2n] \star x[2n]$. Is $y_1[n]$ related to $y[n]$? Should it be? Explain.
(c) Find the convolution $y_2[n] = x[n/2] \star x[n/2]$, assuming zero interpolation. Is $y_2[n]$ related to $y[n]$? Should it be? Explain.
(d) Find the convolution $y_3[n] = x[n/2] \star x[n/2]$, assuming step interpolation. Is $y_3[n]$ related to $y[n]$? Should it be? Explain.
(e) Find the convolution $y_4[n] = x[n/2] \star x[n/2]$, assuming linear interpolation. Is $y_4[n]$ related to $y[n]$? Should it be? Explain.

7.21 (Linear Interpolation) Consider a system that performs linear interpolation by a factor of N. One way to construct such a system, as shown, is to perform up-sampling by N (zero interpolation between signal samples) and pass the up-sampled signal through a filter with impulse response $h[n]$ whose output $y[n]$ is the linearly interpolated signal.

$$x[n] \longrightarrow \boxed{\text{up-sample (zero interpolate) by } N} \longrightarrow \boxed{\text{filter}} \longrightarrow y[n]$$

(a) What should $h[n]$ be for linear interpolation by a factor of N?
(b) Let $x[n] = 4\text{tri}(0.25n)$. Find $y_1[n] = x[n/2]$ by linear interpolation.
(c) Find the system output $y[n]$ for $N = 2$. Does $y[n]$ equal $y_1[n]$?

7.22 (Causality) Argue that the impulse response $h[n]$ of a causal system must be zero for $n < 0$. Based on this result, if the input to a causal system starts at $n = n_0$, when does the response start?

7.23 (Numerical Convolution) The convolution $y(t)$ of two analog signals $x(t)$ and $h(t)$ may be approximated by sampling each signal at intervals t_s to obtain the signals $x[n]$ and $h[n]$, and folding and shifting the samples of one function past the other in steps of t_s (to line up the samples). At each instant kt_s, the convolution equals the sum of the product samples multiplied by t_s. This is equivalent to using the rectangular rule to approximate the area. If $x[n]$ and $h[n]$ are convolved using the sum-by-column method, the columns make up the product, and their sum multiplied by t_s approximates $y(t)$ at $t = kt_s$.

PGR2026-00033
Krisp Technologies EX1024 Page 206

(a) Let $x(t) = \text{rect}(t/2)$ and $h(t) = \text{rect}(t/2)$. Find $y(t) = x(t) \star h(t)$ and compute $y(t)$ at intervals of $t_s = 0.5$ s.

(b) Sample $x(t)$ and $h(t)$ at intervals of $t_s = 0.5$ s to obtain $x[n]$ and $h[n]$. Compute $y[n] = x[n] \star h[n]$ and the convolution estimate $y_R(nt_s) = t_s y[n]$. Do the values of $y_R(nt_s)$ match the exact result $y(t)$ at $t = nt_s$? If not, what are the likely sources of error?

(c) Argue that the trapezoidal rule for approximating the convolution is equivalent to subtracting half the sum of the two end samples of each column from the discrete convolution result and then multiplying by t_s. Use this rule to obtain the convolution estimate $y_T(nt_s)$. Do the values of $y_T(nt_s)$ match the exact result $y(t)$ at $t = nt_s$? If not, what are the likely sources of error?

(d) Obtain estimates based on the rectangular rule and trapezoidal rule for the convolution $y(t)$ of $x(t) = 2\,\text{tri}(t)$ and $h(t) = \text{rect}(t/2)$ by sampling the signals at intervals of $t_s = 0.5$ s. Which rule would you expect to yield a better approximation, and why?

7.24 (Convolution) Let $x[n] = \text{rect}(n/2)$ and $h[n] = \text{rect}(n/4)$.

(a) Find $f[n] = x[n] \star x[n]$ and $g[n] = h[n] \star h[n]$.

(b) Express these results as $f[n] = A\,\text{tri}(n/M)$ and $g[n] = B\,\text{tri}(n/K)$ by selecting appropriate values for the constants $A,\ M,\ B$ and K.

(c) Generalize the above results to show that $\text{rect}(n/2N) \star \text{rect}(n/2N) = (2N+1)\text{tri}(\frac{n}{2N+1})$.

7.25 (Impulse Response of Difference Algorithms) Two systems to compute the forward and backward difference are described by

Forward difference: $y_F[n] = x[n+1] - x[n]$ Backward difference: $y_B[n] = x[n] - x[n-1]$

(a) What is the impulse response of each system?

(b) Which of these systems is stable? Which of these systems is causal?

(c) Find the impulse response of their parallel connection. Is the parallel system stable? Is it causal?

(d) What is the impulse response of their cascade? Is the cascaded system stable? Is it causal?

7.26 (System Response) Find the response of the following filters to the unit step $x[n] = u[n]$, and to the alternating unit step $x[n] = (-1)^n u[n]$, using convolution concepts.

(a) $h[n] = \delta[n] - \delta[n-1]$ (differencing operation)

(b) $h[n] = \{\underset{\Downarrow}{0.5},\ 0.5\}$ (2-point average)

(c) $h[n] = \frac{1}{N}\sum_{k=0}^{N-1} \delta[n-k],\ N = 3$ (moving average)

(d) $h[n] = \frac{2}{N(N+1)}\sum_{k=0}^{N-1} (N-k)\delta[n-k],\ N = 3$ (weighted moving average)

(e) $y[n] + \frac{N-1}{N+1} y[n-1] = \frac{2}{N+1} x[n],\ N = 3$ (exponential average)

7.27 (Eigensignals) Which of the following can be the eigensignal of an LTI system?

(a) $x[n] = 0.5^n u[n]$ **(b)** $x[n] = e^{-jn\pi/2}$ **(c)** $x[n] = e^{jn\pi} + e^{jn\pi/2}$

(d) $x[n] = \cos(n\pi/2)$ **(e)** $x[n] = j^n$ **(f)** $x[n] = (\sqrt{j})^n + (\sqrt{j})^{-n}$

7.28 (Interconnected Systems) Consider two systems described by

$$h_1[n] = \delta[n] + \alpha\delta[n-1] \qquad h_2[n] = (0.5)^n u[n]$$

Find the response to the input $x[n] = (0.5)^n u[n]$ if

PGR2026-00033
Krisp Technologies EX1024 Page 207

(a) The two systems are connected in parallel with $\alpha = 0.5$.
(b) The two systems are connected in parallel with $\alpha = -0.5$.
(c) The two systems are connected in cascade with $\alpha = 0.5$.
(d) The two systems are connected in cascade with $\alpha = -0.5$.

7.29 (Systems in Cascade and Parallel) Consider the realization of Figure P7.29.

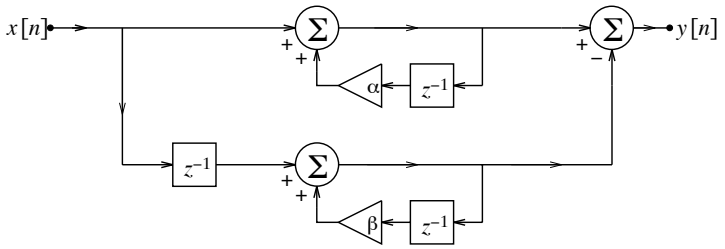


Figure P7.29 System realization for Problem 7.29

(a) Find its impulse response if $\alpha \neq \beta$. Is the overall system FIR or IIR?
(b) Find its impulse response if $\alpha = \beta$. Is the overall system FIR or IIR?
(c) Find its impulse response if $\alpha = \beta = 1$. What does the overall system represent?

7.30 (Cascading) The impulse response of two cascaded systems equals the convolution of their impulse responses. Does the step response $s_C[n]$ of two cascaded systems equal $s_1[n] \star s_2[n]$, the convolution of their step responses? If not, how is $s_C[n]$ related to $s_1[n]$ and $s_2[n]$?

7.31 (Cascading) System 1 is a squaring circuit, and system 2 is an exponential averager described by $h[n] = (0.5)^n u[n]$. Find the output of each cascaded combination. Will their output be identical? Should it be? Explain.

(a) $2(0.5)^n u[n] \longrightarrow \boxed{\text{system 1}} \longrightarrow \boxed{\text{system 2}} \longrightarrow y[n]$

(b) $2(0.5)^n u[n] \longrightarrow \boxed{\text{system 2}} \longrightarrow \boxed{\text{system 1}} \longrightarrow y[n]$

7.32 (Cascading) System 1 is an IIR filter with the difference equation $y[n] = 0.5y[n-1] + x[n]$, and system 2 is a filter with impulse response $h[n] = \delta[n] - \delta[n-1]$. Find the output of each cascaded combination. Will their output be identical? Should it be? Explain.

(a) $2(0.5)^n u[n] \longrightarrow \boxed{\text{system 1}} \longrightarrow \boxed{\text{system 2}} \longrightarrow y[n]$

(b) $2(0.5)^n u[n] \longrightarrow \boxed{\text{system 2}} \longrightarrow \boxed{\text{system 1}} \longrightarrow y[n]$

7.33 (Cascading) System 1 is an IIR filter with the difference equation $y[n] = 0.5y[n-1] + x[n]$, and system 2 is a filter with impulse response $h[n] = \delta[n] - (0.5)^n u[n]$.

(a) Find the impulse response $h_P[n]$ of their parallel connection.
(b) Find the impulse response $h_{12}[n]$ of the cascade of system 1 and system 2.
(c) Find the impulse response $h_{21}[n]$ of the cascade of system 2 and system 1.
(d) Are $h_{12}[n]$ and $h_{21}[n]$ identical? Should they be? Explain.
(e) Find the impulse response $h_I[n]$ of a system whose parallel connection with $h_{12}[n]$ yields $h_P[n]$.

PGR2026-00033
Krisp Technologies EX1024 Page 208

7.34 (Cascading) System 1 is a lowpass filter described by $y[n] = 0.5y[n-1] + x[n]$, and system 2 is described by $h[n] = \delta[n] - 0.5\delta[n-1]$.

(a) What is the output of the cascaded system to the input $x[n] = 2(0.5)^n u[n]$?
(b) What is the output of the cascaded system to the input $x[n] = \delta[n]$?
(c) How are the two systems related?

7.35 (Periodic Convolution) Consider a system whose impulse response is $h[n] = (0.5)^n u[n]$. Show that one period of its periodic extension with period N is given by $h_{\text{pe}}[n] = \dfrac{h[n]}{1-(0.5)^N},\ 0 \le n \le N-1$. Use this result to find the response of this system to the following periodic inputs.

(a) $x[n] = \cos(n\pi)$ **(b)** $x[n] = \{\overset{\Downarrow}{1},\ 1,\ 0,\ 0\}$, with $N = 4$
(c) $x[n] = \cos(0.5n\pi)$ **(d)** $x[n] = (0.5)^n,\ 0 \le n \le 3$, with $N = 4$

7.36 (Convolution in Practice) Often, the convolution of a long sequence $x[n]$ and a short sequence $h[n]$ is performed by breaking the long signal into shorter pieces, finding the convolution of each short piece with $h[n]$, and "gluing" the results together. Let $x[n] = \{1,\ 1,\ 2,\ 3,\ 5,\ 4,\ 3,\ 1\}$ and $h[n] = \{4,\ 3,\ 2,\ 1\}$.

(a) Split $x[n]$ into two equal sequences $x_1[n] = \{1,\ 1,\ 2,\ 3\}$ and $x_2[n] = \{5,\ 4,\ 3,\ 1\}$.
(b) Find the convolution $y_1[n] = h[n] \star x_1[n]$.
(c) Find the convolution $y_2[n] = h[n] \star x_2[n]$.
(d) Find the convolution $y[n] = h[n] \star x[n]$.
(e) How can you find $y[n]$ from $y_1[n]$ and $y_2[n]$? (*Hint:* Shift $y_2[n]$ and use superposition. This forms the basis for the *overlap-add* method of convolution discussed in Chapter 16.)

7.37 (Correlation) Find the correlation $r_{xh}[n]$ of the following signals.

(a) $x[n] = \alpha^n u[n]$ $\quad h[n] = \alpha^n u[n]$
(b) $x[n] = n\alpha^n u[n]$ $\quad h[n] = \alpha^n u[n]$
(c) $x[n] = \text{rect}(n/2N)$ $\quad h[n] = \text{rect}(n/2N)$

7.38 (Mean and Variance from Autocorrelation) The mean value m_x of a random signal $x[n]$ (with nonzero mean value) may be computed from its autocorrelation function $r_{xx}[n]$ as $m_x^2 = \lim\limits_{|n|\to\infty} r_{xx}[n]$. The variance of $x[n]$ is then given by $\sigma_x^2 = r_{xx}(0) - m_x^2$. Find the mean, variance, and average power of a random signal whose autocorrelation function is $r_{xx}[n] = 10\left(\dfrac{1+2n^2}{2+5n^2}\right)$.

COMPUTATION AND DESIGN

dtcongui A GUI for the Visualization of Discrete Convolution

The graphical user interface `dtcongui` allows you to visualize a movie of the convolution or correlation of two discrete-time signals.
You can select the signals and the operation (convolution or correlation).
The interface displays a movie of how the product function and the convolution or correlation results change as one function slides past the other.
Use the mouse to manually drag one function past the other and see the results change.
To explore this routine, type `dtcongui` at the MATLAB prompt.

PGR2026-00033
Krisp Technologies EX1024 Page 209

7.39 (Nonrecursive Forms of IIR Filters) An FIR filter may always be exactly represented in recursive form, but we can only approximately represent an IIR filter by an FIR filter by truncating its impulse response to N terms. The larger the truncation index N, the better is the approximation. Consider the IIR filter described by $y[n] - 0.8y[n-1] = x[n]$. Find its impulse response $h[n]$ and truncate it to 20 terms to obtain $h_A[n]$, the impulse response of the approximate FIR equivalent. Would you expect the greatest mismatch in the response of the two filters to identical inputs to occur for lower or higher values of n?

(a) Use the MATLAB routine `filter` to find and compare the step response of each filter up to $n = 15$. Are there any differences? Should there be? Repeat by extending the response to $n = 30$. Are there any differences? For how many terms does the response of the two systems stay identical, and why?

(b) Use the MATLAB routine `filter` to find and compare the response to $x[n] = 1,\ 0 \le n \le 10$ for each filter up to $n = 15$. Are there any differences? Should there be? Repeat by extending the response to $n = 30$. Are there any differences? For how many terms does the response of the two systems stay identical, and why?

7.40 (Convolution of Symmetric Sequences) The convolution of sequences that are symmetric about their midpoint is also endowed with symmetry (about its midpoint). Use the MATLAB command `conv` to find the convolution of the following sequences and establish the type of symmetry (about the midpoint) in the convolution.

(a)	$x[n] = \sin(0.2n\pi)$,	$-10 \le n \le 10$	$h[n] = \sin(0.2n\pi)$,	$-10 \le n \le 10$
(b)	$x[n] = \sin(0.2n\pi)$,	$-10 \le n \le 10$	$h[n] = \cos(0.2n\pi)$,	$-10 \le n \le 10$
(c)	$x[n] = \cos(0.2n\pi)$,	$-10 \le n \le 10$	$h[n] = \cos(0.2n\pi)$,	$-10 \le n \le 10$
(d)	$x[n] = \mathrm{sinc}(0.2n)$,	$-10 \le n \le 10$	$h[n] = \mathrm{sinc}(0.2n)$,	$-10 \le n \le 10$

7.41 (Extracting Periodic Signals Buried in Noise) Extraction of periodic signals buried in noise requires autocorrelation (to identify the period) and cross-correlation (to recover the signal itself).

(a) Generate the signal $x[n] = \sin(0.1n\pi),\ 0 \le n \le 499$. Add some uniform random noise (with a noise amplitude of 2 and a mean of 0) to obtain the noisy signal $s[n]$. Plot each signal. Can you identify any periodicity from the plot of $x[n]$? If so, what is the period N? Can you identify any periodicity from the plot of $s[n]$?

(b) Obtain the periodic autocorrelation $r_{px}[n]$ of $x[n]$ and plot. Can you identify any periodicity from the plot of $r_{px}[n]$? If so, what is the period N? Is it the same as the period of $x[n]$?

(c) Use the value of N found above (or identify N from $x[n]$ if not) to generate the 500-sample impulse train $i[n] = \sum \delta[n - kN],\ 0 \le n \le 499$. Find the *periodic* cross-correlation $y[n]$ of $s[n]$ and $i[n]$. Choose a normalizing factor that makes the peak value of $y[n]$ unity. How is the normalizing factor related to the signal length and the period N?

(d) Plot $y[n]$ and $x[n]$ on the same plot. Is $y[n]$ a close match to $x[n]$? Explain how you might improve the results.

PGR2026-00033
Krisp Technologies EX1024 Page 210

Chapter 8

FOURIER SERIES

8.0 Scope and Objectives

Our journey into the frequency domain begins with the Fourier series, which describes *periodic signals* by combinations of harmonic signals or sinusoids. This representation unfolds a perspective of periodic signals in the frequency domain in terms of their frequency content, or **spectrum**. This chapter begins with the Fourier series of periodic signals in three related forms, examines the effects of symmetry, and introduces the concept of the spectrum. It develops some useful properties of Fourier series, describes some applications of Fourier series to system analysis, and concludes with a brief historical perspective.

8.1 Fourier Series: A First Look

When we combine periodic signals with commensurate frequencies, we obtain another periodic signal. This is the process of *synthesis*. What we now attempt is the *analysis*, or separation, of a periodic signal into its periodic components. Even though the choice of periodic components is not unique, we choose harmonic signals or sinusoids because they are among the simplest periodic signals and provide a unique link with the frequency domain. From a systems viewpoint, everlasting harmonics of the form $e^{j\omega t}$ are also eigensignals of linear systems. The **Fourier series** (FS) describes a periodic signal $x_p(t)$ as a sum (linear combination), in the right mix, of *harmonics* (or sinusoids) at the fundamental frequency f_0 of $x_p(t)$ and its multiples kf_0. The choice of harmonic signals results in other advantages, too. It allows a simple, consistent, and unique scheme to find the coefficients (the right proportion or weight) for each component. In fact, a sum of sinusoids describing a periodic signal is called a *Fourier series* only if we select the coefficients according to this scheme. The Fourier series also yields the smallest *mean square error* in power between $x_p(t)$ and its series representation up to any number of terms or harmonics. It is a least squares fit to a periodic signal. Finally, the existence conditions for Fourier series (the so-called Dirichlet conditions, to be discussed later) are weak enough to assert that all periodic signals encountered in practice can be described by a Fourier series. Indeed, any exceptions are no more than mathematical curiosities.

8.1.1 The Three Forms of a Fourier Series

The **trigonometric form** of the Fourier series of a periodic signal $x_p(t)$ is simply a linear combination of sines and cosines at multiples of its fundamental frequency $f_0 = 1/T$:

$$x_p(t) = a_0 + \sum_{k=1}^{\infty} a_k \cos(2\pi k f_0 t) + b_k \sin(2\pi k f_0 t) \tag{8.1}$$

The constant term a_0 accounts for any dc offset in $x_p(t)$, and a_0, a_k, and b_k are called the **trigonometric Fourier series coefficients**. There is a pair of terms (a sine and cosine) at each harmonic frequency kf_0.

PGR2026-00033
Krisp Technologies EX1024 Page 211

The **polar form** combines each sine-cosine pair at the frequency kf_0 into a single sinusoid:

$$x_p(t) = c_0 + \sum_{k=1}^{\infty} c_k \cos(2\pi k f_0 t + \theta_k) \tag{8.2}$$

Here, $c_0 = a_0$, and c_k and θ_k are called the **polar coefficients** and are related to the trigonometric coefficients a_k and b_k. This relationship is best found by comparing the phasor representation of the time-domain terms

$$c_k \angle \theta_k = a_k - jb_k \tag{8.3}$$

The **exponential form** invokes Euler's relation to express each sine-cosine pair at the frequency kf_0 by complex exponentials at $\pm kf_0$:

$$x_p(t) = \sum_{k=-\infty}^{\infty} X[k] e^{j2\pi k f_0 t} \tag{8.4}$$

Here, the index k ranges from $-\infty$ to ∞ and $X[0] = a_0$. To see how this form arises, we invoke Euler's relation to give

$$a_k \cos(2\pi k f_0 t) + b_k \sin(2\pi k f_0 t) = 0.5a_k(e^{j2\pi k f_0 t} + e^{-j2\pi k f_0 t}) - j0.5b_k(e^{j2\pi k f_0 t} - e^{-j2\pi k f_0 t}) \tag{8.5}$$

Combining terms at kf_0 and $-kf_0$, we obtain

$$a_k \cos(2\pi k f_0 t) + b_k \sin(2\pi k f_0 t) = 0.5(a_k - jb_k)e^{j2\pi k f_0 t} + 0.5(a_k + jb_k)e^{-j2\pi k f_0 t} \tag{8.6}$$

The relation between the trigonometric and exponential forms is thus

$$X[k] = 0.5(a_k - jb_k), \qquad k \geq 1 \tag{8.7}$$

The coefficient $X[-k]$ is simply the *complex conjugate* of $X[k]$:

$$X[-k] = 0.5(a_k + jb_k) = X^*[k], \qquad k \geq 1 \tag{8.8}$$

REVIEW PANEL 8.1

Three Forms of the Fourier Series for a Periodic Signal $x_p(t)$

Trigonometric	Polar	Exponential
$a_0 + \sum_{k=1}^{\infty} a_k \cos(2\pi k f_0 t) + b_k \sin(2\pi k f_0 t)$	$c_0 + \sum_{k=1}^{\infty} c_k \cos(2\pi k f_0 t + \theta_k)$	$\sum_{k=-\infty}^{\infty} X[k] e^{j2\pi k f_0 t}$

The connection between the three forms of the coefficients is much like the connection between the three forms of a complex number, and the Argand diagram is a useful way of showing these connections.

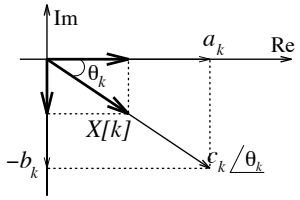
$$X[k] = 0.5(a_k - jb_k) = 0.5c_k \angle \theta_k, \qquad k \geq 1 \tag{8.9}$$

The **conjugate symmetry** of $X[k]$ also implies that a_k is an even symmetric function of k and that b_k is an odd symmetric function of k.

REVIEW PANEL 8.2

How Are the Fourier Series Coefficients Related?

$$X[0] = a_0 = c_0$$
$$X[k] = 0.5(a_k - jb_k)$$
$$X[k] = 0.5c_k \angle \theta_k = 0.5c_k e^{j\theta_k}$$



PGR2026-00033
Krisp Technologies EX1024 Page 212

8.1.2 Evaluating the Coefficients

To find the Fourier series coefficients, we need look at only one period of $x_p(t)$, since a representation that describes $x_p(t)$ over one period guarantees an identical representation over successive periods and, therefore, for the entire signal. The key to finding the Fourier series coefficients is *orthogonality*. Two signals $f(t)$ and $g(t)$ are called **orthogonal** over the interval $\alpha \le t \le \beta$ if the area of their product $f(t)g^*(t)$ over $\alpha \le t \le \beta$ equals zero:

$$\int_\alpha^\beta f(t)g^*(t)\,dt = 0 \qquad \text{(for } f(t) \text{ and } g(t) \text{ to be orthogonal)} \tag{8.10}$$

The conjugation is required only for complex valued signals. Sinusoids and complex exponentials are orthogonal over their common period $T = 1/f_0$ with

$$\int_T \cos(2\pi m f_0 t)\cos(2\pi n f_0 t)\,dt = \int_T \sin(2\pi m f_0 t)\sin(2\pi n f_0 t)\,dt = \int_T e^{j2\pi m f_0 t}e^{-j2\pi n f_0 t}\,dt = 0 \tag{8.11}$$

To invoke orthogonality, we integrate the Fourier series over one period T to give

$$\int_T x_p(t)\,dt = \int_T a_0\,dt + \cdots + \int_T a_k\cos(2\pi k f_0 t)\,dt + \cdots + \int_T b_k\sin(2\pi k f_0 t)\,dt + \cdots \tag{8.12}$$

All but the first term of the right-hand side integrate to zero, and we obtain

$$a_0 = \frac{1}{T}\int_T x_p(t)\,dt \tag{8.13}$$

The dc offset a_0 thus represents the average value of $x_p(t)$.

If we multiply $x_p(t)$ by $\cos(2\pi k f_0 t)$ and then integrate over one period T, we get

$$\int_T x_p(t)\cos(2\pi k f_0 t)\,dt = \int_T a_0\cos(2\pi k f_0 t)\,dt + \cdots + \int_T a_k\cos(2\pi k f_0 t)\cos(2\pi k f_0 t)\,dt + \cdots + \int_T b_k\sin(2\pi k f_0 t)\cos(2\pi k f_0 t)\,dt + \cdots$$

Once again, by orthogonality, all terms of the right-hand side are zero, except the one containing the coefficient a_k, and we obtain

$$\int_T x_p(t)\cos(2\pi k f_0 t)\,dt = \int_T a_k\cos^2(2\pi k f_0 t)\,dt = a_k\int_T 0.5[1+\cos(4\pi k f_0 t)]\,dt = 0.5Ta_k \tag{8.14}$$

This gives the coefficient a_k as

$$a_k = \frac{2}{T}\int_T x_p(t)\cos(2\pi k f_0 t)\,dt \tag{8.15}$$

Multiplying $x_p(t)$ by $\sin(2\pi k f_0 t)$ and integrating over one period, we similarly obtain

$$b_k = \frac{2}{T}\int_T x_p(t)\sin(2\pi k f_0 t)\,dt \tag{8.16}$$

REVIEW PANEL 8.3

Trigonometric Fourier Series Coefficients

$$a_0 = \frac{1}{T}\int_T x(t)\,dt \qquad a_k = \frac{2}{T}\int_T x(t)\cos(2\pi k f_0 t)\,dt \qquad b_k = \frac{2}{T}\int_T x(t)\sin(2\pi k f_0 t)\,dt$$

PGR2026-00033
Krisp Technologies EX1024 Page 213

The polar coefficients c_k and θ_k, or the exponential coefficients $X[k]$, may now be found using the appropriate relations. In particular, a formal expression for evaluating $X[k]$ directly is also found as follows:

$$X[k] = 0.5a_k - j0.5b_k = \frac{1}{T}\int_T x_p(t)\cos(2\pi k f_0 t)\,dt - j\frac{1}{T}\int_T x_p(t)\sin(2\pi k f_0 t)\,dt \tag{8.17}$$

We now invoke Euler's relation, $\cos(2\pi k f_0 t) - j\sin(2\pi k f_0 t) = e^{j2\pi k f_0 t}$, to obtain

$$X[k] = \frac{1}{T}\int_T x_p(t)e^{-j2\pi k f_0 t}\,dt \tag{8.18}$$

The beauty of these results is that each coefficient (for any value of k) is unique, is computed independently, and does not affect any others.

REVIEW PANEL 8.4

The Exponential Fourier Series Coefficients Display Conjugate Symmetry

$$X[0] = \frac{1}{T}\int_T x(t)\,dt \qquad X[k] = \frac{1}{T}\int_T x(t)e^{-j2\pi k f_0 t}dt \qquad X[-k] = X^*[k]$$

EXAMPLE 8.1 (Fourier Series Computations)

(a) Let $x(t) = A\cos(2\pi f_0 t + \phi)$. This sinusoid is at a single frequency f_0 (which is also the fundamental frequency). Its polar Fourier series coefficients are thus $c_1 = A$ and $\theta_1 = \phi$. Its exponential Fourier series coefficients are $X[1] = 0.5Ae^{j\phi}$ and $X[-1] = 0.5Ae^{-j\phi}$.

(b) Let $x_p(t) = 3 + \cos(4\pi t) - 2\sin(20\pi t)$. Its fundamental frequency is $f_0 = \text{GCD}(2,10) = 2$ Hz. The signal contains a dc term (for $k = 0$), a second ($k = 2$) harmonic, and a tenth ($k = 10$) harmonic. We thus have $a_0 = 3$, $a_2 = 1$, and $b_{10} = -2$. All other Fourier series coefficients are zero.

(c) Consider the signal $x(t)$ shown in Figure E8.1.

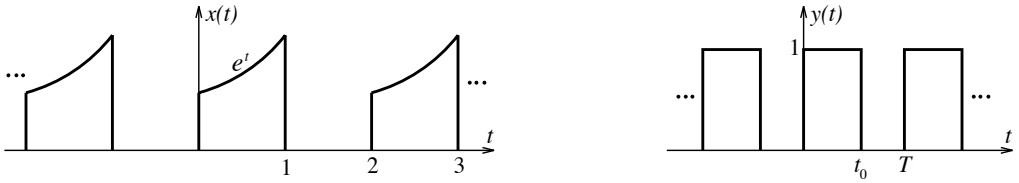


Figure E8.1 The periodic signals $x(t)$ and $y(t)$ for Example 8.1(c and d)

Its exponential Fourier series coefficients are given by

$$X[k] = \frac{1}{T}\int_0^T x(t)e^{-j2\pi k f_0 t}\,dt = \frac{1}{2}\int_0^1 e^t e^{-j2\pi k f_0 t}\,dt = \frac{1}{2}\int_0^1 e^{t(1-jk\pi)}\,dt$$

Integrating and simplifying, we obtain

$$X[k] = \left.\frac{0.5e^{t(1-jk\pi)}}{1-jk\pi}\right|_0^1 = \frac{0.5(e^{1-jk\pi}-1)}{1-jk\pi}$$

PGR2026-00033
Krisp Technologies EX1024 Page 214

(d) Consider the periodic signal $y(t)$ of Figure E8.1.

Its average value is the average area per period, and thus $Y[0] = At_0/T$.

The coefficients $Y[k]$ are given by

$$Y[k] = \frac{1}{T}\int_0^{t_0} e^{-j2\pi k f_0 t}\,dt = \left.\frac{e^{-j2\pi k f_0 t}}{-j2\pi k f_0 T}\right|_0^{t_0} = \frac{1 - e^{-j2\pi k f_0 t_0}}{j2\pi k}$$

If we factor out $e^{-j\pi k f_0 t_0}$ and use Euler's relation, we obtain

$$Y[k] = e^{-j\pi k f_0 t_0}\left[\frac{e^{j\pi k f_0 t_0} - e^{-j\pi k f_0 t_0}}{j2\pi k}\right] = e^{-j\pi k t_0/T}\left[\frac{2\sin(\pi k f_0 t_0)}{2\pi k}\right]$$

This result may also be expressed in terms of a sinc function:

$$Y[k] = e^{-j\pi k t_0/T}\left[\frac{\sin(\pi k t_0/T)}{\pi k}\right] = \frac{t_0}{T}\,\mathrm{sinc}(k\tfrac{t_0}{T})e^{-j\pi k t_0/T}$$

Comment: Note how the Fourier series coefficients of $y(t)$ depend only on the *duty ratio* t_0/T.

8.2 Simplifications Through Signal Symmetry

The Fourier series of a periodic signal $x_p(t)$ with no symmetry contains both odd (sine) and even (dc and cosine) components. If $x_p(t)$ is even symmetric, it must be made up of only even symmetric terms (dc and cosines). Hence, $b_k = 0$, and the $X[k]$ must be purely real with $X[k] = a_k/2$. If $x_p(t)$ is odd symmetric, it must contain only odd symmetric terms (sines). Hence, $a_0 = a_k = 0$, and the $X[k]$ must be purely imaginary with $X[k] = -jb_k/2$. If $x_p(t)$ is half-wave symmetric, it must contain only half-wave symmetric terms. *Only odd-indexed harmonics* (at $f_0, 3f_0, 5f_0, \ldots$) *display half-wave symmetry*. However, even-indexed harmonics (at $2f_0$ $4f_0, \ldots$) show an even number of cycles over the fundamental period T, and each half-period is identical to the next half-period and not an inverted replica as required by half-wave symmetry. Consequently, *even-indexed harmonics cannot be half-wave symmetric*, and a half-wave symmetric signal contains only *odd-indexed harmonics*, and a_0, a_k, b_k, c_k, and $X[k]$ vanish for even k.

The evaluation of the nonzero coefficients for a symmetric signal may be simplified considerably by using symmetric limits $(-0.5T,\ 0.5T)$ in the defining integrals and invoking the effects of symmetry. If $x(t)$ is even symmetric, so is $x(t)\cos(k\omega_0 t)$. To find a_k for an even symmetric signal, we may integrate $x(t)\cos(k\omega_0 t)$ only over $(0,\ 0.5T)$ and multiply by $4/T$ to find a_k. Likewise, to find b_k for an odd symmetric signal, we may integrate $x(t)\sin(k\omega_0 t)$ (which is even symmetric) only over $(0,\ 0.5T)$ and multiply by $4/T$. If $x(t)$ is half-wave symmetric, we can also integrate over $(0,\ 0.5T)$ and multiply by $4/T$, provided we evaluate the result only for odd k (odd harmonics). In the exponential form, the expression for $X[k]$ shows no simplifications except for half-wave symmetry, where we can use the limits $(0,\ 0.5T)$ but only for odd k.

If *both even and half-wave symmetry* are present, only the a_k (k odd) are nonzero and can be found by integrating $x(t)\cos(k\omega_0 t)$ over $(0,\ 0.25T)$ and multiplying by $8/T$. Likewise, for *both odd and half-wave symmetry*, only the b_k (k odd) are nonzero and may be found by integrating $x(t)\sin(k\omega_0 t)$ over $(0,\ 0.25T)$ and multiplying by $8/T$.

In all simplifications through symmetry, note that the lower integration limit is always zero.

PGR2026-00033
Krisp Technologies EX1024 Page 215

REVIEW PANEL 8.5

Effects of Signal Symmetry on the Fourier Series Coefficients

Even Symmetry in $x_p(t)$
No sine terms ($b_k = 0$)
Real $X[k]$; phase $\theta_k = 0$ or π

$$a_k = \frac{4}{T}\int_0^{T/2} x(t)\cos(2\pi k f_0 t)\,dt$$

Odd Symmetry in $x_p(t)$
No dc or cosine terms ($a_0 = a_k = 0$)
Imaginary $X[k]$; phase $\theta_k = \frac{\pi}{2}$ or $-\frac{\pi}{2}$

$$b_k = \frac{4}{T}\int_0^{T/2} x(t)\sin(2\pi k f_0 t)\,dt$$

Half-Wave Symmetry in $x_p(t)$
No dc and even-k terms
$X[k] = 0$ for k even

$$\begin{matrix} a_k \\ b_k \end{matrix} = \frac{4}{T}\int_0^{T/2} x(t) \begin{matrix} \cos(2\pi k f_0 t) \\ \sin(2\pi k f_0 t) \end{matrix} dt \quad (k \text{ odd})$$

8.2.1 Removing the DC Offset to Reveal Hidden Symmetry

Adding a dc offset to a periodic signal changes only its dc value and all other coefficients remain unchanged. Since odd symmetry and half-wave symmetry are destroyed by a nonzero dc offset, it is best to check for such symmetry only after removing the dc offset a_0 from the signal. This idea is illustrated in Figure 8.1.

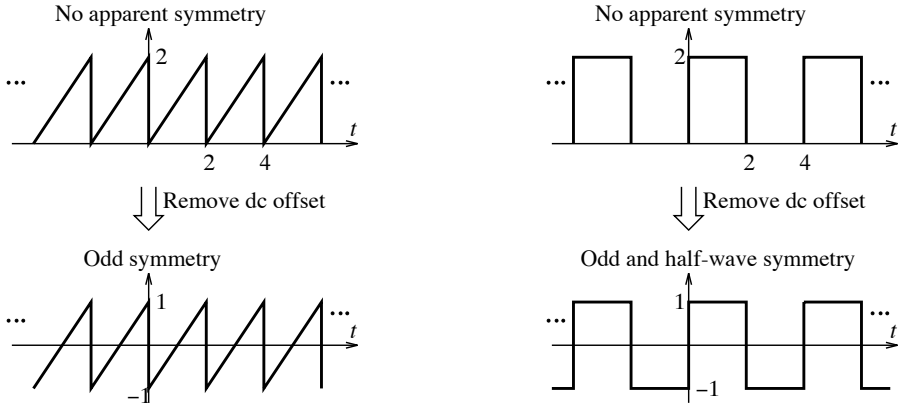


Figure 8.1 Revealing hidden symmetry by removing the dc offset

If such **hidden symmetry** is present, some of the coefficients will be zero and need not be computed. We can then pick the symmetric (zero-offset) version (using simplifications through symmetry) or the original signal (but using no such simplifications) to evaluate the remaining coefficients.

REVIEW PANEL 8.6

Useful Aids and Checks in Fourier Series Computations

$$e^{\pm jk\pi} = \cos(k\pi) = \begin{cases} 1, & k \text{ even} \\ -1, & k \text{ odd} \end{cases} \qquad \sin(k\pi) = 0 \qquad \begin{matrix} \sin(k\pi/2) = 0, & k \text{ even} \\ \cos(k\pi/2) = 0, & k \text{ odd} \end{matrix}$$

The coefficients a_k or $\mathrm{Re}\{X[k]\}$ are *even symmetric* functions of k.
The coefficients b_k or $\mathrm{Im}\{X[k]\}$ are *odd symmetric* functions of k.

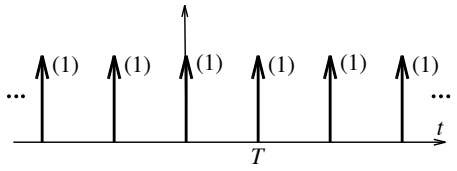
PGR2026-00033
Krisp Technologies EX1024 Page 216

EXAMPLE 8.2 (Computing the Fourier Series Coefficients)

(a) (An Impulse Train) From Figure E8.2A, and by the sifting property of impulses, we get

$$X[k] = \frac{1}{T}\int_{-T/2}^{T/2} \delta(t)e^{-j2k\pi f_0 t}\,dt = \frac{1}{T}$$

All the coefficients of an impulse train are constant!

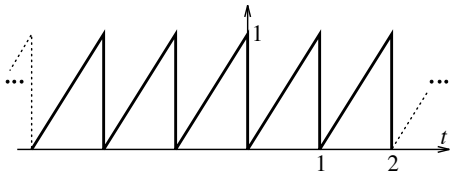


$$\Longleftrightarrow \boxed{X[0] = \frac{1}{T}\ (\text{all } k)}$$

Figure E8.2A Impulse train and its Fourier series coefficients for Example 8.2(a)

(b) (A Sawtooth Signal) From Figure E8.2B, $T = 1$ ($f_0 = 1$) and $x_p(t) = t$, $0 \le t \le 1$. By inspection, $a_0 = 0.5$. Note that $x_p(t)$ has hidden odd symmetry, and thus $a_k = 0,\ k \neq 0$. We find the coefficients $b_k,\ k > 0$, as follows:

$$b_k = 2\int_0^1 t\sin(2\pi kt)\,dt = \frac{2}{(2\pi k)^2}\left[\sin(2\pi kt) - 2\pi kt\cos(2\pi kt)\right]\Big|_0^1 = \frac{-1}{k\pi}$$



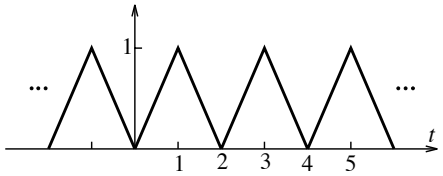
$$\Longleftrightarrow \boxed{X[0] = 0.5 \quad X[k] = \frac{j}{2\pi k}\ (k \neq 0)}$$

Figure E8.2B Periodic sawtooth and its Fourier series coefficients for Example 8.2(b)

(c) (A Triangular Pulse Train) From Figure E8.2C, we compute $a_0 = 0.5$. Since $x_p(t)$ has even symmetry, $b_k = 0,\ k \neq 0$. With $T = 2$ (or $f_0 = 0.5$) and $x_p(t) = t,\ 0 \le t \le 1$, we compute $a_k,\ k > 0$, as follows:

$$\begin{aligned} a_k &= \frac{4}{T}\int_0^{T/2} x_p(t)\cos(k\omega_0 t)\,dt = 2\int_0^1 t\cos(k\pi t)\,dt \\ &= \frac{2}{(\pi k)^2}\left[\cos(\pi kt) + \pi kt\sin(\pi kt)\right]\Big|_0^1 = \frac{2[\cos(k\pi) - 1]}{k^2\pi^2} = \frac{-4}{k^2\pi^2} \quad (k \text{ odd}) \end{aligned}$$

Note that $a_k = 0$ for even k, implying hidden half-wave symmetry (*hidden* because $a_0 \neq 0$).



$$\Longleftrightarrow \boxed{X[0] = 0.5 \qquad X[k] = -\frac{2}{\pi^2 k^2}\ (k \text{ odd})}$$

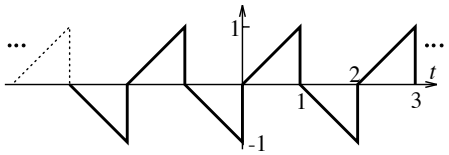
Figure E8.2C Triangular pulse train and its Fourier series coefficients for Example 8.2(c)

PGR2026-00033
Krisp Technologies EX1024 Page 217

(d) (A Half-Wave Symmetric Sawtooth Signal) From Figure E8.2D, $x_p(t)$ has half-wave symmetry. Thus, $a_0 = 0$, and the coefficients a_k and b_k are nonzero only for odd k.
With $T = 2$ ($f_0 = 0.5$) and $x_p(t) = t$, $0 \le t \le 1$, we compute a_k and b_k *only for odd* k as follows:

$$\begin{aligned} a_k &= \frac{4}{T}\int_0^{T/2} x_p(t)\cos(2\pi k f_0 t)\,dt = 2\int_0^1 t\cos(k\pi t)\,dt \\ &= \frac{2}{(\pi k)^2}\left[\cos(\pi kt) + \pi kt\sin(\pi kt)\right]\Big|_0^1 = \frac{2[\cos(k\pi)-1]}{k^2\pi^2} = \frac{-4}{k^2\pi^2} \quad (k \text{ odd}) \end{aligned}$$

$$b_k = 2\int_0^1 t\sin(k\pi t)\,dt = \frac{2}{(\pi k)^2}\left[\sin(\pi kt) - kt\cos(\pi kt)\right]\Big|_0^1 = -2\frac{\cos(k\pi)}{k\pi} = \frac{2}{k\pi} \quad (k \text{ odd})$$



$$\Longleftrightarrow \quad \boxed{X[0] = 0 \qquad X[k] = -\frac{2}{\pi^2 k^2} - j\frac{1}{\pi k} \ (k \text{ odd})}$$

Figure E8.2D Half-wave symmetric sawtooth and its Fourier series coefficients for Example 8.2(d)

(e) (A Trapezoidal Signal) From Figure E8.2E, this signal has both odd and hidden half-wave symmetry. Thus, $a_0 = 0 = a_k$, and the b_k are nonzero only for odd k. With $T = 8$ ($f_0 = \frac{1}{8}$); and $x_p(t) = t,\ 0 \le t \le 1$, and $x_p(t) = 1,\ 1 \le t \le 2$; we use simplifications due to both types of symmetry to obtain

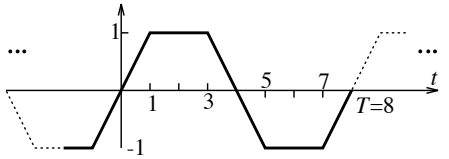
$$b_k = \frac{8}{T}\int_0^{T/4} x_p(t)\sin(\tfrac{k\pi t}{4})\,dt = \int_0^1 t\sin(\tfrac{k\pi t}{4})\,dt + \int_1^2 \sin(\tfrac{k\pi t}{4})\,dt = I_1 + I_2$$

Now, simplifying the two integrals I_1 and I_2 *for odd* k *only*, we find

$$\begin{aligned} I_1 &= \frac{1}{(k\pi/4)^2}\left[\sin(\tfrac{k\pi t}{4}) - \tfrac{k\pi t}{4}\cos(\tfrac{k\pi t}{4})\right]\Big|_0^1 = \frac{16\sin(k\pi/4)}{k^2\pi^2} - \frac{4\cos(k\pi/4)}{k\pi} \\ I_2 &= \frac{-\cos(k\pi t/4)}{(k\pi/4)}\Big|_1^2 = \frac{-4[\cos(k\pi/2) - \cos(k\pi/4)]}{k\pi} = \frac{4\cos(k\pi/4)}{k\pi} \quad (k \text{ odd}) \end{aligned}$$

Then

$$b_k = I_1 + I_2 = \frac{16\sin(k\pi/4)}{k^2\pi^2} \quad (k \text{ odd})$$



$$\Longleftrightarrow \quad \boxed{X[0] = 0 \qquad X[k] = -j\frac{8\sin(k\pi/4)}{\pi^2 k^2} \ (k \text{ odd})}$$

Figure E8.2E Trapezoidal pulse train and its Fourier series coefficients for Example 8.2(e)

(f) (A Half-Wave Rectified Sine) From Figure E8.2F, this signal has no symmetry. Its average value equals $a_0 = 1/\pi$. With $T = 2$ ($f_0 = 0.5$) and $x_p(t) = \sin(\pi t)$, $0 \le t \le 1$, we compute the coefficient a_k

PGR2026-00033
Krisp Technologies EX1024 Page 218

as follows:

$$a_k = \int_0^1 \sin(\pi t)\cos(k\pi t)\,dt = \frac{1}{2}\int_0^1 \big\{\sin[(1+k)\pi t] + \sin[(1-k)\pi t]\big\}\,dt$$

Upon simplification, using $\cos(1 \pm k)\pi = -\cos(k\pi)$, this yields

$$a_k = \frac{-1}{2\pi}\left[\frac{\cos[(1+k)\pi t]}{1+k} + \frac{\cos[(1-k)\pi t]}{1-k}\right]\Bigg|_0^1 = \frac{1+\cos(k\pi)}{\pi(1-k^2)} = \frac{2}{\pi(1-k^2)} \quad (k \text{ even})$$

Similarly, we compute b_k as follows:

$$\begin{aligned} b_k &= \int_0^1 \sin(\pi t)\sin(k\pi t)\,dt = \frac{1}{2}\int_0^1 \big\{\cos[(1-k)\pi t] - \cos[(1+k)\pi t\big\}\,dt \\ &= \frac{1}{2\pi}\left[\frac{\sin[(1-k)\pi t]}{1-k} + \frac{\sin[(1+k)\pi t]}{1+k}\right]\Bigg|_0^1 = 0 \end{aligned}$$

This may seem plausible to the unwary, but if $b_k = 0$, the signal $x(t)$ must show even symmetry. It does not! The catch is that for $k = 1$ the first term has the indeterminate form $0/0$. We must thus evaluate b_1 separately, either by l'Hôpital's rule,

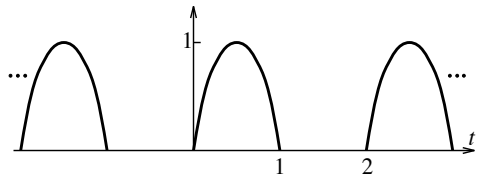
$$b_1 = \frac{1}{2\pi}\lim_{k\to 1}\left[\frac{\sin[(1-k)\pi]}{1-k}\right] = \frac{1}{2\pi}\lim_{k\to 1}\left[\frac{-\pi\cos[(1-k)\pi]}{-1}\right] = 0.5$$

or directly from the defining relation with $k = 1$,

$$b_1 = \int_0^1 \sin(\pi t)\sin(\pi t)\,dt = 0.5\int_0^1 [1-\cos(2\pi t)]\,dt = 0.5$$

Thus, $b_1 = 0.5$ and $b_k = 0,\ k > 1$. By the way, we must also evaluate a_1 separately (it equals zero).

Comment: *Indeterminate forms typically arise for signals with sinusoidal segments.*



$$\Longleftrightarrow \quad \boxed{X[0] = \frac{1}{\pi} \quad X[1] = \frac{-j}{4} \quad X[k] = \frac{1}{\pi(1-k^2)} \quad (k \text{ even})}$$

Figure E8.2F Half-wave rectified sine and its Fourier series coefficients for Example 8.2(f)

8.3 Parseval's Relation and the Power in Periodic Signals

Given the periodic signal $x_p(t)$, it is easy enough to evaluate the signal power in the time domain as

$$P = \frac{1}{T}\int_T x_p^2(t)\,dt \tag{8.19}$$

Since the Fourier series describes $x_p(t)$ as a sum of sinusoids at different frequencies, its signal power also equals the sum of the power in each sinusoid

$$P = c_0^2 + \sum_{k=1}^{\infty} 0.5c_k^2 \tag{8.20}$$

PGR2026-00033
Krisp Technologies EX1024 Page 219

Equivalent formulations in terms of c_k or $X[k]$ are obtained by recognizing that $c_k^2 = a_k^2 + b_k^2 = 4|X[k]|^2$ and $|X[k]|^2 = X[k]X^*[k]$ to give

$$P = a_0^2 + \sum_{k=1}^{\infty} 0.5(a_k^2 + b_k^2) = \sum_{k=-\infty}^{\infty} |X[k]|^2 = \sum_{k=-\infty}^{\infty} X[k]X^*[k] \tag{8.21}$$

The equivalence of the time-domain and frequency-domain expressions for the signal power forms the so-called **Parseval's relation**:

$$P = \frac{1}{T}\int_T x_p^2(t)\,dt = \sum_{k=-\infty}^{\infty} |X[k]|^2 \tag{8.22}$$

We point out that Parseval's relation holds only if one period of $x_p(t)$ is square integrable (an energy signal). Therefore, it does not apply to signals such as the impulse train.

It is far easier to obtain the *total* power using $x_p(t)$. However, finding the power in or up to a given number of harmonics requires the harmonic coefficients. The power P_N up to the Nth harmonic is simply

$$P_N = c_0^2 + \sum_{k=1}^{N} \tfrac{1}{2}c_k^2 = a_0^2 + \sum_{k=1}^{N} \tfrac{1}{2}[a_k^2 + b_k^2] = \sum_{k=-N}^{N} |X[k]|^2 \tag{8.23}$$

This is a sum of positive quantities and is always less than the total signal power. As N increases, P_N approaches the true signal power P.

REVIEW PANEL 8.7

Parseval's Theorem: We Can Find the Signal Power from $x_p(t)$ or Its Spectrum $X[k]$

$$P \quad = \quad \frac{1}{T}\int_T x^2(t)\,dt \quad = \quad \sum_{k=-\infty}^{\infty} |X[k]|^2$$

EXAMPLE 8.3 (Using Parseval's Relation)

Consider the periodic signal $x(t)$ shown in Figure E8.3.

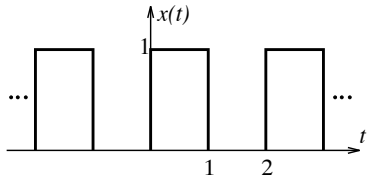


Figure E8.3 The periodic signal of Example 8.3

The signal power in $x(t)$ is

$$P = \frac{1}{T}\int_0^T x^2(t)\,dt = \frac{1}{2}\int_0^1 dt = 0.5\,\text{W}$$

The exponential Fourier series coefficients of $x(t)$ are

$$X[k] = 0.5\,\text{sinc}(0.5k)e^{-jk\pi/2} \qquad |X[k]| = 0.5\,\text{sinc}(0.5k) = \frac{\sin(k\pi/2)}{k\pi}$$

PGR2026-00033
Krisp Technologies EX1024 Page 220

The coefficients are zero for even k. To compute the power up to the second harmonic ($k = 2$), we use

$$P_2 = \sum_{k=-2}^{2} |x[k]|^2 = \frac{1}{\pi^2} + 0.25 + \frac{1}{\pi^2} = 0.25 + \frac{2}{\pi^2} = 0.4526\,\text{W}$$

To compute the total power from the harmonics, we would have to sum the infinite series:

$$P = \sum_{k=-\infty}^{\infty} |x[k]|^2 = 0.25 + 2\sum_{k\ \text{odd}}^{\infty} \frac{1}{k^2\pi^2} = 0.25 + \frac{2}{\pi^2}\sum_{k\ \text{odd}}^{\infty} \frac{1}{k^2} = 0.25 + \frac{2}{\pi^2}\left(\frac{\pi^2}{8}\right) = 0.5\,\text{W}$$

The summation in the above expression equals $\pi^2/8$ to give $P = 0.5\,\text{W}$, as before. Finding the *total* power from the harmonics is tedious! It is far better to find the total power directly from $x(t)$.

8.4 The Spectrum of Periodic Signals

The terms **spectral analysis** or **harmonic analysis** are often used to describe the analysis of a periodic signal $x_p(t)$ by its Fourier series. The quantities a_k, b_k, c_k, θ_k, or $X[k]$ describe the **spectral coefficients** of $x_p(t)$. They may be plotted against the harmonic index k, or kf_0 (hertz), or $k\omega_0$ (radians per second), as shown in Figure 8.2, and are called **spectra** or **spectral plots**.

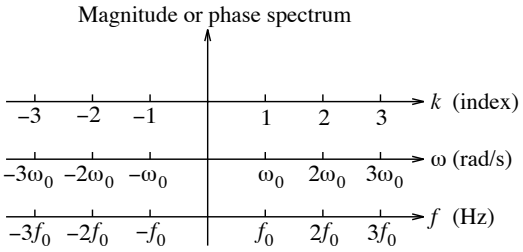


Figure 8.2 Various ways of plotting the Fourier series spectra

The **magnitude spectrum** and **phase spectrum** describe plots of the magnitude and phase of each harmonic. They are plotted as discrete signals and sometimes called **line spectra**. One-sided spectra refer to plots of c_k and θ_k for $k \geq 0$ (positive frequencies). Two-sided spectra refer to plots of $|X[k]|$ and θ_k for all k (all frequencies, *positive and negative*).

For real periodic signals, the $X[k]$ display conjugate symmetry. As a result, the two-sided magnitude $|X[k]|$ shows even symmetry, and the two-sided phase spectrum shows reversed phase at negative indices (or negative frequencies). The magnitude is usually sketched as a positive quantity. An example is shown in Figure 8.3.

REVIEW PANEL 8.8

Periodic Signals Have Discrete Magnitude and Phase Spectra

The magnitude and phase may be plotted against k or $f = kf_0$ (Hz) or $\omega = k\omega_0$ (radians/s).
One-sided spectra: Plot c_k and θ_k, $k \geq 0$. **Two-sided spectra:** Plot $|X[k]|$ and $\angle X[k]$, *all* k.
The two-sided magnitude spectrum is *even symmetric*.
The two-sided phase spectrum is *odd symmetric* (if the dc value is non-negative).

PGR2026-00033
Krisp Technologies EX1024 Page 221

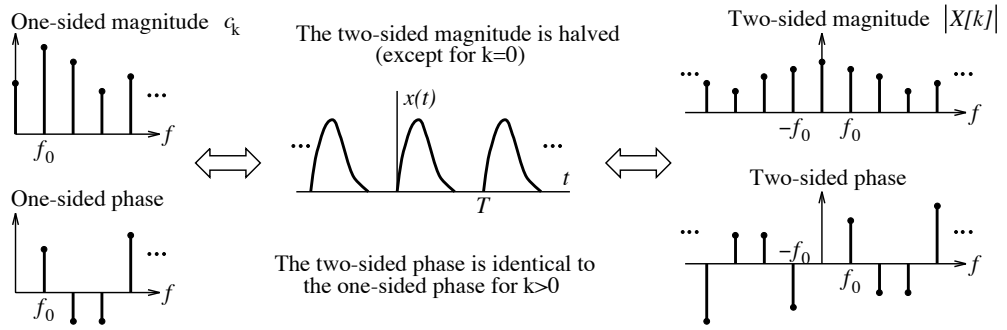


Figure 8.3 One-sided and two-sided Fourier series spectra

For real signals, the $X[k]$ are purely real if $x_p(t)$ is even symmetric or purely imaginary if $x_p(t)$ is odd symmetric. In such cases, it is more meaningful to sketch the **amplitude spectrum** as the real (or imaginary) part of $X[k]$ (which may include sign changes). An example is shown in Figure 8.4.

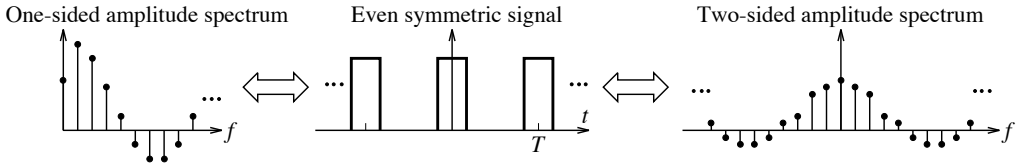


Figure 8.4 The amplitude spectrum

REVIEW PANEL 8.9

The Amplitude Spectrum Is Useful for Symmetric Periodic Signals

Even symmetry in $x_p(t)$: Plot a_k (one-sided) or $X[k]$ (two-sided). The phase is zero.
Odd symmetry in $x_p(t)$: Plot b_k (one-sided) or $\text{Im}\{X[k]\}$ (two-sided). The phase is $\pm 90°$.

8.4.1 Signal Symmetry from Its Fourier Series Spectra

The odd or even symmetry of a periodic signal may be discerned from its phase spectrum. If the phase of *every* Fourier series coefficient is $90°$ or $-90°$, the time signal contains only sine terms and has odd symmetry (if the dc component is zero), or hidden odd symmetry (if the dc component is nonzero). If the phase of *every* Fourier series coefficient is $0°$ or $\pm 180°$, the time signal contains only cosines (and perhaps a dc component) and has even symmetry. For half-wave symmetry, the dc component must be zero, and the magnitude and phase spectra must show harmonics only at odd values of the index k.

REVIEW PANEL 8.10

Recognizing Signal Symmetry from Magnitude and Phase Spectra

Half-wave symmetry: *Zero dc value* and no even-k harmonics in *magnitude or phase spectrum.*
Odd symmetry: *Zero dc value* and phase of *all harmonics* is $90°$ or $-90°$.
Even symmetry: Phase of *all harmonics* is *zero* or $\pm 180°$.

PGR2026-00033
Krisp Technologies EX1024 Page 222

Identifying the kth Harmonic

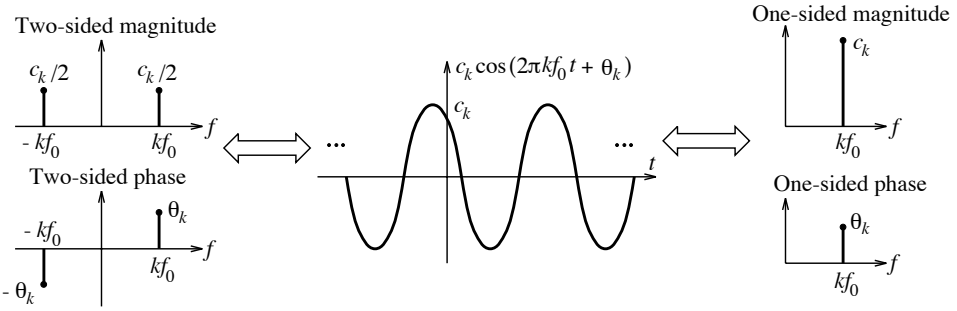
If we write $x_p(t) = \sum_{k=0}^{\infty} x_k(t)$, the quantity $x_k(t)$ describes the kth time-domain harmonic with

$$x_k(t) = c_k \cos(2\pi k f_0 t + \theta_k) = 2|X[k]|\cos(2\pi k f_0 t + \theta_k) \qquad \text{(the } k\text{th harmonic)} \tag{8.24}$$

We can identify $x_k(t)$ from the magnitude and phase spectrum. If the spectra are one-sided (plots of c_k), the kth harmonic is simply $x_k(t) = c_k \cos(2\pi k f_0 t + \theta_k)$. If the spectra are two-sided (plots of $|X[k]|$), the kth harmonic is $x_k(t) = 2|X[k]|\cos(2\pi k f_0 t + \theta_k)$, where θ_k is the phase at the frequency $f = kf_0$.

REVIEW PANEL 8.11

Finding the kth Time-Domain Harmonic $x_k(t)$ from Its One-Sided or Two-Sided Spectra



EXAMPLE 8.4 (Signal Symmetry and Measures from Fourier Series Spectra)

(a) Consider the signal $x(t)$ whose two-sided spectra are shown in Figure E8.4.

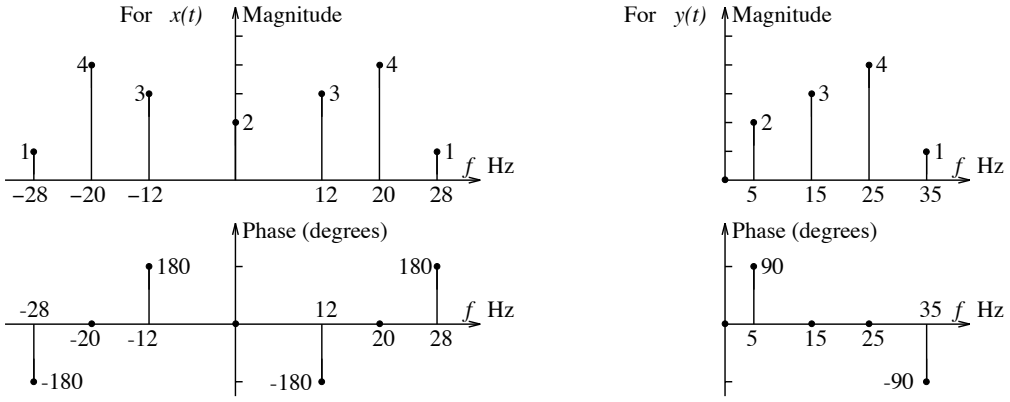


Figure E8.4 The spectra of the periodic signals $x(t)$ and $y(t)$ of Example 8.4(a and b)

The harmonic frequencies are 12 Hz, 20 Hz, and 28 Hz. The fundamental frequency is given by $f_0 = \text{GCD}(12,\ 20,\ 28) = 4$ Hz. The time period is $T = 1/f_0 = 0.25$ s. Only the dc and odd-indexed harmonics are present. Thus, the signal shows hidden (not true) half-wave symmetry. The phase of the harmonics is either zero or $\pm 180°$, implying only cosine terms. Thus, we also have even symmetry.

Since the spectrum is two-sided, the total signal power and the rms value are found as

$$P = \sum |X[k]|^2 = 1 + 16 + 9 + 4 + 9 + 16 + 1 = 56\,\text{W} \qquad x_{\text{rms}} = \sqrt{P} = \sqrt{56} = 7.4833$$

PGR2026-00033
Krisp Technologies EX1024 Page 223

We may also write the Fourier series for the signal by inspection as either
$x(t) = 2 + 6\cos(24\pi t - 180°) + 8\cos(40\pi t) + 2\cos(56\pi t + 180°)$ or
$x(t) = 2 - 6\cos(24\pi t) + 8\cos(40\pi t) - 2\cos(56\pi t)$.

(b) Consider the signal $y(t)$ whose one-sided spectra are shown in Figure E8.4.

The harmonic frequencies are 5 Hz, 15 Hz, 25 Hz, and 35 Hz. Thus, the fundamental frequency is $f_0 = 5$ Hz, and the time period is $T = 1/f_0 = 0.2$ s. Only the odd-indexed harmonics are present. Since the dc value is zero, the signal shows half-wave symmetry. The phase of the harmonics is either zero (cosine terms) or $\pm 90°$ (sine terms). Thus, both sines and cosines are present, and the signal is neither odd symmetric nor even symmetric.

Since the spectrum is one-sided, the total signal power and the rms value are found as

$$P = c_0^2 + 0.5\sum c_k^2 = 0.5(4 + 9 + 16 + 1) = 15\,\text{W} \qquad y_{\text{rms}} = \sqrt{P} = \sqrt{15} = 3.8730$$

We may also write the Fourier series for the signal by inspection as either
$y(t) = 2\cos(10\pi t + 90°) + 3\cos(30\pi t) + 4\cos(50\pi t) + \cos(70\pi t - 90°)$ or
$y(t) = -2\sin(10\pi t) + 3\cos(30\pi t) + 4\cos(50\pi t) + \sin(70\pi t)$.

8.5 Properties of Fourier Series

The Fourier series provides a link between the time domain and the frequency domain, and the connection may be displayed symbolically by the transform pair $x_p(t) \Leftrightarrow X[k]$. We now describe how various operations on $x_p(t)$ affect the spectral coefficients $X[k]$.

8.5.1 Time Shift

A time shift changes $x_p(t)$ to $y_p(t) = x_p(t \pm \alpha)$. We may then write

$$y_p(t) = \sum_{k=-\infty}^{\infty} X[k]e^{j2\pi k f_0(t\pm\alpha)} = \sum_{k=-\infty}^{\infty} X[k]e^{\pm j2\pi k f_0\alpha}e^{j2\pi k f_0 t} \tag{8.25}$$

This represents a Fourier series with coefficients $Y[k] = X[k]e^{\pm j2\pi k f_0\alpha}$. Thus,

$$x_p(t-\alpha) \Longleftrightarrow X[k]e^{-j2\pi k f_0\alpha} \tag{8.26}$$

The magnitude spectrum shows no change. The term $\exp(-j2\pi k f_0\alpha)$ introduces an additional phase of $-2k\pi f_0\alpha$, which varies linearly with k (or frequency). It changes the phase of each harmonic in proportion to its frequency kf_0.

8.5.2 Time Scaling

A time scaling of $x_p(t)$ to $y_p(t) = x_p(\alpha t)$ results in the Fourier series

$$y_p(t) = x_p(\alpha t) = \sum_{k=-\infty}^{\infty} X[k]e^{j2\pi k f_0\alpha t} \tag{8.27}$$

PGR2026-00033
Krisp Technologies EX1024 Page 224

Time compression by α changes the fundamental frequency of the scaled signal $y_p(\alpha t)$ from f_0 to αf_0. The spectral coefficients $Y[k]$ are identical to $X[k]$ for any k, but located at the frequencies $k\alpha f_0$. *The more compressed the time signal, the farther apart are its harmonics in the frequency domain.*

8.5.3 Spectrum Zero Interpolation

For integer scale factors, we also have a useful result. A *zero interpolation* of the spectral coefficients $X[k]$ to $X[k/N]$ leads to an N-fold signal compression. Note that the period of the compressed signal is assumed to be the same as that of the original. In other words, one period of the new signal contains N copies of the original (with a fundamental frequency that is N times the original). In operational notation,

$$x_p(Nt) \ \text{(signal compression by}N\text{)} \quad \Longleftrightarrow \quad X[k/N] \ \text{(zero interpolation by } N\text{)} \tag{8.28}$$

Figure 8.5 illustrates this result for compression by 2.

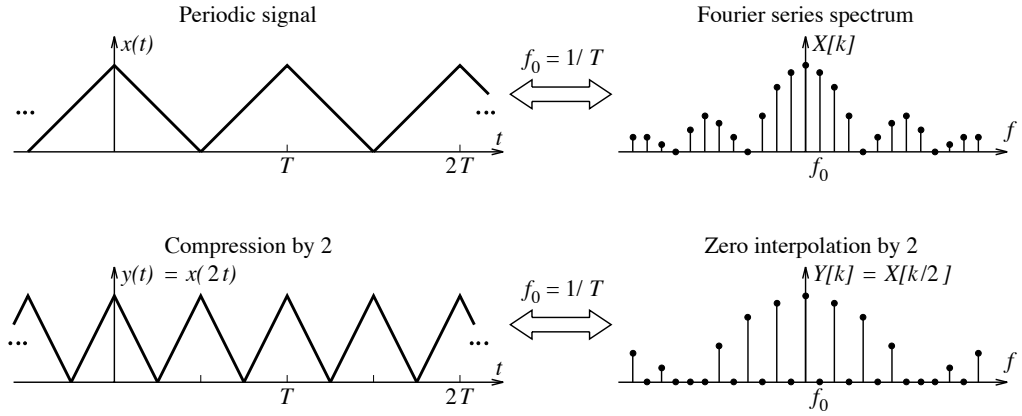


Figure 8.5 Signal compression leads to zero interpolation of the spectrum

8.5.4 Folding

The folding operation means $\alpha = -1$ in the scaling property and gives

$$x_p(-t) \Longleftrightarrow X[-k] = X^*[k] \tag{8.29}$$

The magnitude spectrum remains unchanged. However, the phase is reversed. This also implies that the a_k remain unchanged, but the b_k change sign. Since $x_p(t) + x_p(-t) \Leftrightarrow X^*[k] + X[k] = 2\text{Re}\{X[k]\}$, we see that the Fourier series coefficients of the even part of $x_p(t)$ equal $\text{Re}\{X[k]\}$ and the Fourier series coefficients of the odd part of $x_p(t)$ equal $\text{Im}\{X[k]\}$.

8.5.5 Derivatives

The derivative of $x_p(t)$ is found by differentiating each term in its series:

$$\frac{dx_p(t)}{dt} = \sum_{k=-\infty}^{\infty} \frac{d}{dt}\left(X[k]e^{j2\pi k f_0 t}\right) = \sum_{k=-\infty}^{\infty} j2\pi k f_0 X[k] e^{j2\pi k f_0 t} \tag{8.30}$$

PGR2026-00033
Krisp Technologies EX1024 Page 225

In operational notation,

$$x_p'(t) \Longleftrightarrow j2\pi k f_0 X[k] \tag{8.31}$$

The dc term $X[0]$ vanishes, whereas all other $X[k]$ are multiplied by $j2\pi k f_0$. The magnitude of each harmonic is scaled by $2\pi k f_0$, in proportion to its frequency, and the spectrum has significantly larger high-frequency content. The derivative operation enhances the sharp details and features in any time function and contributes to its high-frequency spectrum.

8.5.6 Integration

Integration of a periodic signal whose average value is zero ($X[0] = 0$) gives

$$\int_0^t x_p(t)\,dt \Longleftrightarrow \frac{X[k]}{j2\pi k f_0},\ (k \neq 0) \quad + \quad \text{constant} \tag{8.32}$$

The integrated signal is also periodic. If we keep track of $X[0]$ separately, integration and differentiation may be thought of as inverse operations. The coefficients $\frac{X[k]}{j2\pi k f_0}$ imply faster convergence and reduce the high-frequency components. Since integration is a smoothing operation, the smoother a time signal, the smaller its high-frequency content.

REVIEW PANEL 8.12

The Effect of Signal Operations on the Magnitude and Phase Spectrum of $x_p(t)$

Time scale: $x_p(\alpha t)$ has Fourier series coefficients $X[k]$, *but* at the harmonic frequencies $k f_0 \alpha$.

Time shift: $x_p(t-\alpha)$ has magnitude $|X[k]|$, *but* its phase is $\theta_k - 2\pi k f_0 \alpha$ (a *linear phase change*).

Derivative: $x_p'(t)$ has magnitude $2\pi k f_0 |X[k]|$; its phase is $\theta_k + 90°$.

EXAMPLE 8.5 (Fourier Series Properties)

The Fourier series coefficients of the signal $x(t)$ shown in Figure E8.5 are

$$X[0] = \frac{1}{\pi} \qquad X[1] = -j0.25 \qquad X[k] = \frac{1}{\pi(1-k^2)} \quad (k \text{ even})$$

Use properties to find the Fourier series coefficients of $y(t)$, $f(t)$, and $g(t)$.

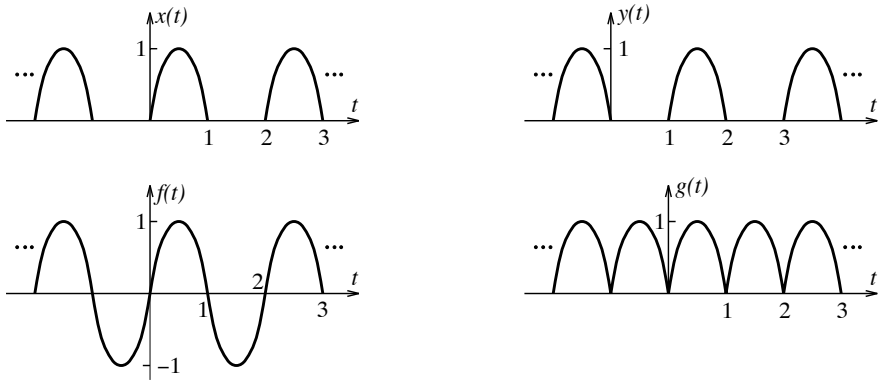


Figure E8.5 The periodic signals for Example 8.5

PGR2026-00033
Krisp Technologies EX1024 Page 226

(a) We note that $y(t) = x(t - 0.5T)$, and thus $Y[k] = X[k]e^{-j\pi k f_0 T} = X[k]e^{-j\pi k} = (-1)^k X[k]$. The coefficients change sign only for odd k, and we have

$$Y[0] = X[0] = \frac{1}{\pi} \qquad Y[1] = -X[1] = j0.25 \qquad Y[k] = X[k] = \frac{1}{\pi(1-k^2)} \quad (k \text{ even})$$

(b) The sinusoid $f(t)$ may be described as $f(t) = x(t) - y(t)$. Its Fourier series coefficients may be written as $F[k] = X[k] - Y[k]$, and are given by

$$F[0] = 0 \qquad F[1] = -j0.25 - j0.25 = -j0.5 \qquad F[k] = 0 \quad (k \neq 1)$$

As expected, the Fourier series coefficients describe a pure sine wave.

(c) The full-wave rectified sine $g(t)$ may be described as $g(t) = x(t) + y(t)$. Its Fourier series coefficients $G[k] = X[k] - Y[k]$ are given by

$$G[0] = \frac{2}{\pi} \qquad G[1] = 0 \qquad G[k] = \frac{2}{\pi(1-k^2)} \quad (k \text{ even})$$

Note that the time period of $g(t)$ is in fact half that of $x(t)$ or $y(t)$. Since only even-indexed coefficients are present, the Fourier series coefficients may also be written as

$$G[k] = \frac{2}{\pi(1-4k^2)} \quad (\text{all } k)$$

8.5.7 Fourier Series Coefficients by the Derivative Method

The derivative property provides an elegant way of finding the Fourier series coefficients without the need for involved integration. It is based on the fact that derivatives of piecewise signals sooner or later yield impulse trains whose coefficients are quite easy to obtain (by inspection or by the sifting property). To implement this method, we take enough derivatives of a signal $x_p(t)$ to obtain another signal (such as an impulse train) $x^{(m)}(t)$ whose coefficients $X_m[k]$ are easy to obtain. The coefficients of $x_p(t)$ and $x^{(m)}(t)$ are then related by the derivative property. Note that the dc value $X[0]$ must be evaluated separately (because the dc component disappears after the very first derivative). Any special cases must also be evaluated separately.

EXAMPLE 8.6 (The Derivative Method)

(a) Consider the rectangular pulse train shown in Figure E8.6A.

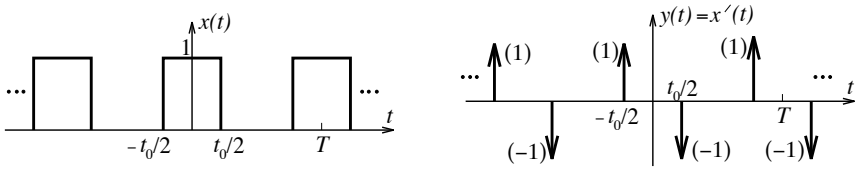


Figure E8.6A The signals for Example 8.6(a)

Its first derivative $y(t) = x'(t)$ contains only impulses. We use the sifting property to find the $Y[k]$ as

$$Y[k] = \frac{1}{T}\int_{-T/2}^{T/2} [\delta(t - 0.5t_0) - \delta(t + 0.5t_0)]e^{-j2\pi k f_0 t}\, dt = \frac{1}{T}\left[e^{j\pi k f_0 t_0} - e^{-j\pi k f_0 t_0}\right] = \frac{2j\sin(\pi k f_0 t_0)}{T}$$

PGR2026-00033
Krisp Technologies EX1024 Page 227

The coefficients of $x(t)$ and $y(t)$ are related by $X[k] = Y[k]/j2\pi kf_0$. Thus,

$$X[k] = \frac{Y[k]}{j2\pi kf_0} = \frac{\sin(\pi kf_0t_0)}{\pi kf_0T} = \frac{t_0}{T}\frac{\sin(\pi kf_0t_0)}{\pi kf_0t_0} = \frac{t_0}{T}\,\text{sinc}(kf_0t_0)$$

(b) Consider the triangular pulse train shown in Figure E8.6B.

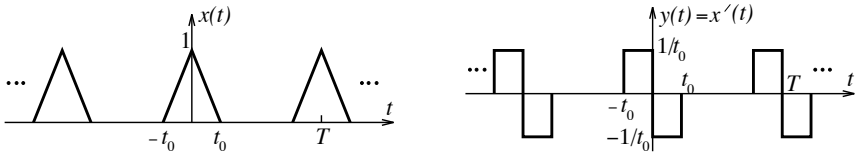


Figure E8.6B The signals for Example 8.6(b)

Its first derivative $y(t) = x'(t)$ contains rectangular pulses of width t_0 and height $1/t_0$. We use the result of the previous part and the shifting property to find the $Y[k]$ as

$$Y[k] = \frac{1}{T}\,\text{sinc}(kf_0t_0)\left[e^{j\pi kf_0t_0} - e^{-j\pi kf_0t_0}\right] = \frac{2j\sin(\pi kf_0t_0)\text{sinc}(kf_0t_0)}{T}$$

The coefficients of $x(t)$ and $y(t)$ are related by $X[k] = Y[k]/j2\pi kf_0$. Thus,

$$X[k] = \frac{Y[k]}{j2\pi kf_0} = \frac{\sin(\pi kf_0t_0)\text{sinc}(kf_0t_0)}{\pi kf_0T} = \frac{t_0}{T}\frac{\sin(\pi kf_0t_0)}{\pi kf_0t_0}\,\text{sinc}(kf_0t_0) = \frac{t_0}{T}\,\text{sinc}^2(kf_0t_0)$$

(c) Consider the full-wave rectified signal shown in Figure E8.6C.

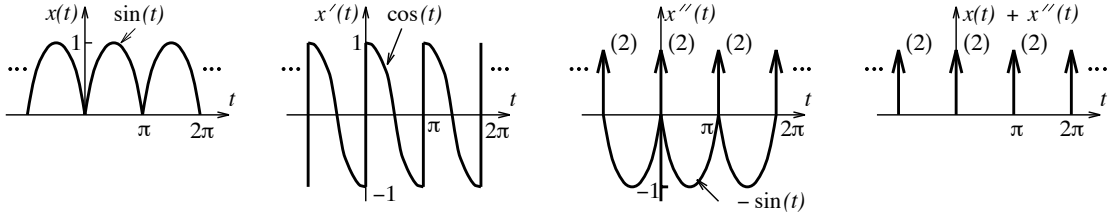


Figure E8.6C The signals for Example 8.6(c)

Its fundamental frequency is $\omega_0 = 2\pi/T = 2$ rad/s. Over one period, $x(t) = \sin(t)$. Two derivatives result in impulses, and the sum $y(t) = x(t) + x''(t)$ yields an impulse train with strengths of 2 units whose Fourier series coefficients are simply $Y[k] = 2/\pi$.

Since $Y[k] = X[k] + (jk\omega_0)^2X[k] = X[k](1 - 4k^2)$, we have

$$X[k] = \frac{2}{\pi(1 - 4k^2)}$$

REVIEW PANEL 8.13

Fourier Series Coefficients of Two Periodic Pulse Trains with Period T



PGR2026-00033
Krisp Technologies EX1024 Page 228

8.5.8 Multiplication and Convolution

Multiplication and convolution are duals. Multiplication in one domain corresponds to convolution in the other. In particular, the multiplication of two signals leads to the *discrete convolution* of their spectra. If the spectra are described by an infinite series, the convolution is not easily amenable to a closed-form solution. Discrete convolution requires equally spaced samples for both sequences. If the signals have different periods, we must use their fundamental frequency as the spacing between the spectral samples. This means that missing values in $X[k]$ and $Y[k]$ at multiples of the fundamental frequency are treated as zero.

Duality suggests that the product of two spectra corresponds to the *periodic convolution* of their periodic time signals:

$$x_p(t) \circledast y_p(t) \Longleftrightarrow X[k]Y[k] \tag{8.33}$$

Only if the spectral coefficients are located at like harmonic frequencies will their product sequence be nonzero. In other words, the periods of $x_p(t)$ and $y_p(t)$ must be identical (or commensurate).

REVIEW PANEL 8.14

Convolution and Multiplication Are Duals

$$x_p(t) \circledast y_p(t) \Longleftrightarrow X[k]Y[k] \qquad\qquad x_p(t)y_p(t) \Longleftrightarrow X[k] \star Y[k]$$

EXAMPLE 8.7 (Convolution)

(a) (Frequency-Domain Convolution)

Let $x(t) = 2 + 4\sin(6\pi t)$ and $y(t) = 4\cos(2\pi t) + 6\sin(4\pi t)$. Their spectral coefficients are given by $X[k] = \{j2, \overset{\Downarrow}{2}, -j2\}$ and $Y[k] = \{j3, 2, \overset{\Downarrow}{0}, 2, -j3\}$. Since the fundamental frequencies of $x(t)$ and $y(t)$ are 6π and 2π, the fundamental frequency of the product is $\omega_0 = \text{GCD}(6\pi, 2\pi) = 2\pi$. We insert zeros at the missing harmonics 2π and 4π in $X[k]$ to get $X[k] = \{j2, 0, 0, \overset{\Downarrow}{2}, 0, 0, -j2\}$. The sequence $Y[k]$ has no missing harmonics. The spectral coefficients $Z[k]$ of the product $x(t)y(t)$ are

$$Z[k] = \{j2, 0, 0, \overset{\Downarrow}{2}, 0, 0, -j2\} \star \{j3, 2, \overset{\Downarrow}{0}, 2, -j3\}$$

We find that $Z[k] = \{-6, j4, 0, j10, 10, \overset{\Downarrow}{0}, 10, -j10, 0, -j4, -6\}$.
The periodic signal corresponding to the spectral coefficients $Z[k]$ is

$$z(t) = x(t)y(t) = 20\cos(2\pi t) + 20\sin(4\pi t) + 8\sin(8\pi t) - 12\cos(10\pi t)$$

You can confirm this result by direct multiplication of $x(t)$ and $y(t)$.

(b) (Time-Domain Convolution)

Consider two rectangular pulse trains $x(t)$ and $y(t)$ with identical periods T and identical heights A, but different pulse widths t_1 and t_2. Their Fourier series coefficients are described by:

$$X[k] = \frac{At_1}{T}\,\text{sinc}(kf_0t_1)e^{-jk\pi f_0 t_1} \qquad\qquad Y[k] = \frac{At_2}{T}\,\text{sinc}(kf_0t_2)e^{-jk\pi f_0 t_2}$$

The Fourier coefficients of the signal $z(t) = x(t) \circledast y(t)$ are simply

$$Z[k] = X[k]Y[k] = \frac{A^2t_1t_2}{T^2}\,\text{sinc}(kf_0t_1)\text{sinc}(kf_0t_2)e^{-jk\pi f_0(t_1+t_2)}$$

PGR2026-00033
Krisp Technologies EX1024 Page 229

The signal $z(t)$ represents the periodic convolution of $x(t)$ and $y(t)$. We use the wraparound method of periodic convolution to study three cases, as illustrated in Figure E8.7B.

1. $A = 2,\ t_1 = t_2 = 1$, and $T = 2$: There is no wraparound, and $z(t)$ describes a triangular waveform. The Fourier coefficients simplify to

$$Z[k] = \tfrac{1}{4}\,\mathrm{sinc}^2(\tfrac{k}{2})e^{-jk\pi} = \tfrac{1}{4}\,\mathrm{sinc}^2(\tfrac{k}{2}), \qquad (k \text{ odd})$$

   This real result reveals the half-wave and even nature of $z(t)$.

2. $A = 1,\ t_1 = 2,\ t_2 = 1$, and $T = 4$: There is no wraparound, and $z(t)$ describes a trapezoidal waveform. The Fourier coefficients simplify to

$$Z[k] = \tfrac{1}{8}\,\mathrm{sinc}(\tfrac{k}{2})\mathrm{sinc}(\tfrac{k}{4})e^{-j3k\pi/4}$$

   We find $Z[0] = \frac{1}{8}$. Since $\mathrm{sinc}(k/2) = 0$ for even k, the signal $z(t)$ should show only hidden half-wave symmetry. It does.

3. $A = 1,\ t_1 = t_2 = 2$, and $T = 3$: We require wraparound past 3 units, and the Fourier coefficients of $z(t)$ simplify to

$$Z[k] = \tfrac{4}{9}\,\mathrm{sinc}^2(\tfrac{2}{3}k)e^{-j4k\pi/3}$$

   The signal has no symmetry, but since $\mathrm{sinc}(\frac{2}{3}k) = 0$ for $k = 3, 6, \ldots$, the third harmonic and its multiples are absent.

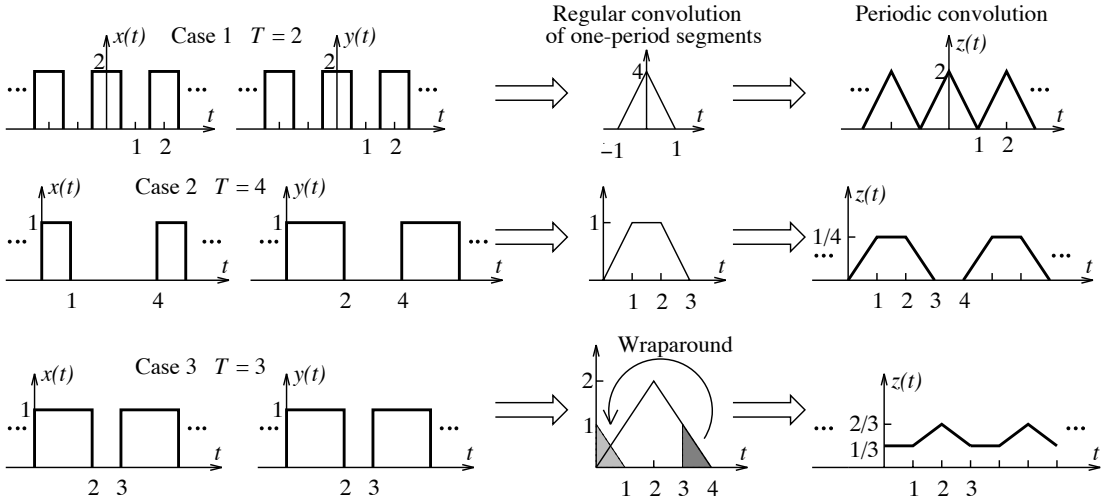


Figure E8.7B The signals and their convolution for Example 8.7(b)

8.6 Signal Reconstruction and the Gibbs Effect

The peaks in the harmonics used in signal reconstruction occur at different times and lead to oscillations (maxima and minima) or ripples in the reconstruction. The number of peaks in the ripples provides a visual indication of the number of nonzero harmonics N used in the reconstruction. An example is shown in Figure 8.6(a). Can you count the 20 peaks in the reconstruction of the sawtooth waveform by its first 20 nonzero harmonics?

PGR2026-00033
Krisp Technologies EX1024 Page 230

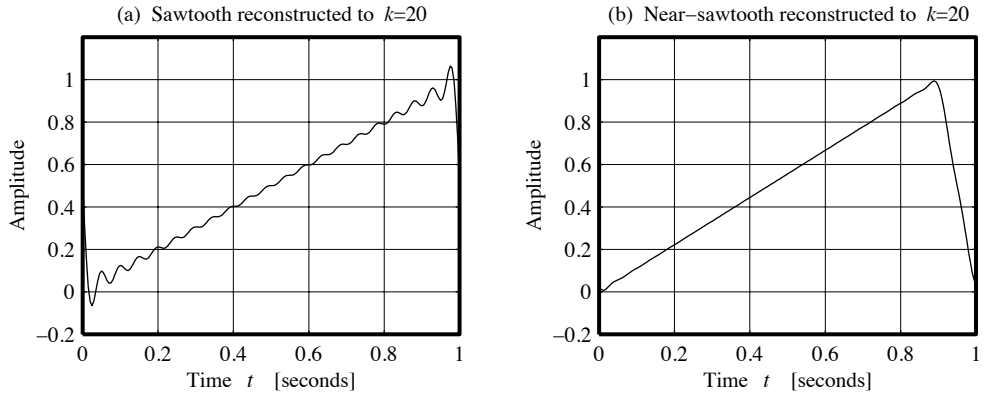


Figure 8.6 Reconstruction of a sawtooth and near-sawtooth by its harmonics

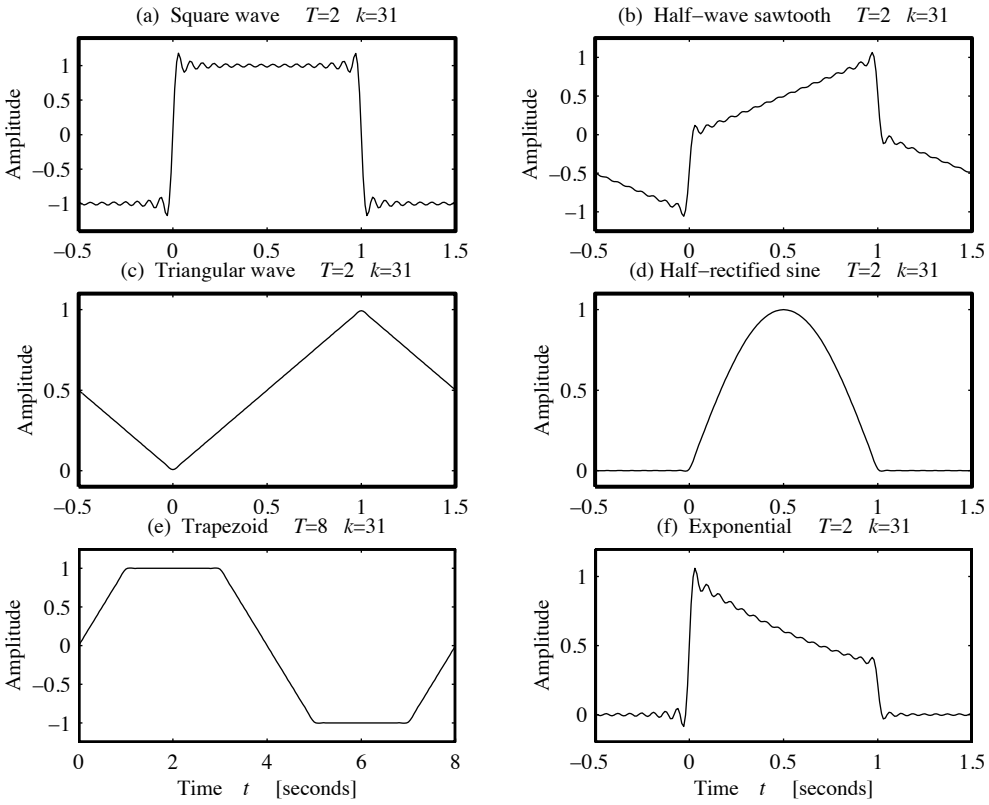


Figure 8.7 Reconstruction of various periodic signals by their harmonics

As N increases, we would intuitively expect the oscillations to die out and lead to perfect reconstruction as $N \to \infty$. This is not what always happens. The same intuition also tells us that we can never get perfect reconstruction for signals with jumps because we simply cannot approximate a discontinuous signal by continuous signals (sinusoids). These intuitive observations are indeed confirmed by Figure 8.7, which shows reconstructions of various signals, including those with jumps.

PGR2026-00033
Krisp Technologies EX1024 Page 231

8.6.1 The Convergence Rate

The Fourier series coefficients $X[k]$ of a periodic signal $x_p(t)$ usually decay with the harmonic index k as $1/k^n$. A larger n implies faster convergence, fewer terms in a truncated series for a given error, and a smoother signal. We can establish n without computing $X[k]$ by using the derivative property. Since the coefficients of an impulse train show no dependence on k, if $x_p(t)$ is differentiated n times before it first shows impulses, $X[k]$ must be proportional to $1/(jk\omega_0)^n$, and thus decay as $1/k^n$. A measure of smoothness may be ascribed to $x_p(t)$ in terms of the number of derivatives n it takes for the waveform to first result in impulses. In Figure 8.7, for example, the trapezoidal and half-rectified sine need two derivatives before yielding impulses, and show a convergence rate of $1/k^2$. Both are smoother than the sawtooth and square waves, which need only one derivative to produce impulses, and show a convergence rate of $1/k$. We stress that the convergence rate is a valid indicator only for large k.

REVIEW PANEL 8.15

The Convergence Rate: What It Means and How to Find It

The convergence rate indicates how rapidly the reconstruction approaches the original signal.
The convergence rate is $1/k^n$ if n derivatives are required for impulses to *first* appear.
A larger n (faster rate) means a smoother signal. Signals with jumps show poor convergence ($1/k$).

8.6.2 The Gibbs Effect

The convergence rate is intimately tied to reconstruction from a finite number of harmonics N. The faster the rate at which the coefficients decay, the less the impact the higher harmonics have on the reconstruction. For signals with jumps, the harmonics decay as $1/k$, and we observe some remarkable anomalies. The oscillations and overshoot persist near the discontinuities but become compressed (their frequency increases) as N is increased. The peak overshoot moves closer to the discontinuity. As $N \to \infty$, the overshoot narrows into spikes, but its magnitude remains constant at about 9% of the magnitude of the discontinuity. The persistence of overshoot and imperfect reconstruction for signals with jumps describes the **Gibbs phenomenon** or **Gibbs effect**. Even when $N = \infty$, the reconstruction contains spikes in the form of vertical lines that extend upward and downward by 9% of the discontinuity, as shown in Figure 8.8. The total power does, however, equal the power in the actual signal, as predicted by Parseval's relation, because the area of the spikes is zero and does not contribute to the power. At a discontinuity, the reconstruction (Fourier series) converges (sums) to the midpoint and is said to converge to the actual function in the *mean squared* sense.

The Gibbs effect is not just the appearance of overshoot for finite N. Rather, it is the lack of disappearance of overshoot as $N \to \infty$. If N is small, the overshoot may differ from 9% or be absent altogether. The Gibbs effect occurs only for waveforms with jump discontinuities whose convergence is $1/k$. For signals that converge as $1/k^2$ or faster, the Gibbs effect is absent. For such signals, we get better reconstruction with fewer terms and exact reconstruction with $N = \infty$.

REVIEW PANEL 8.16

The Gibbs Effect: Perfect Reconstruction Is Impossible for Signals with Jumps

Reconstructed signal shows overshoot and undershoot (about 9% of jump) near each jump location.
At each jump location, the reconstructed value equals the midpoint value of the jump.
The Gibbs effect is *present for signals with jumps* (whose $X[k]$ decay as $1/k$ or slower).
The Gibbs effect is *absent for signals with no jumps* (whose $X[k]$ decay as $1/k^2$ or faster).

PGR2026-00033
Krisp Technologies EX1024 Page 232

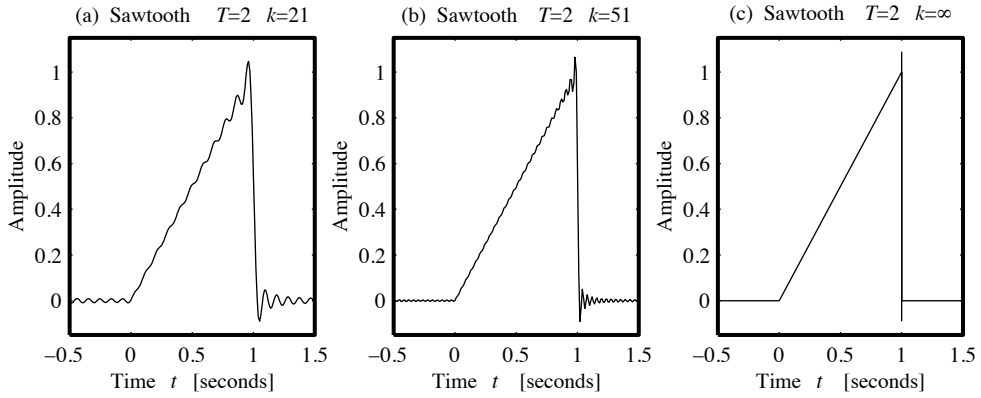


Figure 8.8 Illustrating the Gibbs effect for a sawtooth signal

EXAMPLE 8.8 (Convergence Rate and the Gibbs Effect)

(a) What is the convergence rate if $X[k] = A/k + jB/k^2$? For large k, B/k^2 is much smaller than A/k, and $X[k]$ thus decays as $1/k$ (and not $1/k^2$). For small k, of course, $X[k]$ will be affected by both terms.

(b) A true sawtooth pulse exhibits the Gibbs effect as shown in Figure 8.8. A triangular pulse with a very steep slope approximating a sawtooth does not, as shown in Figure 8.6(b), and leads to perfect reconstruction for large enough k.

8.6.3 Smoothing by Spectral Windows

Using a finite number of harmonics to reconstruct a periodic signal $x_p(t)$ is equivalent to multiplying its spectral coefficients $X[k]$ by a rectangular window $W_D[k] = \text{rect}(n/2N)$. The reconstructed signal $x_R(t)$ may be described by

$$x_R(t) = \sum_{k=-\infty}^{\infty} \text{rect}(\tfrac{n}{2N}) X[k] e^{j2k\pi f_0 t} = \sum_{k=-N}^{N} X[k] e^{j2k\pi f_0 t} \tag{8.34}$$

The signal $x_R(t)$ also equals the *periodic convolution* of $x_p(t)$ and $w_D(t)$, the signal corresponding to the rectangular spectral window given by

$$w_D(t) = \sum_{k=-N}^{N} \text{rect}(\tfrac{n}{2N}) e^{j2k\pi f_0 t} = \sum_{k=-N}^{N} e^{j2k\pi f_0 t} = (2N+1)\frac{\text{sinc}[(2N+1)f_0 t]}{\text{sinc}(f_0 t)} \tag{8.35}$$

The signal $w_D(t)$ is called the **Dirichlet kernel** and shows variations similar to a sinc function. Its periodic convolution with $x_p(t)$ is what leads to overshoot and oscillations in the reconstructed signal $x_R(t) = x_p(t) \circledast w_D(t)$ as illustrated in Figure 8.9. In other words, the Gibbs effect arises due to the *abrupt truncation* of the signal spectrum by the rectangular window.

PGR2026-00033
Krisp Technologies EX1024 Page 233

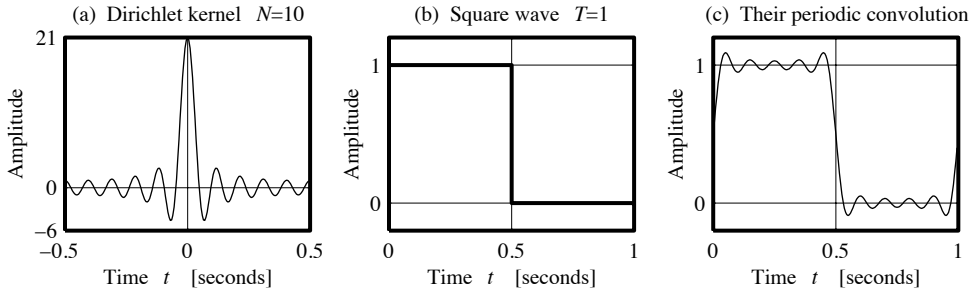


Figure 8.9 Periodic convolution of a signal with jumps and the Dirichlet kernel shows the Gibbs effect

The use of a *tapered* spectral window $W[k]$ (whose coefficients or weights typically decrease with increasing $|k|$) yields much better convergence and a much smoother reconstruction. The reconstruction $x_W(t)$ is given by

$$x_W(t) = x_p(t) \circledast w(t) = \sum_{k=-N}^{N} W[k]X[k]e^{j2k\pi f_0 t} \tag{8.36}$$

Table 8.1 lists some useful spectral windows. Note that all windows display even symmetry and are positive, with $W[0] = 1$; all windows (except the rectangular or boxcar window) are tapered; and for all windows (except the boxcar and Hamming), the end samples equal zero.

Table 8.1 Some Common Spectral Windows

Entry	Window Type	Expression for $W[k]$ Over $-N \le k \le N$
1	Boxcar	$\text{rect}\left(\frac{k}{2N}\right) = 1$
2	Bartlett	$\text{tri}\left(\frac{k}{N}\right) = 1 - \frac{\lvert k\rvert}{N}$
3	von Hann (Hanning)	$0.5 + 0.5\cos\left(\frac{k\pi}{N}\right)$
4	Hamming	$0.54 + 0.46\cos\left(\frac{k\pi}{N}\right)$
5	Cosine	$\cos\left(\frac{k\pi}{2N}\right)$
6	Riemann	$\text{sinc}\left(\frac{k}{N}\right)$
7	Blackman	$0.42 + 0.5\cos\left(\frac{k\pi}{N}\right) + 0.08\cos\left(\frac{2k\pi}{N}\right)$

Every window (except the rectangular) provides a reduction, or elimination, of overshoot. The smoothed reconstruction of a square wave using some of these windows is shown in Figure 8.10. Details of how the smoothing comes about are described later.

REVIEW PANEL 8.17

For Signals with Jumps, the Gibbs Effect Is Due to Abrupt Spectrum Truncation

To reduce/eliminate Gibbs effect in the reconstruction, we must use *tapered* spectral windows.

PGR2026-00033
Krisp Technologies EX1024 Page 234

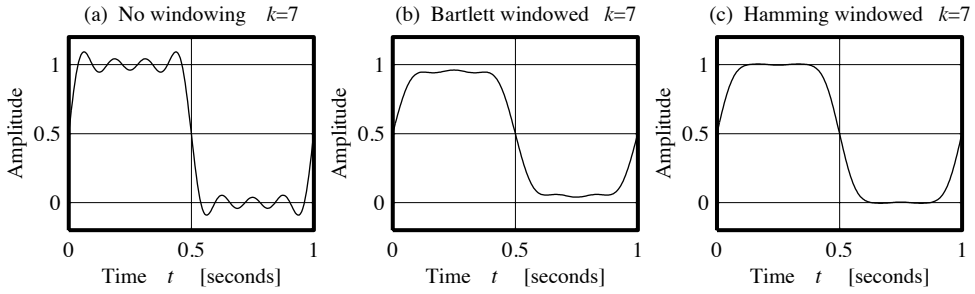


Figure 8.10 The use of tapered windows leads to a smoother reconstruction free of overshoot and oscillations

8.7 System Response to Periodic Inputs

The Fourier series allows us to evaluate the *steady-state response* of an LTI system, due to a periodic input $x_p(t)$. We have seen in Chapter 6 that the response of an LTI system to the harmonic $e^{jk\omega_0 t}$ is also a harmonic $H(k\omega_0)e^{jk\omega_0 t}$ at the input frequency. Its magnitude and phase are determined by the system function $H(\omega)$ evaluated at $\omega = k\omega_0$. This input-output relationship may be symbolically described by

$$e^{jk\omega_0 t} \quad \Longrightarrow \quad H(k\omega_0)e^{jk\omega_0 t} \tag{8.37}$$

With $H(k\omega_0) = H_k \angle \phi_k$, the output may be written

$$e^{jk\omega_0 t} \quad \Longrightarrow \quad H_k e^{j(k\omega_0 t + \phi_k)} \tag{8.38}$$

In other words, to obtain the output, we multiply the input magnitude by H_k and augment the input phase by ϕ_k. Euler's relation and superposition (or the fact that $\cos\alpha = \text{Re}[e^{j\alpha}]$) allows us to generalize this result for sinusoids:

$$c_k \cos(jk\omega_0 t + \theta_k) \quad \Longrightarrow \quad c_k H_k \cos(k\omega_0 t + \phi_k + \theta_k) \tag{8.39}$$

The response $y_p(t)$ of an LTI system to a periodic input $x_p(t)$ with period T and fundamental frequency $\omega_0 = 2\pi/T$ may be found as the superposition of the response due to each of the harmonics that make up the input (by evaluating $H(\omega)$ successively at each input frequency $k\omega_0$). Note that $y_p(t)$ is also periodic with period T since it contains the same harmonic frequencies as the input, but with different magnitudes and phase. For a half-wave symmetric input, the output is also half-wave symmetric because the odd-indexed harmonics are missing in both the input and the output. The convergence rate of $y_p(t)$, and other type of symmetry (or lack thereof), depends on both $x_p(t)$ and the system function. The closed form for the response using Fourier series analysis is very difficult to obtain. We could of course use convolution to find the exact time-domain result, but it lacks the power to untangle the frequency information for which the Fourier series is so well suited.

REVIEW PANEL 8.18

The Response to Periodic Inputs by Phasor Analysis and Superposition

Input: $x_k(t) = c_k \cos(2k\pi f_0 t + \theta_k)$ **Transfer function:** Evaluate $H(f)$ at $f = kf_0$ as $H_k \angle \phi_k$.
Output: $y_k(t) = c_k H_k \cos(k\omega_0 t + \theta_k + \phi_k)$; by superposition $x_p(t) = \sum x_k(t)$ and $y_p(t) = \sum y_k(t)$.

PGR2026-00033
Krisp Technologies EX1024 Page 235

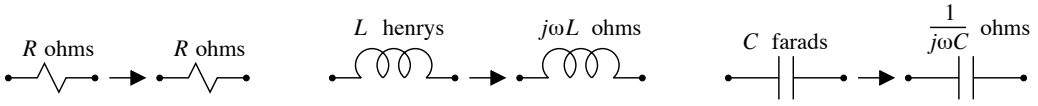
8.7.1 Response of Electric Circuits to Periodic Inputs

Here is how we find the response of an electric circuit to the sinusoid $c_k \cos(k\omega_0 t + \theta_k)$:

1. Replace sources and system variables by their phasor form and circuit elements by their impedance.
2. Set up equations based on Kirchhoff's laws for the transformed circuit to obtain $H(\omega)$.
3. Evaluate $H(\omega)$ (at the input frequency $k\omega_0$) to obtain $H(k\omega_0) = H_k \angle \phi_k$.
4. Compute the output as $c_k H_k \cos(k\omega_0 t + \theta_k + \phi_k)$.
5. For a periodic input (or combinations of sinusoids), find and superpose the *time-domain outputs*.

REVIEW PANEL 8.19

How Circuit Elements Are Replaced in a Transformed Circuit ($\omega = 2\pi f$)



EXAMPLE 8.9 (System Response to Periodic Inputs)

(a) The periodic sawtooth $v(t) = 4 - \sum_{k=1}^{\infty} \frac{10}{k} \sin(3kt)$ volts is applied to the RC circuit shown in Figure E8.9A, with $R = 1\ \Omega$ and $C = \frac{1}{3}F$. What is the output through the third harmonic?

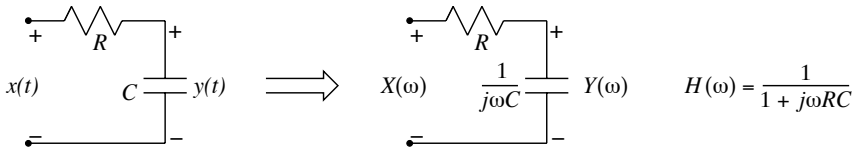


Figure E8.9A The circuit for Example 8.9(a)

The system function is given by

$$H(\omega) = \frac{V(\omega)}{V_{\text{in}}(\omega)} = \frac{1}{1 + j\omega/3}$$

Now, $H(0) = 1$. So, the dc response equals $c_0 H(0) = 4$ volts.

The system function at the kth harmonic ($\omega = k\omega_0$) is $H(k\omega_0) = \dfrac{1}{1 + jk\omega_0/3} = \dfrac{1}{1 + jk}$.

With $c_k = 10/k$ and $\theta_k = 90°$, the response to the kth harmonic is

$$V_k = H(k\omega_0) c_k \angle 90° = \frac{j10}{k(1 + jk)}$$

For $k = 1,\ k = 2$, and $k = 3$, we successively compute

$$V_1 = 5\sqrt{2}\angle 45° = 7.0711\angle 45° \qquad V_2 = \sqrt{5}\angle 26.6° = 2.2361\angle 26.6° \qquad V_3 = \tfrac{1}{3}\sqrt{10}\angle 18.4° = 1.0541\angle 18.4°$$

The time-domain response through the third harmonic may then be written as

$$v_R(t) = 4 + 7.0711\cos(3t + 45°) + 2.2361\cos(6t + 26.6°) + 1.0541\cos(9t + 18.4°)$$

The response $V_k = \dfrac{j10}{k(1 + jk)}$ implies a convergence rate of $1/k^2$.

Comment: Unlike the input, the output cannot contain jumps or exhibit the Gibbs effect.

PGR2026-00033
Krisp Technologies EX1024 Page 236

(b) Consider the system shown in Figure E8.9B(1). Sketch $G[k]$ and $g(t)$ and find the output $y(t)$.

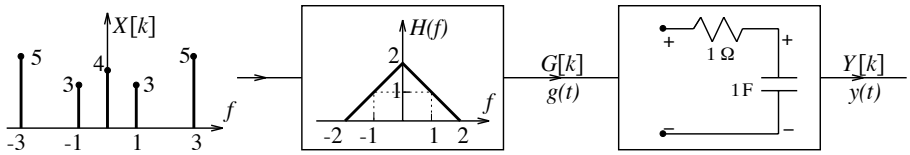


Figure E8.9B(1) The system for Example 8.9(b)

The output of the filter produces the spectrum shown in Figure E8.9B(2).
The signal $g(t)$ is $g(t) = 12 + 6\cos(2\pi t)$.

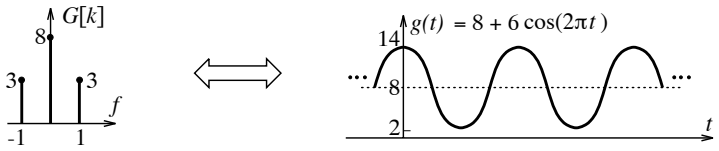


Figure E8.9B(2) The filter output of the system for Example 8.9(b)

The transfer function of the RC filter is $T(f) = \dfrac{1}{1 + j2\pi f}$.

Since the input comprises components only at dc ($f = 0$) and $f = 1$ Hz, we evaluate $T(0) = 1$ and $T(1) = \frac{1}{1+j2\pi} = 0.1572\angle -80.96°$. Then the output $y(t)$ is given by

$$y(t) = 8 + (6)(0.1572)\cos(2\pi t - 80.96°) = 8 + 0.9431\cos(2\pi t - 80.96°)$$

(c) Consider the system shown in Figure E8.9C(1). Sketch $g(t)$ and find the output spectrum $Y[k]$.

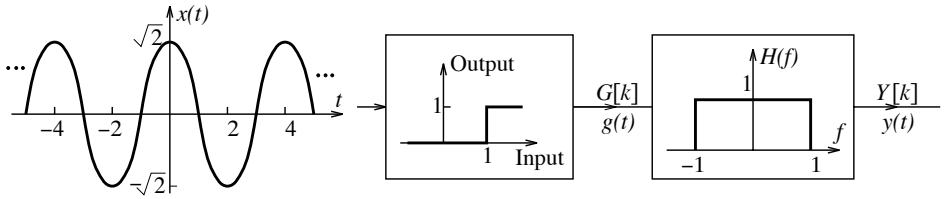


Figure E8.9C(1) The system for Example 8.9(c)

The output of the nonlinear system is unity whenever the input exceeds or equals unity. Now, since $x(t) = \sqrt{2}\cos(0.5\pi t) = 1$ at $t = \pm 0.5$, the output $g(t)$ describes a rectangular pulse train, as shown in Figure E8.9C(2).

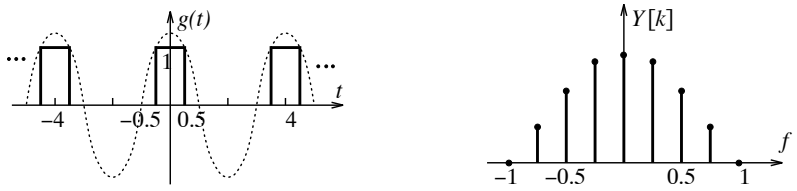


Figure E8.9C(2) The output $g(t)$ of the system for Example 8.9(c)

The Fourier series coefficients of $g(t)$ are $G[k] = \text{sinc}(k/4)$. Since $f_0 = 1/T = 0.25$ Hz, the spectrum $Y[k]$ of the ideal filter contains harmonics only up to $|k| = 4$ (or 1 Hz) as shown.

PGR2026-00033
Krisp Technologies EX1024 Page 237

8.8 Application-Oriented Examples

Applications of the Fourier series typically depend on a frequency-domain description of periodic signals and steady-state system analysis. We now describe several such applications.

8.8.1 DC Power Supplies

Conceptually a dc power supply involves feeding a periodic signal to a (full-wave or half-wave) rectifier and lowpass filter. The rectifier ensures that its output contains a nonzero dc component (even if its input may not). Ideally, the filter should pass only this dc component in the rectified signal. In practice, the dc output is contaminated by a small amount of **ripple** due to the presence of harmonics. The ripple is defined as the ratio of the magnitude of the largest nonzero harmonic output to the dc output. Since harmonic magnitudes often decay with frequency, the largest harmonic is typically the first nonzero harmonic. The design of the lowpass filter is based on how much ripple is allowed.

EXAMPLE 8.10 (Design of a DC Power Supply)
We design a dc power supply with a ripple R of less than 1%, using the input $A\sin(\omega_0 t)$ to a half-wave rectifier, followed by an RC lowpass filter with $H(\omega) = \dfrac{1}{1+j\omega\tau}$, as shown in Figure E8.10.

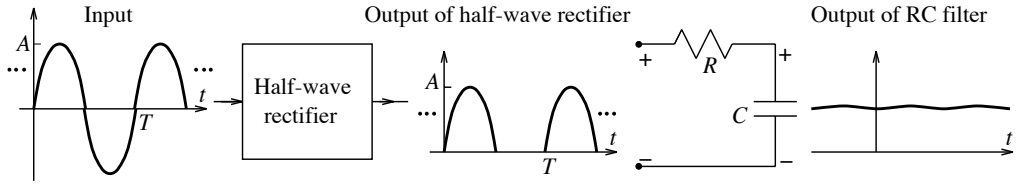


Figure E8.10 The dc power supply for Example 8.10

The Fourier series coefficients of the rectifier output are

$$c_0 = \frac{A}{\pi} \qquad c_1 = 0.5A \qquad c_k = \frac{2A}{\pi(1-k^2)} \quad (k \text{ even})$$

The largest harmonic is c_1 at the fundamental frequency ω_0. The magnitude of the filter output (the capacitor voltage) due to the component at ω_0 equals

$$|V_1| = |c_1 H(\omega_0)| = \left|\frac{c_1}{1+j\omega_0\tau}\right| \approx \left|\frac{c_1}{\omega_0\tau}\right|, \quad \omega_0\tau \gg 1$$

The dc output is simply $V_0 = c_0$. For $R < 1\%$, we require $|V_0/V_1| > 100$ and thus

$$\left|\frac{c_0\omega_0\tau}{c_1}\right| > 100 \qquad \text{or} \qquad \tau > \left|\frac{100c_1}{c_0\omega_0}\right| = \frac{25}{f_0}$$

The time constant τ can now be used to select values for R and C (if f_0 is specified).

8.8.2 Harmonic Distortion

Ideally, an amplifier with a gain A should amplify a pure cosine wave $\cos(\omega_0 t)$ to $A\cos(\omega_0 t)$. When the output is contaminated by spurious harmonics during amplification, there is **harmonic distortion**. This

PGR2026-00033
Krisp Technologies EX1024 Page 238

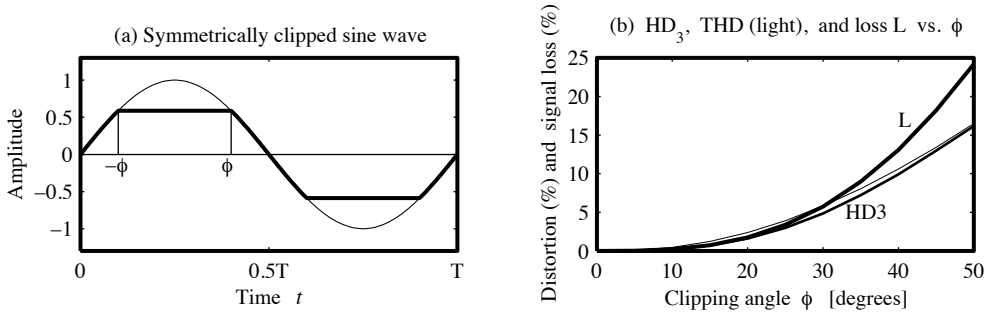


Figure 8.11 Symmetric clipping of a sine wave leads to harmonic distortion

can happen, for example, if the amplifier reaches saturation and clips the input sinusoid to levels below its peak value, as shown in Figure 8.11(a).

The **clipping angle** ϕ serves as a measure of the distortion. The larger its value, the more severe is the distortion. The Fourier series of the clipped signal contains harmonics other than the desired fundamental, and the magnitude of the fundamental is reduced from the ideal case. This reduction results in what is often called **signal loss**. It is the difference between the gain A of an ideal amplifier and the (amplified) level of the fundamental component c_1 at the output, expressed as a ratio

$$L = \left| \frac{A - c_1 A}{A} \right| = |1 - c_1| \tag{8.40}$$

Total harmonic distortion (THD) is a measure of the power contained in the unwanted harmonics as compared to the desired output. It is defined as the square root of the ratio of the power P_u in the unwanted harmonics and the power P_1 in the desired component (the fundamental). Since the unwanted power P_u is simply the difference between the total ac power P_{AC} and the power P_1 in the fundamental, we have

$$\text{THD} = \left(\frac{P_u}{P_1} \right)^{1/2} = \left(\frac{P_{\text{AC}} - P_1)}{P_1} \right)^{1/2} \qquad \text{(total harmonic distortion)} \tag{8.41}$$

Note that P_u describes only the power in the harmonics and excludes the dc component (even if present). The square root operation is used because it has been customary to deal with rms values instead of power. We can also find the distortion HD_k due only to the kth harmonic as

$$\text{HD}_k = \left(\frac{P_k}{P_1} \right)^{1/2} = \left| \frac{c_k}{c_1} \right| \qquad (k\text{th harmonic distortion}) \tag{8.42}$$

EXAMPLE 8.11 (Harmonic Distortion)

(a) The input to an amplifier is $x(t) = \cos(10\pi t)$. The gain of the amplifier is 10. Its response is $y(t) = 8\cos(10\pi t) - 0.08\cos(30\pi t) + 0.06\cos(50\pi t)$. What is the harmonic distortion and signal loss?

The third harmonic distortion (HD_3) equals $|\frac{c_3}{c_1}| = \frac{0.08}{8} = 1\%$.

The total harmonic distortion (THD) equals $\sqrt{\frac{P_U}{P_1}} = \left[\frac{(0.06)^2 + (0.08)^2}{64} \right]^{1/2} = 1.25\%$.

The signal loss L equals $|\frac{10-8}{10}| = 20\%$.

PGR2026-00033
Krisp Technologies EX1024 Page 239

(b) The Fourier series coefficients of the symmetrically clipped sine of Figure 8.11 can be found to be

$$c_1 = \frac{1}{\pi}\left[\sin 2\phi + \pi - 2\phi\right] \qquad c_k = \frac{2}{k\pi}\left[\frac{\sin(k+1)\phi}{k+1} - \frac{\sin(k-1)\phi}{k-1}\right] \quad (k \text{ odd})$$

What is the harmonic distortion and signal loss?

No second harmonic distortion occurs, since all even harmonics are absent.
The third harmonic distortion HD_3 and the signal loss L are given by

$$\text{HD}_3 = \left|\frac{c_3}{c_1}\right| = \left|\frac{\frac{1}{6}\sin 4\phi - \frac{1}{3}\sin 2\phi}{\sin 2\phi + \pi - 2\phi}\right| \qquad L = |1 - c_1| = \frac{2\phi - \sin 2\phi}{\pi}$$

With $\phi = \frac{\pi}{8}$, for example, we obtain $\text{HD}_3 \approx 2.25\%$ and $L \approx 2.49\%$.
Figure 8.11(b) shows how HD_3 and L vary nonlinearly with the clipping angle ϕ.

8.9 The Dirichlet Kernel and the Gibbs Effect

Reconstruction of a periodic signal by adding a finite number of harmonics is often called a **partial sum**. Truncation to N harmonics is equivalent to multiplying the Fourier series coefficients $X[k]$ pointwise by a rectangular **spectral window** or frequency-domain window $W_D[k] = \text{rect}(k/2N)$, which equals 1 for $-N \le k \le N$ and is zero elsewhere. The weighted coefficients $X_N[k]$ then equal

$$X_N[k] = X[k]W_D[k] = X[k]\text{rect}\left(\tfrac{k}{2N}\right) \tag{8.43}$$

The signal $w_D(t)$ corresponding to $W_D[k]$ is called the **Dirichlet kernel**. It represents the sum of harmonics of unit magnitude at multiples of f_0, and can be written as the summation

$$w_D(t) = \sum_{k=-N}^{N} e^{-j2\pi k f_0 t} \tag{8.44}$$

From tables (at the end of this book), the closed form for $w_D(t)$ is

$$w_D(t) = \frac{\sin[(N+\frac{1}{2})\pi f_0 t]}{\sin(\pi f_0 t)} = (2N+1)\frac{\text{sinc}[(2N+1)f_0 t]}{\text{sinc}(f_0 t)} = \frac{M\,\text{sinc}(M f_0 t)}{\text{sinc}(f_0 t)} \tag{8.45}$$

where $M = 2N + 1$. This kernel is periodic with period $T = 1/f_0$ and has some very interesting properties, as Figure 8.12 illustrates. Over one period,

1. Its area equals T, it attains a maximum peak value of M at $t = 0$, and its value at $t = 0.5T$ is -1.
2. It shows N maxima, a positive mainlobe of width $2T/M$, and decaying positive and negative sidelobes of width T/M, with $2N$ zeros at $kT/M,\ k = 1, 2, \ldots, 2N$.
3. The ratio R of the mainlobe height and the peak sidelobe magnitude stays nearly constant (between 4 and 4.7) for finite M, and $R \to 1.5\pi \approx 4.71$ (or 13.5 dB) for very large M.
4. Increasing M increases the mainlobe height and compresses the sidelobes. As $M \to \infty$, $w_D(t)$ approaches an impulse with strength T.

PGR2026-00033
Krisp Technologies EX1024 Page 240

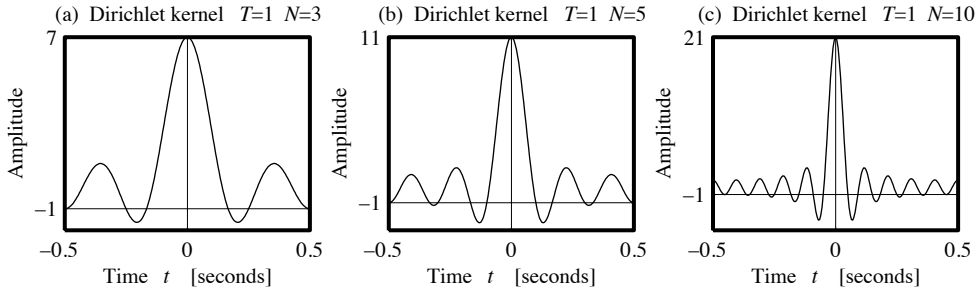


Figure 8.12 The Dirichlet kernel

Since $X_N[k] = X[k]W_D[k]$, the reconstructed signal $x_N(t)$ equals the *periodic convolution* $w_D(t) \circledast x_p(t)$:

$$x_N(t) = w_D(t) \circledast x_p(t) = \frac{1}{T} \int_T w_D(\tau) x_p(t - \tau)\, d\tau \tag{8.46}$$

If $x_p(t)$ contains discontinuities, $x_N(t)$ exhibits an overshoot and oscillations near each discontinuity, as shown in Figure 8.13.

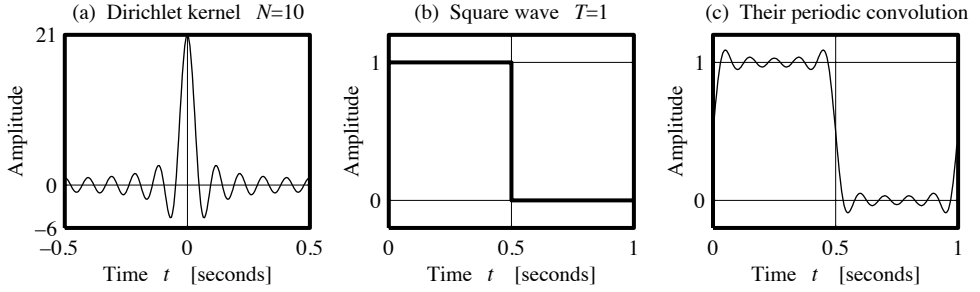


Figure 8.13 The Dirichlet kernel leads to overshoot when reconstructing signals with jumps

With increasing N, the mainlobe and sidelobes become taller and narrower, and the overshoot (and associated undershoot) near the discontinuity becomes narrower but more or less maintains its height.

It can be shown that if a periodic function $x_p(t)$ jumps by J_k at times t_k, the Fourier reconstruction $x_p(t)$, as $N \to \infty$, yields not only the function $x_p(t)$ but also a pair of straight lines at each t_k, which extend equally above and below the discontinuity by an amount A_k, given by

$$A_k = \left[\frac{1}{2} - \frac{1}{\pi}\mathrm{si}\,(\pi)\right] J_k = 0.0895 J_k \tag{8.47}$$

where $\mathrm{si}(x)$ is the **sine integral** defined as

$$\mathrm{si}(x) = \int_0^x \frac{\sin(\xi)}{\xi}\, d\xi \tag{8.48}$$

The lines thus extend by about 9% on either side of the jump J_k, as shown in Figure 8.14. This fraction is independent of both the nature of the periodic signal and the number of terms used, as long as N is large. We are of course describing the Gibbs effect.

PGR2026-00033
Krisp Technologies EX1024 Page 241

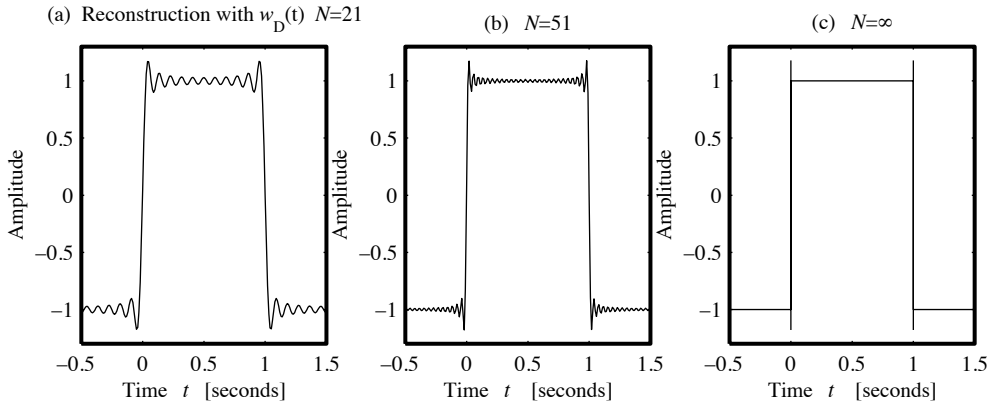


Figure 8.14 The Gibbs effect occurs when reconstructing signals with jumps

8.9.1 Reconstruction and Partial Sums

The reconstruction of a periodic signal with discontinuities using partial sums is what leads to overshoot and the Gibbs effect. Certain operations on the partial sum result in a reduction or elimination of overshoot because their effect is equivalent to the use of tapered windows for smoothing the reconstructed signal. The operations on the partial sum, which we now discuss, are summarized in Table 8.2.

Table 8.2 Smoothing by Operations on the Partial Sum

Operation	**Spectral Window** $W[k]$	**Convolution Kernel** **Note:** $M = 2N + 1$	**Reconstruction at Jumps**
Partial sum	$\text{rect}(k/2N)$	$\dfrac{M\text{sinc}(Mf_0\tau)}{\text{sinc}(f_0\tau)}$	9% overshoot (Gibbs effect)
Arithmetic mean of partial sum	$\text{tri}(k/N)$	$N\left[\dfrac{\text{sinc}(Nf_0t)}{\text{sinc}(f_0t)}\right]^2$	No overshoot Less steep slope
Integration of partial sum	$\text{sinc}(k/N)$	———	1.2% overshoot Steeper slope

The Partial Sum

For a series represented by $x = r_0 + r_1 + r_2 + r_3 + \cdots + r_N + r_{N-1} + \cdots$, the partial sum s_N to N terms is defined by

$$s_N = r_0 + r_1 + r_2 + r_3 + \cdots + r_{N-1} \tag{8.49}$$

PGR2026-00033
Krisp Technologies EX1024 Page 242

8.9.2 Arithmetic Mean of Partial Sums

The **sum by arithmetic means** to N terms is defined as the average (arithmetic mean) of the intermediate partial sums:

$$\hat{x}_N = \frac{1}{N}(s_0 + s_1 + \cdots + s_{N-1}) = \frac{1}{N}\sum_{k=0}^{N-1} s_k \tag{8.50}$$

This result is valid for any series, convergent or otherwise. It is particularly useful for finding the time signal from its Fourier series even when such a series diverges.

There are two important theorems relating to such a sum. First, if the actual series converges to x, the sum by the method of arithmetic means also converges to the same value. Second, the Fourier series of an absolutely integrable signal $x(t)$ is summable by arithmetic means, and the sum converges uniformly to $x(t)$ at every point of continuity and to the midpoint of discontinuities otherwise.

Writing out the partial sums, we have

$$s_1 = r_0 \qquad s_2 = r_0 + r_1 \qquad s_3 = r_0 + r_1 + r_2 \quad \ldots \quad s_N = r_0 + r_1 + r_2 + \cdots + r_{N-1} \tag{8.51}$$

The sum of these partial sums may then be written

$$\hat{s}_N = s_1 + \cdots + s_N = Nr_0 + (N-1)r_1 + (N-2)r_2 + \cdots + r_{N-1} \tag{8.52}$$

The arithmetic mean of the partial sums then equals

$$\hat{x}_N = \frac{1}{N}[Nr_0 + (N-1)r_1 + (N-2)r_2 + \cdots + r_{N-1}] = \sum_{k=0}^{N-1} \frac{N-k}{N} r_k \tag{8.53}$$

This clearly reveals a *triangular* weighting on the individual terms r_k. For a periodic signal $x_p(t)$ with $r_k = X[k]e^{j2\pi k f_0 t}$, the arithmetic mean $x_N(t)$ of the partial sums may be written, by analogy, as

$$x_N(t) = \sum_{k=-(N-1)}^{N-1} \frac{N-|k|}{N} X[k]e^{j2\pi k f_0 t} = \sum_{k=-N}^{N} W_F[k]X[k]e^{j2\pi k f_0 t} \tag{8.54}$$

In this result, $W_F[k]$ describes a tapered triangular window, called the Bartlett window, whose weights decrease linearly with $|k|$. With $W_F[\pm N] = 0$, we may write

$$W_F[k] = \mathrm{tri}\left(\frac{k}{N}\right) = 1 - \frac{|k|}{N}, \qquad -N \le k \le N \tag{8.55}$$

The reconstructed signal $x_N(t)$ is

$$x_N(t) = \frac{1}{T}\int_{-T/2}^{T/2} w_F(\tau)x_p(t-\tau)\,d\tau \tag{8.56}$$

where $w_F(t)$ can be shown to equal

$$w_F(t) = \frac{1}{N}\left[\frac{\sin[(\frac{1}{2}N\omega_0 t)}{\sin(\frac{1}{2}\omega_0 t)}\right]^2 = N\left[\frac{\mathrm{sinc}(Nf_0 t)}{\mathrm{sinc}(f_0 t)}\right]^2 \tag{8.57}$$

This is the so-called **Fejer kernel**. It has a peak value of N and describes a periodic signal corresponding to the triangular (Bartlett) spectral window $W_F[k]$. It is entirely positive, and, for a given N, its sidelobes are much smaller than those of a Dirichlet kernel (of the rectangular window). These properties result in a reconstruction that is not only smoother but also *completely free of overshoot* for periodic signals with discontinuities. An example is shown in Figure 8.15.

Even though partial sums lead to a poor reconstruction with overshoot for signals with jumps, the arithmetic means of these partial sums provide a smoother reconstruction without overshoot.

PGR2026-00033
Krisp Technologies EX1024 Page 243

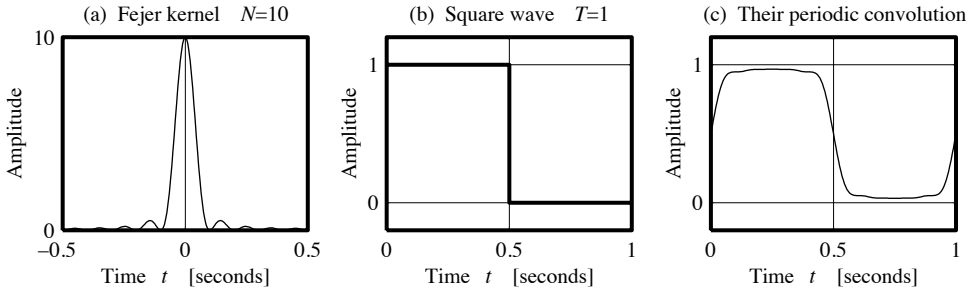


Figure 8.15 The Fejer kernel leads to a smooth reconstruction of signals with jumps

8.9.3 Integration of Partial Sums

The reconstruction based on integrating rather than averaging the partial sums is described by

$$x_N(t) = \frac{N}{T} \sum_{k=-N}^{N} \int_{t-\frac{T}{2N}}^{t+\frac{T}{2N}} X[k] e^{j2\pi k f_0 t} dt \tag{8.58}$$

Carrying out the integration, and noting that $\omega_0 T = 2\pi$, we get

$$x_N(t) = \frac{N}{T} \sum_{k=-N}^{N} X[k] e^{j2\pi k f_0 t} \left[\frac{e^{jk\pi/N} - e^{-jk\pi/N}}{j2\pi k f_0} \right] \tag{8.59}$$

A final simplification, using Euler's relation, yields

$$x_N(t) = N \sum_{k=-N}^{N} X[k] e^{j2\pi k f_0 t} \left[\frac{\sin(k\pi/N)}{k\pi} \right] = \sum_{k=-N}^{N} X[k] \operatorname{sinc}\left(\tfrac{k}{N}\right) e^{j2\pi k f_0 t} \tag{8.60}$$

The quantity $W_L[k] = \operatorname{sinc}(k/N)$ describes the **Lanczos window**, which has a sinc taper. Its kernel, $w_L(t)$, reduces the reconstruction overshoot to less than 1.2% (as compared with 9% for the Dirichlet kernel) but does not eliminate it (as the Fejer kernel does). The reconstruction, however, shows a much steeper slope at the discontinuities as compared with the Fejer kernel.

8.10 The Fourier Series, Orthogonality, and Least Squares

The geometric interpretation of orthogonality for two or three dimensions is obvious. The vectors are mutually perpendicular. The mathematical interpretation is based on the so-called **dot product** or **inner product** for two vectors $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$, denoted $\vec{\mathbf{X}}.\vec{\mathbf{Y}}$ or $< \vec{\mathbf{X}} \mid \vec{\mathbf{Y}} >$, and defined by

$$\vec{\mathbf{X}}.\vec{\mathbf{Y}} = < \vec{\mathbf{X}} \mid \vec{\mathbf{Y}} > = x_1 y_1 + x_2 y_2 + \cdots + x_N y_N = \sum_{i=1}^{N} x_i y_i = \|X\| \|Y\| \cos(\theta_{xy}) \tag{8.61}$$

Here θ_{xy} is the angle subtended by the two vectors, and $\|X\|$ and $\|Y\|$ represent the **length** of the vectors $\vec{\mathbf{X}}$ and $\vec{\mathbf{Y}}$, respectively. The inner product of two vectors equals zero if the angle they subtend is 90°, whereas the inner product of a vector with itself simply equals the square of its length.

PGR2026-00033
Krisp Technologies EX1024 Page 244

8.10.1 The Projection Theorem

The best approximation $\vec{\mathbf{X}}$ to a vector $\mathbf{X}$ is found by minimizing the square of the difference $\|\mathbf{X} - \vec{\mathbf{X}}\|^2$, which is of course related to the difference in the length of the two vectors. A fundamental idea in least squares estimation arises from an intuitive geometric result known as the **projection theorem**. In essence, the projection theorem is a generalization of the fact that the shortest distance from a point to a line (or plane) is the perpendicular dropped from the point to that line (or plane), as illustrated in Figure 8.16. If the point represents the tip of a vector $\vec{\mathbf{X}}$ in the x_1x_2 Cartesian plane and the line is directed along the x_1 axis, the best estimate of $\vec{\mathbf{X}}$ along x_1 is its **projection** $\widehat{\mathbf{X}}$ onto the x_1 axis since this clearly minimizes the error between $\vec{\mathbf{X}}$ and its estimate $\widehat{\mathbf{X}}$. Put another way, the best estimate requires the error vector to be orthogonal to x_1. Similarly, the best two-dimensional estimate in the x_1x_2 plane of a three-dimensional vector $\vec{\mathbf{X}}$ is its projection $\widehat{\mathbf{X}}$ onto the x_1x_2 plane. This is equivalent to making the error vector orthogonal to both x_1 and x_2. This result can be generalized to higher dimensions and implies that *the error vector must be made orthogonal to every basis vector* to achieve the smallest possible error.

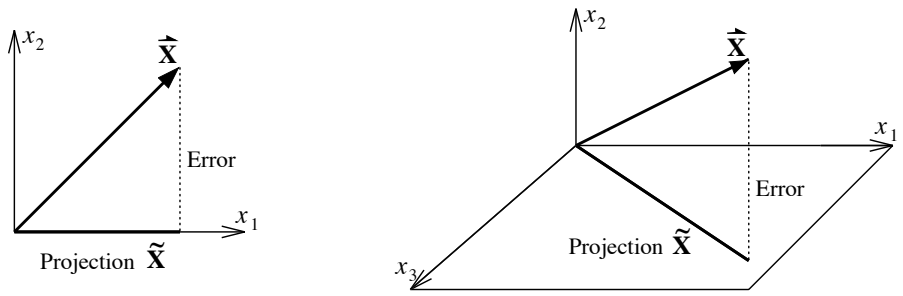


Figure 8.16 Illustrating the projection theorem

To extend this idea to function spaces, let $x(t)$ be a periodic signal with period T approximated by $\hat{x}(t)$ in terms of the orthogonal set of basis functions $\phi_k(t) = e^{j2\pi k f_0 t}$ as

$$\hat{x}(t) = \sum_{k=-N}^{N} \alpha_k \phi_k(t) \tag{8.62}$$

In analogy with vectors, this relation suggests that the coefficients α_k describe the projections of $x(t)$ on the basis functions $\phi_k(t)$. For a least squares solution, the α_k must therefore be chosen such that the error $x(t) - \hat{x}(t)$ is orthogonal to every basis function $\phi_k(t)$. Since the $\phi_k(t)$ are complex, we require

$$\int_T [x(t) - \hat{x}(t)]\phi_k^*(t)\, dt = 0, \qquad k = 0, 1, \ldots, N \tag{8.63}$$

Substitution for $\hat{x}(t)$ and separation of terms results in

$$\int_T [x(t) - \hat{x}(t)]\phi_k^*(t)\, dt = \int_T x(t)\phi_k^*(t)\, dt - \int_T \left[\sum_m \alpha_m \phi_m(t)\right] \phi_k^*(t)\, dt = 0 \tag{8.64}$$

Due to orthogonality, the products $\phi_m\phi_k^*,\ m \neq k$, integrate to zero to give

$$\int_T x(t)\phi_k^*(t)\, dt - \alpha_k \int_T \phi_k(t)\phi_k^*(t)\, dt = 0 \tag{8.65}$$

PGR2026-00033
Krisp Technologies EX1024 Page 245

The second integral equals T and represents the energy E_k in $\phi_k(t)$, and we have

$$\alpha_k = \frac{1}{T}\int_T x(t)\phi_k^*(t)\,dt = \frac{1}{T}\int_T x(t)e^{-j2\pi k f_0 t}\,dt \tag{8.66}$$

The α_k correspond exactly to the Fourier series coefficients $X[k]$ and yield a least squares estimate for $x(t)$. Of course, the least squares method may also be implemented at a brute force level if we minimize the mean squared error by setting its derivative (with respect to α_k) to zero. This approach yields the same results.

REVIEW PANEL 8.20

The Fourier Series Is the Best Least Squares Fit to a Periodic Signal $x(t)$

It describes $x(t)$ as a sum of sinusoids at its fundamental frequency $f_0 = 1/T$ and multiples kf_0 whose weights (magnitude and phase) are selected to minimize the *mean square error*.

8.11 Existence, Convergence, and Uniqueness

To formalize the notion of existence, convergence, and uniqueness of the Fourier series, we must address and answer several questions. Under what conditions will the Fourier series for a periodic signal $x(t)$ exist? If it does, will the series converge? If it does converge, will it converge for every t? If it does converge for every t, will it converge to $x(t)$ for every t? These questions deal with the existence, convergence, and uniqueness of Fourier series.

8.11.1 Existence

The idea of existence may be justified by defining the coefficients of the series by the proposed Fourier relations and then showing that the series with these coefficients does indeed represent $x(t)$. We start from the opposite end and assume the coefficients can be found using the prescribed relations. Otherwise, the existence of Fourier series is very difficult to prove in a formal, rigorous sense. This approach requires that the Fourier series coefficients $|X[k]|$ be finite. Thus,

$$|X[k]| \le \int_T |x(t)\|e^{j2\pi k f_0 t}|\,dt < \infty \tag{8.67}$$

Since $\left|e^{j2\pi k f_0 t}\right| = 1$, we have

$$\int_T |x(t)|\,dt < \infty \tag{8.68}$$

Thus, $x(t)$ must be absolutely integrable over one period. A consequence of absolute integrability is the **Riemann-Lebesgue theorem**, which states that the coefficients $X[k]$ approach zero as $k \to \infty$:

$$\lim_{k\to\infty}\int_T x(t)e^{j2\pi k f_0 t}dt = 0 \tag{8.69}$$

8.11.2 Convergence

This result also leads us toward the idea of convergence of the series with the chosen coefficients. Convergence to the actual function $x(t)$ for every value of t is known as **uniform convergence**. This requirement is satisfied for every continuous function $x(t)$ since we are trying to represent a function in terms of sines and cosines, which are themselves continuous. We face problems when reconstructing functions with jumps. If

PGR2026-00033
Krisp Technologies EX1024 Page 246

we require convergence to the midpoint of the jump at the discontinuity, we sidetrack such problems. In fact, this is exactly the condition that obtains for functions with jump discontinuities.

Even though our requirement calls for $x(t)$ to be absolutely integrable over one period, it includes square integrable (energy) signals for which

$$\int_T |x(t)|^2 \, dt < \infty \tag{8.70}$$

Obviously, every continuous or piecewise continuous signal, over one period, is square integrable. For such a signal, it turns out that the Fourier series also converges to $x(t)$ *in the mean* (or *in the mean squared sense*) in that, as more terms are added to the series $x_N(t)$ truncated to N terms, the mean squared error decreases and approaches zero as $N \to \infty$:

$$\int_T |x(t) - x_N(t)|^2 \, dt \to 0 \qquad \text{as} \qquad N \to \infty \tag{8.71}$$

A consequence of this result is that if $x_N(t)$ converges in the mean to $x(t)$, it cannot converge to any other function. In other words, $x(t)$ is *uniquely* represented by $x_N(t), N \to \infty$, even though it may not converge pointwise to $x(t)$ at every value of t.

8.11.3 Uniqueness

Uniqueness means that if two periodic signals $x(t)$ and $y(t)$ possess the same Fourier series, then they are equal. Here equality implies that the two functions could be different at a finite set of points (such as points of discontinuity) over one period. For signals that are square integrable over one period, convergence in the mean implies uniqueness. An important result in this regard is the converse of Parseval's relation, known as the **Riesz-Fischer theorem**, which, loosely stated, tells us that a series with finite total power ($P < \infty$) given in terms of the summation

$$P = \sum_{k=-\infty}^{\infty} |X[k]|^2 < \infty \tag{8.72}$$

must represent the Fourier series of a square integrable function.

8.11.4 The Dirichlet Conditions

The **Dirichlet conditions** describe *sufficient* conditions for the existence of Fourier series. A periodic function $x(t)$ will possess a unique Fourier series if, over one period,

1. $x(t)$ is absolutely integrable.
2. $x(t)$ has a finite number of maxima and minima and a finite number of finite discontinuities.

If these conditions are satisfied, the Fourier series will converge to $x(t)$ pointwise for every value of t where $x(t)$ is continuous and piecewise differentiable (uniform convergence) or to the midpoint of the discontinuity at each jump discontinuity otherwise (nonuniform convergence). If, in addition, $x(t)$ is square integrable over one period, the series will converge to $x(t)$ in a mean squared sense. The mean squared error will approach zero as the number of terms in the series approaches infinity.

We stress that the Dirichlet conditions are *sufficient* conditions that guarantee a Fourier series if satisfied. They do not preclude the existence of Fourier series for signals that violate some or all of these conditions, however. The periodic impulse train is such an example. Other classical examples of functions violating the Dirichlet conditions include

- $x(t) = \sin(2\pi/t),\ 0 < t \leq 1$, and $x(t) = \cos[\tan(\pi t)],\ 0 < t \leq 1$, each of which possesses an infinite number of maxima and minima.

PGR2026-00033
Krisp Technologies EX1024 Page 247

- $x(t) = \tan(t),\ 0 \le t < \pi/2$, with an infinite discontinuity per period.
- Dirichlet's own example, $x(t) = 1$ (for rational t) or 0 (for irrational t) over $(-\pi, \pi)$, which has an infinite number of finite discontinuities due to the infinite number of rational and irrational values t can attain over any finite interval.

Such functions, however, are mostly of theoretical interest. More stringent formulations of sufficient conditions are based on how the operation of integration is interpreted. These involve concepts beyond the scope of this text, such as the theory of infinite sets, and lead to sufficient conditions that are no more useful in practice than the Dirichlet conditions themselves.

REVIEW PANEL 8.21

The Dirichlet Existence Conditions Are Satisfied by Most Periodic Signals

Dirichlet conditions: *Over one period,* a periodic signal $x(t)$ must be *absolutely integrable* and contain a finite number of maxima/minima and a finite number of finite discontinuities.

Since its inception, the theory of Fourier series has been an area of intense mathematical activity and research and has resulted in many advances, most of little concern from a practical standpoint and most well beyond our scope. But the convergence problem remains unsolved. No set of *necessary and sufficient* conditions has yet been formulated.

8.12 A Historical Perspective

Jean Baptiste Joseph Fourier is dead. The man whose name adorns the two most celebrated techniques in spectral analysis was, in the words of his contemporary, Jacobi, a man who believed that the principal aim of mathematics was "public utility and the explanation of natural phenomena." Lest this description pass as an unqualified tribute to Fourier, Jacobi went on to add that Fourier did not "perceive the end of science as the honor of the human mind." What did he mean by that?

8.12.1 Prologue

To read between Jacobi's lines, we must realize that Fourier was more an engineer or physicist than a mathematician. If we bear in mind the two key ingredients of the Fourier series, harmonically related sinusoids and the independence of coefficients, we find that in this specific context, three facts emerge:

1. Fourier's own memoirs indicate quite a lack of mathematical rigor in the manner in which he obtained his results concerning Fourier series.
2. Fourier certainly made no claim to originality in obtaining the formulas for calculating the coefficients, and such formulas were, in fact, described by others years earlier.
3. Fourier was not the first to propose the use of harmonically related sinusoids to represent functions. This, too, was suggested well before his time.

What, then, is Fourier's claim to fame in the context of the series named after him? To be honest, his single contribution was to unify these ideas, look beyond the subtle details, and proclaim that any *arbitrary* function could be expressed as an *infinite* sum of harmonically related sinusoids. That Fourier realized the enormity of his hypothesis is evident from the fact that he spent his last years basking in the glory of his achievements, in the company of many a sycophant. But it is due more to the achievements of other celebrated mathematicians, both contemporaries and successors, who sought to instill mathematical rigor by filling in and clarifying the details of his grand vision and, in so doing, invented concepts that now form the very fabric of modern mathematical analysis, that Fourier's name is immortalized.

PGR2026-00033
Krisp Technologies EX1024 Page 248

8.12.2 The Eighteenth Century

Our historical journey into the development of Fourier series begins in the eighteenth century, a time of little rivalry and even lesser distinction between mathematicians and physicists. The concept of a function was a mathematical expression, a single mathematical formula that described its extent. A triangular waveform would not have been regarded as one function but two, and an arbitrary graph might not be represented by a function at all. The mid-eighteenth century saw the study of vibrating strings, which is subject to a partial differential equation, occupy the minds of three famous personalities of the day. The solution in functional form, assuming the initial string shape to be described by a single expression, was first proposed by d'Alembert in 1748 and extended a year later by Euler, arguably the most prolific mathematician in history, to include arbitrary shapes having different analytical expressions over different parts. It was Bernoulli who, in 1753, first suggested a trigonometric sine series as a general solution for string motion when it starts from rest. To put the problem in mathematical perspective, the one-dimensional differential equation governing the string motion y may be written as $\partial^2 f/\partial t^2 = c^2 \partial^2 f/\partial x^2$. The solution d'Alembert and Euler proposed was $f = \psi(x + ct) + \psi(x - ct)$ and involves traveling waves. Bernoulli's solution was of the form $f = K_1 \sin x \cos ct + K_2 \sin 2x \cos 2ct + \cdots$, when the string starts from rest. This solution, being the most general, he said, must include the solutions of Euler and d'Alembert. Euler objected on the grounds that such a series could represent an arbitrary function of a single variable in terms of such a series of sines only if the function were both odd and periodic.

In 1767, Lagrange, then only 23 years of age but respected by both Euler and d'Alembert, obtained a solution tantalizingly close to the form with which we are now familiar, if only to transform it into the functional form of Euler. His only objective at the time was to defend Euler's claim against the critics. The solution that Lagrange transformed to fit the form given by Euler involved a time dependence that, when set to zero, would have led to a series of sines describing the string displacement, with coefficients given by the integral form for b_k as we now recognize it. Had Lagrange taken this one step more, the fame and glory would have been all his.

Clairaut in 1757, and Euler two decades later (on returning to the problem of vibrating strings in 1777), proposed evaluation of the coefficients of trigonometric series whose existence was already demonstrated by other means, more or less in the form we know them now. But they were still skeptical and believed that such results could be applied only to functions as then understood (a single analytical expression).

8.12.3 The Nineteenth Century

The advent of the nineteenth century saw the emergence of a new and sometimes bitter rivalry between the proponents of mathematics and physics as the two fields became more well defined and the gulf between them widened. This was the time when Fourier burst on the scene. It was during the course of his studies in heat conduction, his first and lasting love (he even felt a sheer physical need for heat, always dressing heavily and living in excessively warm quarters), that he took the decisive step of suggesting a trigonometric series representation for an entirely arbitrary function, whether or not such a function could be described by mathematical expressions over its entire extent. He also realized that, given a function over a prescribed range, its value outside that range is in no way determined and this foreshadowed the concept of periodic extension central to the series representation.

These ideas were contained in Fourier's first paper on the theory of heat conduction, competed in draft form in 1804–1805 (while in Grenoble) and subsequently presented to the Paris Academy in 1807. The referees encouraged Fourier to continue his study, even though his results generated as much skepticism and controversy as interest and excitement. In a letter rebutting his critics (written probably to Lagrange), Fourier apologized for not acknowledging previous studies by Euler and d'Alembert (due to his inability to refer to those works, he alleged). Yet, he also showed his familiarity with such studies when he claimed that they were inadequate because they did not consider the range over which such expansions were valid while

PGR2026-00033
Krisp Technologies EX1024 Page 249

denying their very existence for an arbitrary function. In any event, the propagation of heat was made the subject of the *grand prix de mathematiques* for 1812. It is curious that there was only one other candidate besides Fourier! Whether this topic was selected because of Fourier's goading and the implied challenge to his critics or because of the genuine interest of the referees in seeing his work furthered remains shrouded in mystery. Fourier submitted his prize paper late in 1811, and even though he won the prize, the judges—the famous trio of Laplace, Lagrange, and Legendre among them—were less than impressed with the rigor of Fourier's methods and decided to withhold publication in the Academy's memoirs.

Remember that 1812 was also a time of great turmoil in France. Napoleon's escape from Elba in 1815 and the second restoration led to a restructuring of the Academy to which Fourier was finally nominated in 1816 amidst protests and elected only a year later. And the year 1822 saw his election to the powerful post of Secretary of the Academy and with it, the publication of his treatise on heat conduction, which contained without change, his first paper of 1807. That he had resented criticism of his work all along became evident when, a couple of years later, he caused publication in the Academy's memoirs of his prize-winning paper of 1811, exactly as it was communicated. Amidst all the controversy, however, the importance of his ideas did not go unrecognized (Lord Kelvin, a physicist himself, called Fourier's work a great mathematical poem) and soon led to a spate of activity by Poisson, Cauchy, and others. Cauchy, in particular, took the first step toward lending mathematical credibility to Fourier's work by proposing in 1823 a definition of an integral as the limit of a sum and giving it a geometric interpretation. But it was Dirichlet's decade of dedicated work that truly patched the holes, sanctified Fourier's method with rigorous proofs, and set the stage for further progress. He not only extended the notion of a function, giving a definition that went beyond geometric visualization, but more important he provided seminal results on the sufficient conditions for the convergence of the series (1829) and the criterion of absolute integrability (1837), which we encountered toward the end of this chapter.

Motivated by Dirichlet, Riemann after completing his doctorate started work on his probationary essay on trigonometric series in 1851 in the hope of gaining an academic post at the University of Göttingen. This essay, completed in 1853, was unfortunately published only in 1867, a year after he died. Meanwhile, his probationary lecture on *the hypotheses that lie at the foundations of geometry* on June 10, 1854, shook its very foundations and paved the way for his vision of a new differential geometry (which was to prove crucial to the development of the theory of relativity). But even his brief excursion into Fourier series resulted in two major advances. First, Riemann widened the concept of the integral of a continuous function beyond Cauchy's definition as the limit of a sum to include functions that were neither continuous nor contained a finite number of discontinuities. Second, and more important, he proved that, for bounded integrable functions, the Fourier coefficients tend to zero as the harmonic index increases without limit, suggesting that convergence at a point depends only on the behavior of the function in the vicinity of that point.

The issues of convergence raised by Riemann in his essay—and the problems of convergence in general—motivated several personalities who made major contributions, Heine and Cantor among them. Heine in 1870 established uniform convergence (first conceived in 1847) at all points except at discontinuities for functions subject to the Dirichlet conditions. Cantor's studies, on the other hand, probed the foundations of analysis and led him in 1872 to propose the abstract theory of sets.

There were other developments in the nineteenth century, too. In 1875 du Bois-Raymond showed that if a trigonometric series converges to an integrable function $x(t)$, then that series must be the Fourier series for $x(t)$. And a few years later in 1881, Jordan introduced the concept of functions of bounded variation leading to his own convergence conditions. Interestingly enough, Parseval, who proposed a theorem for summing series of products in 1805 that now carries his name, was not directly involved in the study of Fourier series, and his theorem seems to have been first used in this context nearly a century later in 1893. The continuous form, to be encountered in the chapter on Fourier transforms and often referred to as the energy relation,

PGR2026-00033
Krisp Technologies EX1024 Page 250

was first used by Rayleigh in connection with black body radiation. Both forms were later (much later, in the twentieth century, in fact) generalized by Plancherel.

The Gibbs effect was apparently first observed and described by the English mathematician Wilbraham. However, it was made more widely known by the American physicist Michelson (the famous Michelson-Morley experiment of 1887 is dubbed as the greatest experiment in physics that ever failed because its negative results led to a firm foundation for the theory of relativity). Michelson, who had developed a harmonic analyzer in 1898, could not get perfect-square wave reconstruction (without overshoot) by adding its 80 harmonics. Convinced that it was not an artifact of the device itself, he described his problem in a letter to the scientific journal *Nature*. A year later in 1899, Gibbs (who is remembered more for his work in thermodynamics leading to the phase rule and, lest we forget, for introducing the notation $\cdot$ and $\times$ to represent the dot product and cross-product of vectors) offered an explanation of the effect that now bears his name in a letter to the same journal. The story goes that Michelson had written to Gibbs to seek an explanation for his observations. It is curious that the collected works of Gibbs (which include his letters to Michelson) make no reference to such correspondence. Gibbs' letter to *Nature* dealt with a sawtooth waveform, did not contain any proof, and received scant attention at the time. Only in 1906 did Bôcher demonstrate how the Gibbs effect actually occurs for any waveform with jump discontinuities.

In his studies on Fourier series smoothing, Fejer was led to propose summation methods usually reserved for divergent series. In 1904 he showed that summation by the method of arithmetic means results in a smoothing effect at the discontinuities and the absence of the Gibbs phenomenon. This was to be the forerunner of the concept of spectral windows that are now so commonly used in spectral analysis.

8.12.4 The Twentieth Century

As we move into the twentieth century, we cross another frontier in the development of Fourier methods, led this time by Lebesgue. Following in the footsteps of Borel, who in 1898 formulated the concept of measure of a set, Lebesgue in 1902 went on to reformulate the notion of the integral on this basis and in 1903 extended Riemann's theorem to include unbounded functions, provided integration was understood in the Lebesgue sense. In 1905 he presented a new sufficient condition for convergence that included all the previously known ones. He even settled the nagging problem of term-by-term integration of Fourier series, showing that it no longer depends on the uniform convergence of the series.

There have been many other developments, mostly of purely mathematical interest, as a result of Lebesgue's fundamental work: the conjecture of Lusin in 1915 that the Fourier series of a square integrable function converges except on a set of measure zero, proved finally in 1966 by Carleson; the Riesz-Fischer theorem, proving the converse of the Parseval's theorem; and others. All these developments have finally given Fourier series a respectability that mathematicians have craved but engineers hardly cared about.

This, then, is where our saga ends. Even though there are certainly other aspects that deserve attention, we mention some of these only in passing. On the theoretical side are the Fourier integral (transform) itself; Schwartz's concept of generalized functions, which we have been using so casually (in the form of impulse functions) in this book; and Wiener's exposition of generalized harmonic analysis. On the practical side, the advent of high-speed computing in the 1950s and the development of fast algorithms for Fourier analysis, starting with the work of Danielson and Lanczos in 1942 and of Cooley and Tukey in 1965, has led to a proliferation of Fourier-based methods in almost every conceivable discipline. The end is nowhere in sight.

8.12.5 Epilogue

In the words of Fourier, "The profound study of nature is the most fertile source of mathematical discoveries." Be that as it may, it surely would have been difficult, even for Fourier himself, to predict the household recognition of his name within the scientific world that would follow from his own study of a tiny aspect of that nature. Tiny or not, let us also remember that every time we describe a definite integral by the notation $\int_a^b$, we are using yet another one of his gifts. What else is there to say but *viva Fourier!*

PGR2026-00033
Krisp Technologies EX1024 Page 251

CHAPTER 8 PROBLEMS

DRILL AND REINFORCEMENT

8.1 (Fourier Series Concepts) Express each of the following periodic signals by its Fourier series in all three forms, identify the time period and the harmonics present, compute the signal power and the rms value, and, for parts (a) and (b), sketch the one-sided and two-sided spectra.

(a) $x(t) = 4 + 2\sin(4\pi t + \frac{\pi}{4}) + 3\sin(16\pi t) + 4\cos(16\pi t)$

(b) $x(t) = \sum_{\substack{k=-4 \\ k\neq 0}}^{4} \frac{6}{k}\sin(\frac{k\pi}{2})e^{-jk\pi/3}e^{jk6\pi t}$

(c) $x(t) = 2 - 4\cos(8\pi t + 0.25\pi) + 10\sin(12\pi t - \frac{\pi}{3})$

(d) $x(t) = [\cos(t) + 2\cos(2t)]^2$

8.2 (Fourier Series Concepts) Consider the periodic signal $x(t) = \sum_{k=1}^{\infty} \frac{6}{k}\sin^2(\frac{k\pi}{2})\cos(1600k\pi t)$.

(a) Identify the fundamental frequency f_0 and time period T.

(b) Identify the Fourier series coefficients a_k, b_k, c_k, θ_k, and $X[k]$.

(c) Identify any symmetry (true and hidden) present in $x(t)$.

8.3 (Fourier Series Concepts) Consider the periodic signal $x(t) = \sum_{k=1}^{\infty} \frac{6}{k}\sin(\frac{k\pi}{2})\sin(100k\pi t + k\frac{\pi}{3})$.

(a) Identify the fundamental frequency f_0 and time period T.

(b) Identify the Fourier series coefficients a_k, b_k, c_k, θ_k and $X[k]$.

(c) Identify any symmetry (true and hidden) present in $x(t)$.

8.4 (Fourier Series Coefficients) Sketch a few periods of each of the following periodic signals described over one period, and find the indicated Fourier series coefficient.

(a) $X[k]$ for $x(t) = e^{-t},\ 0 \le t \le 1$, with $T = 1$

(b) a_k for $x(t) = \text{rect}(t - 0.5)$, with $T = 2$

(c) b_k for $x(t) = (1 + t),\ 0 \le t \le 1$, with $T = 1$

8.5 (Symmetry) The magnitude and phase spectra of two periodic signals are shown in Figure P8.5.

(a) Identify the harmonics present in the Fourier series of each signal.

(b) Identify the symmetry (hidden or otherwise) in each periodic signal (if any).

(c) Write out the Fourier series of each signal in polar form.

(d) Find the signal power and the rms value of each signal.

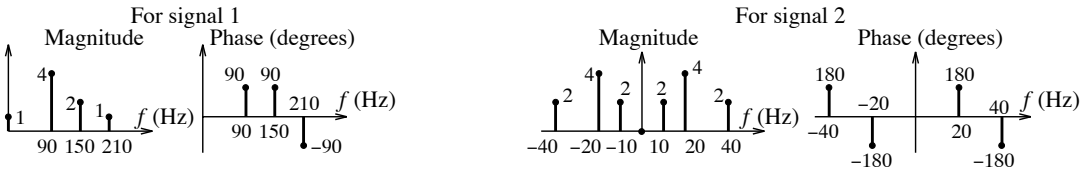


Figure P8.5 The spectra for Problem 8.5

PGR2026-00033
Krisp Technologies EX1024 Page 252

8.6 (Symmetry) A periodic signal $x(t)$ is described by $x(t) = t,\ 0 \le t \le 1$, over a portion of its time period T. Sketch this periodic signal over $-2T \le t \le 2T$ and indicate what symmetry it possesses about $t = \frac{T}{2}$ and $t = \frac{T}{4}$ for the following cases.

(a) $x(t)$ possesses only even symmetry, and $T = 2$.
(b) $x(t)$ possesses only odd symmetry, and $T = 2$.
(c) $x(t)$ possesses even and half-wave symmetry, and $T = 4$.
(d) $x(t)$ possesses odd and half-wave symmetry, and $T = 4$.

8.7 (Fourier Series) For each periodic signal shown in Figure P8.7,

(a) Compute the Fourier series coefficients a_k, b_k, and $X[k]$ and, where appropriate, simplify for odd k and even k. Evaluate special or indeterminate cases separately, if necessary.
(b) Compute the signal power in the fundamental component.
(c) Compute the signal power up to the fourth harmonic.
(d) Compute the total signal power.
(e) Identify the convergence rate.

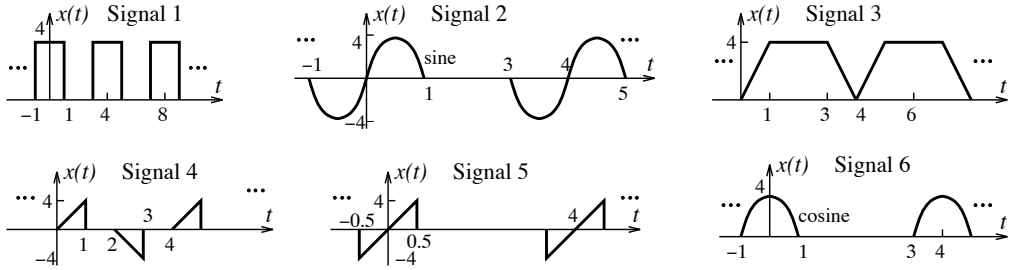


Figure P8.7 The periodic signals for Problem 8.7

8.8 (Properties) The magnitude and phase spectra of a periodic signal $x(t)$ are shown in Figure P8.8.

(a) Write out the Fourier series in polar form and simplify where appropriate.
(b) Sketch the magnitude and phase spectra of $f(t) = x(2t)$.
(c) Sketch the magnitude and phase spectra of $g(t) = x(t - \frac{1}{6})$.
(d) Sketch the magnitude and phase spectra of $h(t) = x'(t)$.

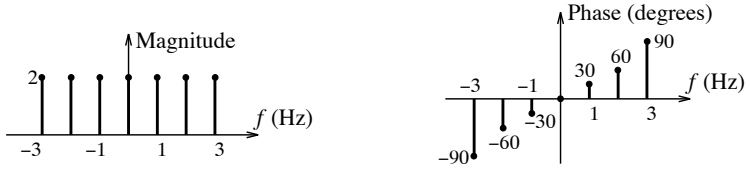


Figure P8.8 The spectra for Problem 8.8

8.9 (Properties) Let $X[k]$ be the exponential Fourier series coefficients of a periodic signal $x(t)$. Find the Fourier series coefficients of the following:

(a) $f(t) = x(2t)$ **(b)** $g(t) = x(-t)$ **(c)** $h(t) = x(-2t)$ **(d)** $y(t) = 2 + x(2t)$

8.10 (Derivative Method) Sketch each periodic signal $x(t)$ described over one period (with $T = 1$) and find its exponential Fourier series coefficients $X[k]$, using the derivative method.

(a) $x(t) = \text{rect}(2t)$ **(b)** $x(t) = e^t \text{rect}(t - 0.5)$ **(c)** $x(t) = t\,\text{rect}(t)$ **(d)** $x(t) = \text{tri}(2t)$

PGR2026-00033
Krisp Technologies EX1024 Page 253

8.11 (Convergence and the Gibbs Effect) Consider a periodic signal with time period $T = 2$ whose one period is described by $x(t) = \text{tri}(2t)$. Answer the following without computing its Fourier series coefficients.

(a) What is the convergence rate? Will its Fourier series reconstruction show the Gibbs effect? If so, what is the peak overshoot at each discontinuity?

(b) What value will its Fourier series converge to at $t = 0,\ t = 0.25,\ t = 0.5$, and $t = 1$?

8.12 (Convergence and the Gibbs Effect) Consider a periodic signal with time period $T = 2$ whose one period is described by $x(t) = 6\,\text{rect}(t - 0.5)$. Answer the following without computing its Fourier series coefficients.

(a) What is the convergence rate? Will its Fourier series reconstruction show the Gibbs effect? If so, what is the peak overshoot at each discontinuity?

(b) What value will its Fourier series converge to at $t = 0,\ t = 0.5,\ t = 1$, and $t = 1.5$?

8.13 (Convergence and the Gibbs Effect) Consider a periodic signal with time period $T = 2$ whose one period is described by $x(t) = 6\,\text{rect}(t)\text{sgn}(t)$. Answer the following without computing its Fourier series coefficients.

(a) What is the convergence rate? Will its Fourier series reconstruction show the Gibbs effect? If so, what is the peak overshoot at each discontinuity?

(b) What value will its Fourier series converge to at $t = 0,\ t = 0.5,\ t = 1$, and $t = 1.5$?

8.14 (Convergence and the Gibbs Effect) Consider a periodic signal with time period $T = 2$ whose one period is described by $x(t) = 4e^{-t}\text{rect}(t - 0.5)$. Answer the following without computing its Fourier series coefficients.

(a) What is the convergence rate? Will its Fourier series reconstruction show the Gibbs effect? If so, what is the peak overshoot at each discontinuity?

(b) What value will its Fourier series converge to at $t = 0,\ t = 0.5,\ t = 1$, and $t = 1.5$?

8.15 (Modulation) A periodic signal $x(t)$ is described by $x(t) = 2 + \sum_{k=1}^{4} \frac{6}{k} \sin^2(\frac{k\pi}{2})\cos(1600k\pi t)$.

(a) Sketch the two-sided spectrum of $x(t)$.

(b) Sketch the two-sided spectrum of the modulated signal $y(t) = x(t)\cos(1600\pi t)$.

(c) Find the signal power in $x(t)$ and in the modulated signal $y(t)$.

8.16 (System Analysis) The periodic signal $x(t) = |\sin(250\pi t)|$ is applied to an ideal filter as shown:

$$x(t) \longrightarrow \boxed{\text{ideal filter}} \longrightarrow y(t)$$

Sketch the output spectrum of the filter and the time-domain output $y(t)$ if

(a) The ideal filter blocks all frequencies past 200 Hz.

(b) The ideal filter passes only frequencies between 200 and 400 Hz.

(c) The ideal filter blocks all frequencies past 400 Hz.

8.17 (System Analysis) A periodic signal $x(t)$ is described by $x(t) = 2 + \sum_{k=1}^{\infty} \frac{6}{k} \sin^2(\frac{k\pi}{2})\cos(1600k\pi t)$.

This signal forms the input to the following system:

$$x(t) \longrightarrow \boxed{\text{ideal filter}} \longrightarrow \boxed{\text{RC lowpass filter } (\tau = 1 \text{ ms})} \longrightarrow y(t)$$

Sketch the output spectrum of the ideal filter and the time-domain output $y(t)$ if

PGR2026-00033
Krisp Technologies EX1024 Page 254

(a) The ideal filter blocks all frequencies past 200 Hz.
(b) The ideal filter passes only frequencies between 200 Hz and 2 kHz.
(c) The ideal filter blocks all frequencies past 2 kHz.

8.18 (System Analysis) The signal $x(t) = |10\sin(t)|$ volts is applied to each circuit shown in Figure P8.18. Assume that $R = 1\ \Omega$, $C = 1$ F, and $L = 1$ H. For each circuit,

(a) Find the Fourier series coefficients c_0 and c_k for the signal $x(t)$.
(b) Find the dc component of the output $y(t)$.
(c) Find the fundamental component of the output $y(t)$.
(d) Find the power in $y(t)$ up to (and including) the second harmonic.

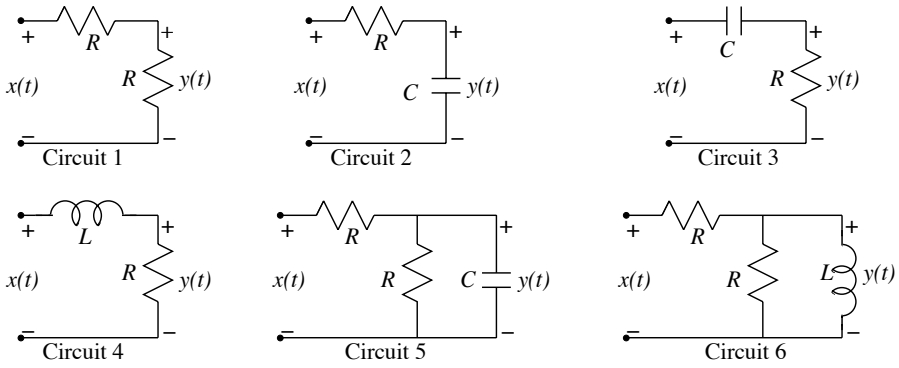


Figure P8.18 The circuits for Problem 8.18

8.19 (Application) The input to an amplifier is $x(t) = \cos(10\pi t)$, and the output of the amplifier is given by $y(t) = 10\cos(10\pi t) + 2\cos(30\pi t) + \cos(50\pi t)$.

(a) Compute the third harmonic distortion.
(b) Compute the total harmonic distortion.

8.20 (Application) The signal $x(t) = \sin(10\pi t)$ is applied to the system whose output is $y(t) = x^3(t)$.

(a) Which harmonics are present in the output $y(t)$?
(b) Compute the total harmonic distortion and the third harmonic distortion (if any) in $y(t)$.

8.21 (Application) A square wave with zero dc offset and period T is applied to an RC circuit with time constant τ. The output is the capacitor voltage. Without extensive computations,

(a) Sketch the filter output if $\tau \approx 100\ T$.
(b) Sketch the filter output if $\tau \approx 0.001\ T$.

8.22 (Design) We wish to design a dc power supply using the following scheme:

$$x(t) \longrightarrow \boxed{\text{full-wave rectifier}} \longrightarrow \boxed{\text{RC lowpass filter}} \longrightarrow y(t)$$

The input is a symmetric triangular wave with $T = 0.02$ s and zero dc offset. What value of τ, the filter time constant, will ensure a ripple of less than 1% in the filter output $y(t)$? What value of C is required if $R = 1$ kΩ?

PGR2026-00033
Krisp Technologies EX1024 Page 255

8.23 (Design) We wish to design a dc power supply using the following scheme:

$$x(t) \longrightarrow \boxed{\text{half-wave rectifier}} \longrightarrow \boxed{\text{RC lowpass filter}} \longrightarrow y(t)$$

The input is a pure sine wave with $T = 0.02$ s. What value of τ, the filter time constant, will ensure a ripple of less than 1% in the filter output $y(t)$? What value of C is required if $R = 1$ kΩ?

REVIEW AND EXPLORATION

8.24 (Fourier Series Coefficients) Consider a periodic signal $x(t)$ described by

$$x(t) = \alpha_0 + \sum_{k=1}^{\infty} \alpha_k \cos[k\omega_0(t + t_0)]$$

What are the Fourier series coefficients a_k and b_k of this signal?

8.25 (Fourier Series Coefficients) If $x(t)$ is a periodic signal whose shifted version $y(t) = x(t - t_0)$ is symmetric, we can first compute the Fourier series coefficients of its symmetric version $y(t)$, express $x(t)$ as $x(t) = y(t + t_0)$, and use the result of the previous problem to extract the Fourier series coefficients of $x(t)$. Use this approach to find the Fourier series coefficients of the following periodic signals described over one period.

(a) $x(t) = -1 + t,\ -1 \le t \le 3$, with $T = 4$

(b) $x(t) = \text{tri}[0.5(t - 1)],\ -1 \le t \le 3$, with $T = 4$

8.26 (Fourier Series) Sketch the periodic signal $x(t)$ whose nonzero Fourier series coefficients are

$$b_k = \int_0^2 12t \sin(0.5k\pi t)\, dt$$

8.27 (Fourier Series) Sketch the periodic signal $x(t)$ whose nonzero Fourier series coefficients are

$$a_k = \int_0^1 12t \cos(0.5k\pi t)\, dt \quad (k \text{ odd})$$

8.28 (Modulation) The signal $x(t) = \cos(2\pi f_C t)$ is modulated by an even symmetric periodic square wave $s(t)$ with time period T and one period described by $s(t) = \text{rect}(t/\tau)$. Find and sketch the spectrum of the modulated signal if $f_C = 1$ MHz, $T = 1$ ms, and $\tau = 0.1$ ms.

8.29 (Application) The signal $x(t) = \sin(10\pi t)$ is applied to the system whose output is $y(t) = |x(t)|$.

(a) Sketch the output $y(t)$. Which harmonics are present in $y(t)$?

(b) Compute the total harmonic distortion and the third harmonic distortion (if any) in $y(t)$.

8.30 (Application) The signal $x(t) = \sin(10\pi t)$ is applied to the system whose output is $y(t) = \text{sgn}[x(t)]$.

(a) Sketch the output $y(t)$. Which harmonics are present in $y(t)$?

(b) Compute the total harmonic distortion and the third harmonic distortion (if any) in $y(t)$.

8.31 (Application) Consider two periodic signals $g(t)$ and $h(t)$ described by

$$g(t) = 2 + \sum_{k=1}^{\infty} \frac{4}{k} \cos(1600k\pi t) \qquad h(t) = \sum_{k=1}^{\infty} \frac{8}{k^2} \sin(800k\pi t)$$

PGR2026-00033
Krisp Technologies EX1024 Page 256

(a) Each is passed through an ideal lowpass filter that blocks frequencies past 1 kHz to obtain the filtered signals $x(t)$ and $y(t)$. Sketch the two-sided spectra of $x(t)$ and $y(t)$.

(b) The filtered signals are multiplied to obtain the signal $w(t) = x(t)y(t)$. Sketch the two-sided spectra of $w(t)$.

(c) The signal $w(t)$ is passed through an ideal lowpass filter that blocks frequencies past 1 kHz to obtain the filtered $z(t)$. Sketch $z(t)$ and its two-sided spectra.

8.32 (Fourier Series) The Fourier series coefficients of a periodic signal $x(t)$ with period $T = 2$ are zero for $k \geq 3$. It is also known that $x(t) = x(-t)$ and $x(t) = -x(t-1)$. The signal power in $x(t)$ is 4. Find an expression for $x(t)$.

8.33 (Fourier Series) A periodic signal $x(t)$ with period $T = 0.1$ s is applied to an ideal lowpass filter that blocks all frequencies past 15 Hz. Measurements of the filter output $y(t)$ indicate that $y_{\text{av}} = 2$, $y_{\text{rms}} = 3$, and $y(0) = 4$. Find an expression for $y(t)$.

8.34 (System Analysis) A rectangular pulse train with a duty ratio of 0.5 is described over one period by $x(t) = \text{rect}(t)$. This signal is applied to a system whose differential equation is $y'(t) + y(t) = x(t)$.

(a) What type of symmetry (if any) is present in the input $x(t)$?

(b) What type of symmetry (if any) will be present in the output $y(t)$?

(c) Compute the dc component and second harmonic component of $y(t)$.

(d) Will the Fourier series reconstruction of $y(t)$ exhibit the Gibbs effect? Should it? Explain.

8.35 (Design) For a series RLC circuit excited by a voltage source, the voltage across the resistor peaks at the resonant frequency $\omega_0 = \frac{1}{\sqrt{LC}}$. The sharpness of the frequency response increases with the *quality factor* $Q = \omega_0 L/R$. Specify the Q of an RLC circuit that is excited by a half-wave symmetric square wave pulse train and produces a resistor voltage that is essentially a sinusoid at the 25th harmonic of the input with a contribution of less than 5% due to any other harmonics.

8.36 (Design) A periodic square wave signal with zero dc offset and $T = 1$ ms is applied to an RC lowpass filter with $R = 1$ kΩ.

(a) Find C such that the output phase differs from the input phase by exactly 45° at 5 kHz.

(b) The half-power frequency is defined as the frequency at which the output power equals half the input power. How is this related to the time constant of the RC filter? What is the half-power frequency of the RC filter designed in part (a)?

8.37 (Convergence) The Fourier series of a periodic signal is $x(t) = 2 + \sum_{k\,\text{odd}}^{\infty} \frac{4}{k\pi} \sin(k\pi t)$.

(a) Will the reconstruction of $x(t)$ exhibit the Gibbs effect?

(b) What value does the series converge to at $t = 0$?

(c) What value does the series converge to at $t = 0.5T$?

(d) What value does the series converge to at $t = 0.25T$?

8.38 (Gibbs Effect) A periodic signal with convergence rate $\propto 1/k$ is applied to a series RC circuit.

(a) Will the capacitor voltage exhibit the Gibbs effect? Should it? Explain.

(b) Will the resistor voltage exhibit the Gibbs effect? Should it? Explain.

PGR2026-00033
Krisp Technologies EX1024 Page 257

8.39 (Reconstruction at a Discontinuity) If t_- and t_+ denote instants on either side of a discontinuity in a signal $x(t)$, the mean square error in the Fourier series reconstruction $x_n(t)$ may be written as $\epsilon = [x_n(t) - x(t_-)]^2 + [x_n(t) - x(t_+)]^2$. Its minimum with respect to $x_n(t)$ occurs when $\dfrac{d\epsilon}{dx_n} = 0$. Show that this results in $x_n(t) = \frac{1}{2}[x(t_-) + x(t_+)]$ and implies that the reconstructed value converges to the midpoint at a discontinuity.

8.40 (Closed Forms for Infinite Series) Parseval's theorem provides an interesting approach to finding closed-form solutions to infinite series. Starting with the Fourier series coefficients of each signal, use Parseval's theorem to generate the following closed-form results.

(a) $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$, using a sawtooth periodic signal

(b) $\displaystyle\sum_{k=1,\text{odd}}^{\infty} \frac{1}{k^4} = \frac{\pi^4}{96}$, using a triangular periodic signal

(c) $\displaystyle\sum_{k=1}^{\infty} \frac{1}{(1-4k^2)^2} = -0.5 + \frac{\pi^2}{16}$, using a full-wave rectified sine

8.41 (Spectral Bounds) Starting with $x^{(n)}(t) \Leftrightarrow (j2\pi f_0)^n X[k]$ and the fundamental theorem of calculus, $|\int x(\alpha)\, d\alpha| \leq \int |x(\alpha)|\, d\alpha$, and noting that $|e^{\pm j\theta}| = 1$, prove that

$$|X[k]| \leq \frac{1}{T|2\pi f_0|^n} \int_T |x_p^{(n)}(t)|\, dt$$

This result sets the nth bound on the spectrum in terms of the *absolute area* under the nth derivative of $x_p(t)$. Since the derivatives of an impulse possess zero area, the number of nonzero bounds that can be found equals the number of times we can differentiate before only derivatives of impulses occur. Use this result to find all the nonzero spectral bounds (starting with $n = 0$) for the following periodic signals defined over one period.

(a) $x(t) = \text{rect}(t)$, with $T = 2$

(b) $x(t) = \text{tri}(t)$, with $T = 2$

(c) $x(t) = |\sin(t)|$, with $T = \pi$

8.42 (Smoothing Kernels) The periodic signal (Dirichlet kernel) $d(t)$ corresponding to $D[k] = \text{rect}(\frac{k}{2N})$ is $d(t) = M\frac{\text{sinc}(Mf_0t)}{\text{sinc}(f_0t)}$, where $M = 2N + 1$. Use this result (and the convolution property) to find the periodic signal $f(t)$ (the Fejer kernel) corresponding to the triangular window $F[k] = \text{tri}(\frac{k}{M})$.

8.43 (Significance of Fourier Series) To appreciate the significance of Fourier series, consider a periodic sawtooth signal $x(t) = t,\ 0 < t < 1$, with time period $T = 1$ whose nonzero Fourier series coefficients are $a_0 = 0.5$ and $b_k = \frac{-1}{k\pi},\ k > 0$. Its Fourier series up to the first harmonic is $x_1(t) \approx \frac{1}{2} - \frac{1}{\pi}\sin(2\pi t)$.

(a) Find the power P_T in $x(t)$, the power P_1 in $x_1(t)$, and the power error $P_T - P_1$.

(b) Approximate $x(t)$ by $y_1(t) = A_0 + B_1 \sin(2\pi t)$ as follows. Pick two time instants over $(0, 1)$—say $t_1 = \frac{1}{4}$ and $t_2 = \frac{1}{2}$—and substitute into the expression for $y_1(t)$ to obtain:

$$y_1(t_1) = A_0 + B_1 \sin(2\pi t_1) \qquad y_1(\tfrac{1}{2}) = \tfrac{1}{2}A_0 + B_1 \sin(\pi) = A_0$$
$$y_1(t_2) = A_0 + B_1 \sin(2\pi t_2) \qquad y_1(\tfrac{1}{4}) = A_0 + B_1 \sin(\tfrac{\pi}{2}) = A_0 + B_1$$

PGR2026-00033
Krisp Technologies EX1024 Page 258

(c) Solve for A_0 and B_1, find the power P_y in $y_1(t)$, and find the power error $P_T - P_y$. Does $x_1(t)$ or $y_1(t)$ have a smaller power error?

(d) Start with $t_1 = \frac{1}{6}$, $t_2 = \frac{1}{4}$, and recompute A_0 and B_1. Why are they different? Compute the power error. Is the power error less than for part (c)? Is there a unique way to choose t_1 and t_2 to yield the smallest power error?

(e) If we want to extend this method to an approximation with many more harmonics and coefficients, what problems do you expect?

(f) Argue that the Fourier series method of computing the coefficients is better.

8.44 (Orthogonal and Orthonormal Signals) A set of signals ϕ_k satisfying $\int_a^b \phi_j \phi_k^* = 0$, $j \neq k$, is said to be **orthogonal** over the interval $(a,\ b)$. If, in addition, the energy $E_k = \int_a^b |\phi_k|^2 = 1$ in each ϕ_k, the set is called **orthonormal**.

(a) Show that an even symmetric signal is always orthogonal to any odd symmetric signal over the symmetric duration $-\alpha \leq t \leq \alpha$.

(b) Let $x(t) = \cos(2\pi t)$ and $y(t) = \cos(4\pi t)$. Are they orthogonal over $0 \leq t \leq 1$?

(c) Let $x(t) = \cos(2\pi t)$ and $y(t) = \cos(4\pi t)$. Are they orthogonal over $0 \leq t \leq 0.25$?

8.45 (Orthogonal Sets) A set of functions $\{\phi_k\}$ is orthogonal only if $\int_a^b \phi_j \phi_k^* = 0$, $j \neq k$, for *every* pair of signals in the set. Of the following sets, which are orthogonal and which are also orthonormal?

(a) $\{1,\ \ t,\ \ t^2\},\ \ -1 \leq t \leq 1$

(b) $\{e^{-t/2}u(t),\ \ (1-t)e^{-t/2}u(t),\ \ (1-2t+0.5t^2)e^{-t/2}u(t)\}$

(c) $\{e^{-t^2/2},\ \ te^{-t^2/2}\},\ \ -\infty < t < \infty$

(d) $\{e^{-t},\ \ 2e^{-t} - 3e^{-2t},\ \ 3e^{-t} - 12e^{-2t} + 10e^{-3t}\},\ \ t \geq 0$

8.46 (Generalized Fourier Series) A generalized Fourier series describes an energy signal $x(t),\ a \leq t \leq b$, by a sum of orthogonal *basis signals* $\phi_k(t),\ k = 0, 1, \ldots$ over the same interval $a \leq t \leq b$ as

$$x(t) = \sum_{k=0}^{\infty} \alpha_k \phi_k(t)$$

(a) Show that the relation for finding α_k is $\alpha_k = \dfrac{1}{E_k} \displaystyle\int_a^b x(t)\phi_k^*(t)\, dt$, where E_k is the energy in $\phi_k(t)$.

(b) Show that the energy in the generalized Fourier series equals $\sum |\alpha_k|^2 E_k$.

8.47 (Generalized Fourier Series) Let $x(t) = te^{-t/2}u(t)$.

(a) Obtain a three-term generalized Fourier series approximation $x(t) \approx \alpha_0\phi_0 + \alpha_1\phi_1 + \alpha_2\phi_2$ in terms of the basis functions

$$\phi_0(t) = e^{-t/2}u(t) \qquad \phi_1(t) = (1-t)e^{-t/2}u(t) \qquad \phi_2(t) = (1-2t+0.5t^2)e^{-t/2}u(t)$$

(b) Compute the energy error in the series representation for $x(t)$.

8.48 (Generalized Fourier Series) Let $x(t) = 12e^{-3t}u(t)$.

(a) Obtain a three-term generalized Fourier series approximation $x(t) \approx \alpha_0\phi_0 + \alpha_1\phi_1 + \alpha_2\phi_2$ in terms of the basis functions

$$\{\phi_0(t) = e^{-t} \qquad \phi_1(t) = 2e^{-t} - 3e^{-2t} \qquad \phi_2(t) = 3e^{-t} - 12e^{-2t} + 10e^{-3t}\},\ \ t \geq 0$$

(b) Compute the energy error in the series representation for $x(t)$.

PGR2026-00033
Krisp Technologies EX1024 Page 259

COMPUTATION AND DESIGN

fssyngui A GUI for Fourier Series Synthesis

The graphical user interface `fssyngui` allows you to visualize a periodic signal from its exponential Fourier series coefficients $X[k]$.
You can select the time period and the number of harmonics used.
The interface displays the reconstructed signal and its Fourier series spectrum.
To explore this routine, type `fssyngui` at the MATLAB prompt.

fsgui A GUI for Fourier Series Reconstruction and Smoothing

The graphical user interface `fsgui` allows you to reconstruct a periodic signal by its harmonics and observe how various windows result in a smoothed reconstruction.
You can select the periodic signal, its parameters (time period, duty ratio, etc.), the number of harmonics used, and the smoothing window.
The interface displays the reconstructed signal and its Fourier series spectrum.
To explore this routine, type `fsgui` at the MATLAB prompt.

fsrgui A GUI for System Response to Periodic Signals

The graphical user interface `fsrgui` allows you to visualize the response of a system to a periodic signal.
You can select the periodic signal, its parameters (time period, duty ratio, etc.), and the system parameters.
The interface displays the input and response up to a desired number of harmonics.
To explore this routine, type `fsrgui` at the MATLAB prompt.

8.49 (What Did Michelson See?) As described in the historical perspective (Section 8.12), Gibbs' explanation for the effect that bears his name was prompted by a letter from Michelson who failed to perfectly reconstruct a square wave from its 80 harmonics using his harmonic analyzer.

(a) Use MATLAB to reconstruct a square wave to 80 harmonics and describe what Michelson might have observed.

(b) The explanation offered by Gibbs was for the reconstruction of a sawtooth wave. Use MATLAB to reconstruct a sawtooth wave to 80 harmonics and describe what Gibbs might have observed.

8.50 (Numerical Approximation of Fourier Series Coefficients) An approximation of $x(t)$ to M harmonics may be expressed as

$$x(t) \approx \sum_{k=-M}^{M} X[k] e^{j2k\pi t/T}$$

A numerical approach to estimating $X[k]$ is based on sampling $x(t)$ at N uniformly spaced intervals $t = nt_s,\ 0 \le n \le N-1$, where $t_s = T/N$, to generate N equations whose solution yields the $X[k]$. Argue that $N = 2M+1$ such equations are needed to find all the coefficients $X[k],\ -M \le k \le M$. We wish to use this approach for finding the Fourier series coefficients of

(a) $x(t) = t,\ 0 \le t \le 1$, with $T = 1$ **(b)** $x(t) = \text{tri}(t),\ -1 \le t \le 1$, with $T = 2$

How many harmonics M can we estimate using $N = 5$ and $N = 11$? Compute the $X[k]$ with these values of N and compare with the exact results. Repeat for $N = 21$. What are the effects of increasing N on the accuracy of the results?

PGR2026-00033
Krisp Technologies EX1024 Page 260

8.51 (Numerical Approximation of Fourier Series Coefficients) There is yet another way by which we can approximate the Fourier series coefficients of a periodic signal $x(t)$. We sample $x(t)$ at N uniformly spaced intervals $t = nt_s,\ 0 \le n \le N-1$, where $t_s = T/N$. Argue that the integral for computing the coefficients may be approximated by the summation

$$X[k] = \frac{1}{T}\int_0^T x(t)e^{-j2k\pi t/T}dt \approx \frac{1}{N}\sum_{n=0}^{N-1} x(nt_s)e^{-j2k\pi n/N}$$

We wish to use this approach for finding the Fourier series coefficients of

(a) $x(t) = t,\ 0 \le t \le 1$, with $T = 1$ **(b)** $x(t) = \text{tri}(t),\ -1 \le t \le 1$, with $T = 2$

Compute the $X[k],\ 0 \le k \le N-1$, for $N = 6$ and $N = 10$ and compare with the exact results. Comment on how the values of $X[k],\ k > 0.5N$, are related to $X[N-k]$. Can you explain why they are related? What are the effects of increasing N on the accuracy of $X[k]$ for the first few harmonics?

PGR2026-00033
Krisp Technologies EX1024 Page 261

Chapter 9

THE FOURIER TRANSFORM

9.0 Scope and Objectives

The Fourier transform (FT) provides a frequency-domain description of time-domain signals and may be regarded as an extension of the Fourier series (FS) as applied to aperiodic signals. This chapter develops the Fourier transform as a consequence of transforming a periodic signal to an aperiodic one by stretching the time period without limit and provides a means of going back and forth between the Fourier series and Fourier transform results. It describes the properties of the Fourier transform and explores the effect of operations on the spectrum of a signal. It concludes with some applications of the Fourier transform to signal and system analysis, including an introduction to filters.

9.1 Introduction

The approach we adopt to develop the Fourier transform serves to unite the representations for periodic functions and their aperiodic counterparts, provided the ties that bind the two are also construed to be the very ones that separate them and are therefore understood in their proper context.

Recall that a periodic signal $x_p(t)$ with period T and its exponential Fourier series coefficients $X[k]$ are related by

$$x_p(t) = \sum_{k=-\infty}^{\infty} X[k]e^{j2\pi kf_0 t} \qquad X[k] = \frac{1}{T}\int_{-T/2}^{T/2} x_p(t)e^{-j2\pi kf_0 t}\,dt \tag{9.1}$$

If the period T of a periodic signal $x_p(t)$ is stretched without limit, the periodic signal no longer remains periodic but becomes a single pulse $x(t)$ corresponding to one period of $x_p(t)$. The transition from a periodic to an aperiodic signal also represents a transition from a power signal to an energy signal.

The harmonic spacing $f_0 = 1/T$ approaches zero, and its Fourier series spectrum becomes a continuous curve. In fact, if we replace f_0 by an infinitesimally small quantity $df \to 0$, the discrete frequency kf_0 may be replaced by the continuous frequency f. The factor $1/T$ in the integral relation means that the coefficients $X[k]$ approach zero and are no longer a useful indicator of the spectral content of the aperiodic signal $x(t)$. However, if we eliminate the dependence of $X[k]$ on the offending factor $1/T$ in the integral and work with $TX[k]$, as follows,

$$TX[k] = \int_{-T/2}^{T/2} x_p(t)e^{-j2\pi kf_0 t}\,dt \tag{9.2}$$

the integral on the right-hand side often exists as $T \to \infty$ (even though $TX[k]$ is in indeterminate form), and we obtain meaningful results. Further, since $kf_0 \to f$, the integral describes a function of f. As a result, we

PGR2026-00033
Krisp Technologies EX1024 Page 262

define $TX[k] = X(f)$ and obtain

$$X(f) = \lim_{T\to\infty} TX[k] = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}\,dt \tag{9.3}$$

This relation describes the **Fourier transform** $X(f)$ of the signal $x(t)$ and may also be written in terms of the frequency variable ω as

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}\,dt \qquad \text{(the } \omega\text{-form)} \tag{9.4}$$

The Fourier transform provides a frequency-domain representation of the aperiodic signal $x(t)$.

9.1.1 The Inverse Fourier Transform

A periodic signal $x_p(t)$ can be reconstructed from its spectral coefficients $X[k]$, using

$$x_p(t) = \sum_{k=-\infty}^{\infty} X[k]e^{j2\pi kf_0 t} \tag{9.5}$$

If $T \to \infty$, resulting in the aperiodic signal $x(t)$, it is the quantity $TX[k]$ that describes its spectrum $X(f)$, and we must modify the above expression (multiply and divide by T) to give

$$x_p(t) = \sum_{k=-\infty}^{\infty} TX[k]e^{j2\pi kf_0 t}\frac{1}{T} = \sum_{k=-\infty}^{\infty} TX[k]e^{j2\pi kf_0 t}f_0 \tag{9.6}$$

As $T \to \infty$ and $kf_0 \to f$, the summation tends to an integration over $(-\infty, \infty)$. With $f_0 \to df \to 0$, we obtain

$$x(t) = \int_{-\infty}^{\infty} X(f)e^{j2\pi ft}\,df \tag{9.7}$$

This is the **inverse Fourier transform**, which allows us to obtain $x(t)$ from its spectrum $X(f)$. The inverse transform relation may also be written in terms of the variable ω (by noting that $d\omega = 2\pi\,df$) to give

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)e^{j\omega t}\,d\omega \qquad \text{(from the } \omega\text{-form)} \tag{9.8}$$

Unlike the Fourier series, it does matter whether we use the f-form or ω-form, especially when using the inverse transform relation! We prefer the f-form because the Fourier transform and its inverse are almost symmetric and yield easily memorized forms for many signals. However, we shall work out some examples in the ω-form and provide all properties in both forms.

The signal $x(t)$ and its Fourier transform $X(f)$ or $X(\omega)$ form a unique transform pair, and their relationship is shown symbolically using a double arrow:

$$x(t) \Longleftrightarrow X(f) \qquad \text{or} \qquad x(t) \Longleftrightarrow X(\omega) \tag{9.9}$$

REVIEW PANEL 9.1

The Fourier Transform and Inverse Fourier Transform

Transform: $X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}dt$ **Inverse transform:** $x(t) = \int_{-\infty}^{\infty} X(f)e^{j2\pi ft}df$

PGR2026-00033
Krisp Technologies EX1024 Page 263

9.1.2 Connections: Sampling and Periodic Extension

Turning things around, the continuous spectrum of an aperiodic signal $x(t)$ sampled at the locations kf_0 describes a periodic signal $x_p(t)$ with period $T = 1/f_0$, which arises from replications of $x(t)$. Formally, *sampling in the frequency domain corresponds to periodic extension in the time domain. The sample spacing f_0 in the frequency domain is reciprocally related to the period T on the time domain through $T = 1/f_0$.* Only if $x(t)$ is time-limited with width $t_0 < T$ does the periodic extension correspond to a simple periodic *continuation* of $x(t)$. This concept forms an important link between the two domains.

REVIEW PANEL 9.2

Relating the Fourier Transform and the Fourier Series (with $T = 1/f_0$)

Series to transform: $X(f) = TX[k]|_{kf_0 \Rightarrow f}$ **Transform to series:** $X[k] = \frac{1}{T}X(f)\big|_{f \Rightarrow kf_0}$

EXAMPLE 9.1 (Connections Between Fourier Series and Fourier Transforms)

(a) The coefficients $X[k]$ of a train of the rectangular pulse train shown in Figure E9.1A are

$$X[k] = \frac{1}{T}At_0 \operatorname{sinc}(kf_0t_0)$$

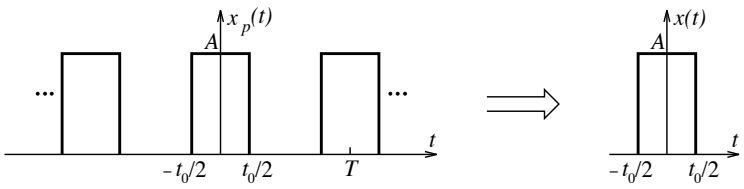


Figure E9.1A The rectangular pulse train for Example 9.1(a)

The Fourier transform of the single pulse $x(t)$ that represents one period of $x_p(t)$ is then

$$X(f) = TX[k]_{kf_0=f} = At_0 \operatorname{sinc}(ft_0)$$

Note the equivalence of the envelope (a sinc function), the removal of T, and the change to the continuous variable f (from kf_0) as we go from the series to the transform.

(b) Suppose the Fourier transform of $x(t) = \text{tri}(t)$ is $X(f) = \text{sinc}^2(f)$. The Fourier series coefficients of its corresponding periodic extension $x_p(t)$ with period T are

$$X[k] = \frac{X(kf_0)}{T} = \frac{\text{sinc}^2(kf_0)}{T}$$

The signals $x_p(t)$ for three choices of T are shown in Figure E9.1B and yield:

For $x_1(t)$ with $T = 2$: A triangular wave with $X_1[k] = \frac{1}{2}\,\text{sinc}^2(\frac{k}{2})$

For $x_2(t)$ with $T = 1$: A constant with $X_2[k] = \text{sinc}^2(k)$ or $X_2[0] = 1$ and $X_2[k] = 0,\ k \neq 0$

For $x_3(t)$ with $T = 1.5$: A triangular form with $X_3[k] = \frac{2}{3}\,\text{sinc}^2(\frac{2}{3}k)$

PGR2026-00033
Krisp Technologies EX1024 Page 264

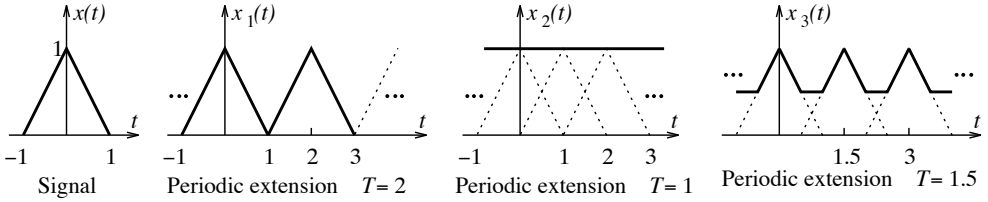


Figure E9.1B Periodic extensions of $x(t) = \text{tri}(t)$ for Example 9.1(b)

9.1.3 Magnitude and Phase Spectra

The Fourier transform $X(f)$ is, in general, complex and may be represented in any of the following forms:

$$X(f) = \text{Re}\{X(f)\} + j\text{Im}\{X(f)\} = |X(f)|\angle\phi(f) = |X(f)|e^{j\phi(f)} \tag{9.10}$$

For real signals, $X(f)$ is *conjugate symmetric* with $X(-f) = X^*(f)$. This means that the magnitude $|X(f)|$ or $\text{Re}\{X(f)\}$ displays even symmetry and the phase $\phi(f)$ or $\text{Im}\{X(f)\}$ displays odd symmetry. It is customary to plot the magnitude and phase of $X(f)$ as two-sided functions.

REVIEW PANEL 9.3

Nonperiodic Signals Have Continuous Spectra

The spectra can be plotted against f (Hz) or ω (radians/s).
Real signals: **Magnitude** spectrum has *even symmetry*; **phase** spectrum has *odd symmetry*.

The phase spectrum may be restricted to values in the *principal range* $(-\pi, \pi)$. Sometimes, it is more convenient to *unwrap* the phase (by adding/subtracting multiples of 2π) and plot it as a *monotonic* function.

The Fourier transform $X(f)$ of a real, even symmetric signal $x(t)$ is always a real and even symmetric function of f, and of the form $X(f) = A(f)$. The Fourier transform $X(f)$ of a real, odd symmetric signal $x(t)$ is always imaginary and odd symmetric in f, and of the form $X(f) = jA(f)$. For such signals, it is convenient to plot just the **amplitude spectrum** $A(f)$.

REVIEW PANEL 9.4

Effect of Signal Symmetry on the Fourier Transform of Real-Valued Signals

Even symmetry in $x(t)$: The Fourier transform $X(f)$ is real and even symmetric.
Odd symmetry in $x(t)$: The Fourier transform $X(f)$ is imaginary and odd symmetric.
No symmetry in $x(t)$: $\text{Re}\{X(f)\}$ is even symmetric, and $\text{Im}\{X(f)\}$ is odd symmetric.

REVIEW PANEL 9.5

The Amplitude Spectrum Is Useful for Real Symmetric Signals

Even symmetry in $x(t)$: Plot $\text{Re}\{X(f)\}$ (which is even symmetric) versus f. The phase is zero.
Odd symmetry in $x(t)$: Plot $\text{Im}\{X(f)\}$ (which is odd symmetric) versus f. The phase is 90°.

PGR2026-00033
Krisp Technologies EX1024 Page 265

EXAMPLE 9.2 (Signals and Their Fourier Transform and Spectra)
Consider the signal $x(t) = \delta(t + 0.5) - \delta(t - 0.5)$. Its Fourier transform is

$$X(f) = \int_{-\infty}^{\infty} [\delta(t + 0.5) + \delta(t - 0.5)]e^{-j2\pi ft}\, dt = e^{j\pi f} - e^{-j\pi f} = j2\sin(\pi f)$$

We observe that $x(t)$ has odd symmetry and $X(f)$ is purely imaginary. The signal $x(t)$ and its amplitude spectrum $A(f) = 2\sin(\pi f)$ are plotted in the Figure E9.2A(1).

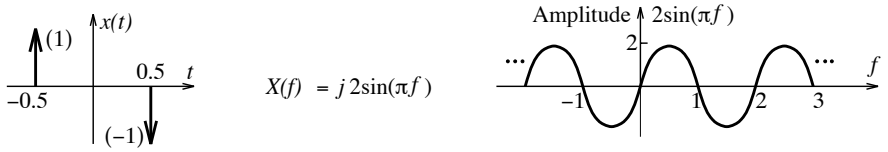


Figure E9.2A(1) The signal for Example 9.2 and its amplitude spectrum

The magnitude spectrum $|X(f)|$ and phase spectrum is sketched in Figure E9.2A(2). The sign changes in the amplitude account for the phase jumps of $\pm\pi$. The unwrapped phase is obtained by adding or subtracting multiples of 2π at the phase jumps to make the phase a monotonic function.

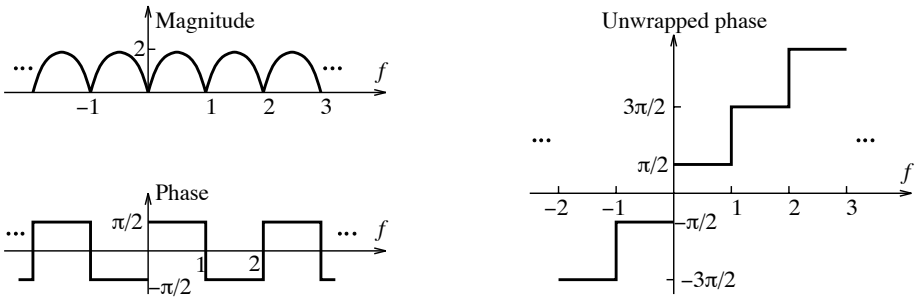


Figure E9.2A(2) Magnitude and phase spectra of the signal for Example 9.2

9.2 Fourier Transform Pairs and Properties

Table 9.1 presents the Fourier transforms of some useful signals in both the f-form and ω-form. For energy signals, the conversion from $X(f)$ to $X(\omega)$ is straightforward (with $\omega = 2\pi f$). However, the transform of many power signals or signals that are not absolutely integrable (such as the constant and the sinusoid) almost invariably includes impulses. As a result, we must also use the scaling property of impulse functions, $\delta(f) = \delta(\omega/2\pi) = 2\pi\delta(\omega)$, when converting $X(f)$ to $X(\omega)$. This is what leads to the ubiquitous factors of 2π in the Fourier transform of the constant and the sinusoid in the ω-form of Table 9.1. Finally, we remark that the transform of signals that grow exponentially or faster does not exist. The reason for this lies in the nature of convergence of the Fourier transform, which we discuss toward the end of this chapter.

REVIEW PANEL 9.6

Differences Between the f-Form and ω-Form of the Fourier Transform

If the Fourier transform contains no impulses: $H(f)$ and $H(\omega)$ are related by $\omega = 2\pi f$.
If it contains impulses: Replace $\delta(f)$ by $2\pi\delta(\omega)$ (and $2\pi f$ by ω elsewhere) to get $H(\omega)$.

PGR2026-00033
Krisp Technologies EX1024 Page 266

Table 9.1 Some Useful Fourier Transform Pairs

Entry	$x(t)$	$X(f)$	$X(\omega)$
1	$\delta(t)$	1	1
2	$\text{rect}(t)$	$\text{sinc}(f)$	$\text{sinc}\left(\frac{\omega}{2\pi}\right)$
3	$\text{tri}(t)$	$\text{sinc}^2(f)$	$\text{sinc}^2\left(\frac{\omega}{2\pi}\right)$
4	$\text{sinc}(t)$	$\text{rect}(f)$	$\text{rect}\left(\frac{\omega}{2\pi}\right)$
5	$\cos(2\pi\alpha t)$	$0.5[\delta(f+\alpha)+\delta(f-\alpha)]$	$\pi[\delta(\omega+2\pi\alpha)+\delta(\omega-2\pi\alpha)]$
6	$\sin(2\pi\alpha t)$	$j0.5[\delta(f+\alpha)-\delta(f-\alpha)]$	$j\pi[\delta(\omega+2\pi\alpha)-\delta(\omega-2\pi\alpha)]$
7	$e^{-\alpha t}u(t)$	$\frac{1}{\alpha+j2\pi f}$	$\frac{1}{\alpha+j\omega}$
8	$te^{-\alpha t}u(t)$	$\frac{1}{(\alpha+j2\pi f)^2}$	$\frac{1}{(\alpha+j\omega)^2}$
9	$e^{-\alpha\lvert t\rvert}$	$\frac{2\alpha}{\alpha^2+4\pi^2f^2}$	$\frac{2\alpha}{\alpha^2+\omega^2}$
10	$e^{-\pi t^2}$	$e^{-\pi f^2}$	$e^{-\omega^2/4\pi}$
11	$\text{sgn}(t)$	$\frac{1}{j\pi f}$	$\frac{2}{j\omega}$
12	$u(t)$	$0.5\delta(f)+\frac{1}{j2\pi f}$	$\pi\delta(\omega)+\frac{1}{j\omega}$
13	$e^{-\alpha t}\cos(2\pi\beta t)u(t)$	$\frac{\alpha+j2\pi f}{(\alpha+j2\pi f)^2+(2\pi\beta)^2}$	$\frac{\alpha+j\omega}{(\alpha+j\omega)^2+(2\pi\beta)^2}$
14	$e^{-\alpha t}\sin(2\pi\beta t)u(t)$	$\frac{2\pi\beta}{(\alpha+j2\pi f)^2+(2\pi\beta)^2}$	$\frac{2\pi\beta}{(\alpha+j\omega)^2+(2\pi\beta)^2}$
15	$\sum_{n=-\infty}^{\infty}\delta(t-nT)$	$\frac{1}{T}\sum_{k=-\infty}^{\infty}\delta\left(f-\frac{k}{T}\right)$	$\frac{2\pi}{T}\sum_{k=-\infty}^{\infty}\delta\left(\omega-\frac{2\pi k}{T}\right)$
16	$x_p(t)=\sum_{k=-\infty}^{\infty}X[k]e^{j2\pi kf_0t}$	$\sum_{k=-\infty}^{\infty}X[k]\delta(f-kf_0)$	$\sum_{k=-\infty}^{\infty}2\pi X[k]\delta(\omega-k\omega_0)$

PGR2026-00033
Krisp Technologies EX1024 Page 267

Table 9.2 Operational Properties of the Fourier Transform

Property	$x(t)$	$X(f)$	$X(\omega)$
Similarity	$X(t)$	$x(-f)$	$2\pi x(-\omega)$
Time Scaling	$x(\alpha t)$	$\frac{1}{\lvert\alpha\rvert}X\left(\frac{f}{\alpha}\right)$	$\frac{1}{\lvert\alpha\rvert}X\left(\frac{\omega}{\alpha}\right)$
Folding	$x(-t)$	$X(-f)$	$X(-\omega)$
Time Shift	$x(t-\alpha)$	$e^{-j2\pi f\alpha}X(f)$	$e^{-j\omega\alpha}X(\omega)$
Frequency Shift	$e^{j2\pi\alpha t}x(t)$	$X(f-\alpha)$	$X(\omega-2\pi\alpha)$
Convolution	$x(t)\star h(t)$	$X(f)H(f)$	$X(\omega)H(\omega)$
Multiplication	$x(t)h(t)$	$X(f)\star H(f)$	$\frac{1}{2\pi}X(\omega)\star H(\omega)$
Modulation	$x(t)\cos(2\pi\alpha t)$	$0.5[X(f+\alpha)+X(f-\alpha)]$	$0.5[X(\omega+2\pi\alpha)+X(\omega-2\pi\alpha)]$
Derivative	$x'(t)$	$j2\pi fX(f)$	$j\omega X(\omega)$
Times-t	$-j2\pi tx(t)$	$X'(f)$	$2\pi X'(\omega)$
Integration	$\int_{-\infty}^{t}x(t)\,dt$	$\frac{1}{j2\pi f}X(f)+0.5X(0)\delta(f)$	$\frac{1}{j\omega}X(\omega)+\pi X(0)\delta(\omega)$
Conjugation	$x^*(t)$	$X^*(-f)$	$X^*(-\omega)$
Correlation	$x(t)\star\star\, y(t)$	$X(f)Y^*(f)$	$X(\omega)Y^*(\omega)$
Autocorrelation	$x(t)\star\star\, x(t)$	$X(f)X^*(f)=\lvert X(f)\rvert^2$	$X(\omega)X^*(\omega)=\lvert X(\omega)\rvert^2$

Fourier Transform Theorems	
Central ordinates	$x(0)=\int_{-\infty}^{\infty}X(f)\,df=\frac{1}{2\pi}\int_{-\infty}^{\infty}X(\omega)\,d\omega \qquad X(0)=\int_{-\infty}^{\infty}x(t)\,dt$
Parseval's theorem	$E=\int_{-\infty}^{\infty}x^2(t)\,dt=\int_{-\infty}^{\infty}\lvert X(f)\rvert^2\,df=\frac{1}{2\pi}\int_{-\infty}^{\infty}\lvert X(\omega)\rvert^2\,d\omega$
Plancherel's theorem	$\int_{-\infty}^{\infty}x(t)y^*(t)\,dt=\int_{-\infty}^{\infty}X(f)Y^*(f)\,df=\frac{1}{2\pi}\int_{-\infty}^{\infty}X(\omega)Y^*(\omega)\,d\omega$

PGR2026-00033
Krisp Technologies EX1024 Page 268

The Fourier transform is a linear operation and obeys superposition. Its properties are summarized in Table 9.2 for both the f-form and ω-form. The conversion between the two forms is sometimes not so obvious and may include (or omit) factors of 2π. Our suggestion: *Use one form consistently.*

EXAMPLE 9.3 (Three Basic Fourier Transform Pairs)

(a) (The Unit Impulse) The Fourier transform of $x(t) = \delta(t)$ is found by the sifting property of impulses

$$X(f) = \int_{-\infty}^{\infty} \delta(t) e^{-j2\pi ft}\, dt = 1$$

The spectrum of an impulse is a constant for all frequencies.

(b) (The Decaying Exponential) The Fourier transform of $x(t) = e^{-\alpha t}u(t)$ is

$$X(f) = \int_0^{\infty} e^{-\alpha t} e^{-j2\pi ft}\, dt = \int_0^{\infty} e^{-(\alpha + j2\pi f)t}\, dt = \frac{1}{\alpha + j2\pi f}$$

The magnitude spectrum decays monotonically with frequency.

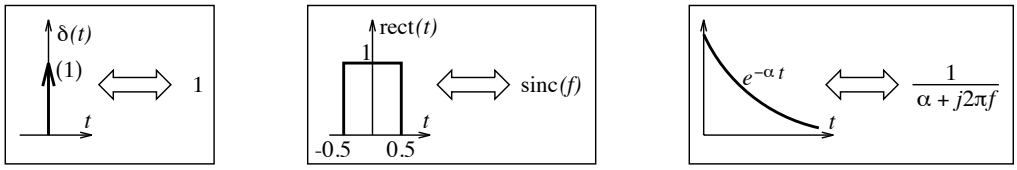
(c) (The rect Function) The signal $x(t) = \text{rect}(t)$ is unity over $(-0.5,\ 0.5)$ and zero elsewhere. We find $X(f)$ by evaluating the defining integral and using Euler's relation to give

$$X(f) = \int_{-1/2}^{1/2} e^{-j2\pi ft}\, dt = \frac{e^{-j2\pi ft}}{-j2\pi f}\bigg|_{-1/2}^{1/2} = \frac{\sin(\pi f)}{\pi f} = \text{sinc}(f)$$

The amplitude spectrum of a rectangular pulse has a sinc form.

REVIEW PANEL 9.7

Three Basic Fourier Transform Pairs



9.2.1 How the Properties Arise

In this section we show how the properties arise and illustrate the computation of Fourier transforms, using the defining relation or its properties. Many of the properties also have duals, and these are listed alongside without proof in what follows.

The **similarity theorem** is a consequence of the fact that the operations embodied in both the direct and inverse transforms involve finding the area of the product of a function and a complex exponential of the form $\pm\exp(j2\pi ft)$ and leads to the following result:

$$\text{If} \quad x(t) \Longleftrightarrow X(f) \quad \text{or} \quad X(\omega) \qquad \text{then} \qquad X(t) \Longleftrightarrow x(-f) \quad \text{or} \quad 2\pi x(-\omega) \tag{9.11}$$

PGR2026-00033
Krisp Technologies EX1024 Page 269

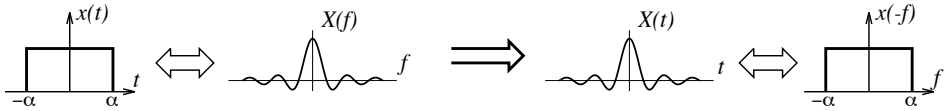
The interchange of time and frequency ($t \to -f$) also includes a sign reversal to account for the sign reversal in the exponential in the direct and inverse transforms. For even symmetric functions, we can simply use $t \to f$. For example, the transform pair $\text{rect}(t) \Leftrightarrow \text{sinc}(f)$ leads to $\text{sinc}(t) \Leftrightarrow \text{rect}(-f) = \text{rect}(f)$. Similarly, the pair $\delta(t) \Leftrightarrow 1$ gives $1 \Leftrightarrow \delta(-f) = \delta(f)$. The transform pair for the decaying exponential gives

$$e^{-\alpha t}u(t) \Longleftrightarrow \frac{1}{\alpha + j2\pi f} \quad \Rightarrow \text{ (similarity) } \Rightarrow \quad \frac{1}{\alpha - j2\pi t} \Longleftrightarrow e^{-\alpha f}u(f)$$

REVIEW PANEL 9.8

The Similarity Theorem: If $x(t) \Longleftrightarrow X(f)$, then $X(t) \Longleftrightarrow x(-f)$

Going from one domain to the other, use $f \leftrightarrow -t$ (use $f \leftrightarrow t$ for *even symmetric* time signals).



Transform pair — New transform pair by symmetry

The **time-scaling property** follows from a change of variable. For $\alpha > 0$, for example, we have

$$\int_{-\infty}^{\infty} x(\alpha t)e^{-j2\pi ft}\,dt \overset{\alpha t=\lambda}{\to} = \frac{1}{\alpha}\int_{-\infty}^{\infty} x(\lambda)e^{-j2\pi f\lambda/a}\,d\lambda = \frac{1}{\alpha}X\left(\frac{f}{\alpha}\right) \tag{9.12}$$

The scaling property is its own dual in that a scaling by α in one domain results in inverse scaling (by $\frac{1}{\alpha}$) and amplitude scaling by $\frac{1}{|\alpha|}$ in the other. For example, compression of $x(t)$ to $x(\alpha t)$ leads to a stretching of $X(f)$ by α and an amplitude reduction by $|\alpha|$. The multiplier $\frac{1}{|\alpha|}$ ensures that the scaled signal and the scaled spectrum possess the same energy.

The **folding property** follows from the scaling property with $\alpha = -1$:

$$x(-t) \Longleftrightarrow X(-f) = X^*(f) \quad \text{or} \quad X(-\omega) = X^*(\omega) \tag{9.13}$$

The **time-shift property** follows from the defining relation and the change of variable $\lambda = t - \alpha$:

$$\int_{-\infty}^{\infty} x(t-\alpha)e^{-j2\pi ft}\,dt = \int_{-\infty}^{\infty} x(\lambda)e^{-j2\pi f(\lambda+\alpha)}d\lambda = e^{-j2\pi f\alpha}X(f) \tag{9.14}$$

The quantity $e^{-j2\pi f\alpha}$ contributes a delay of α units to the time signal.

The **frequency-shift property** is the dual of the time-shift property:

$$x(t)e^{-j2\pi\alpha t} \Longleftrightarrow X(f-\alpha) \tag{9.15}$$

REVIEW PANEL 9.9

The Time-Shift and Frequency-Shift Properties Are Duals

$$x(t-\alpha) \Longleftrightarrow X(f)e^{-j2\pi f\alpha} \qquad x(t)e^{j2\pi t\alpha} \Longleftrightarrow X(f-\alpha)$$

The **convolution property** follows if we interchange the order of integration and use a change of variables ($t - \lambda = \tau$) in the defining integral. We obtain

$$\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} x(t-\lambda)u(t-\lambda)h(\lambda)d\lambda\right]e^{-j2\pi ft}\,dt = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} x(t-\lambda)e^{-j2\pi f(t-\lambda)}\,dt\right]h(\lambda)e^{-j2\pi f\lambda}\,d\lambda = X(f)H(f)$$

PGR2026-00033
Krisp Technologies EX1024 Page 270

Its dual is the **multiplication property**:

$$x(t)h(t) \Longleftrightarrow X(f) \star H(f) \qquad \text{or} \qquad \frac{1}{2\pi}\left[X(\omega) \star H(\omega)\right] \tag{9.16}$$

Convolution in one domain corresponds to multiplication in the other.

REVIEW PANEL 9.10

Convolution and Multiplication Are Duals

$$x(t) \star y(t) \Longleftrightarrow X(f)Y(f) \qquad\qquad x(t)y(t) \Longleftrightarrow X(f) \star Y(f)$$

The **modulation property** follows from $x(t)e^{j2\pi\alpha t} \Leftrightarrow X(f-\alpha)$ (frequency shift) and Euler's relation:

$$x(t)\cos(2\pi\alpha t) = 0.5x(t)[e^{-j2\pi\alpha t} + e^{-j2\pi\alpha t}] \Longleftrightarrow 0.5[X(f+\alpha) + X(f-\alpha)] \tag{9.17}$$

It is a special case of the frequency-domain convolution property. If a signal $x(t)$ is modulated by $\cos(2\pi\alpha t)$, its spectrum $X(f)$ gets shifted to higher frequencies and centered at $f = \pm\alpha$.

REVIEW PANEL 9.11

Modulation Spreads the Signal Spectrum to Higher Frequencies

$$x(t)\cos(2\pi\alpha t) \Longleftrightarrow 0.5[X(f+\alpha) + X(f-\alpha)]$$

The **derivative property** follows from the definition of $X(f)$ if we use integration by parts (assuming that $x(t) \to 0$ as $|t| \to \infty$):

$$\int_{-\infty}^{\infty} x'(t)e^{-j2\pi ft}\,dt = x(t)e^{-j2\pi ft}\Big|_{-\infty}^{\infty} + j2\pi f\int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}\,dt = j2\pi fX(f) \tag{9.18}$$

The dual of the derivative property is the **times-t property**, given by

$$-j2\pi tx(t) \Longleftrightarrow X'(f) \tag{9.19}$$

Multiplication by the variable in one domain results in a derivative in the other.

REVIEW PANEL 9.12

The Derivative and Times-t Properties Are Duals

$$x'(t) \Longleftrightarrow j2\pi fX(f) \qquad\qquad -j2\pi tx(t) \Longleftrightarrow X'(f)$$

The **integration property** follows from the convolution property if we describe the running integral of $x(t)$ as the convolution of $x(t)$ with $u(t)$. With $u(t) \Leftrightarrow 0.5\delta(f) + \frac{1}{j2\pi f}$ (we derive this pair later), we have

$$\int_{-\infty}^{t} x(t)\,dt = x(t) \star u(t) \Longleftrightarrow X(f)\left[0.5\delta(f) + \frac{1}{j2\pi f}\right] = 0.5X(0)\delta(f) + \frac{X(f)}{j2\pi f} \tag{9.20}$$

Since $X(0)$ equals the area (not absolute area) of $x(t)$, this relation holds only if $\int x(t)\,dt$ (or $X(0)$) is finite. The second term disappears if $\int x(t)\,dt = X(0) = 0$. If $X(0) = 0$, integration and differentiation may be regarded as inverse operations.

REVIEW PANEL 9.13

For an Energy Signal $x(t)$, the Derivative and Integration Operations Are Inverses

$$\text{If} \qquad y(t) = \frac{d\,x(t)}{dt} \Longleftrightarrow j2\pi fX(f) = Y(f) \qquad \text{then} \qquad \int_{-\infty}^{t} y(t)\,dt \Longleftrightarrow \frac{Y(f)}{j2\pi f} = X(f)$$

PGR2026-00033
Krisp Technologies EX1024 Page 271

EXAMPLE 9.4 (Some Transform Pairs Using Properties)

(a) For $x(t) = \text{tri}(t) = \text{rect}(t) \star \text{rect}(t)$, use $\text{rect}(t) \Leftrightarrow \text{sinc}(f)$ and convolution (see Figure E9.4A):

$$\text{tri}(t) = \text{rect}(t) \star \text{rect}(t) \Longleftrightarrow \text{sinc}(f)\text{sinc}(f) = \text{sinc}^2(f)$$

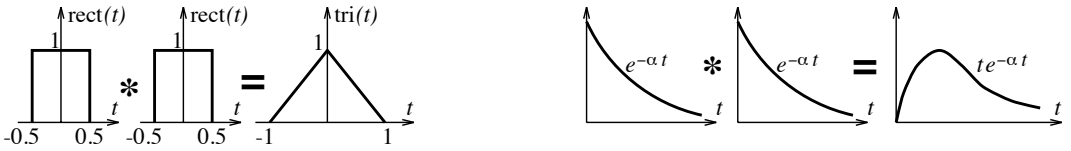


Figure E9.4A The signals for Example 9.4(a and b)

(b) For $x(t) = te^{-\alpha t}u(t)$, start with $e^{-\alpha t} \Leftrightarrow \frac{1}{\alpha + j2\pi f}$ and use convolution (see Figure E9.4A):

$$te^{-\alpha t}u(t) = e^{-\alpha t}u(t) \star e^{-\alpha t}u(t) \Longleftrightarrow \frac{1}{(\alpha + j2\pi f)^2}$$

We could also start with $e^{-\alpha t}u(t) \Leftrightarrow \dfrac{1}{\alpha + j2\pi f}$ and use the times-t property to get

$$te^{-\alpha t}u(t) \Longleftrightarrow \frac{j}{2\pi}\frac{d}{df}\left[\frac{1}{\alpha + j2\pi f}\right] = \frac{1}{(\alpha + j2\pi f)^2}$$

(c) Applying the times-t property successively to the previous result, we obtain

$$t^n e^{-\alpha t}u(t) \Longleftrightarrow \frac{n!}{(\alpha + j2\pi f)^{n+1}}$$

(d) For $x(t) = e^{-\alpha|t|} = e^{\alpha t}u(-t) + e^{-\alpha t}u(t)$, start with $e^{-\alpha t}u(t) \Leftrightarrow \dfrac{1}{\alpha + j2\pi f}$ and use the folding property and superposition:

$$e^{-\alpha|t|} \Longleftrightarrow \frac{1}{\alpha + j2\pi f} + \frac{1}{\alpha - j2\pi f} = \frac{2\alpha}{\alpha^2 + 4\pi^2 f^2}$$

(e) For $x(t) = \text{sgn}(t)$, use the limiting form for $y(t) = e^{-\alpha t}u(t) - e^{\alpha t}u(-t)$ as $\alpha \to 0$ to give

$$e^{-\alpha t}u(t) - e^{\alpha t}u(-t) \Longleftrightarrow \frac{1}{\alpha + j2\pi f} - \frac{1}{\alpha - j2\pi f} = \frac{-4j\pi f}{\alpha^2 + 4\pi^2 f^2} \qquad \text{sgn}(t) \Longleftrightarrow \frac{1}{j\pi f}$$

(f) For $x(t) = u(t)$, start with $x(t) = \int_{-\infty}^{t} y(t)\,dt$, where $y(t) = \delta(t) \Leftrightarrow 1$ and $Y(0) = 1$:

$$u(t) = \int_{-\infty}^{t} y(t)\,dt \Longleftrightarrow \frac{Y(f)}{j2\pi f} + 0.5Y(0)\delta(f) = \frac{1}{j2\pi f} + 0.5\delta(f)$$

We cannot extend this result to find the transform of $r(t)$ from $u(t)$ because $U(0) = \infty$. The Fourier transform of signals that are not absolutely integrable usually contains impulses. We could also have obtained the transform of $u(t)$ using $u(t) = 0.5 + 0.5\,\text{sgn}(t)$.

PGR2026-00033
Krisp Technologies EX1024 Page 272

(g) For $x(t) = \cos(2\pi\alpha t) = 0.5e^{j2\pi\alpha t} + 0.5e^{-j2\pi\alpha t}$, start with $1 \Leftrightarrow \delta(f)$ and use the dual of the time-shift property:

$$\cos(2\pi\alpha t) = 0.5e^{j2\pi\alpha t} + 0.5e^{-j2\pi\alpha t} \Longleftrightarrow 0.5\delta(f-\alpha) + 0.5\delta(f+\alpha)$$

Its magnitude spectrum (see Figure E9.4H(a)) is an impulse pair at $f = \pm\alpha$, with strengths of 0.5.

(h) For $x(t) = \cos(2\pi\alpha t + \theta)$, start with $\cos(2\pi\alpha t) \Leftrightarrow 0.5\delta(f-\alpha) + 0.5\delta(f+\alpha)$, and use the shifting property with $t \to t + \theta/2\pi\alpha$ (and the product property of impulses) to get

$$\cos(2\pi\alpha t + \theta) \Longleftrightarrow 0.5e^{j\theta f/\alpha}[\delta(f-\alpha) + \delta(f+\alpha)] = 0.5e^{j\theta}\delta(f-\alpha) + 0.5e^{-j\theta}\delta(f-\alpha)$$

Its magnitude spectrum is an impulse pair at $f = \pm\alpha$ with strengths of 0.5. Its phase spectrum shows a phase of θ at $f = \alpha$ and $-\theta$ at $f = -\alpha$. The spectra are shown in Figure E9.4H(b). These resemble the two-sided spectrum for its Fourier series coefficients $X[k]$, except that the magnitude spectrum is now plotted as *impulses*.

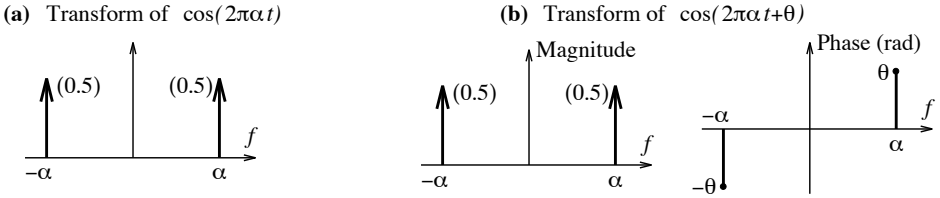


Figure E9.4H The Fourier transforms of $\cos(2\pi\alpha t)$ and $\cos(2\pi\alpha t + \theta)$

(i) For $x(t) = \cos(2\pi\alpha t)u(t)$, start with $u(t) \Leftrightarrow 0.5\delta(f) + \frac{1}{j2\pi f}$ and use modulation to give

$$\cos(2\pi\alpha t)u(t) \Longleftrightarrow 0.25[\delta(f+2\pi\alpha) + \delta(f-2\pi\alpha)] + 0.5\left[\frac{1}{j(f+2\pi\alpha)} + \frac{1}{j(f-2\pi\alpha)}\right]$$

This can be simplified further, if desired.

9.2.2 Fourier Transform of Signals Described Graphically

For signals described graphically, the Fourier transform is best evaluated using familiar transform pairs and operations. Depending on the signal given and the properties we use, we may end up with several different (but functionally equivalent) forms. We may also use the defining relation, if necessary. Whatever else you do, we caution you not to describe signals by step and ramp functions because their transform includes impulses and leads to extremely cumbersome results.

EXAMPLE 9.5 (Transform of Signals Described Graphically)
Find the Fourier transforms of the functions shown in Figure E9.5, using recognizable transform pairs and properties.

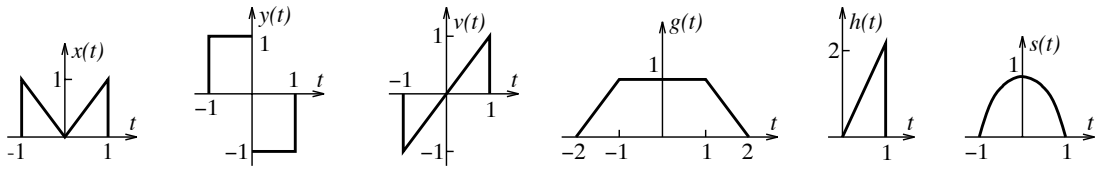


Figure E9.5 The signals for Example 9.5

PGR2026-00033
Krisp Technologies EX1024 Page 273

(a) The signal $x(t)$ is a linear combination of rect and tri functions, $x(t) = \text{rect}(t/2) - \text{tri}(t)$, as shown in Figure E9.5A.

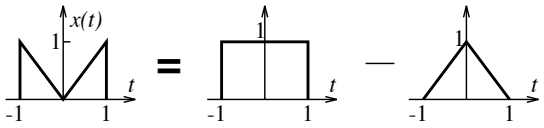


Figure E9.5A Describing the signal for Example 9.5(a)

With $\text{rect}(t) \Leftrightarrow \text{sinc}(f)$ and $\text{tri}(t) \Leftrightarrow \text{sinc}^2(f)$, we have

$$X(f) = 2\,\text{sinc}(2f) - \text{sinc}^2(f)$$

(b) The signal $y(t)$ may be regarded as the derivative of $\text{tri}(t)$, as shown in Figure E9.5B.

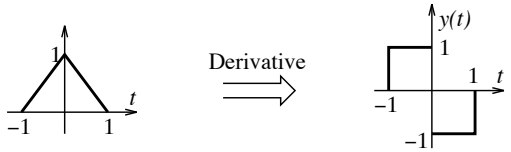


Figure E9.5B. Describing the signal for Example 9.5(b).

Its transform is thus $Y(f) = j2\pi f\,\text{sinc}^2(f)$.

We could also describe it as $y(t) = -\text{rect}(t + 0.5) + \text{rect}(t - 0.5)$ and use $\text{rect}(t) \Leftrightarrow \text{sinc}(f)$ and the shifting property to get

$$Y(f) = -\text{sinc}(f)e^{j\pi f} + \text{sinc}(f)e^{-j\pi f}$$

This result may be simplified further, if desired.

(c) The signal $v(t)$ may be described as $v(t) = t\,\text{rect}(t/2)$, as shown in Figure E9.5C.

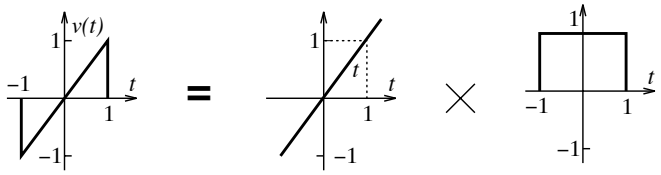


Figure E9.5C Describing the signal for Example 9.5(c)

By the times-t property, $V(f) = \dfrac{j}{2\pi}\dfrac{d}{df}\left[2\,\text{sinc}(2f)e^{j2\pi f}\right]$. This can be simplified, if required.

(d) The transform of the trapezoidal pulse $g(t)$ may be found in several ways. It can be described as the sum of three tri functions, as shown in Figure E9.5D(1).

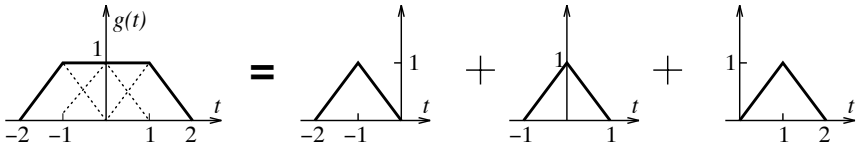


Figure E9.5D(1) First way of describing the signal for Example 9.5(d)

PGR2026-00033
Krisp Technologies EX1024 Page 274

Thus, $g(t) = \text{tri}(t+1) + \text{tri}(t) + \text{tri}(t-1)$. The pair $\text{tri}(t) \Leftrightarrow \text{sinc}^2(f)$ and the shift property gives

$$G(f) = \text{sinc}^2(f)e^{j2\pi f} + \text{sinc}^2(f) + \text{sinc}^2(f)e^{-j2\pi f} = \text{sinc}^2(f)[1 + 2\cos(\pi f)]$$

It may also be described as $g(t) = 2\,\text{tri}(t/2) - \text{tri}(t)$, as shown in Figure E9.5D(2).

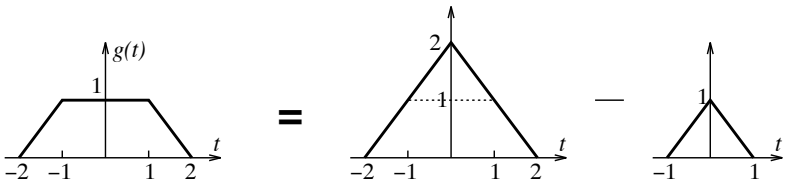


Figure E9.5D(2) Second way of describing the signal for Example 9.5(d)

We then get $G(f) = 2\,\text{sinc}^2(2f) - \text{sinc}^2(f)$

It may also be described as the convolution $g(t) = \text{rect}(t) \star \text{rect}(t/3)$, as shown in Figure E9.5D(3). We then get $G(f) = 3\,\text{sinc}(f)\text{sinc}(3f)$. All these forms are equivalent.

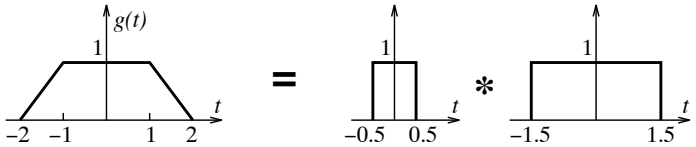


Figure E9.5D(3) Third way of describing the signal for Example 9.5(d)

(e) One way to find the transform of $h(t)$ is to regard it as the sum of two signals, as shown in Figure E9.5E(1), whose transforms we have already found.

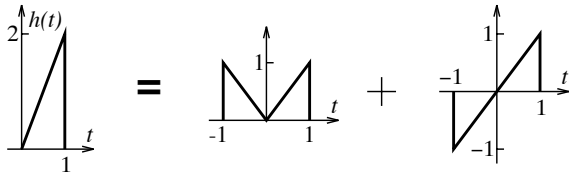


Figure E9.5E(1) Describing the signal for Example 9.5(e)

Superposition gives $H(f) = 2\,\text{sinc}(2f) - \text{sinc}^2(f) + \dfrac{j}{2\pi}\dfrac{d}{df}\left[2\,\text{sinc}(2f)e^{j2\pi f}\right]$

Another method is to take its derivative $h'(t) = 2\,\text{rect}(t-0.5) - 2\delta(t-0.5)$, as shown in Figure E9.5E(2). Since $h'(t) \Leftrightarrow j2\pi fH(f)$, we get $j2\pi fH(f) = 2\,\text{sinc}(f)e^{-j\pi f} - 2e^{-j\pi f}$, or $H(f) = \dfrac{\text{sinc}(f) - 1}{j\pi f}e^{-j\pi f}$.

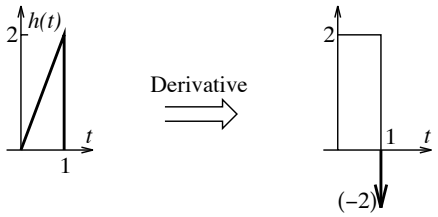


Figure E9.5E(2) Another way of describing the signal for Example 9.5(e)

We can also use the times-t property if we express the signal as $h(t) = 2t\,\text{rect}(t - 0.5)$.

PGR2026-00033
Krisp Technologies EX1024 Page 275

(f) The signal $s(t) = \cos(\pi t)\text{rect}(t)$ may be regarded as the product of $\text{rect}(t)$ and $\cos(\pi t)$, as shown in Figure E9.5F.

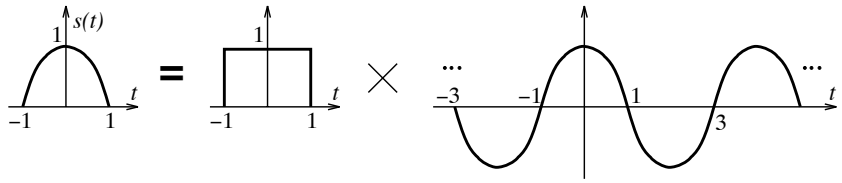


Figure E9.5F Describing the signal for Example 9.5(f)

With $\cos(2\pi\alpha t) \Leftrightarrow 0.5\delta(f+\alpha) + 0.5\delta(f-\alpha)$ and $\text{rect}(t) \Leftrightarrow \text{sinc}(f)$, the convolution property gives

$$S(f) = 0.5[\delta(f+0.5) + \delta(f-0.5)] \star \text{sinc}(f) = 0.5\,\text{sinc}(f+0.5) + 0.5\,\text{sinc}(f-0.5)$$

(g) Consider the tone-burst signal $x(t) = \cos(2\pi f_0 t)\text{rect}(t/t_0)$. To find its spectrum, we use the modulation property and the transform pair $\text{rect}(t/t_0) \Leftrightarrow t_0\,\text{sinc}(ft_0)$ to give

$$X(f) = 0.5[\delta(f+f_0) + \delta(f-f_0)] \star t_0\,\text{sinc}(ft_0) = 0.5t_0\,\text{sinc}[t_0(f+f_0)] + 0.5t_0\,\text{sinc}[t_0(f-f_0)]$$

This is just the sum of two shifted sinc functions. The signal $x(t)$ and its spectrum are shown in Figure E9.5G for $t_0 = 2\,\text{s}$ and $f_0 = 5$ Hz.

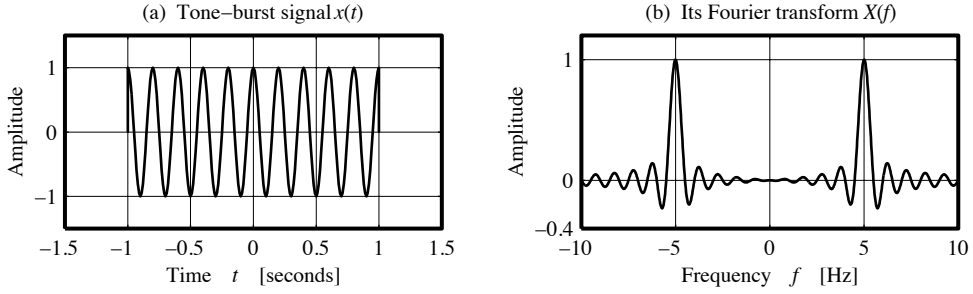


Figure E9.5G The tone-burst signal for Example 9.5(g) and its spectrum

---

9.2.3 Fourier Transform of Periodic Signals

A formal expression for the Fourier transform of a periodic signal $x_p(t)$ is based on its exponential Fourier series representation. We invoke superposition and the transform pair $e^{j2\pi k f_0 t} \Leftrightarrow \delta(f - kf_0)$ to obtain

$$x_p(t) = \sum_{k=-\infty}^{\infty} X[k]e^{j2\pi k f_0 t} \Longleftrightarrow \sum_{k=-\infty}^{\infty} X[k]\delta(f - kf_0) \tag{9.21}$$

The Fourier transform of a periodic signal is an impulse train. The impulses are located at the harmonic frequencies kf_0, and the impulse strengths equal the Fourier series coefficients $X[k]$. The impulse train is not, in general, periodic since the strengths $X[k]$ are different. Note that we can also relate the Fourier transform of a periodic signal $x_p(t)$ to the Fourier transform $X_1(f)$ of its single period $x_1(t)$ if we recognize that $TX[k] = X_1(kf_0)$. Thus,

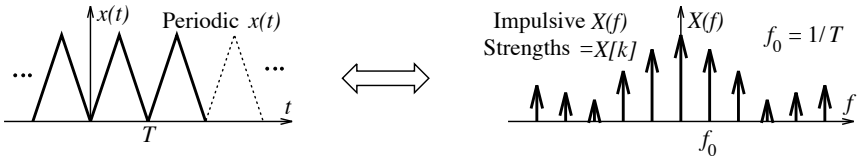
$$x_p(t) \Longleftrightarrow \sum_{k=-\infty}^{\infty} X[k]\delta(f - kf_0) = \frac{1}{T}\sum_{k=-\infty}^{\infty} X_1(kf_0)\delta(f - kf_0) \tag{9.22}$$

PGR2026-00033
Krisp Technologies EX1024 Page 276

This result demonstrates, once again, the important concept that sampling in one domain is a consequence of periodic extension in the other.

REVIEW PANEL 9.14

The Spectrum of a Periodic Signal $x_p(t)$ Is a Train of Impulses at $f = kf_0 = k/T$



The strength of the impulse at $f = kf_0$ equals the Fourier series coefficient $X[k]$.

Recall that we can also find the Fourier series coefficients of a periodic signal with period $T = 1/f_0$ directly from the Fourier transform $X(f)$ of its single period using $X[k] = \frac{1}{T}X(kf_0)$

REVIEW PANEL 9.15

Fourier Series Coefficients from the Fourier Transform $X(f)$ of One Period

$$X[k] = \tfrac{1}{T}X(f)\big|_{f \Rightarrow kf_0}, \text{ with period } T = 1/f_0$$

EXAMPLE 9.6 (Fourier Transform of Periodic Signals)

(a) The Fourier series coefficients of the impulse train $x(t) = \sum \delta(t-n)$ shown in Figure E9.6A are constant with $X[k] = 1$. The Fourier transform of this function is

$$X(f) = \sum_{k=-\infty}^{\infty} \delta(f-k)$$

The signal $x(t)$ and its transform $X(f)$ are identical in form.

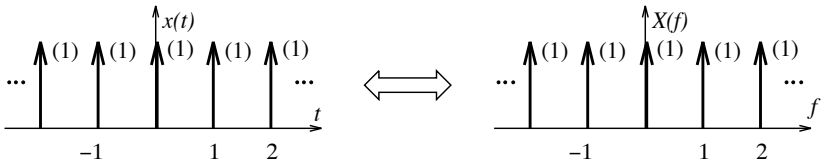


Figure E9.6A The impulse train for Example 9.6(a) and its spectrum

(b) The Fourier transform of one period $x_1(t)$ of the signal $x(t)$ shown in Figure E9.6B is given by $X_1(f) = 12\,\mathrm{sinc}(2f)$. The transform $X(f)$ of the signal $x(t)$ (with $T = 12$, and $f_0 = 1/12$) is thus

$$X(f) = \frac{1}{T}\sum_{k=-\infty}^{\infty} X_1(kf_0)\delta(f-kf_0) = \sum_{k=-\infty}^{\infty} \mathrm{sinc}(\tfrac{k}{6})\delta(f-\tfrac{k}{12})$$

Its amplitude spectrum is an impulse train that shows a sinc form.

PGR2026-00033
Krisp Technologies EX1024 Page 277

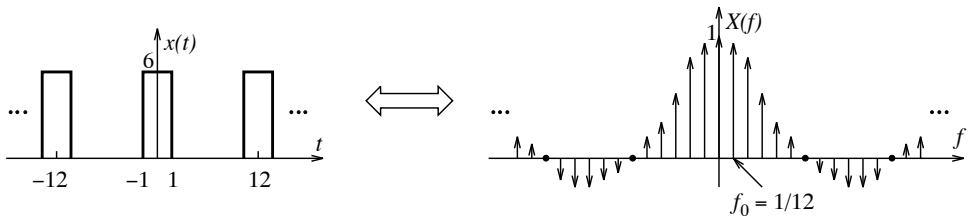


Figure E9.6B The rectangular pulse train for Example 9.6(b) and its spectrum

9.2.4 The Inverse Fourier Transform

Finding the inverse transform of $X(f)$ may be accomplished by using known transform pairs and properties to obtain the required result. It is seldom that we need to resort to the defining relation.

EXAMPLE 9.7 (Finding Inverse Fourier Transforms)

(a) Find the inverse Fourier transform of the two spectra shown in Figure E9.7A.

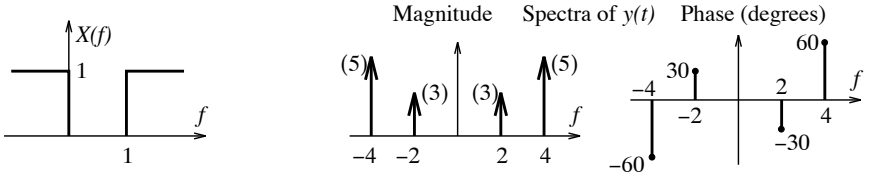


Figure E9.7A The spectra for Example 9.7(a)

For $X(f)$, we write $X(f) = 1 - \text{rect}(f - 0.5)$. Thus, its inverse is $x(t) = \delta(t) - \text{sinc}(t)e^{j\pi t}$.

For $y(t)$, each impulse pair in its magnitude spectrum corresponds to a sinusoid whose phase is read from the phase plot. We thus have $y(t) = 6\cos(4\pi t - 30^\circ) + 10\cos(8\pi t + 60^\circ)$.

(b) Find the inverse transform of $H(f) = \dfrac{j3\pi f}{1 + j\pi f}$.

Rewrite this as $H(f) = 3\dfrac{j2\pi f}{2 + j2\pi f}$.

Now, start with $g(t) = e^{-2t}u(t) \Leftrightarrow \dfrac{1}{2 + j2\pi f}$ and the times-t property $g'(t) \Leftrightarrow j2\pi f G(f)$, to get

$$h(t) = 3\frac{d}{dt}[e^{-2t}u(t)] = 3\delta(t) - 6e^{-2t}u(t)$$

Note that the impulse arises due to the jump in $e^{-2t}u(t)$ at $t = 0$.

9.2.5 Effect of Operations on the Spectrum

A time scaling (compression) of the signal $x(t)$ to $x(\alpha t)$ stretches both the magnitude and phase of $X(f)$ by α and reduces the magnitude $|X(f)|$ by $|\alpha|$. Folding a signal $x(t)$ to $x(-t)$ causes no change in the magnitude spectrum $|X(f)|$, but the phase spectrum of the folded signal shows a sign reversal. Shifting $x(t)$ to $x(t - \alpha)$

PGR2026-00033
Krisp Technologies EX1024 Page 278

causes no change in the magnitude spectrum $X(f)$ but produces a *linear phase change* (with f) in the phase spectrum. In other words, the *phase difference* $-2\pi\alpha f$ varies linearly with f. The magnitude spectrum of a differentiated signal increases in direct proportion to f. The phase is augmented by 90° for all frequencies. Since differentiation enhances sharp details and features in a time function, it follows that the sharp details in a signal are responsible for the high-frequency content of its spectrum.

Similarly, the magnitude spectrum of an integrated signal decreases in inverse proportion to f, and the phase is augmented by 90° for all frequencies. Since integration is a smoothing operation, the smoother a function, the less significant the high-frequency content in its spectrum.

Modulation of $x(t)$ by $\cos(2\pi\alpha t)$ shifts the scaled-by-half transform $0.5X(f)$ by $\pm\alpha$. However, its magnitude equals half the sum of the magnitudes $0.5|X(f+\alpha)| + 0.5|X(f-\alpha)|$ only in special cases (when $x(t)$ is band-limited to $|f| \le \alpha$, for example).

REVIEW PANEL 9.16

The Fourier Spectrum of Time-Scaled, Shifted, or Differentiated Signals

Time Scaling (stretching) a signal by α compresses its spectrum and scales the magnitude by α.
Shift of a signal by α changes *only* its phase spectrum by adding $-2\pi f\alpha$ to the phase.
Differentiating a signal scales its magnitude by $2\pi f$ and adds 90° to its phase.

EXAMPLE 9.8 (Effect of Operations on the Spectrum)
The Fourier transform of a signal $x(t)$ is $X(f) = 2\,\text{rect}(0.25f)$.

(a) Sketch the magnitude and phase spectra for $y(t) = x(2t)$.

The spectrum is compressed in frequency (by 2) and the magnitude is halved (see Figure E9.8A).

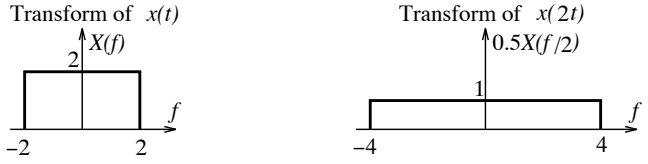


Figure E9.8A The spectrum of $x(2t)$ for Example 9.8(a)

(b) Sketch the magnitude and phase spectra for $y(t) = x(t-2)$.

Only the phase spectrum changes. Since the original phase is zero, the total phase is just $-4\pi f$, as sketched in Figure E9.8B.

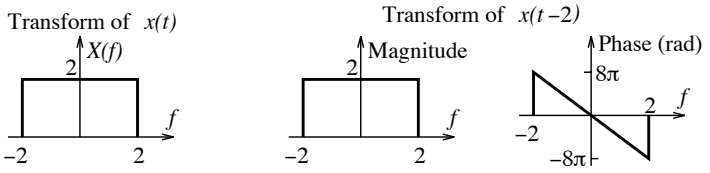


Figure E9.8B The spectrum of $x(t-2)$ for Example 9.8(b)

(c) Sketch the magnitude and phase spectra for $y(t) = x'(t)$.

The magnitude is scaled by $2\pi f$ and 90° is added to the phase, as shown in Figure E9.8C. We have plotted the magnitude and phase to ensure conjugate symmetry.

PGR2026-00033
Krisp Technologies EX1024 Page 279

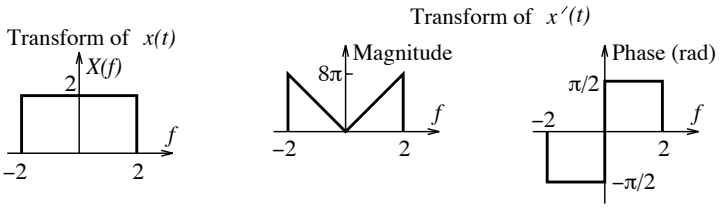


Figure E9.8C The spectrum of $x'(t)$ for Example 9.8(c)

(d) Sketch the magnitude and phase spectra for $y(t) = tx(t)$.

We use the property $-j2\pi tx(t) \Leftrightarrow X'(f)$ or $tx(t) \Leftrightarrow \frac{j}{2\pi}X'(f)$. The magnitude spectrum is differentiated and divided by 2π. A phase of $\pi/2$ is added to the already present phase (see Figure E9.8D). Since the differentiated spectrum contains impulses, the phase plot is discrete as shown, and we have plotted the magnitude and phase to ensure conjugate symmetry.

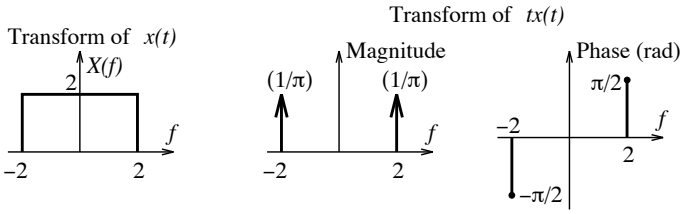


Figure E9.8D The spectrum of $tx(t)$ for Example 9.8(d)

(e) Sketch the spectra for $y(t) = x^2(t)$ and $h(t) = x(t) \star x(t)$.

For $y(t)$, the spectrum is the convolution of two rectangular pulses. The phase is zero.
For $h(t)$, the magnitude spectrum gets squared (to 4). The phase is zero (see Figure E9.8E).

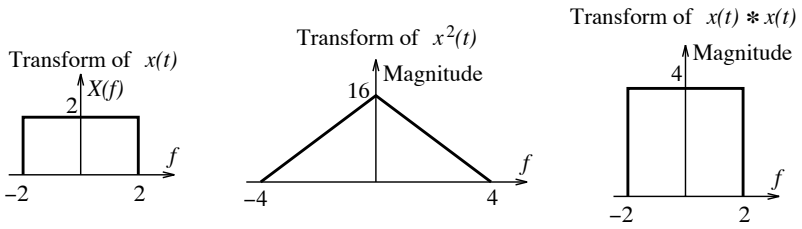


Figure E9.8E The spectra of $x^2(t)$ and $x(t) \star x(t)$ for Example 9.8(e)

(f) Sketch the spectra for $y(t) = x(t)\cos(4\pi t)$ and $h(t) = x(t)\cos(2\pi t)$.

See Figure E9.8F for the spectra of $y(t)$ and $h(t)$.
For $y(t)$, the spectrum is halved, centered at $f = \pm 2$ Hz, and added. The phase is zero.
Comment: The spectrum is identical to the transform of $x(2t)$. Does that mean $x(2t) = x(t)\cos(4\pi t)$? Of course, it does! Try it.

For $h(t)$, the spectrum is halved, centered at $f = \pm 1$ Hz, and added. The phase is zero.

PGR2026-00033
Krisp Technologies EX1024 Page 280

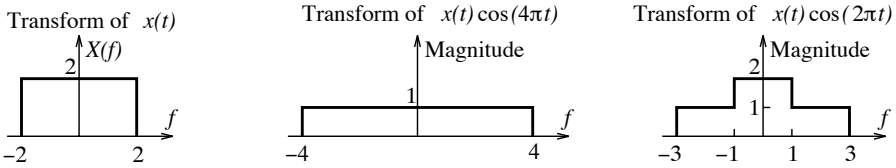


Figure E9.8F The spectra of $x(t)\cos(4\pi t)$ and $x(t)\cos(2\pi t)$ for Example 9.8(f)

9.2.6 Parseval's Theorem

The Fourier transform is an *energy-conserving* relation, and the energy may be found from the time signal $x(t)$ or its spectrum $|X(f)|$:

$$E = \int_{-\infty}^{\infty} x^2(t)\,dt = \int_{-\infty}^{\infty} |X(f)|^2\,df \tag{9.23}$$

We use $|X(f)|$ (and not $X(f)$) to compute the signal energy because of its complex nature. This relation, called **Parseval's theorem**, or **Rayleigh's theorem**, applies only to energy signals. It is the counterpart of and may be derived from Parseval's relation for the power in a periodic signal $x_p(t)$:

$$P = \frac{1}{T}\int_{-T/2}^{T/2} x_p^2(t)\,dt = \sum_{k=-\infty}^{\infty} |X[k]|^2 \tag{9.24}$$

Since $x(t)$ corresponds to one period of $x_p(t)$ as $T \to \infty$, the energy in $x(t)$ may be expressed as

$$E = \int_{-\infty}^{\infty} x^2(t)\,dt = \lim_{T\to\infty}\int_{-T/2}^{T/2} x_p^2(t)\,dt = \lim_{T\to\infty}\sum_{k=-\infty}^{\infty} T|X[k]|^2 = \lim_{T\to\infty}\sum_{k=-\infty}^{\infty} |TX[k]|^2 f_0 \tag{9.25}$$

As $T \to \infty$, $f_0 \to df$ and $|TX[k]| \to |X(f)|$, and we obtain

$$E = \int_{-\infty}^{\infty} |X(f)|^2\,df \tag{9.26}$$

Once again, if we use $\omega = 2\pi f$ instead of f, we get

$$E = \int_{-\infty}^{\infty} x^2(t)\,dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} |X(\omega)|^2\,d\omega = \frac{1}{\pi}\int_{0}^{\infty} |X(\omega)|^2\,d\omega \qquad \text{(the } \omega\text{-form)} \tag{9.27}$$

REVIEW PANEL 9.17

Parseval's Relation: The Energy Can Be Found from $x(t)$ or Its Spectrum $|X(f)|$

$$E = \int_{-\infty}^{\infty} x^2(t)\,dt = \int_{-\infty}^{\infty} |X(f)|^2\,df$$ **Note:** $|X(f)|$ is the *magnitude spectrum*

.

PGR2026-00033
Krisp Technologies EX1024 Page 281

EXAMPLE 9.9 (Parseval's Theorem)

(a) What is the signal energy of $x(t) = e^{-\alpha t}u(t)$ in the frequency band $-0.5 \le \omega \le 0.5$ rad/s?

The signal energy is $E = \displaystyle\int_{-\infty}^{\infty} x^2(t)\,dt = \frac{1}{2\alpha}$.

The magnitude spectrum is $|X(\omega)| = \left|\dfrac{1}{\alpha + j\omega}\right| = \dfrac{1}{(\alpha^2 + \omega^2)^{1/2}}$.

The energy E_B in the band $(-\omega_B,\ \omega_B)$ is

$$E_B = \frac{1}{\pi}\int_0^{\omega_B} \frac{1}{\alpha^2 + \omega^2}\,d\omega = \frac{\tan^{-1}(\omega/\alpha)}{\alpha\pi}\bigg|_0^{\omega_B} = \frac{\tan^{-1}(\omega_B/\alpha)}{\alpha\pi}$$

The energy E_0 in the range $|\omega| \le 0.5$ is $E_0 = \dfrac{1}{\pi\alpha}\tan^{-1}(0.5/\alpha)$.

(b) Let $x(t) = e^{-\alpha t}u(t)$. What is ω_B if the frequency band $(-\omega_B \le \omega \le \omega_B)$ is to contain 50% of the total signal energy?

We require $E_B = 0.5E$. Using the results of part (a) (with $E = \frac{1}{2\alpha}$),

$$E_B = \frac{0.25}{\alpha} = \frac{1}{\alpha\pi}\tan^{-1}(\omega_B/\alpha) \qquad \text{or} \qquad \omega_B = \alpha\tan(\pi/4) = \alpha$$

Thus, 50% of the total energy of $x(t) = e^{-\alpha t}u(t)$ is contained in the frequency range $|\omega| \le \alpha$.

(c) Evaluate the signal energy in $x(t) = 8\,\text{sinc}(4t)\cos(2\pi t)$.

With $8\,\text{sinc}(4t) \Leftrightarrow 2\,\text{rect}(f/4)$, the modulation property gives

$$X(f) = \text{rect}\Big(\frac{f-1}{4}\Big) + \text{rect}\Big(\frac{f+1}{4}\Big)$$

The magnitude spectra $|X(f)|$ and $|X(f)|^2$ are shown in Figure E9.9C.

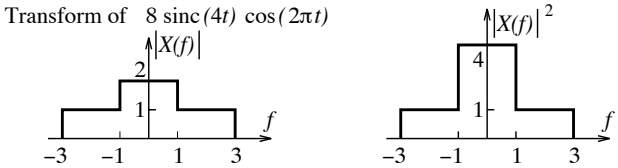


Figure E9.9C The spectra of $x(t) = 8\,\text{sinc}(4t)\cos(2\pi t)$ for Example 9.9(c)

By Parseval's relation, we find the energy as the area of $|X(f)|^2$ graphically to give $E = 12$ J.

(d) What fraction of the total energy is contained in the central lobe of the spectrum of $x(t) = \text{rect}(t)$?

The total energy in $x(t)$ is simply $E = \int x^2(t)\,dt = 1$.

Since $X(f) = \text{sinc}(f)$, whose central lobe extends over $|f| \le 1$, we seek

$$E_B = \int_{-1}^{1} \text{sinc}^2(f)\,df = 2\int_0^1 \text{sinc}^2(f)\,df$$

This integral can only be evaluated numerically and yields $E_B = 0.9028$ J. Thus, about 90% of the energy is concentrated in the central lobe. This result is true of any rectangular pulse shape.

PGR2026-00033
Krisp Technologies EX1024 Page 282

9.2.7 Other Fourier Transform Theorems

An extension of Parseval's relation, called **Plancherel's relation**, applies to the product of two energy signals and reads

$$\int_{-\infty}^{\infty} x_1(t)x_2^*(t)\,dt = \int_{-\infty}^{\infty} X_1(f)X_2^*(f)\,df \quad \text{or} \quad \frac{1}{2\pi}\int_{-\infty}^{\infty} X_1(\omega)X_2^*(\omega)\,d\omega \tag{9.28}$$

REVIEW PANEL 9.18

Plancherel's Relation Generalizes Parseval's Relation to Two Energy Signals

$$\int_{-\infty}^{\infty} x(t)y^*(t)\,dt \;=\; \int_{-\infty}^{\infty} X(f)Y^*(f)\,df$$

The **central ordinate theorems** follow directly from the defining relations for $x(t)$ and $X(f)$ by setting $t = 0$ or $f = 0$.

$$X(0) = \int_{-\infty}^{\infty} x(t)\,dt \qquad x(0) = \int_{-\infty}^{\infty} X(f)\,df \quad \text{or} \quad \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)\,d\omega \tag{9.29}$$

For real signals, we need use only $\text{Re}\{X(f)\}$ to find $x(0)$. For such signals, $\text{Im}\{X(f)\}$ is odd and integrates to zero between $(-\infty, \infty)$. If $x(t)$ possesses a discontinuity at the origin, we actually obtain $x(0)$ as the *average value* at the discontinuity. The central ordinate theorems form a useful check in problem solving.

REVIEW PANEL 9.19

The Central Ordinate Theorems Are Duals

$$X(0) \;=\; \text{area of } x(t) = \int_{-\infty}^{\infty} x(t)\,dt \qquad\qquad x(0) \;=\; \text{area of } X(f) = \int_{-\infty}^{\infty} X(f)\,df$$

The **time-limited/band-limited theorem** asserts that *no signal can be both time-limited and band-limited simultaneously*. A time-limited function $x_L(t)$ may be regarded as the product of an infinite-duration function $x(t)$ and a rect function that restricts its view over a finite duration. The spectrum $X_L(f)$ is just the convolution of $X(f)$ and a sinc function of infinite duration and is thus also of infinite extent. Likewise, a band-limited transform (confined to a finite frequency range $|f| < f_B$) corresponds to a time signal of infinite extent. Of course, a signal may be of infinite extent in both domains.

REVIEW PANEL 9.20

The Time-Limited/Band-Limited Theorem

No signal can be simultaneously time-limited and have a band-limited spectrum.

EXAMPLE 9.10 (Fourier Transform Theorems)

(a) Evaluate the integrals $I = \int_{-\infty}^{\infty} \text{sinc}^2(f)e^{-j\pi f/2}\,df$ and $J = \int_{-\infty}^{\infty} \text{sinc}(f)e^{-j\pi f}\,df$.

The integrand in I is the Fourier transform of $x(t) = \text{tri}(t - 0.25)$.
The integrand in J is the Fourier transform of $y(t) = \text{rect}(t - 0.5)$.
The signals $x(t)$ and $y(t)$ are sketched in Figure E9.10A.

PGR2026-00033
Krisp Technologies EX1024 Page 283

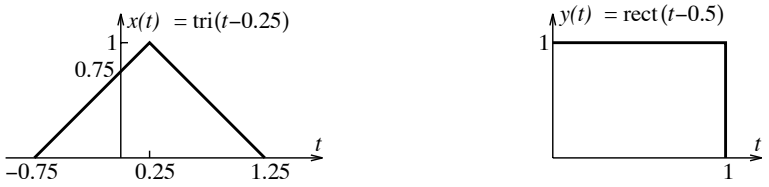


Figure E9.10A The spectra for Example 9.10(a)

By the central ordinate theorem, we find $I = \int X(f)\,df = x(0) = 0.75$ (from the figure).
Similarly, $J = \int Y(f)\,df = y(0) = 0.5$ (from the figure). Note that $y(0)$ equals the average of the discontinuity in $y(t)$ at $t = 0$. You can confirm this by direct (but tedious) evaluation of the integral.

(b) Refer to the system shown in Figure E9.10B. What is the value of $Y(0)$?

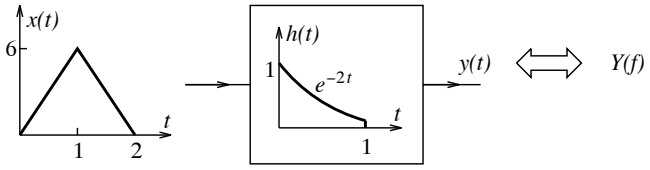


Figure E9.10B The system for Example 9.10(b)

The output $y(t)$ is the convolution of $x(t)$ and $h(t)$. By the central ordinate theorem, $Y(0)$ equals the area of $y(t)$. The area property of convolution says that the area of $y(t)$ equals the product of the area of $x(t)$ (which is 6) and the area of $h(t)$ (which is 0.5). Thus, $Y(0) = (6)(0.5) = 3$.

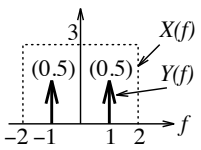
(c) Evaluate $I = \displaystyle\int_{-\infty}^{\infty} \text{sinc}^4(f)\,df$.

With $x(t) = \text{tri}(t) \Leftrightarrow X(f) = \text{sinc}^2(f)$, Parseval's theorem suggests that the integral I corresponds to the energy in tri(t) (a triangular pulse of width 2 and unit height) and thus $I = \frac{2}{3}$.

(d) Evaluate $I = \displaystyle\int_{-\infty}^{\infty} 12\,\text{sinc}(4t)\cos(2\pi t)\,dt$.

Let $x(t) = 12\,\text{sinc}(4t)$ and $y^*(t) = y(t) = \cos(2\pi t)$.
Then $X(f) = 3\,\text{rect}(f/4)$ and $Y(f) = Y^*(f) = 0.5[\delta(f-1) + \delta(f+1)]$.
Since $I = \int x(t)y^*(t)\,dt = \int X(f)Y^*(f)\,df$, we use Plancherel's theorem to obtain

$$I = \int_{-\infty}^{\infty} X(f)Y^*(f) = \int_{-\infty}^{\infty} 1.5[\delta(f-1) + \delta(f+1)]\,df = 3$$

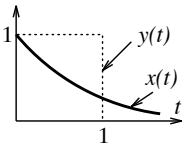


PGR2026-00033
Krisp Technologies EX1024 Page 284

(e) Evaluate $I = \displaystyle\int_{-\infty}^{\infty} \frac{\text{sinc}(f)e^{j\pi f}}{1+j2\pi f}\,df$.

Let $X(f) = \dfrac{1}{1+j2\pi f}$ and $Y^*(f) = \text{sinc}(f)e^{j\pi f}$ or $Y(f) = \text{sinc}(f)e^{-j\pi f}$.
Then $x(t) = e^{-t}u(t)$ and $y(t) = y^*(t) = \text{rect}(t-0.5)$.
Since $I = \int X(f)Y^*(f)\,df = \int x(t)y^*(t)\,dt$, we use Plancherel's relation to obtain

$$I = \int_{-\infty}^{\infty} x(t)y^*(t)\,dt = \int_0^1 e^{-t}\,dt = 1 - e^{-1} = 0.6321$$



9.2.8 Existence, Convergence, and Uniqueness

The *Dirichlet conditions* for Fourier series establish sufficient conditions for the existence of Fourier series and are based on one period of a periodic signal. The same conditions for existence are also valid for Fourier transforms, provided they are now applied to the entire extent of a nonperiodic signal. These conditions may be restated as follows: A function $x(t)$ will possess a unique Fourier transform if

1. $x(t)$ is absolutely integrable.
2. $x(t)$ has a finite number of maxima and minima and a finite number of finite discontinuities.

If the Dirichlet conditions are satisfied, then, as $f \to \infty$, the Fourier transform of $x(t)$ will converge to

1. $x(t)$ wherever it is continuous and piecewise differentiable (uniform convergence) or
2. $0.5[x(t^-) + x(t^+)]$ (the midpoint) at each jump discontinuity (nonuniform convergence).

If, in addition, $x(t)$ is an energy signal, the transform will converge to $x(t)$ in a mean squared sense, and the mean squared error will approach zero as $f \to \infty$ (as we include more frequencies to reconstruct $x(t)$).

The Dirichlet conditions impose much more severe restrictions for the existence of Fourier transforms because they apply to a function *over its entire extent*. They guarantee a Fourier transform if satisfied. However, they are only *sufficient* conditions and do not preclude the existence of transforms for functions if violated. For example, the signals $u(t)$, $r(t)$, and $\cos(2\pi f_0 t)$ are not absolutely integrable but still possess a Fourier transform (that almost invariably contains impulses). We can state that the Fourier transform of absolutely integrable signals always exists, the Fourier transform of signals that are not absolutely integrable but do not grow exponentially almost always includes impulses, and the Fourier transform of signals that grow exponentially or faster (such as $e^{\pm\alpha t}$ or $e^{\pm\alpha t^2}$) does not exist.

9.3 System Analysis Using the Fourier Transform

Conceptually, the Fourier transform can be used for the analysis of relaxed LTI systems by resorting to the fact that the convolution operation in time corresponds to multiplication in frequency.

9.3.1 The Transfer Function

The response $y(t)$ of a relaxed system (with zero initial conditions) whose impulse response is $h(t)$ to an input $x(t)$ is the convolution $y(t) = x(t) \star h(t)$. In the frequency domain, this relation translates to

$$Y(f) = X(f)H(f) \quad \text{or} \quad Y(\omega) = X(\omega)H(\omega) \tag{9.30}$$

PGR2026-00033
Krisp Technologies EX1024 Page 285

The quantity $H(f)$ or $H(\omega)$ defines the system **transfer function**. The transfer function also equals the Fourier transform of the impulse response $h(t)$. The input-output relations in the time domain and frequency domain are illustrated in Figure 9.1.

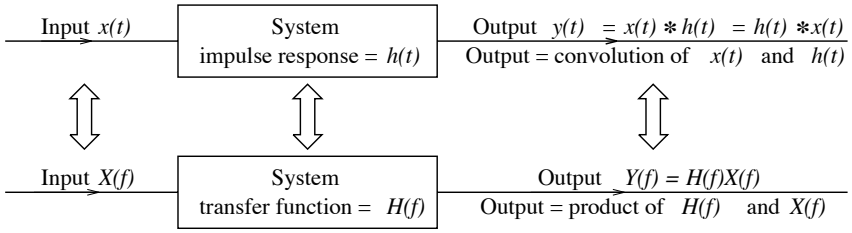


Figure 9.1 System input-output relations in the time domain and frequency domain

REVIEW PANEL 9.21

The Transfer Function Is a Frequency-Domain Description of a Relaxed LTI System

The transfer function $H(f)$ equals the transform of the system impulse response $h(t)$.
It is also defined as $H(f) = Y(f)/X(f)$, the ratio of the *transformed output* and *transformed input*.

A relaxed LTI system is also described by a differential equation of the form

$$A_n\frac{d^n y}{dt^n} + \cdots + A_2\frac{d^2 y}{dt^2} + A_1\frac{dy}{dt} + A_0 y = B_m\frac{d^m x}{dt^m} + \cdots + B_2\frac{d^2 x}{dt^2} + B_1\frac{dx}{dt} + B_0 x \tag{9.31}$$

If $x(t) \Leftrightarrow X(\omega)$ and $y(t) \Leftrightarrow Y(\omega)$, the Fourier transform and the derivative property (in the ω-form) gives

$$[(j\omega)^n A_n + \cdots + (j\omega)^2 A_2 + (j\omega)A_1 + A_0]Y(\omega) = [(j\omega)^m B_m + \cdots + (j\omega)^2 B_2 + (j\omega)B_1 + B_m]X(\omega) \tag{9.32}$$

The transfer function $H(\omega)$ is then given by the ratio $Y(\omega)/X(\omega)$ as

$$H(\omega) = \frac{Y(\omega)}{X(\omega)} = \frac{B_m(j\omega)^m + \cdots + B_2(j\omega)^2 + B_1(j\omega) + B_0}{A_n(j\omega)^n + \cdots + A_2(j\omega)^2 + A_1(j\omega) + A_0} \tag{9.33}$$

For an LTI system, the transfer function is a ratio of polynomials in $j\omega$ (or $j2\pi f$).

9.3.2 System Representation in Various Forms

A relaxed LTI system may be described in various forms: by its differential equation, by its impulse response, or by its transfer function. Given one form, it is fairly easy to obtain the others using time-domain methods and/or frequency-domain transformations. For example, given the impulse response $h(t)$, we can find the transfer function from its Fourier transform $H(f)$. If we express the transfer function as $H(f) = Y(f)/X(f)$, cross-multiplication and inverse transformation lead to the required system differential equation.

REVIEW PANEL 9.22

Three Representations of a Relaxed LTI System

Differential equation: From $H(f) = \frac{Y(f)}{X(f)}$ by cross-multiplication and inverse transform (IFT)

Transfer function: $H(f) = \frac{Y(f)}{X(f)}$ **Impulse response:** $h(t) = \text{IFT}\{H(f)\}$

PGR2026-00033
Krisp Technologies EX1024 Page 286

EXAMPLE 9.11 (Various Ways of System Representation)

(a) Let $h(t) = e^{-t}u(t) + e^{-2t}u(t)$. Then $H(\omega) = \dfrac{1}{1+j\omega} + \dfrac{1}{2+j\omega} = \dfrac{3+2j\omega}{2+3j\omega+(j\omega)^2}$.

(b) Let $y''(t) + 3y'(t) + 2y = 2x'(t) + 3x(t)$. Then $[(j\omega)^2 + 3j\omega + 2]Y(\omega) = (2j\omega + 3)X(\omega)$.
Since $H(\omega) = \dfrac{Y(\omega)}{X(\omega)}$, we obtain $H(\omega) = \dfrac{3+2j\omega}{2+3j\omega+(j\omega)^2}$.

(c) Let $H(\omega) = \dfrac{Y(\omega)}{X(\omega)}$. Then cross-multiplication gives $[(j\omega)^2 + 3j\omega + 2]Y(\omega) = (2j\omega + 3)X(\omega)$.
Its inverse transform gives $y''(t) + 3y'(t) + 2y(t) = 2x'(t) + 3x(t)$.

9.3.3 The Transfer Function of Electric Circuits

Given an electric circuit, we can find its transfer function in one of two ways. We can set up its differential equation using Kirchhoff's laws and transform it. A simpler method is to start with the transform of the constitutive relations for the circuit elements themselves. This leads to the concept of generalized impedance, as follows:

$$v_R(t) = Ri_R(t) \qquad V_R(\omega) = RI_R(\omega) \qquad Z_R = \frac{V_R(\omega)}{I_R(\omega)} = R$$

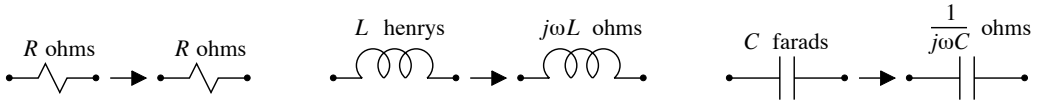
$$v_L(t) = L\frac{d\,i_L(t)}{dt} \qquad V_L(\omega) = j\omega LI_L(\omega) \qquad Z_L = \frac{V_L(\omega)}{I_L(\omega)} = j\omega L$$

$$i_C(t) = C\frac{d\,v_C(t)}{dt} \qquad I_C(\omega) = j\omega CV_C(\omega) \qquad Z_C = \frac{V_C(\omega)}{I_C(\omega)} = \frac{1}{j\omega C}$$

The impedances Z_R, Z_C, and Z_L have units of ohms. We can transform a circuit directly to the frequency domain if we replace circuit elements by their impedances, and sources and system variables by their Fourier transforms. Equations based on Kirchhoff's laws assume a simpler algebraic form and allow the computation of the transfer function or system response.

REVIEW PANEL 9.23

Transformation of Circuit Elements to the Frequency Domain ($\omega = 2\pi f$)



EXAMPLE 9.12 (Transfer Function of Electric Circuits)

Find the transfer function $H(\omega)$ of the RLC circuit shown in Figure E9.12, assuming $i(t)$ as the output.

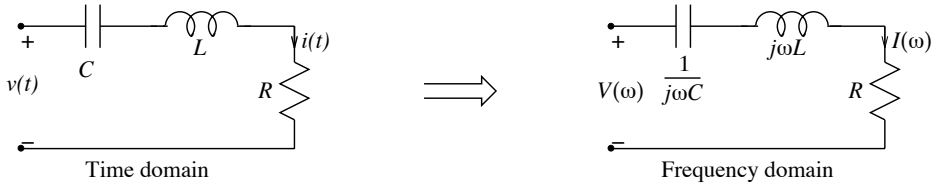


Figure E9.12 The circuit for Example 9.12

PGR2026-00033
Krisp Technologies EX1024 Page 287

Let $v(t) = \delta(t)$. Then $V(\omega) = 1$. We transform the circuit, as shown in Figure E9.12, and find

$$H(\omega) = \frac{I(\omega)}{V(\omega)} = \frac{1}{2 + j\omega + 3/j\omega} = \frac{j\omega}{3 + 2j\omega + (j\omega)^2}$$

9.3.4 The Transfer Function and System Analysis

The transfer function $H(\omega)$ allows evaluation of the zero-state response of a relaxed LTI system. It is also referred to as the **frequency response** of the system since a plot (or expression) of $|H(\omega)|$ (its magnitude or **gain**) and $\angle H(\omega)$ (its phase) versus ω provides an indication of its magnitude and phase spectrum in the frequency domain. We point out that although the Fourier transform is ideally suited for signal analysis, it is not as well suited to system analysis for several reasons. We are limited only to relaxed systems (zero initial conditions), inputs that are not energy signals usually involve impulses, and some signals (such as growing exponentials) cannot be handled at all. The Laplace transform (discussed in later chapters), which is an extension of the Fourier transform, overcomes all these limitations and is the method of choice for system analysis in the transformed domain. It would be fair to say that the *Fourier transform is better suited for signal analysis, whereas the Laplace transform is better suited for system analysis*.

REVIEW PANEL 9.24

Finding the Zero-State Response of LTI Systems

Input: $x(t) \Longleftrightarrow X(f)$ **Transfer function:** Evaluate $H(f)$ from $h(t)$ or system equation.
Zero-state output: Evaluate $Y(f) = X(f)H(f)$ and find its inverse Fourier transform $y(t)$.

EXAMPLE 9.13 (System Response Using Fourier Transforms)

(a) An LTI system is described by $H(\omega) = \dfrac{4}{2 + j\omega}$.

Find its response $y(t)$ if the input is $x(t) = 3e^{-2t}u(t)$.

The response $Y(\omega)$ is given by

$$Y(\omega) = H(\omega)X(\omega) = Y(\omega) = \left(\frac{4}{2 + j\omega}\right)\left(\frac{3}{2 + j\omega}\right) = \frac{12}{(2 + j\omega)^2}$$

Inverse transformation then yields $y(t) = 12te^{-2t}u(t)$.

(b) An LTI system is described by $H(f) = \dfrac{4}{2 + j2\pi f}$.

Find its response $y(t)$ if the input is $x(t) = u(t)$.

The response $Y(f)$ is given by

$$Y(f) = H(f)X(f) = \left[\frac{4}{2 + j2\pi f}\right]\left[\frac{1}{j2\pi f} + 0.5\delta(f)\right]$$

We separate terms, use the product property of impulses, $f(x)\delta(x) = f(0)\delta(x)$, and simplify:

$$Y(f) = \frac{2}{j\pi f(2 + j2\pi f)} + \frac{2\delta(f)}{2 + j2\pi f} = \frac{2}{j\pi f(2 + j2\pi f)} + \delta(f)$$

PGR2026-00033
Krisp Technologies EX1024 Page 288

The first term has no recognizable inverse, but (using partial fractions) we can write it as

$$\frac{2}{j\pi f(2+j2\pi f)} = \frac{1}{j\pi f} - \frac{2}{2+j2\pi f}$$

The response $Y(f)$ now equals

$$Y(f) = \frac{1}{j\pi f} - \frac{2}{2+j2\pi f} + \delta(f) = 2\left[\frac{1}{j2\pi f} + 0.5\delta(f)\right] - \frac{2}{2+j2\pi f}$$

We recognize its inverse transform as $y(t) = 2u(t) - 2e^{-2t}u(t)$.

Comment: The process is tedious due to the presence of impulses. We inverted $\delta(f) + \frac{1}{j\pi f}$ to obtain the term $2u(t)$. Term-by-term inversion gives $y(t) = 1 + \text{sgn}(t) - 2e^{-2t}u(t)$, an equivalent result.

9.3.5 The Steady-State Response to Harmonic Inputs

The response of an LTI system to a harmonic input is also a harmonic at the input frequency and describes the *steady-state* response. The quantity $H(\omega)$ allows us to find this steady-state response and is also called the **steady-state transfer function**. The magnitude and phase of the output depend on the system transfer function $H(\omega)$ at the input frequency ω_0. To find the steady-state response $y_{ss}(t)$ to the sinusoidal input $x(t) = A\cos(\omega_0 t + \theta)$, we evaluate $H(\omega)$ at the input frequency ω_0 as $H(\omega_0) = K\angle\phi$. The steady-state output is then given by $y_{ss}(t) = KA\cos(\omega_0 t + \theta + \phi)$. In effect, the transfer function magnitude K multiplies the input magnitude A, and the transfer function phase ϕ adds to the input phase θ.

REVIEW PANEL 9.25

Finding the Steady-State Response of an LTI System to a Sinusoidal Input

Input: $x(t) = A\cos(\omega_0 t + \theta)$ **Transfer function:** Evaluate $H(\omega)$ at ω_0 as $K\angle\phi$.

Steady-state output: $y_{ss}(t) = KA\cos(\omega_0 t + \theta + \phi)$

EXAMPLE 9.14 (Steady-State Response to Harmonic Inputs)

Let $H(\omega) = \dfrac{4}{2+j2\omega}$. Let the input be $x(t)$ be $x(t) = 3\sin(2t)$. Find the *steady-state* response $y_{ss}(t)$.

We evaluate $H(\omega)$ at the input frequency $\omega = 2$ to get $H(\omega) = \dfrac{4}{2+j2} = 1.4142\angle -45°$.

The steady-state output is then $y_{ss}(t) = 3(1.4142)\sin(2t - 45°) = 4.2426\sin(2t - 45°)$.

Comment: The formal alternative is to find the *total response* and then drop the transient terms. This method is not only tedious but also requires the product property of impulses for simplification of results. Stay away from this alternative if all you need is the *steady-state* response.

9.4 Frequency Response of Filters

The frequency response $H(\omega)$ represents the transfer function for harmonic (and dc) inputs and is useful primarily for stable systems whose natural (transient) response does indeed decay with time. For LTI systems, $H(\omega)$ is a ratio of polynomials in $j\omega$ and possesses *conjugate symmetry* with $H(-\omega) = H^*(\omega)$. Its magnitude or **gain** $|H(\omega)|$ is even symmetric in ω, and its phase $\theta(\omega)$ is odd symmetric in ω.

PGR2026-00033
Krisp Technologies EX1024 Page 289

A frequency-selective filter is a device that passes a certain range of frequencies and blocks the rest. The range of frequencies passed defines the **passband**, and the range of frequencies blocked defines the **stopband**. The band-edge frequencies are called the **cutoff frequencies**. Ideally, a filter should have perfect transmission in the passband and perfect rejection in the stopband. Perfect transmission implies a transfer function with constant gain and linear phase over the passband. If the filter has a *constant gain* $|H(\omega)|$, the output is an amplitude-scaled replica of the input. Similarly, if the output undergoes a *linear phase shift with frequency*, the output is a time-shifted replica of the input. If the gain is not constant over the required frequency range, we have **amplitude distortion**. If the phase shift is not linear with frequency, we have **phase distortion** as the signal undergoes different delays for different frequencies.

The phase and group delay of a system whose transfer function is $H(\omega) = |H(\omega)|\angle\theta(\omega)$ are defined as

$$t_p = -\frac{\theta(\omega)}{\omega} \quad \text{(phase delay)} \qquad t_g = -\frac{d\,\theta(\omega)}{d\omega} \quad \text{(group delay)} \tag{9.34}$$

If $\theta(\omega)$ varies linearly with frequency, t_p and t_g are not only constant but also equal. For LTI systems (with rational transfer functions), the phase $\theta(\omega)$ is a transcendental function but *the group delay is always a rational function* of ω^2 and is much easier to work with in many filter applications.

REVIEW PANEL 9.26

Phase Delay and Group Delay

Phase delay $t_p = -\frac{\theta(\omega)}{\omega}$: The delay at a single frequency in $\cos[\omega(t - t_p)] = \cos(\omega t + \theta)$.

Group delay $t_g = -\frac{d\theta(\omega)}{d\omega}$: The delay for a group of frequencies in a signal.

9.4.1 Ideal Filters

A filter transfer function with constant gain and linear phase is referred to as **ideal** or **distortionless**. The transfer function $H_{\rm LP}(f)$ and impulse response $h_{\rm LP}(t)$ of an ideal lowpass filter (LPF) with unit gain, zero phase, and cutoff frequency f_C may be written as

$$H_{\rm LP}(f) = \text{rect}(f/2f_C) \qquad h_{\rm LP}(t) = 2f_C\,\text{sinc}(2f_Ct) \quad \text{(ideal LPF)} \tag{9.35}$$

Other filter types include highpass filters that block low frequencies (below f_C, say), bandpass filters that pass only a range of frequencies (of width $2f_C$ centered at f_0, say), and bandstop filters that block a range of frequencies (of width $2f_C$ centered at f_0, say). The magnitude spectrum of these filters is shown in Figure 9.2. The transfer function and impulse response of these filters can be generated from the lowpass form, using properties of the Fourier transform.

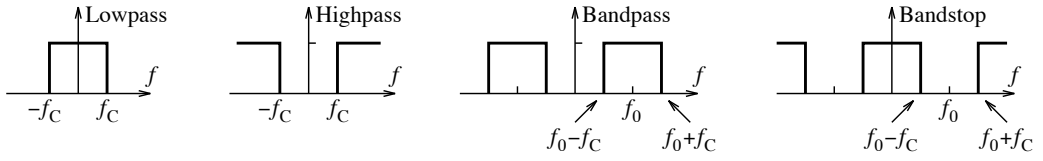


Figure 9.2 The spectrum of ideal filters

For an ideal highpass filter (HPF), we have $H_{\rm HP}(f) = 1 - H_{\rm LP}(f)$. Thus,

$$H_{\rm HP}(f) = 1 - \text{rect}(f/2f_C) \qquad h_{\rm HP}(t) = \delta(t) - 2f_C\,\text{sinc}(2f_Ct) \quad \text{(ideal HPF)} \tag{9.36}$$

PGR2026-00033
Krisp Technologies EX1024 Page 290

For an ideal bandpass filter (BPF), we use the modulation property to get

$$H_{\text{BP}}(f) = \text{rect}\left(\frac{f+f_0}{2f_C}\right) + \text{rect}\left(\frac{f-f_0}{2f_C}\right) \qquad h_{\text{BP}}(t) = 4f_C\,\text{sinc}(2f_Ct)\cos(2\pi f_0 t) \quad \text{(ideal BPF)} \tag{9.37}$$

The transfer function and impulse response of an ideal bandstop filter (BSF) follow as

$$H_{\text{BS}}(f) = 1 - \text{rect}\left(\frac{f+f_0}{2f_C}\right) + \text{rect}\left(\frac{f-f_0}{2f_C}\right) \qquad h_{\text{BS}}(t) = \delta(t) - 4f_C\,\text{sinc}(2f_Ct)\cos(2\pi f_0 t) \tag{9.38}$$

Ideal filters are unstable because their impulse response (which contains a sinc function) is not absolutely integrable. The infinite extent of the sinc function also makes such filters *noncausal* and not physically realizable. Truncating the impulse response (in an effort to make it causal and/or stable) leads to some undesirable results. One-sided truncation leads to *infinite* spikes at the band edges. Two-sided truncation produces overshoot and oscillations near band edges (Gibbs effect). Physically realizable filters cannot display a region of constant gain, or linear phase, over their passband.

In practice, we often require the step response of lowpass filters to have a monotonic form, free of overshoot. For a *monotonic* step response $s(t)$, its derivative (the impulse response) must be entirely positive ($s'(t) = h(t) > 0$). An ideal lowpass filter can never have a monotonic step response because its impulse response (a sinc function) is not entirely positive. Even though ideal filters are unrealizable, they form the yardstick by which the design of practical filters is measured.

REVIEW PANEL 9.27

Ideal Filters Are Noncausal, Unstable, and Physically Unrealizable

1. Ideal filters possess constant gain and linear phase in the passband.
2. Their impulse response (with a sinc form) makes them noncausal and unstable.
3. The step response of ideal filters cannot be monotonic and shows overshoot and *ringing*.

9.4.2 An RC Lowpass Filter

Consider the RC circuit shown in Figure 9.3. If we assume that the output is the capacitor voltage, its transfer function may be written as

$$H(f) = \frac{Y(f)}{X(f)} = \frac{1/j2\pi fC}{R + (1/j2\pi fC)} = \frac{1}{1 + j2\pi fRC} = \frac{1}{1 + j2\pi f\tau} \tag{9.39}$$

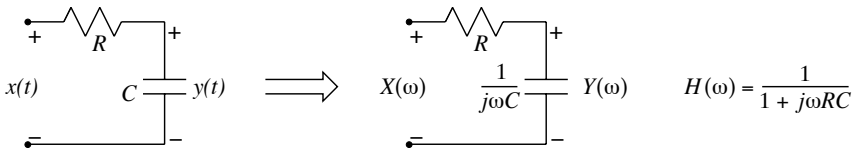


Figure 9.3 An RC lowpass filter

The quantity $\tau = RC$ is the circuit time constant. The magnitude $|H(f)|$ and phase $\theta(f)$ of the transfer function are sketched in Figure 9.4 and given by

$$|H(f)| = \frac{1}{\sqrt{1 + 4\pi^2 f^2 \tau^2}} \qquad \theta(f) = -\tan^{-1}(2\pi f\tau) \tag{9.40}$$

PGR2026-00033
Krisp Technologies EX1024 Page 291

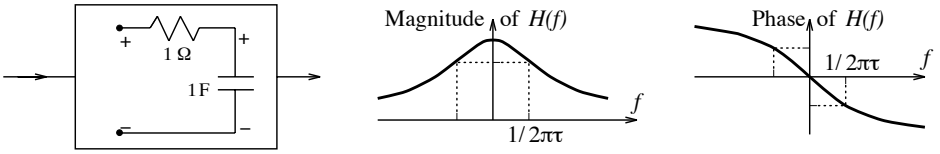


Figure 9.4 Frequency response of the RC lowpass filter

The system is called a *lowpass* filter because $|H(f)|$ decays monotonically with positive f, leading to a reduced output amplitude at higher frequencies. At $f = \frac{1}{2\pi\tau}$, the magnitude equals $1/\sqrt{2}$ (or 0.707) and the phase equals $-45°$. The frequency $f = \frac{1}{2\pi\tau}$ is called the **half-power frequency** because the output power of a sinusoid at this frequency is only half the input power. The frequency range $0 \le f \le \frac{1}{2\pi\tau}$ defines the **half-power bandwidth** over which the magnitude is less than or equal to $1/\sqrt{2}$ (or 0.707) times the peak magnitude.

The time-domain performance of this system is measured by its impulse response $h(t) = e^{-t/\tau}u(t)$, or by its step response $s(t)$, plotted in Figure 9.5, and described by

$$h(t) = e^{-t/\tau}u(t) \qquad s(t) = (1 - e^{-t/\tau})u(t) \tag{9.41}$$

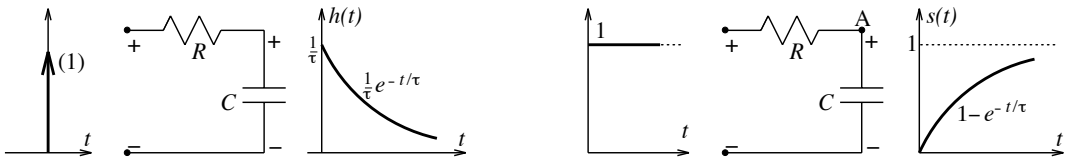


Figure 9.5 Impulse response $h(t)$ and step response $s(t)$ of the RC lowpass filter

The step response rises smoothly to unity and is within 1% of the final value in about 5τ. Other performance measures include the rise time t_r (the time taken to rise from 10% to 90% of the final value), the delay time t_d (the time taken to rise to 50% of the final value), and the settling time $t_{P\%}$ (the time taken to settle to within $P\%$ of its final value). Commonly used measures are the 5% settling time and the 2% settling time. Exact expressions for these measures are found to be

$$t_r = \tau \ln 9 \qquad t_d = \tau \ln 2 \qquad t_{5\%} = \tau \ln 20 \qquad t_{2\%} = \tau \ln 50 \tag{9.42}$$

A smaller time constant τ implies a faster rise time and a larger bandwidth. The phase delay and group delay of the RC lowpass filter are given by

$$t_p = -\frac{\theta(f)}{2\pi f} = \frac{1}{2\pi f}\tan^{-1}(2\pi f\tau) \qquad t_g = -\frac{1}{2\pi}\frac{d\,\theta(f)}{df} = \frac{\tau}{1 + 4\pi^2 f^2 \tau^2} \tag{9.43}$$

REVIEW PANEL 9.28

Frequency Response of an RC Lowpass Filter with Time Constant $\tau = RC$

$h(t) = \frac{1}{\tau}e^{-t/\tau}u(t)$ $\quad H(f) = \frac{1}{1 + j2\pi f\tau}$ $\quad$ **Half-Power (or 3-dB) frequency:** $f_H = \frac{1}{2\pi\tau}$ Hz

At f_H: Magnitude $|H(f_H)| = \frac{1}{\sqrt{2}} = 0.707$ $\quad H_{\text{dB}} = 20\log|H(f_H)| = -3$ dB $\quad$ Phase $= -45°$

PGR2026-00033
Krisp Technologies EX1024 Page 292

EXAMPLE 9.15 (Ideal and Real Filters)

(a) Consider the system of Figure E9.15A. Sketch $G(f)$ and find the system outputs $g(t)$ and $y(t)$.

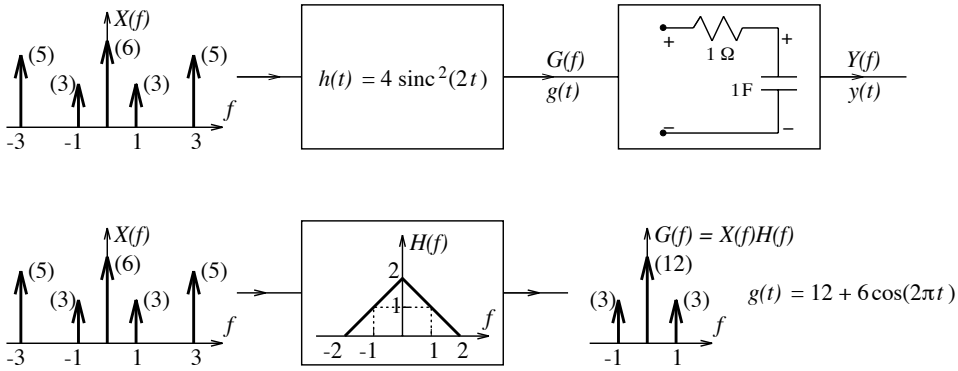


Figure E9.15A The system for Example 9.15(a)

We have $H(f) = 2\,\text{rect}(f/2)$. This produces the output $G(f) = H(f)X(f)$, as sketched. The signal $g(t)$ is thus $g(t) = 12 + 6\cos(2\pi t)$.

The transfer function of the RC filter is $T(f) = \dfrac{1}{1 + j2\pi f}$. Since the input contains components at dc ($f = 0$) and $f = 1$ Hz, we compute $T(0) = 1$ and $T(1) = \dfrac{1}{1 + j2\pi} = 0.1572\angle -80.96°$.

The output $y(t)$ is given by $y(t) = 12 + (6)(0.1572)\cos(2\pi t - 80.96°) = 12 + 0.9431\cos(2\pi t - 80.96°)$.

(b) Refer to the cascaded system of Figure E9.15B. Will the outputs $g(t)$ and $w(t)$ be equal? Explain.

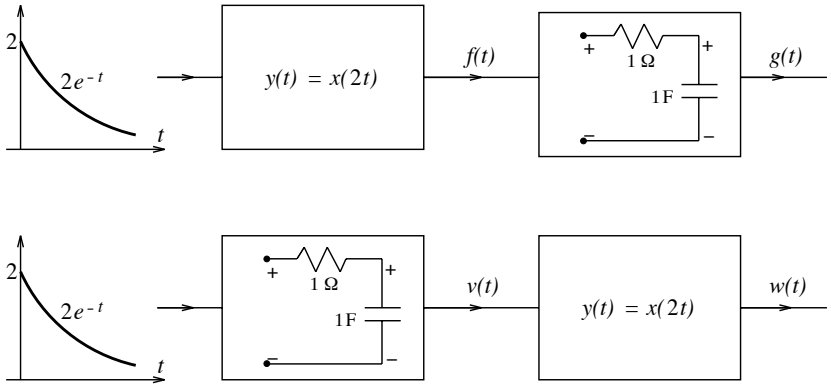


Figure E9.15B The system for Example 9.15(b)

The transfer function of the RC circuit is $T(f) = \dfrac{1}{1 + j2\pi f}$.

For the first system, the output $f(t)$ is $f(t) = 2e^{-2t}u(t)$. Using its Fourier transform gives

$$G(f) = \frac{2}{(1 + j2\pi f)(2 + j2\pi f)} = \frac{2}{1 + j2\pi f} - \frac{2}{2 + j2\pi f} \qquad g(t) = 2(e^{-t} - e^{-2t})u(t)$$

PGR2026-00033
Krisp Technologies EX1024 Page 293

For the second system, the outputs $V(f)$, $v(t)$, and $w(t)$ are

$$V(f) = \frac{2}{(1+j2\pi f)^2} \qquad v(t) = 2te^{-t}u(t) \qquad w(t) = v(2t) = 4te^{-2t}u(t)$$

Clearly, $w(t)$ and $g(t)$ are not equal. The reason is that the order of cascading is unimportant *only* for LTI systems, and the scaling block represents a time-varying system.

(c) Refer to the system shown in Figure E9.15C. Find the signal energy in $g(t)$ and $y(t)$. Also find the output $y(t)$ at $t = 0$.

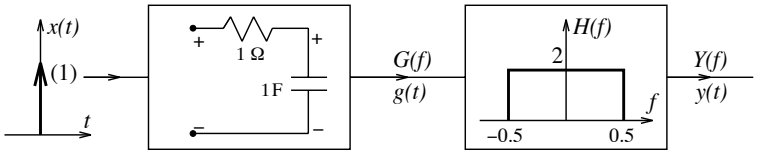


Figure E9.15C The system for Example 9.15(c)

The transform of the input is $X(f) = 1$, and the transfer function of the RC filter is $T(f) = \dfrac{1}{1+j2\pi f}$. The output of the RC filter is $G(f) = \dfrac{1}{1+j2\pi f}$. Thus, $g(t) = e^{-t}u(t)$. Its signal energy is

$$E_g = \int_0^\infty e^{-t}\,dt = 1$$

Now, $Y(f) = 2G(f),\ -0.5 \le f \le 0.5$. By Parseval's theorem, the signal energy in $y(t)$ is

$$E_y = \int_{-1/2}^{1/2} |Y(f)|^2\,df = \int_{-1/2}^{1/2} \frac{4}{1+4\pi^2 f^2}\,df = \frac{4\tan^{-1}\pi}{\pi} = 1.6076$$

By the central ordinate theorem, $y(0)$ equals the area of $Y(f)$. We need compute only the area of $\mathrm{Re}[Y(f)]$ because its odd part is imaginary and integrates to zero. Thus,

$$y(0) = \int_{-1/2}^{1/2} \mathrm{Re}[Y(f)]\,df = \int_{-1/2}^{1/2} \frac{2}{1+2\pi^2 f^2}\,df = \frac{2\tan^{-1}\pi}{\pi} = 0.8038$$

(d) (A Twin-T Notch Filter)

Analyze the twin-T notch filter shown in Figure E9.15D(1).

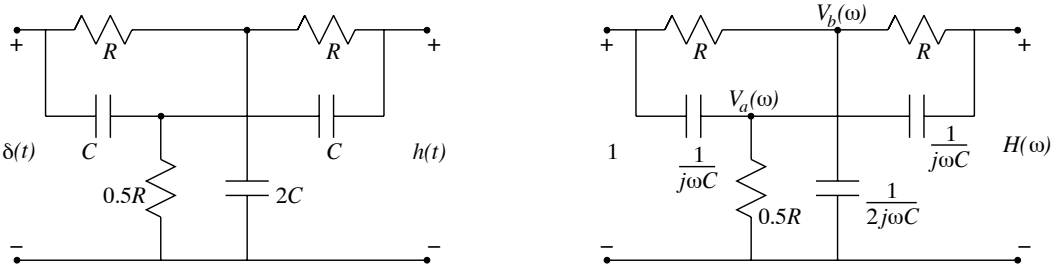


Figure E9.15D(1) The twin-T circuit for Example 9.15(d)

PGR2026-00033
Krisp Technologies EX1024 Page 294

With the input $V_1(f) = 1$, the node equations and algebraic simplification result in

$$\begin{aligned} j\omega C(V_a - 1) + \frac{2V_a}{R} + j\omega C(V_a - H) &= 0 \\ \frac{V_b - 1}{R} + \frac{V_b - H}{R} + 2j\omega CV_b &= 0 \\ \frac{H - V_b}{R} + j\omega C(H - V_a) &= 0 \end{aligned} \qquad H(\omega) = \frac{1/R^2C^2 - \omega^2}{-\omega^2 + (1/R^2C^2) + (j4\omega/RC)}$$

Note that $H(0) = 1 = H(\infty)$. Also, $H(\omega) = 0$ when $\omega = \frac{1}{RC}$. These results confirm the bandstop nature of this filter. The frequency $\omega_0 = \frac{1}{RC}$ is called the **notch frequency** or **null frequency**.

The stopband is defined as the region between the half-power frequencies ω_1 and ω_2. At the half-power frequencies, $|H(\omega)| = 1\sqrt{2}$ or $|H(\omega)|^2 = 0.5$, and we have

$$|H(\omega)|^2 = \frac{(1/R^2C^2 - \omega^2)^2}{(1/R^2C^2 - \omega^2)^2 + (16\omega^2/R^2C^2)} = 0.5$$

Rearranging and simplifying this equation, we obtain

$$\left(\frac{1}{R^2C^2} - \omega^2\right)^2 = \frac{16\omega^2}{R^2C^2} \quad \text{or} \quad \frac{1}{R^2C^2} - \omega^2 = \pm\frac{4\omega}{RC}$$

The solution of the two quadratics yields the two positive frequencies

$$\omega_1 = \omega_0(\sqrt{5} - 2) \qquad \omega_2 = \omega_0(\sqrt{5} + 2)$$

The width of the stopband thus equals $\Delta\omega = \omega_2 - \omega_1 = 4\omega_0$ rad/s. The magnitude and phase spectra of this filter are sketched in Figure E9.15D(2).

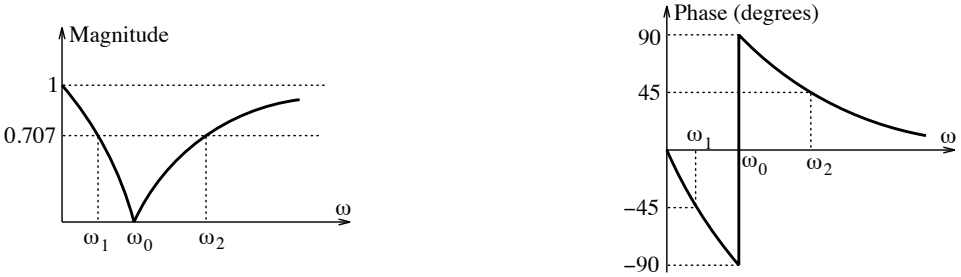


Figure E9.15D(2) The spectra of the twin-T circuit for Example 9.15(d)

The phase of $H(\omega)$ equals 0° at both $\omega = 0$ and $\omega = \infty$. There is a phase reversal from −90° to 90° at $\omega = \omega_0$. The phase equals −45° at ω_1 and +45° at ω_2.

9.4.3 The Central Limit Theorem Revisited

The central limit theorem asserts that the repeated convolution of energy signals approaches a Gaussian form. The Fourier transform of a Gaussian signal is also a Gaussian signal. Since convolution transforms to a product, the product of the transforms of n energy signals must result in a Gaussian form as $n \to \infty$.

REVIEW PANEL 9.29

The Central Limit Theorem: Frequency-Domain Version

The product of the Fourier transforms of n energy signals approaches a Gaussian shape for large n.

PGR2026-00033
Krisp Technologies EX1024 Page 295

EXAMPLE 9.16 (The Central Limit Theorem)
For a cascade of n systems, each with an identical impulse response $h(t) = e^{-t}u(t)$, the impulse response $h_n(t)$ of the composite system is just the repeated convolution of $h(t)$ leading to the exact result and its Gaussian approximation for large n and given by

$$h_n(t) = e^{-t}u(t) \star \cdots \star e^{-t}u(t) = \frac{t^2}{n!}e^{-t}u(t) \approx \frac{1}{\sqrt{2\pi n}} \exp\left[-\frac{(t-n)^2}{2n}\right]$$

Since $e^{-t}u(t) \Leftrightarrow \frac{1}{1+j2\pi f}$, the transform $H_n(f)$ of $h_n(t)$ is given by the repeated *product* of $H(f)$ with itself:

$$H_n(f) = [H(f)]^n = \frac{1}{(1+j2\pi f)^n}$$

To see that $H_n(f)$ also approaches a Gaussian form as $n \to \infty$, we take logarithms and use the expansion $\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \cdots$ to obtain

$$\ln[H_n(f)] = -n \ln(1+j2\pi f) = -j2\pi nf + \tfrac{1}{2}(2\pi f)^2 n - \tfrac{1}{3}j(2\pi f)^3 n + \cdots$$

If we retain terms only up to the second order in f (which is equivalent to assuming that f decays faster than $1/\sqrt{n}$), we have

$$\ln[H_n(f)] \approx -j2n\pi f - 2n\pi^2 f^2 \quad \text{or} \quad H_n(f) \approx e^{-2n\pi^2 f^2} e^{-j2n\pi f}$$

This describes the Fourier transform of the Gaussian approximation to $h_n(t)$. The term $e^{-j2n\pi f}$ accounts for the delay of n units.

9.5 Energy and Power Spectral Density

Parseval's theorem for energy signals tells us that the total signal energy equals the area of the squared magnitude spectrum $|X(f)|^2$. The quantity $R_{xx}(f) = |X(f)|^2$ is called the **energy spectral density** and describes the distribution of energy with frequency. The energy spectral density is measured in joules per hertz (JHz^{-1}) and may be thought of as the energy associated with a 1-Hz frequency bin centered at f hertz. The signal energy over a small frequency band Δf equals $|X(f)|^2\Delta f$. Note that $R_{xx}(f)$ is always a real, nonnegative, and even function of f.

If we express the spectral density as $R_{xx}(f) = |X(f)|^2 = X(f)X(-f)$, its inverse transform may be written as the convolution $x(t) \star x(-t)$ or the autocorrelation $r_{xx}(t) = x(t) \star\star x(-t) = x(t) \star x(-t)$. In other words, the spectral density and the autocorrelation function form a transform pair.

$$\text{(autocorrelation)} \quad r_{xx}(t) \Longleftrightarrow R_{xx}(f) \quad \text{(spectral density)} \tag{9.44}$$

This is the celebrated **Wiener-Khintchine theorem**. Application of the central ordinate theorem yields

$$r_{xx}(0) = \int_{-\infty}^{\infty} R_{xx}(f)\, df = \int_{-\infty}^{\infty} |X(f)|^2\, df = E \tag{9.45}$$

Thus, $r_{xx}(0)$ equals the energy in $x(t)$.

PGR2026-00033
Krisp Technologies EX1024 Page 296

REVIEW PANEL 9.30

The Spectral Density Is the Fourier Transform of Autocorrelation Function

For power signals (and periodic signals), we must use averaged measures consistent with power (and not energy). This leads to the concept of **power spectral density** (PSD) that corresponds to a continuous function whose area equals the total signal power. The PSD is measured in watts per hertz (WHz^{-1}) and may be thought of as the average power associated with a 1-Hz frequency bin centered at f hertz. For example, the spectrum of a periodic signal with period T is a train of impulses at the locations $f = kf_0$ (where $f_0 = 1/T$) with strengths $X[k]$. The total signal power equals $\sum |X[k]|^2$. To express its PSD as a continuous function of frequency, we describe the signal power $|X[k]|^2$ of each harmonic at $f = kf_0$ by an impulse whose strength equals $|X[k]|^2$:

$$R_{xx}(f) = \sum_{k=-\infty}^{\infty} |X[k]|^2 \delta(f - kf_0) \tag{9.46}$$

In other words, the PSD of a periodic signal is a train of impulses at $f = kf_0$ with strengths $|X[k]|^2$ whose total area (the sum of the impulse strengths) equals the total signal power.

REVIEW PANEL 9.31

The PSD of Periodic Signals Is an Impulse Train: $R_{xx}(f) = \sum_{k=-\infty}^{\infty} |X[k]|^2 \delta(f - kf_0)$

It is also possible to describe the PSD of a power signal $x(t)$ in terms of a limiting form of its truncated version $x_T(t)$, as $T \to \infty$. Truncation ensures that $x_T(t)$ represents an energy signal and its Fourier transform $X_T(f)$ yields the total energy via Parseval's energy relation:

$$\int_{-\infty}^{\infty} x_T^2(t)\, dt = \int_{-\infty}^{\infty} |X_T(f)|^2\, df \tag{9.47}$$

Using the limits $(-T, T)$ on the left-hand side allows us to replace $x_T(t)$ by its original version $x(t)$, and dividing both sides by $2T$, we get

$$\frac{1}{2T} \int_{-T}^{T} x^2(t)\, dt = \frac{1}{2T} \int_{-\infty}^{\infty} |X_T(f)|^2\, df \tag{9.48}$$

Taking limits as $T \to \infty$ and interchanging the integration and limiting operation on the right-hand side yields

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x^2(t)\, dt = \lim_{T\to\infty} \frac{1}{2T} \int_{-\infty}^{\infty} |X_T(f)|^2\, df = \int_{-\infty}^{\infty} \left[\lim_{T\to\infty} \frac{1}{2T} |X_T(f)|^2 \right] df \tag{9.49}$$

Since the left-hand side is just the signal power in $x(t)$, the integrand on the right-hand side must represent $R_{xx}(f)$ if it is to yield the power when integrated. We thus express the PSD as

$$R_{xx}(f) = \lim_{T\to\infty} \frac{|X_T(f)|^2}{2T} \tag{9.50}$$

For power signals, $r_{xx}(t)$ and $R_{xx}(f)$ form a transform pair and play a role analogous to the signal $x(t)$ and its spectrum $X(f)$. For example, $r_{xx}(0)$ equals the average power in $x(t)$. The PSD is also referred to as the **power spectrum** and is often used to characterize random signals. It is important to realize that the PSD is not a unique indicator of the underlying time signal because many time signals can yield the same power spectral density.

PGR2026-00033
Krisp Technologies EX1024 Page 297

9.5.1 White Noise and Colored Noise

A signal with zero mean whose PSD is constant (with frequency) is called **white noise**. The autocorrelation function of such a signal is an impulse. White noise implies infinite power. In theory, white noise inputs to a system are useful in characterizing system performance. If a signal $x(t)$ forms the input to an LTI system with transfer function $H(f)$, the system response is given by $Y(f) = H(f)X(f)$. This result can be expressed in terms of the PSD as

$$|Y(f)|^2 = |X(f)|^2|H(f)|^2 \tag{9.51}$$

The quantity $|H(f)|^2$ is called the power transfer function and can be computed as the output of the system if the input is white noise input with a PSD of unity. Since physical measurements are restricted to finite bandwidths, we often use inputs whose PSD is also band-limited or *colored* in some way. In particular, a signal whose PSD is constant only over a finite frequency range is called **band-limited white noise**.

REVIEW PANEL 9.32

The PSD of White Noise Is Constant. Its Autocorrelation Function Is an Impulse.

Just as the autocorrelation function yields the PSD as its Fourier transform, the cross-correlation function yields the cross-spectral density. We have

$$r_{xy}(t) = x(t) \star y(-t) = X(f)Y(-f) = R_{xy}(f) \qquad r_{yx}(t) = y(t) \star x(-t) = Y(f)X(-f) = R_{yx}(f) \tag{9.52}$$

Unlike the PSD, the cross-spectral densities need not be real.

Non-Parametric System Identification

One application of cross-spectral density functions is in transfer function estimation. For a system with impulse response $h(t)$, the input $x(t)$ yields the output $y(t) = x(t) \star h(t)$, and we have

$$r_{yx}(t) = y(t) \star x(-t) = h(t) \star x(t) \star x(-t) = h(t) \star r_{xx}(t) \tag{9.53}$$

The Fourier transform of both sides yields

$$R_{yx}(f) = H(f)R_{xx}(f) \qquad H(f) = \frac{R_{yx}(f)}{R_{xx}(f)} \tag{9.54}$$

This relation allows us to identify an unknown transfer function $H(f)$ as the ratio of the cross-spectral density $R_{yx}(f)$ and the PSD $R_{xx}(f)$ of the input $x(t)$. If $x(t)$ is a white noise signal with a PSD of unity, we get $R_{xx}(f) = 1$ and $H(f) = R_{yx}(f)$. This approach is termed **non-parametric system identification** because it presupposes no model for the system.

9.6 Time-Bandwidth Measures

The scaling property of the Fourier transform tells us that the narrower a time signal, the more spread out its spectrum. If we rearrange the two central ordinate relations, we can write

$$T_0 = \frac{\int_{-\infty}^{\infty} x(t)\,dt}{x(0)} \qquad B_0 = \frac{\int_{-\infty}^{\infty} X(f)\,df}{2X(0)} \qquad T_0B_0 = 0.5 \tag{9.55}$$

The term T_0 may be regarded as a measure defining time duration and the term B_0 as a measure of bandwidth. This result tells us that the time-bandwidth product is a constant. The larger the bandwidth, the smaller is

PGR2026-00033
Krisp Technologies EX1024 Page 298

the time duration. Time-bandwidth relations quantify this fundamental concept and assert that measures of duration in the time domain are inversely related to measures of bandwidth in the frequency domain and their product is a constant. Measures of duration and bandwidth are especially useful for lowpass and bandpass filters. The duration usually refers to the duration of the impulse response. Since the impulse response and step response are related, the rise time of the step response is often used as a practical measure of impulse-response duration.

REVIEW PANEL 9.33

Duration and Bandwidth Measures Are Useful for Lowpass and Bandpass Systems

Time domain: Duration of the impulse response $h(t)$ or rise-time of the step response $s(t)$
Frequency domain: Bandwidth of the transfer function $H(f)$

The commonly used **10%–90% rise time** T_r is defined as the time required for the step response $s(t)$ to rise from 10% to 90% of its final value:

$$T_r = s(t_{90}) - s(t_{10}) \tag{9.56}$$

The **bandwidth** is a measure of the frequency spread of the spectrum of a signal and corresponds to the range of frequencies that contains a significant fraction of the total signal energy. We shall use definitions based on the *one-sided* bandwidth that includes only positive frequencies.

REVIEW PANEL 9.34

Bandwidth Measures (in this book) **Are Defined for the One-Sided Bandwidth**

For a signal band-limited to B, the **absolute bandwidth** B_{abs} is just its frequency extent B. For signals whose spectrum shows sidelobes, we sometimes define a **null bandwidth** as the half-width of the central lobe of the spectrum. A practical measure of bandwidth is the **half-power bandwidth** or **3-dB bandwidth** $B_{3\text{dB}}$, the range of positive frequencies over which the magnitude exceeds $1/\sqrt{2}$ times its maximum.

REVIEW PANEL 9.35

Definitions of the 10%–90% Rise Time and Half-Power (3-dB) Bandwidth

10%–90% rise time T_r: Time for the step response to rise from 10% to 90% of its final value
3-dB bandwidth $B_{3\text{dB}}$: Positive frequency range for which $|H(f)| \geq \frac{|H(f)|_{\max}}{\sqrt{2}} \approx 0.707|H(f)|_{\max}$

EXAMPLE 9.17 (Time-Bandwidth Relations)

(a) Consider an ideal filter with $H(f) = \text{rect}(f)$ and $h(t) = \text{sinc}(t)$.

Using measures for duration and bandwidth based on the the central ordinate theorems, we find

$$T_0 = \frac{\int_{-\infty}^{\infty} \text{sinc}(t)\,dt}{\text{sinc}(0)} = 1 \qquad B_0 = \frac{\int_{-\infty}^{\infty} H(f)\,dt}{2H(0)} = 0.5 \qquad T_0 B_0 = 0.5$$

It should come as no surprise to you that measures based on the central ordinate theorems will yield the result $T_0 B_0 = 0.5$ for any Fourier transform pair for which the required quantities exist.

PGR2026-00033
Krisp Technologies EX1024 Page 299

(b) Consider an RC lowpass filter with time constant τ. Its transfer function $H(f)$, impulse response $h(t)$, and step response $s(t)$ are given by

$$H(f) = \frac{1}{1 + j2\pi f\tau) \qquad h(t) = \frac{1}{\tau}e^{-t/\tau}u(t) \qquad s(t) = (1 - e^{-t/\tau})u(t)$$

The half-power bandwidth $B_{3\text{dB}}$ is the range of positive frequencies for which $|H(f)| \geq \sqrt{1/2}$. This corresponds to a bandwidth of $B_{3\text{dB}} = \frac{1}{2\pi\tau}$. The 10%–90% rise time T_r is computed by finding

$$s(t_{10}) = 1 - e^{-t_{10}/\tau} = 0.1 \qquad s(t_{90}) = 1 - e^{-t_{90}/\tau} = 0.9$$

We obtain $t_{10} = \tau \ln(10/9)$, $t_{90} = \tau \ln 10$, and $T_r = t_{90} - t_{10} = \tau \ln 9$. Thus,

$$T_r B_{3\text{dB}} = \frac{\tau \ln 9}{2\pi\tau} \approx 0.35$$

This result is often used as an approximation for higher-order systems also.

9.6.1 Energy-Based Measures

Energy-based measures for bandwidth are based on the concept of equivalent bandwidth. The **equivalent bandwidth** is defined as the bandwidth of an ideal filter with the same energy as the actual filter. The *one-sided* equivalent bandwidth B_{eq} is given by

$$B_{\text{eq}} = \frac{\int_0^\infty |H(f)|^2\, df}{|H(0)|^2} = \frac{\int_{-\infty}^\infty |H(f)|^2\, df}{2|H(0)|^2} \tag{9.57}$$

The equivalent bandwidth is also called the **equivalent noise bandwidth** and is a useful measure for characterizing the performance of practical filters. The equivalent bandwidth approaches the 3-dB bandwidth as the filter spectrum approaches the spectrum of an ideal filter.

EXAMPLE 9.18 (Equivalent Bandwidth)

(a) For a first-order RC lowpass filter with time constant τ, the transfer function $H(f)$ and its characteristics are given by

$$H(f) = \frac{1}{1 + j2\pi f\tau} \qquad H(0) = 1 \qquad |H(f)|^2 = \frac{1}{1 + 4\pi^2 f^2\tau^2}$$

We then find its equivalent bandwidth B_{eq} as

$$B_{\text{eq}} = \frac{\int_{-\infty}^\infty |H(f)|^2\, df}{2|H(0)|^2} = \frac{1}{2}\int_{-\infty}^\infty \frac{1}{1 + 4\pi^2 f^2\tau^2}\, df = \frac{1}{4\tau}$$

The 3-dB bandwidth $B_{3\text{dB}}$ of this filter and the ratio $B_{\text{eq}}/B_{3\text{dB}}$ are given by

$$B_{3\text{dB}} = \frac{1}{2\pi\tau} \qquad \frac{B_{\text{eq}}}{B_{3\text{dB}}} = \frac{\pi}{2} = 1.57$$

PGR2026-00033
Krisp Technologies EX1024 Page 300

(b) The nth-order Butterworth lowpass filter with a 3-dB bandwidth of $B_{3\text{dB}} = \frac{1}{2\pi}$ is defined by the magnitude-squared function $|H(f)|^2 = \dfrac{1}{1+(2\pi f)^{2n}}$. With $|H(0)|^2 = 1$ and $\displaystyle\int_0^\infty \frac{1}{1+x^n}\,dx = \frac{1}{\text{sinc}(n)}$, its equivalent bandwidth B_{eq} is given by

$$B_{\text{eq}} = \frac{\int_0^\infty |H(f)|^2\,df}{|H(0)|^2} = \int_0^\infty \frac{1}{1+(2\pi f)^{2n}}\,df = \frac{1}{2\pi\,\text{sinc}(0.5/n)} = \frac{1}{4n\sin(0.5\pi/n)}$$

We find that B_{eq} equals 0.25 for $n = 1$, $\sqrt{2}/8$ for $n = 2$, and 1/6 for $n = 3$. We also find the ratio $B_{\text{eq}}/B_{3\text{dB}}$ equals 1.57 for $n = 1$, 1.111 for $n = 2$, and 1.047 for $n = 3$, and approaches unity with increasing n.

Energy-based measures for duration and the time-bandwidth product rely on the following definitions for **equivalent duration** and **equivalent bandwidth**

$$T_{\text{eq}} = \frac{\{\int_{-\infty}^{\infty} h(t)\,dt\}^2}{\int_{-\infty}^{\infty} h^2(t)\,dt} \qquad B_{\text{eq}} = \frac{\int_{-\infty}^{\infty} |H(f)|^2\,df}{2|H(0)|^2} \qquad T_{\text{eq}}B_{\text{eq}} = 0.5 \tag{9.58}$$

The equality $T_{\text{eq}}B_{\text{eq}} = 0.5$ holds for any transform pair for which the required quantities exist. To accommodate complex-valued functions, we often replace $h(t)$ by $|h(t)|$ and $H(f)$ by $|H(f)|$. In such a case, it turns out that $T_{\text{eq}}B_{\text{eq}} \geq 0.5$.

9.6.2 Moment-Based Measures

Other measures of duration and bandwidth rely on moments. A set of measures based on the moments of $h(t)$ and $H(f)$ is defined by

$$T_M = 2\left[\frac{\int_{-\infty}^{\infty} t^2|h(t)|\,dt}{\int_{-\infty}^{\infty} |h(t)|\,dt}\right]^{1/2} \qquad B_M = \left[\frac{\int_{-\infty}^{\infty} f^2|H(f)|\,df}{\int_{-\infty}^{\infty} |H(f)|\,df}\right]^{1/2} \qquad T_M B_M \geq \frac{1}{\pi} \tag{9.59}$$

Another set of measures based on the moments of $|h(t)|^2$ and $|H(f)|^2$ yields the **rms duration** and **rms bandwidth** defined by

$$T_{\text{rms}} = 2\left[\frac{\int_{-\infty}^{\infty} t^2|h(t)|^2\,dt}{\int_{-\infty}^{\infty} |h(t)|^2\,dt}\right]^{1/2} \qquad B_{\text{rms}} = \left[\frac{\int_{-\infty}^{\infty} f^2|H(f)|^2\,df}{\int_{-\infty}^{\infty} |H(f)|^2\,df}\right]^{1/2} \qquad T_{\text{rms}}B_{\text{rms}} \geq \frac{1}{2\pi} \tag{9.60}$$

In the moment-based measures, the equality in the time-bandwidth product holds only for a Gaussian signal.

EXAMPLE 9.19 (More Time-Bandwidth Relations)

(a) Consider an ideal filter with $H(f) = \text{rect}(f)$ and $h(t) = \text{sinc}(t)$.

Using the equivalent duration and bandwidth, we get

$$T_{\text{eq}} = \frac{\left[\int_{-\infty}^{\infty} \text{sinc}(t)\,dt\right]^2}{\int_{-\infty}^{\infty} \text{sinc}^2(t)\,dt} = 1 \qquad B_{\text{eq}} = \frac{\int_{-\infty}^{\infty} |H(f)|^2\,dt}{2|H(0)|^2} = 0.5 \qquad T_{\text{eq}}B_{\text{eq}} = 0.5$$

PGR2026-00033
Krisp Technologies EX1024 Page 301

(b) Consider an RC lowpass filter with time constant τ. Its transfer function $H(f)$ and impulse response $h(t)$ are given by

$$H(f) = \frac{1}{1 + j2\pi f\tau} \qquad h(t) = \frac{1}{\tau}e^{-t/\tau}u(t)$$

From Example 9.18(a), its equivalent bandwidth is $B_{\text{eq}} = \frac{1}{4\tau}$. Its equivalent duration is given by

$$T_{\text{eq}} = \frac{\left[\int_{-\infty}^{\infty} \frac{1}{\tau}e^{-t/\tau}\,dt\right]^2}{\int_{-\infty}^{\infty} \frac{1}{\tau^2}e^{-2t/\tau}\,dt} = 2\tau$$

Once again, the duration-bandwidth product is $T_{\text{eq}}B_{\text{eq}} = 0.5$.

(c) Consider the cascade of two identical RC lowpass filters with time constant $\tau = 1$. The transfer function $H(f)$ and impulse response $h(t)$ of the cascaded system are given by

$$H(f) = \frac{1}{(1 + j2\pi f)^2} \qquad h(t) = te^{-t}u(t)$$

Using rms measures, its rms bandwidth is given by

$$B_{\text{rms}} = \left[\frac{\int_{-\infty}^{\infty} f^2|H(f)|^2\,df}{\int_{-\infty}^{\infty} |H(f)|^2\,df}\right]^{1/2} = \left[\frac{\int_0^{\infty} \frac{f^2}{(1+4\pi^2f^2)^2}\,df}{\int_0^{\infty} \frac{1}{(1+4\pi^2f^2)^2}\,df}\right]^{1/2} = 1$$

Its rms duration is given by

$$T_{\text{rms}} = 2\left[\frac{\int_{-\infty}^{\infty} t^2|h(t)|^2\,dt}{\int_{-\infty}^{\infty} |h(t)|^2\,dt}\right]^{1/2} = 2\left[\frac{\int_0^{\infty} t^4e^{-2t}\,dt}{\int_0^{\infty} t^2e^{-2t}\,dt}\right]^{1/2} = 2\left(\frac{0.75}{0.25}\right)^{1/2} = 3.4641$$

The time-bandwidth product is $T_{\text{rms}}B_{\text{rms}} = 3.4641$ and satisfies the inequality $T_{\text{rms}}B_{\text{rms}} \geq \frac{1}{2\pi}$.

9.6.3 Concluding Remarks

The preceding examples reveal that the time-bandwidth product satisfies the inequality $T_rB \geq C$, where the constant C depends on the definitions used. This important concept represents the equivalent of the celebrated *uncertainty principle* of quantum mechanics. We cannot simultaneously make both time and frequency measures arbitrarily small. The more localized a signal is in time, the more spread out it is in frequency. In fact, we also have an analogy for discrete-time signals. Sampling a band-limited signal without loss of information requires a sampling interval t_s smaller than $\frac{1}{2B}$, where B is the signal bandwidth. This implies $t_sB > 2$, and tells us that we cannot simultaneously make the bandwidth and sampling interval arbitrarily small.

REVIEW PANEL 9.36

We Cannot Simultaneously Make the Duration and Bandwidth Arbitrarily Small

The **duration-bandwidth product** satisfies the inequality $T_rB_f \geq C$.

PGR2026-00033
Krisp Technologies EX1024 Page 302

CHAPTER 9 PROBLEMS

DRILL AND REINFORCEMENT

9.1 (Fourier Transforms from the Definition) Sketch each signal $x(t)$ and, starting with the defining relation, find its Fourier transform $X(f)$.

(a) $x(t) = \text{rect}(t - 0.5)$ **(b)** $x(t) = 2t\,\text{rect}(t)$ **(c)** $x(t) = te^{-2t}u(t)$ **(d)** $x(t) = e^{-2|t|}$

9.2 (Fourier Transform in Sinc Forms) Consider the signals $x(t)$ sketched in Figure P9.2. Express each signal by linear combinations of rect and/or tri functions and evaluate the Fourier transform $X(f)$ in terms of sinc functions.

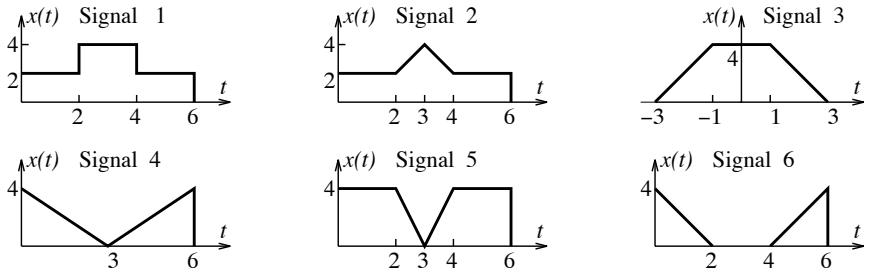


Figure P9.2 Figure for Problem 9.2

9.3 (Fourier Transforms) Sketch each signal $x(t)$, find its Fourier transform $X(f)$, and sketch its magnitude and phase spectrum.

(a) $x(t) = e^{-2|t-1|}$ **(b)** $x(t) = te^{-2|t|}$
(c) $x(t) = e^{-2t}\cos(2\pi t)u(t)$ **(d)** $x(t) = u(1 - |t|)$
(e) $x(t) = u(1 - |t|)\text{sgn}(t)$ **(f)** $x(t) = \text{sinc}(t) \star \text{sinc}(2t)$
(g) $x(t) = \cos^2(2\pi t)\sin(2\pi t)$ **(h)** $x(t) = (1 - e^{-t})u(t)$

9.4 (Properties) The Fourier transform of $x(t)$ is $X(f) = \text{rect}(f/2)$. Use properties to find and sketch the magnitude and phase of the Fourier transform of the following:

(a) $d(t) = x(t - 2)$ **(b)** $f(t) = x'(t)$ **(c)** $g(t) = x(-t)$
(d) $h(t) = tx(t)$ **(e)** $p(t) = x(2t)$ **(f)** $r(t) = x(t)\cos(2\pi t)$
(g) $s(t) = x^2(t)$ **(h)** $v(t) = tx'(t)$ **(i)** $w(t) = x(t) \star x(t)$

9.5 (Properties) Consider the transform pair $x(t) \Leftrightarrow X(f)$ where $x(t) = te^{-2t}u(t)$. Without evaluating $X(f)$, find the time signals corresponding to the following:

(a) $Y(f) = X(2f)$ **(b)** $D(f) = X(f - 1) + X(f + 1)$ **(c)** $G(f) = X'(f)$
(d) $H(f) = fX'(f)$ **(e)** $M(f) = j2\pi fX(2f)$ **(f)** $P(f) = X(f/2)\cos(4\pi f)$
(g) $R(f) = X^2(f)$ **(h)** $S(f) = (1 - 4\pi^2 f^2)X(f)$ **(i)** $V(f) = X(f) \star X(f)$

9.6 (Fourier Transforms) Argue for or against the following statements.

(a) Real signals have a real Fourier transform.
(b) Two-sided signals have a two-sided Fourier transform.
(c) The Fourier transform of a real, even symmetric signal is always a function of f^2.

PGR2026-00033
Krisp Technologies EX1024 Page 303

9.7 (Fourier Transforms) Compute the Fourier transform of the following signals and plot their amplitude spectrum and phase spectrum.

(a) $x(t) = \delta(t+1) + \delta(t-1)$ **(b)** $x(t) = \delta(t+1) + \delta(t) + \delta(t-1)$
(c) $x(t) = \delta(t+1) - \delta(t-1)$ **(d)** $x(t) = \delta(t+1) - \delta(t) + \delta(t-1)$

9.8 (Inverse Transforms) Find the inverse Fourier transforms of the following:

(a) $u(1-|f|)$ **(b)** $\dfrac{j2\pi f}{1+4\pi^2 f^2}$
(c) $\dfrac{j2\pi f}{1+6j\pi f - 8\pi^2 f^2}$ **(d)** $\dfrac{e^{-4j\pi f}}{1+4\pi^2 f^2}$
(e) $\text{sinc}(0.25f)\cos(2\pi f)$ **(f)** $8\,\text{sinc}(2f)\text{sinc}(4f)$
(g) $\dfrac{4\,\text{sinc}(2f)}{1+j2\pi f}$ **(h)** $\dfrac{\cos(\pi f)e^{-j\pi f}}{1+j2\pi f}$

9.9 (Fourier Transforms and Convolution) Sketch each signal $y(t)$, express it as a convolution of two functions $x(t)$ and $h(t)$, and find its Fourier transform $Y(f)$ as the product of $X(f)$ and $H(f)$.

(a) $y(t) = r(t+4) - 2r(t) + r(t-4)$
(b) $y(t) = r(t+4) - r(t+2) - r(t-2) - r(t-4)$
(c) $y(t) = te^{-2t}u(t)$

9.10 (Fourier Transforms and Convolution) Given the Fourier transforms $X_1(f)$ and $X_2(f)$ of two signals $x_1(t)$ and $x_2(t)$, compute the Fourier transform $Y(f)$ of the product $x_1(t)x_2(t)$.

(a) $X_1(f) = \text{rect}(f/4)$ $X_2(f) = \text{rect}(f/4)$
(b) $X_1(f) = \text{rect}(f/2)$ $X_2(f) = X_1(f/2)$
(c) $X_1(f) = \delta(f) + \delta(f+4) + \delta(f-4)$ $X_2(f) = \delta(f+2) + \delta(f-2)$
(d) $X_1(f) = \text{tri}(f)$ $X_2(f) = \delta(f+1) + \delta(f) + \delta(f-1)$

9.11 (Modulation) Sketch the spectrum of the modulated signal $y(t) = x(t)m(t)$ if

(a) $X(f) = \text{rect}(f)$ $m(t) = \cos(\pi t)$
(b) $X(f) = \text{rect}(0.25f)$ $m(t) = \cos(2\pi t)$
(c) $X(f) = \text{tri}(f)$ $m(t) = \cos(10\pi t)$
(d) $X(f) = \text{tri}(f)$ $m(t) = 0.5\displaystyle\sum_{k=-\infty}^{\infty} \delta(t - 0.5k)$

9.12 (Parseval's Theorem) A series RC circuit with time constant $\tau = 1$ is excited by a voltage input $x(t)$. The output is the capacitor voltage. Find the total energy in the input and output. Then, compute the signal energy in the output over $|f| \le \frac{1}{2\pi}$ Hz and over $|f| \le 1$ Hz.

(a) $x(t) = \delta(t)$ **(b)** $x(t) = e^{-t}u(t)$

9.13 (Fourier Transform of Periodic Signals) Find the Fourier transform $X(f)$ of each periodic signal $x(t)$ and sketch the magnitude spectrum $|X(f)|$ and phase spectrum $\angle X(f)$.

(a) $x(t) = 3 + 2\cos(10\pi t)$ **(b)** $x(t) = 8\sum \delta(t - 2k)$
(c) $x(t) = 3\cos(10\pi t) + 4\sin(10\pi t) + 6\cos(20\pi t + \frac{\pi}{4})$ **(d)** $x(t) = |\sin(\pi t)|$

9.14 (Application) A signal $x(t)$ is passed through an ideal lowpass filter as shown:

$$x(t) \longrightarrow \boxed{\text{ideal lowpass filter}} \longrightarrow y(t)$$

PGR2026-00033
Krisp Technologies EX1024 Page 304

The filter blocks frequencies above 5 Hz. Find $X(f)$ and $Y(f)$ and sketch their spectra for the following $x(t)$. Also compute and sketch the output $y(t)$.

(a) $x(t) = -4 + 2\sin(4\pi t) + 6\cos(12\pi t) + 3\sin(16\pi t) + 4\cos(16\pi t)$
(b) $x(t) = 8 - 8\cos^2(6\pi t)$
(c) $x(t) = |\sin(3\pi t)|$

9.15 (Frequency Response of Filters) Sketch the magnitude spectrum and phase spectrum of each of the following filters and find their response to the input $x(t) = \cos(\pi t)$.

(a) $h(t) = 8\,\text{sinc}[8(t-1)]$ **(b)** $H(f) = \text{tri}(f/6)e^{-j\pi f}$

9.16 (System Response) The transfer function of a system is $H(\omega) = \dfrac{16}{4 + j\omega}$. Find the time-domain response $y(t)$ for the following inputs.

(a) $x(t) = 4\cos(4t)$ **(b)** $x(t) = 4\cos(4t) - 4\sin(4t)$ **(c)** $x(t) = \delta(t)$
(d) $x(t) = \epsilon^{-4t}u(t)$ **(e)** $x(t) = 4\cos(4t) - 4\sin(2t)$ **(f)** $x(t) = u(t)$

9.17 (System Formulation) Set up the system differential equation for each of the following systems.

(a) $H(\omega) = \dfrac{3}{2 + j\omega}$ **(b)** $H(\omega) = \dfrac{1 + j2\omega - \omega^2}{(1-\omega^2)(4-\omega^2)}$

(c) $H(\omega) = \dfrac{2}{1 + j\omega} - \dfrac{1}{2 + j\omega}$ **(d)** $H(\omega) = \dfrac{2j\omega}{1 + j\omega} - \dfrac{j\omega}{2 + j\omega}$

9.18 (Frequency Response) Find the frequency response $H(f)$ and impulse response $h(t)$ of each system.

(a) $y''(t) + 3y'(t) + 2y(t) = 2x'(t) + x(t)$ **(b)** $y'(t) + 3y(t) = 2x(t)$
(c) $y''(t) + 4y'(t) + 4y(t) = 2x'(t) + x(t)$ **(d)** $y(t) = 0.2x(t)$

9.19 (System Analysis) Consider the following system:

$$x(t) \longrightarrow \boxed{h(t) = 2e^{-t}u(t) - \delta(t)} \longrightarrow y(t)$$

(a) Find the transfer function $H(f)$ and compute its magnitude and phase.
(b) Let $x(t) = \text{sinc}^2(t)$. Find the frequency domain output $Y(f)$.
(c) Let $x(t) = \cos(t)$. Find the frequency domain output $Y(f)$. Then find its inverse transform $y(t)$. Verify your result by finding $y(t)$, using time-domain convolution.
(d) Let $x(t) = e^{-t}u(t)$. Find $y(t)$ using Fourier transforms, and using time-domain convolution.
(e) Let $x(t) = e^{t}u(-t)$. Find $y(t)$ using Fourier transforms, and using time-domain convolution.

9.20 (Ideal Filters) Express the frequency response $H(f)$ and impulse response $h(t)$ of each filter in terms of the frequency response $H_{\text{LP}}(f)$ and impulse response $h_{\text{LP}}(t)$ of an ideal lowpass filter with cutoff frequency f_C.

(a) A lowpass filter with cutoff frequency $0.25f_C$
(b) A highpass filter with cutoff frequency $2f_C$
(c) A bandpass filter with center frequency $8f_C$ and passband $2f_C$
(d) A bandstop filter with center frequency $8f_C$ and stopband $2f_C$

9.21 (Time-Bandwidth Product) Compute the time-bandwidth product of a lowpass filter whose impulse response is $h(t) = (1/\tau)e^{-t/\tau}u(t)$, using the following:

PGR2026-00033
Krisp Technologies EX1024 Page 305

(a) The measures T_0 and B_0 based on the central ordinate theorems
(b) The measures T_{eq} and B_{eq} based on equivalent duration and bandwidth
(c) The duration $T_r = \int h(t)\,dt/h(t_0)$, where t_0 is the time at which the step response reaches 50% of its final value and bandwidth B_0 as defined by the central ordinate theorems

REVIEW AND EXPLORATION

9.22 (Fourier Transforms from Fourier Series) The Fourier series coefficients $X[k]$ of a periodic pulse train $x_p(t)$, with period T, are listed below along with an expression for $x_p(t)$ over over one period. Use the Fourier series coefficients $X[k]$ to find the Fourier transform $X(f)$ of one period $x(t)$ of each signal and sketch the signal $x(t)$.

(a) $x_p(t) = A\delta(t - t_0), \quad 0 \le t \le T \qquad X[k] = \dfrac{A}{T}e^{-j2\pi k f_0 t_0}$

(b) $x_p(t) = A, \quad |t| \le 0.5t_0 \qquad X[k] = \dfrac{At_0}{T}\,\text{sinc}(kf_0t_0)$

(c) $x_p(t) = A\left(1 - \dfrac{2|t|}{t_0}\right), \quad |t| \le 0.5t_0 \qquad X[k] = \dfrac{At_0}{2T}\,\text{sinc}^2(0.5kf_0t_0)$

(d) $x_p(t) = Ae^{-\alpha t}, \quad 0 \le t \le t_0 \qquad X[k] = \dfrac{A}{T}\left(\dfrac{1 - e^{-t_0(\alpha + j2\pi k f_0)}}{\alpha + j2\pi k f_0}\right)$

9.23 (Fourier Series from the Fourier Transform) For each signal $x(t)$, find the Fourier transform $X(f)$. Then sketch the periodic extension $x_p(t)$ with period T as indicated and find the Fourier series coefficients $X[k]$ of $x_p(t)$ from $X(f)$.

(a) $x(t) = \text{rect}(t)$, with $T = 2$ **(b)** $x(t) = \text{rect}(t - 0.5)$, with $T = 2$
(c) $x(t) = \text{rect}(t)$, with $T = 0.75$ **(d)** $x(t) = \text{tri}(t)$, with $T = 2$
(e) $x(t) = \text{tri}(t - 1)$, with $T = 2$ **(f)** $x(t) = \text{tri}(t)$, with $T = 1.5$
(g) $x(t) = e^{-t}u(t)$, with $T = 2$ **(h)** $x(t) = \text{rect}(t)\text{sgn}(t)$, with $T = 2$

9.24 (Fourier Transforms) Compute the Fourier transform $X(f)$ of the following signals and explain which transforms will show conjugate symmetry.

(a) $x(t) = e^{-t-1}u(t)$ **(b)** $x(t) = e^{-j\pi t}$
(c) $x(t) = e^{-t-j\pi t}u(t)$ **(d)** $x(t) = e^{-|t|-j2\pi t}$

9.25 (Properties) Find the Fourier transform of the following signals.

(a) $x(t) = \text{sgn}(t)\text{rect}(t)$ **(b)** $x(t) = \cos(2\pi t)\sin(2\pi t)$
(c) $x(t) = \cos(2\pi t)\text{sinc}(t)$ **(d)** $x(t) = \sin(2\pi t)\text{sinc}(t)$
(e) $x(t) = \cos(2\pi t)\text{sinc}(2t)$ **(f)** $x(t) = \cos(2\pi t)\text{sinc}^2(t)$
(g) $x(t) = \cos(2\pi t)\text{sinc}^2(2t)$ **(h)** $x(t) = \text{sinc}(2t)\text{sinc}(4t)$
(i) $x(t) = \sin(2\pi t)\text{rect}(t)$ **(j)** $x(t) = \cos(2\pi t)\text{tri}(t)$

9.26 (Properties) Sketch each signal and then find its Fourier transform.

(a) $x(t) = r(t) - r(t - 1) - u(t - 1)$ **(b)** $x(t) = r(t) - r(t - 1) - u(t - 2)$
(c) $x(t) = r(t) - r(t - 1) - r(t - 2) + r(t - 3)$ **(d)** $x(t) = r(t) - 2r(t - 1) + r(t - 2)$

9.27 (Properties) Find the Fourier transform of the following signals. What is the restriction on α for the Fourier transform to exist?

(a) $x(t) = e^{-\alpha t}u(t)$ **(b)** $x(t) = te^{-\alpha t}u(t)$ **(c)** $x(t) = e^{-\alpha|t|}$

PGR2026-00033
Krisp Technologies EX1024 Page 306

9.28 (Properties) Let $x(t) = 4\,\text{sinc}(4t)$. Find its Fourier transform $X(f)$ and use the result to find the Fourier transforms of the following signals.

(a) $y(t) = x(-t)$ **(b)** $g(t) = x(-2t-2)$
(c) $h(t) = e^{-j2\pi t}x'(t)$ **(d)** $p(t) = (t-1)x(t-1)e^{j2\pi t}$
(e) $r(t) = tx'(t)$ **(f)** $s(t) = \int_{-\infty}^{t} x(\xi)\,d\xi$
(g) $w(t) = x(t)\cos^2(2\pi t)$ **(h)** $v(t) = x^2(t)\cos^2(2\pi t)$

9.29 (Convolution) Find the Fourier transform $Y(f)$ of each signal $y(t)$.

(a) $y(t) = e^{-t}u(t) \star e^{-t}u(t)$ **(b)** $y(t) = e^{-t}u(t) \star e^{t}u(-t)$
(c) $y(t) = \text{rect}(t) \star \text{rect}(t)$ **(d)** $y(t) = \text{rect}(t) \star \sum_{k=-\infty}^{\infty} \delta(t-2k)$
(e) $y(t) = \text{rect}(t) \star \sum_{k=-1}^{1} \delta(t-2k)$ **(f)** $y(t) = 0.25\left[\text{rect}(t) \star \sum_{k=-\infty}^{\infty} \delta(t-2k)\right]\text{sinc}^2(t/4)$

9.30 (Inverse Fourier Transforms) Find $x(t)$ corresponding to each $X(f)$.

(a) $X(f) = \dfrac{\cos(\pi f)}{1+j2\pi f}e^{-j\pi f}$ **(b)** $X(f) = \dfrac{1-j2\pi f}{1+j2\pi f}$
(c) $X(f) = \dfrac{\sin(2\pi f)}{1+j2\pi f}e^{-j2\pi f}$ **(d)** $X(f) = u(f+2) - u(f+1) - u(f-1) + u(f-2)$
(e) $X(f) = \dfrac{e^{j\pi f}}{j\pi f - \pi^2 f^2}$ **(f)** $X(f) = [u(f+2) - 2u(f) + u(f-2)]e^{-j2\pi f}$

9.31 (Spectra) Sketch the spectra of the following signals.

(a) $x(t) = \text{sinc}(t-2)$ **(b)** $\text{sinc}^2(t)\cos(8\pi t)$
(c) $x(t) = \cos(\pi t - 0.25\pi)$ **(d)** $x(t) = \cos(\pi t)\text{rect}(t)$
(e) $x(t) = \cos(\pi t)\text{rect}(0.001t)$ **(f)** $x(t) = \cos(\pi t)\text{rect}(1000t)$

9.32 (Inverse Fourier Transforms) For each of the following transforms, find the signal $x(t)$ and state whether it is real, imaginary, or complex.

(a) $X(f) = \cos(\pi f)$ **(b)** $X(f) = \cos(\pi f)u(f)$
(c) $X(f) = \sin(\pi f)$ **(d)** $X(f) = \sin^2(\pi f)$
(e) $X(f) = \sin(\pi f + 0.25\pi)$ **(f)** $X(f) = \text{tri}(f)$
(g) $X(f) = \text{tri}(f-1)$ **(h)** $X(f) = \text{tri}(f-0.5) + \text{tri}(f+0.5)$
(i) $X(f) = \dfrac{1}{4\pi^2 f^2 + 2}$ **(j)** $X(f) = \dfrac{1}{2 + j8\pi f - 8\pi^2 f^2}$

9.33 (Fourier Transform Theorems) Use the various Fourier transform theorems as appropriate to evaluate the following integrals.

(a) $\int_{-\infty}^{\infty} \dfrac{1}{\alpha^2 + x^2}\,dx$ **(b)** $\int_{-\infty}^{\infty} \dfrac{1}{(\alpha^2 + x^2)^2}\,dx$
(c) $\int_{-\infty}^{\infty} \dfrac{\text{sinc}(0.5x/\pi)}{\alpha^2 + x^2}\,dx$ **(d)** $\int_{-\infty}^{\infty} \text{sinc}^2(x)\,dx$
(e) $\int_{-\infty}^{\infty} \text{sinc}^3(x)\,dx$ **(f)** $\int_{-\infty}^{\infty} \text{sinc}^4(x)\,dx$
(g) $\int_{-\infty}^{\infty} e^{-j\pi x}\,\text{sinc}^3(x)\,dx$ **(h)** $\int_{-\infty}^{\infty} j4\pi x e^{-j\pi x}\,\text{sinc}^2(x)\,dx$

PGR2026-00033
Krisp Technologies EX1024 Page 307

9.34 (Central Ordinate Theorems) Use the central ordinate theorem and the Fourier transform $X(f)$ to compute $x(0)$. Verify your results by finding $x(0)$ directly from $x(t)$. For which cases does $x(0)$ equal the initial value $x(0+)$?

(a) $x(t) = \text{rect}(t)$ **(b)** $x(t) = u(t)$ **(c)** $x(t) = e^{-t}u(t)$ **(d)** $x(t) = \text{rect}(t - 0.5)$

9.35 (Periodic Extension) Find the Fourier transform $X(f)$ of each signal $x(t)$ and the Fourier transform $Y(f)$ of its periodic extension $y(t)$ with period $T = 2$.

(a) $x(t) = \delta(t)$ **(b)** $x(t) = \text{rect}(t)$ **(c)** $x(t) = e^{-t}u(t)$

9.36 (Symmetry) Recall that conjugate symmetry of the Fourier transform applies only to real time signals. Identify the real/complex nature and symmetry of the time signals whose Fourier transform is

(a) Real **(b)** Real and even **(c)** Real and odd

9.37 (High-Frequency Decay) The high-frequency behavior of an energy signal may be estimated, without computing $X(f)$, by taking successive derivatives of $x(t)$. If it takes n derivatives to first show impulses, the high-frequency decay rate is proportional to $1/f^n$. The high-frequency decay rate is an indicator of how smooth the time signal $x(t)$ is; the faster the high-frequency decay rate, the smoother is the time signal $x(t)$. Without computing the Fourier transform, predict the high-frequency decay rate of the following signals.

(a) $x(t) = \text{rect}(t)$ **(b)** $x(t) = \text{tri}(t)$ **(c)** $x(t) = [1 + \cos(\pi t)]\text{rect}(0.5t)$
(d) $x(t) = e^{-\alpha t}u(t)$ **(e)** $x(t) = \cos(0.5\pi t)\text{rect}(0.5t)$ **(f)** $x(t) = (1 - t^2)\text{rect}(0.5t)$

9.38 (The Initial Value Theorem) The value of a causal energy signal $x(t)u(t)$ at $t = 0^+$ may be found from its transform $X(f)$, using the *initial value theorem* that states $x(0^+) = \lim_{f\to\infty}[j2\pi fX(f)]$. Find the initial value $x(0^+)$ for the signal corresponding to each $X(f)$, compare with the value $x(0)$ obtained by using the central ordinate theorem, and explain the differences, if any.

(a) $x(t) = u(t)$ **(b)** $x(t) = \text{rect}(t - 0.5)$ **(c)** $x(t) = \text{tri}(t - 1)$ **(d)** $x(t) = e^{-\alpha t}u(t)$

9.39 (System Response) For each circuit of Figure P9.39, let $R = 1\,\Omega$, $C = 1\,\text{F}$, and $L = 1\,\text{H}$ as needed.
(a) Find the transfer function $H(f)$.
(b) Sketch the magnitude and phase of $H(f)$ and identify the filter type.
(c) Find the response if the input is $x(t) = \delta(t)$.
(d) Find the response if the input is $x(t) = u(t)$.
(e) Find the response if the input is $x(t) = e^{-t}u(t)$.

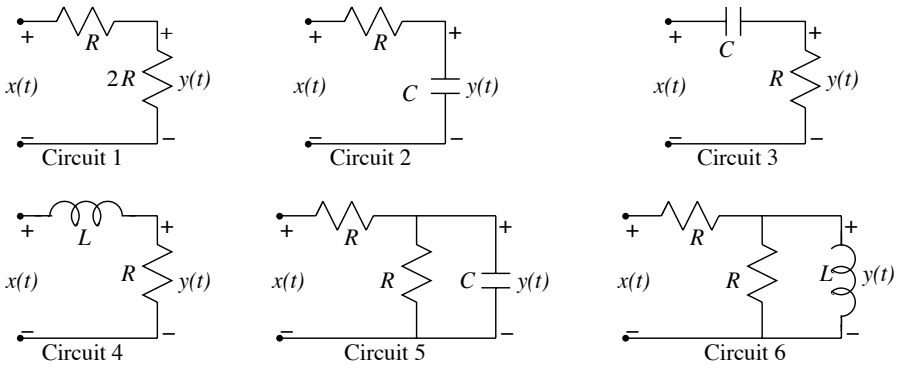


Figure P9.39 Figure for Problem 9.39

PGR2026-00033
Krisp Technologies EX1024 Page 308

9.40 (System Response) The signal $x(t) = 4 + \cos(4\pi t) - \sin(8\pi t)$ forms the input to a filter whose impulse response is $h(t)$, as shown. Find the response $y(t)$.

$$x(t) \longrightarrow \boxed{\text{filter } h(t)} \longrightarrow y(t)$$

(a) $h(t) = \text{sinc}(5t)$ **(b)** $h(t) = \text{sinc}(5t-2)$
(c) $h(t) = \text{sinc}^2(5t-2)$ **(d)** $h(t) = e^{-t}u(t)$
(e) $h(t) = \delta(t) - e^{-t}u(t)$ **(f)** $h(t) = \text{sinc}(t)\cos(8\pi t)$
(g) $h(t) = \text{sinc}^2(t)\cos(5\pi t)$ **(h)** $h(t) = \text{sinc}^2(t)\cos(16\pi t)$

9.41 (System Response) The impulse train input $x(t) = \sum_{n=0}^{\infty} (0.4)^n \delta(t-3n)$ is applied to a filter and the filter output is $y(t)$. Compute $Y(f)$ if the filter impulse response is

(a) $h(t) = \text{sinc}(t)$ **(b)** $h(t) = e^{-2t}u(t)$

9.42 (Application) Each of the following signals is applied to an ideal lowpass filter as shown.

$$x(t) \longrightarrow \boxed{\text{ideal filter}} \longrightarrow y(t)$$

The filter blocks frequencies above 2 Hz. Compute the response $Y(f)$ and (where possible) $y(t)$.

(a) $x(t) = e^{-t}u(t)$ **(b)** $x(t) = e^{-|t|}$ **(c)** $x(t) = \text{sinc}(8t)$
(d) $x(t) = \text{rect}(2t)$ **(e)** $x(t) = |\sin(3\pi t/2)|$ **(f)** $x(t) = \cos^2(3\pi t)$

9.43 (Cascaded Systems) Find the response $y(t)$ and sketch the magnitude and phase of $Y(f)$ for each cascaded system. Do the two systems yield the same output? Should they? Explain.

(a) $x(t) = \cos(\pi t) \longrightarrow \boxed{\text{phase shift of } -\frac{\pi}{2}} \longrightarrow \boxed{\text{squaring circuit}} \longrightarrow y(t)$

(b) $x(t) = \cos(\pi t) \longrightarrow \boxed{\text{squaring circuit}} \longrightarrow \boxed{\text{phase shift of } -\frac{\pi}{2}} \longrightarrow y(t)$

9.44 (Modulation and Filtering Concepts) System 1 is a modulator that multiplies a signal by $\cos(6\pi t)$, and system 2 is an ideal lowpass filter with a cutoff frequency of 3 Hz. Compute the output of all systems in both the time domain and the frequency domain if the input is $x(t) = \text{sinc}^2(2t)$ and the systems are connected in cascade as shown. Do the two configurations produce identical outputs?

(a) $x(t) \longrightarrow \boxed{\text{system 1}} \longrightarrow \boxed{\text{system 2}} \longrightarrow y(t)$

(b) $x(t) \longrightarrow \boxed{\text{system 2}} \longrightarrow \boxed{\text{system 1}} \longrightarrow y(t)$

9.45 (Modulation and Filtering Concepts) System 1 is a modulator that multiplies a signal by $\cos(6\pi t)$, system 2 is an ideal lowpass filter with a cutoff frequency of 3 Hz, and system 3 is a squaring system. Compute the output of all systems in both the time domain and the frequency domain if the input is $x(t) = \text{sinc}(2t)$ and the systems are connected in cascade as shown. Are any of the overall outputs identical?

(a) $x(t) \longrightarrow \boxed{\text{system 1}} \longrightarrow \boxed{\text{system 2}} \longrightarrow \boxed{\text{system 3}} \longrightarrow y(t)$

(b) $x(t) \longrightarrow \boxed{\text{system 3}} \longrightarrow \boxed{\text{system 2}} \longrightarrow \boxed{\text{system 1}} \longrightarrow y(t)$

(c) $x(t) \longrightarrow \boxed{\text{system 2}} \longrightarrow \boxed{\text{system 3}} \longrightarrow \boxed{\text{system 1}} \longrightarrow y(t)$

(d) $x(t) \longrightarrow \boxed{\text{system 3}} \longrightarrow \boxed{\text{system 1}} \longrightarrow \boxed{\text{system 2}} \longrightarrow y(t)$

(e) $x(t) \longrightarrow \boxed{\text{system 1}} \longrightarrow \boxed{\text{system 3}} \longrightarrow \boxed{\text{system 2}} \longrightarrow y(t)$

(f) $x(t) \longrightarrow \boxed{\text{system 2}} \longrightarrow \boxed{\text{system 1}} \longrightarrow \boxed{\text{system 3}} \longrightarrow y(t)$

PGR2026-00033
Krisp Technologies EX1024 Page 309

9.46 (System Response) The transfer function of a system is $H(\omega) = \dfrac{2 + 2j\omega}{4 + 4j\omega - \omega^2}$. Find the time-domain response $y(t)$ and the spectrum $Y(f)$ for the following inputs.

(a) $x(t) = 4\cos(2t)$ **(b)** $x(t) = e^{-t}u(t)$ **(c)** $x(t) = \delta(t)$ **(d)** $x(t) = 2\delta(t) + \delta'(t)$

9.47 (System Response) The transfer function of a filter is given by $H(f) = \text{rect}(0.1f)e^{-j\pi f}$. Find its response $y(t)$ to each input $x(t)$.

(a) $x(t) = 1$
(b) $x(t) = \cos(8\pi t)$
(c) $x(t) = 4 + \cos(\pi t - 0.25\pi) - \cos(40\pi t)$
(d) $x(t) = \text{sinc}[5(t-1)]$

9.48 (System Response) The signal $x(t) = \text{sinc}(0.5t) + \text{sinc}(0.25t)$ is applied to a filter whose impulse response is $h(t) = A\,\text{sinc}(\alpha t)$.

(a) For what values of A and α will the filter output equal $\text{sinc}(0.5t)$?
(b) For what values of A and α will the filter output equal $\text{sinc}(0.25t)$?
(c) For what values of A and α will the filter output equal $x(t)$?
(d) What will be the output if $A = 1$ and $\alpha = 1$?

9.49 (System Response) Find the output of a filter whose impulse response is $h(t) = 4\,\text{sinc}(5t - 8)$ to the following inputs.

(a) $x(t) = 1$ **(b)** $x(t) = \cos(2\pi t)$ **(c)** $x(t) = \cos(12\pi t)$
(d) $x(t) = \text{sinc}(t)$ **(e)** $x(t) = \text{sinc}(12t)$ **(f)** $x(t) = \displaystyle\sum_{k=-\infty}^{\infty} \text{tri}(2t - k)$

9.50 (System Response) Find the output of a filter whose impulse response is $h(t) = 4\,\text{sinc}^2(2t - 1)$ to the following inputs.

(a) $x(t) = 1$ **(b)** $x(t) = \cos(2\pi t)$ **(c)** $x(t) = \cos(12\pi t)$
(d) $x(t) = \text{sinc}(t)$ **(e)** $x(t) = \text{sinc}(12t)$ **(f)** $x(t) = \displaystyle\sum_{k=-\infty}^{\infty} \text{tri}(2t - k)$

9.51 (Integral Equations) Integral equations arise in many contexts. One approach to solving integral equations is by using the convolution property of the Fourier transform. Find the transfer function and impulse response of a filter whose input-output relation is described by the following:

(a) $y(t) = x(t) - 2\displaystyle\int_{-\infty}^{t} y(\lambda)e^{-(t-\lambda)}u(t-\lambda)\,d\lambda$

(b) $y(t) = x(t) + \displaystyle\int_{-\infty}^{t} y(\lambda)e^{-3(t-\lambda)}u(t-\lambda)\,d\lambda$

9.52 (System Representation) Find the transfer function $H(f)$ of each system described by its impulse response $h(t)$, find the differential equation and its order (where possible), and determine if the system is stable and/or causal.

(a) $h(t) = e^{-2t}u(t)$
(b) $h(t) = (1 - e^{-2t})u(t)$
(c) $h(t) = \text{sinc}(t)$
(d) $h(t) = te^{-t}u(t)$
(e) $h(t) = 0.5\delta(t)$
(f) $h(t) = (e^{-t} - e^{-2t})u(t)$
(g) $h(t) = \delta(t) - e^{-t}u(t)$
(h) $h(t) = \text{sinc}^2(t+1)$

PGR2026-00033
Krisp Technologies EX1024 Page 310

9.53 (System Response) Consider the *relaxed* system $y'(t) + \frac{1}{\tau}y(t) = \frac{1}{\tau}x(t)$.
- **(a)** What is the frequency response $H(f)$ of this system?
- **(b)** What is the response $y(t)$ of this system if the input is $x(t) = \cos(2\pi f_0 t)$?
- **(c)** Compute and sketch $y(t)$ if $f_0 = 10$ Hz and $\tau = 1$ s. Repeat for $f_0 = 1$ Hz and $\tau = 1$ ms.
- **(d)** Under what conditions for τ and f_0 would you expect the response to resemble (be a good approximation to) the input?

9.54 (Eigensignals) If the input to a system is its eigensignal, the response has the same form as the eigensignal. Argue for or against the following statements by computing the system response in the frequency domain using the Fourier transform.
- **(a)** Every signal is an eigensignal of the system described by $h(t) = A\delta(t)$.
- **(b)** The signal $x(t) = e^{j2\pi\alpha t}$ is an eigensignal of any LTI system such as that described by the impulse response $h(t) = e^{-\beta t}u(t)$.
- **(c)** The signal $x(t) = e^{j2\pi\alpha t}$ is an eigensignal of any LTI system described by a differential equation such as $y'(t) + \beta y(t) = x(t)$.
- **(d)** The signal $x(t) = \cos(\alpha t)$ is an eigensignal of any LTI system described by a differential equation such as $y'(t) + \beta y(t) = x(t)$.
- **(e)** The signal $x(t) = \text{sinc}(\alpha t)$ is an eigensignal of ideal filters described by $h(t) = \text{sinc}(\beta t),\ \beta > \alpha$.
- **(f)** The signal $x(t) = \cos(2\pi\alpha t)$ is an eigensignal of ideal filters described by $h(t) = \text{sinc}(\beta t),\ \beta > 2\alpha$.
- **(g)** A signal $x(t)$ band-limited to B Hz is an eigensignal of the ideal filter $h(t) = \text{sinc}(\beta t),\ \beta > 2B$.

9.55 (Frequency Response of Filters) The frequency response of a filter is $H(\omega) = \dfrac{8}{8 - \omega^2 + 4j\omega}$.
- **(a)** What is the dc gain A_0 of this filter?
- **(b)** At what frequency does the gain equal $A_0/\sqrt{2}$? At what frequency (approximately) does the gain equal $0.01A_0$?
- **(c)** Compute the response $y(t)$ of this filter to the input $x(t) = 4 + 4\cos(2t) - 4\sin(2t) + \cos(30t)$.

9.56 (Frequency Response of Filters) The frequency response of a filter is $H(\omega) = \dfrac{1}{1 - \omega^2 + \sqrt{2}j\omega}$.
- **(a)** What is the dc gain A_0 of this filter?
- **(b)** At what frequency does the gain equal $A_0/\sqrt{2}$? At what frequency (approximately) does the gain equal $0.01A_0$?
- **(c)** Compute the response $y(t)$ of this filter to the input $x(t) = 4 + 4\cos(t) - 4\sin(t) + \cos(10t)$.

9.57 (Power Spectral Density) Sketch the power spectral density (PSD) of the following:
- **(a)** $x(t) = 8\cos(10\pi t) + 4\sin(20\pi t) + 6\cos(30\pi t + \frac{\pi}{4})$
- **(b)** $x(t) = 4 + 8\cos(10\pi t) + 4\sin(20\pi t)$

9.58 (Power Spectral Density) A transform $X(f)$ can qualify as the PSD of an autocorrelation function only if it satisfies certain properties. For each of the following that qualify as a valid PSD, find the autocorrelation function. For those that do not, explain why they cannot be a valid PSD.

- **(a)** $R(f) = \dfrac{1}{1 + j2\pi f}$
- **(b)** $R(f) = \dfrac{1}{1 + 4\pi^2 f^2}$
- **(c)** $R(f) = \dfrac{1}{1 + 2\pi f + 4\pi^2 f^2}$
- **(d)** $R(f) = \dfrac{3}{4 + 20\pi^2 f^2 + 16\pi^4 f^4}$
- **(e)** $R(f) = \dfrac{1}{1 - 32\pi^2 f^2 + 16\pi^4 f^4}$
- **(f)** $R(f) = \text{sinc}^2(2f)$
- **(g)** $R(f) = \delta(f) + \text{sinc}^2(2f)$
- **(h)** $R(f) = \delta(f) + \dfrac{2\pi f}{1 + 4\pi^2 f^2}$

PGR2026-00033
Krisp Technologies EX1024 Page 311

9.59 (Power Spectral Density) Find the PSD corresponding to the following autocorrelation functions and describe their spectra.

(a) $r_{xx}(t) = A\delta(t)$ (white noise) **(b)** $r_{xx}(t) = Ae^{-\alpha|t|}$

(c) $r_{xx}(t) = A\,\text{sinc}(\alpha t)$ (band-limited white noise) **(d)** $r_{xx}(t) = Ae^{-\alpha t^2}$

9.60 (Dirichlet Kernel) Find the quantities indicated for the following signals.

(a) $x(t) = \delta(t+1) + \delta(t) + \delta(t-1)$. Find $X(f)$.

(b) $d(t) = \sum_{k=-N}^{N} \delta(t-k)$. Show that $D(f) = (2N+1)\dfrac{\text{sinc}[(2N+1)f]}{\text{sinc}(f)}$.
This is the Dirichlet kernel. Does $D(f)$ equal $X(f)$ for $N = 1$?

(c) $y(t) = \sum_{k=-N}^{N} \delta(t-kt_0)$. Find its Fourier transform $Y(f)$.

9.61 (Bandwidth) Find the indicated bandwidth measures for each signal $x(t)$.

(a) $x(t) = \text{sinc}(\alpha t)$ (the absolute bandwidth)

(b) $x(t) = \text{rect}(\alpha t)$ (the null bandwidth)

(c) $x(t) = e^{-\alpha t}u(t)$ (the half-power bandwidth)

(d) $x(t) = e^{-\pi t^2}$ (the rms bandwidth)

(e) $x(t) = e^{-\pi t^2}$ (the equivalent bandwidth)

9.62 (Time-Bandwidth Product) Compute the time-bandwidth product of an ideal lowpass filter assuming that the duration corresponds to the width of the central lobe of its impulse response $h(t)$ and the bandwidth equals the absolute bandwidth.

9.63 (Time-Bandwidth Product) Find the equivalent bandwidth and the time-bandwidth product for a second-order Butterworth filter with a cutoff frequency of 1 rad/s.

9.64 (Time-Bandwidth Product) Compute the indicated time-bandwidth product for each system.

(a) $h(t) = \text{tri}(t)$, using equivalent duration and equivalent bandwidth

(b) $h(t) = e^{-\pi t^2}$, using the moment-based measures T_M and B_M as defined in the text

9.65 (Equivalent Noise Bandwidth) Find the equivalent bandwidth and the half-power bandwidth for

(a) A first-order lowpass filter with time constant τ.

(b) The cascade of two first-order lowpass filters, each with time constant τ.

(c) A second-order Butterworth filter with a cutoff frequency of 1 rad/s.

9.66 (Frequency Response of Oscilloscopes) To permit accurate displays and time measurements, commercial oscilloscopes are designed to have an approximately Gaussian response. This also permits a monotonic step response (with no ringing) and a constant group delay (freedom from distortion). However, a more fundamental reason is based on the fact that an oscilloscope is a complex instrument with many subsystems and the overall response depends on the response of the individual systems. How does this fact account for the Gaussian response of commercial oscilloscopes (think central limit theorem)?

9.67 (Oscilloscope Bandwidth) When displaying step-like waveforms on an oscilloscope, the measured rise time T_m is related to the oscilloscope rise time T_o and the signal rise time T_s by $T_m^2 \approx T_s^2 + T_o^2$.

PGR2026-00033
Krisp Technologies EX1024 Page 312

(a) Pick a reasonable estimate of the rise time-bandwidth product and determine the oscilloscope bandwidth required to measure a signal rise time of 1 ns with an error of 10% or less.
(b) The signal is connected to the oscilloscope through a probe whose time constant is 0.1 ns. Make a reasonable estimate of the probe rise time from its time constant and determine the oscilloscope bandwidth required to measure a signal rise time of 1 ns with an error of 10% or less.

COMPUTATION AND DESIGN

ftprpgui A GUI for Visualization of Fourier Transform Properties

The graphical user interface `ftprpgui` allows you to visualize how the spectrum is affected by operations on the time signal (scaling, time shift, modulation, etc.).
To explore this routine, type `ftprpgui` at the MATLAB prompt.

freqgui A GUI for Visualization of Frequency Response

The graphical user interface `freqgui` allows you to visualize the frequency response.
You can displays results against f or ω, using several options.
To explore this routine, type `freqgui` at the MATLAB prompt.

9.68 (Frequency Response of Butterworth Filters) The gain of an nth-order Butterworth lowpass filter is described by

$$|H(\omega)| = \frac{1}{\sqrt{1 + (\frac{\omega}{\alpha})^{2n}}}$$

(a) Show that the half-power frequency of this filter is $\omega = \alpha$ for any n.
(b) Use MATLAB to plot the magnitude spectrum for $\alpha = 1$ and $n = 2,\ 3,\ 4,\ 5$ on the same graph. At what frequency do these filters have identical magnitudes? Describe how the magnitude spectrum changes as the order n is increased.

9.69 (Performance Measures) Performance measures for lowpass filters include the rise time, the 5% settling time, the half-power bandwidth, and the time-bandwidth product. Plot the step response, magnitude spectrum, and phase spectrum of the following filters. Then, use the ADSP routine `trbw` to numerically estimate each performance measure.

(a) A first-order lowpass filter with $\tau = 1$
(b) The cascade of two first-order filters, each with $\tau = 1$
(c) $y''(t) + \sqrt{2}y'(t) + y(t) = x(t)$ (a second-order Butterworth lowpass filter)
(d) $y'''(t) + 2y''(t) + 2y'(t) + y(t) = x(t)$ (a third-order Butterworth lowpass filter)

9.70 (Steady-State Response in Symbolic Form) The ADSP routine `ssresp` yields a symbolic expression for the steady-state response to sinusoidal inputs (see Chapter 21 for examples of its usage). Find the steady-state response to the input $x(t) = 2\cos(3t - \frac{\pi}{3})$ for each of the following systems.

(a) $y'(t) + \alpha y(t) = 2x(t)$, for $\alpha = 1,\ 2$
(b) $y''(t) + 4y(t) + Cy(t) = x(t)$, for $C = 3,\ 4,\ 5$

PGR2026-00033
Krisp Technologies EX1024 Page 313

Chapter 10

MODULATION

10.0 Scope and Objectives

The Fourier transform finds widespread application in the area of communications. This chapter concentrates on amplitude modulation (AM), where a message signal modulates the amplitude of a high-frequency carrier; phase modulation (PM), where the message modulates the phase of the carrier phase; and frequency modulation (FM), where the message modulates the carrier phase or frequency. It concludes with an introduction to the Hilbert transform and its applications to communication.

10.1 Amplitude Modulation

Communication of information often requires transmission of signals, or *messages*, through space by converting electrical signals to electromagnetic waves using antennas. Efficient radiation requires antenna dimensions comparable to the wavelength of the signal. The wavelength λ of a radiated signal equals c/f, where c is the speed of light (at which electromagnetic waves travel) and f is the signal frequency. Transmission of low-frequency messages, such as speech and music, would require huge antenna sizes measured in kilometers. A practical approach is to shift low-frequency messages to a much higher frequency using *modulation* and to transmit the modulated signal instead. In turn, this requires a means of recovering the original message at the receiving end, a process called *demodulation*.

We start with several amplitude modulation schemes based on modulating the amplitude of a high-frequency carrier $x_C(t) = \cos(2\pi f_C t)$ by a message signal $x_S(t)$ band-limited to a frequency $B \ll f_C$.

10.1.1 DSBSC AM

The simplest amplitude modulation scheme is to multiply the carrier $x_C(t)$ by the message signal $x_S(t)$, as illustrated in Figure 10.1.

The modulated signal $x_M(t)$ and its spectrum $X_M(f)$ are given by

$$x_M(t) = x_S(t)x_C(t) = x_S(t)\cos(2\pi f_C t) \qquad X_M(f) = 0.5[X(f+f_C) + X(f-f_C)] \tag{10.1}$$

The spectrum of the modulated signal shows the message spectrum centered about $\pm f_C$. It exhibits symmetry about $\pm f_C$. The portion for $f > |f_C|$ is called the **upper sideband** (USB) and that for $f < |f_C|$, the **lower sideband** (LSB). Since $x_S(t)$ is band-limited to B, the spectrum occupies a band of $2B$ (centered about $\pm f_C$), which is twice the signal bandwidth.

This scheme is called double-sideband suppressed-carrier (DSBSC) AM. The lowpass signal $x_S(t)$ is band-limited to B and called a **baseband signal**. The signal $x_M(t)$ is also band-limited to $B + f_C$, but it is called a **bandpass signal** because its spectrum does not include the origin.

PGR2026-00033
Krisp Technologies EX1024 Page 314

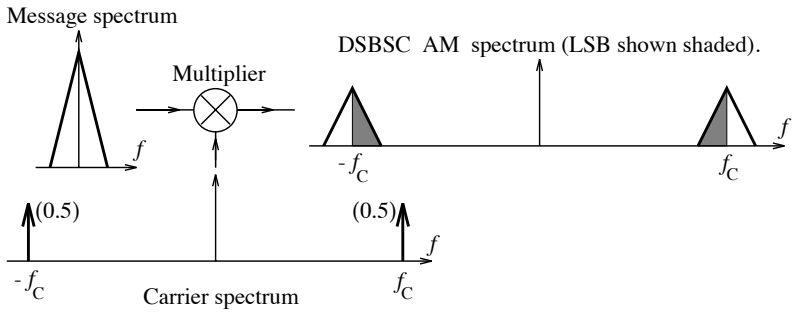


Figure 10.1 DSBSC amplitude modulation

In the time domain, note that the envelope of the modulated signal $x_M(t)$ follows the message $x_S(t)$, provided $x_S(t) \geq 0$, as illustrated in Figure 10.2.

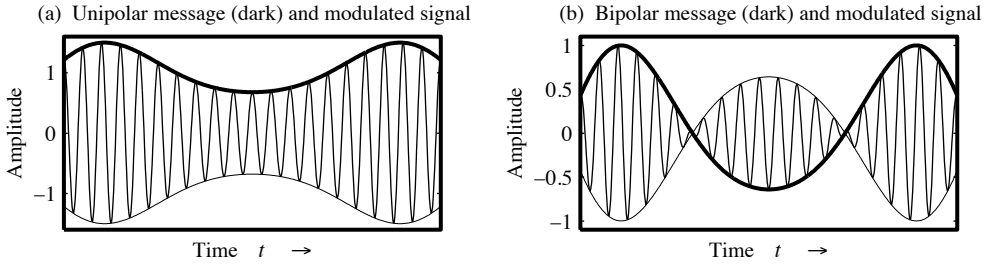


Figure 10.2 The envelope of the modulated signal corresponds to the message signal only if the message amplitude is always positive.

10.1.2 Standard AM

In standard AM, the modulated signal is transmitted along with the carrier signal, as illustrated in Figure 10.3. The message signal $x_S(t)$ modulates the carrier $x_C(t)$ to give the modulated signal $x_M(t)$ described by

$$x_M(t) = [A_C + x_S(t)]\cos(2\pi f_C t) \tag{10.2}$$

With $A_S = |x_S(t)|_{\max}$ and $\beta = A_S/A_C$, we may rewrite this as

$$x_M(t) = A_C \cos(2\pi f_C t) + \frac{\beta A_C}{A_S} x_S(t)\cos(2\pi f_C t) \tag{10.3}$$

Here, β is called the **modulation index**. The spectrum of the modulated signal $x_M(t)$ equals

$$X_M(f) = \frac{A_C}{2}[\delta(f + f_C) + \delta(f - f_C)] + \frac{\beta A_C}{2A_S}[X(f + f_C) + X(f - f_C)] \tag{10.4}$$

The spectrum is identical to that of DSBSC AM, except for the presence of two impulses at $\pm f_C$ due to the carrier component. Its bandwidth equals $2B$ (centered about $\pm f_C$), which is twice the signal bandwidth.

PGR2026-00033
Krisp Technologies EX1024 Page 315

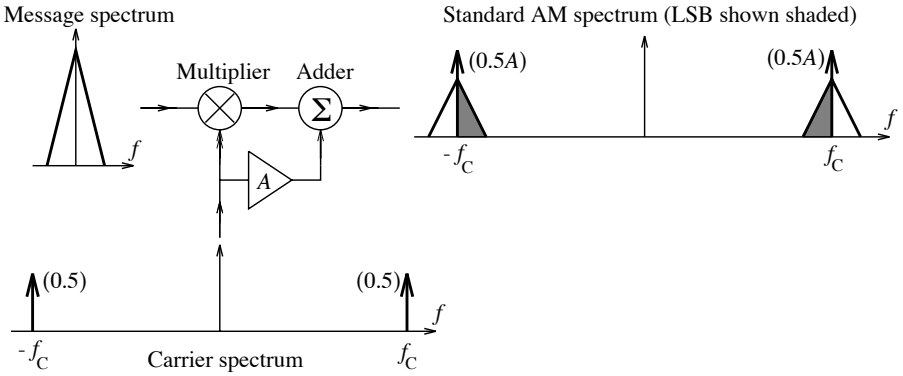


Figure 10.3 Standard AM

REVIEW PANEL 10.1

Suppressed Carrier AM and Transmitted Carrier (or Standard) AM

Suppressed carrier (DSBSC AM): $x_M(t) = x_S(t)\cos(2\pi f_C t)$
Transmitted carrier (standard AM): $x_M(t) = [A_C + x_S(t)]\cos(2\pi f_C t)$

In the time domain, if $x_S(t)$ varies slowly compared with the carrier, the envelope of $x_M(t)$ follows $x_S(t)$, provided $0 < \beta < 1$ (or $|x_S(t)| < A_C$). For $|\beta| > 1$, the envelope no longer resembles $x_S(t)$, and we have **over-modulation** (see Figure 10.2).

REVIEW PANEL 10.2

When Does the Envelope of AM Correspond to the Message Signal $x_s(t)$?

When $x_S(t) \geq 0$ for DSBSC AM or $A_C + x_S(t) \geq 0$ (i.e., $\beta \leq 1$) for standard AM.

10.1.3 Efficiency of Modulation

The **efficiency** of modulation is a measure of the power required to transmit just the message component. The efficiency of DSBSC AM is 100% since no power is expended in transmitting the carrier. For standard AM, however, the message power is only a fraction of the total transmitted power since the modulated signal also includes the carrier. We define the efficiency η as the ratio of the power P_S in the two sidebands and the total power P_M.

REVIEW PANEL 10.3

Modulation Index and Efficiency of Standard AM

Modulation index: $\beta = \dfrac{|x_S(t)|_{\max}}{A_C}$ **Efficiency:** $\eta = \dfrac{\text{power in sidebands}}{\text{total power}}$

Single-tone message: $\eta = \dfrac{\beta^2}{2+\beta^2}$ $\eta_{\max} = \frac{1}{3}$, or 33.33% (when $\beta = 1$)

PGR2026-00033
Krisp Technologies EX1024 Page 316

EXAMPLE 10.1 (Single-Tone Standard AM)
For a single-tone message $x_S(t) = A_S \cos(2\pi f_0 t)$, with $f_0 \ll f_C$, we may write

$$x_M(t) = [A_C + x_S(t)]\cos(2\pi f_C t) = [1 + \beta \cos(2\pi f_0 t)]A_C \cos(2\pi f_C t)$$

where $\beta = A_S/A_C$. Expanding this result, we get

$$x_M(t) = A_C \cos(2\pi f_C t) + 0.5A_C\beta \cos[2\pi(f_C + f_0)t] + 0.5A_C\beta \cos[2\pi(f_C - f_0)t]$$

The spectrum of $x_M(t)$ consists of the carrier component with strength A_C, and two sidebands at $f_C \pm f_0$ with strengths $0.5A_C$ and zero phase.

The power P_M in the modulated signal equals

$$P_M = 0.5A_C^2 + 0.5(0.25A_C^2\beta^2 + 0.25A_C^2\beta^2) = 0.25A_C^2(2 + \beta^2)$$

The total power in the two sidebands is $0.25A_C^2\beta^2$. The efficiency η is thus

$$\eta = \frac{0.25A_C^2\beta^2}{0.25A_C^2(2 + \beta^2)} = \frac{\beta^2}{2 + \beta^2}$$

The maximum possible efficiency occurs when $\beta = 1$ (with $A_S = A_C$) and equals $\frac{1}{3}$ (33.33%). With a modulation index of $\beta = 0.5$, the efficiency drops to $\frac{1}{9}$ (11.11%). For periodic signals with $\beta < 1$, it turns out that $\eta < 1/(1 + p_f^2)$, where p_f is the **peak factor** (the ratio of the peak and rms value). Since $x_{\text{rms}} \leq x_p$, we have $p_f \geq 1$. Large values of p_f imply poor efficiency.

10.1.4 Detection of AM Signals

At the receiver, the message $x_S(t)$ may be recovered using **demodulation** or **detection**. We discuss two detection schemes.

Synchronous (Coherent) Detection

Synchronous or coherent detection entails modulation of the modulated signal by $\cos(2\pi f_C t)$, followed by lowpass filtering. The demodulation scheme for DSBSC AM is shown in Figure 10.4.

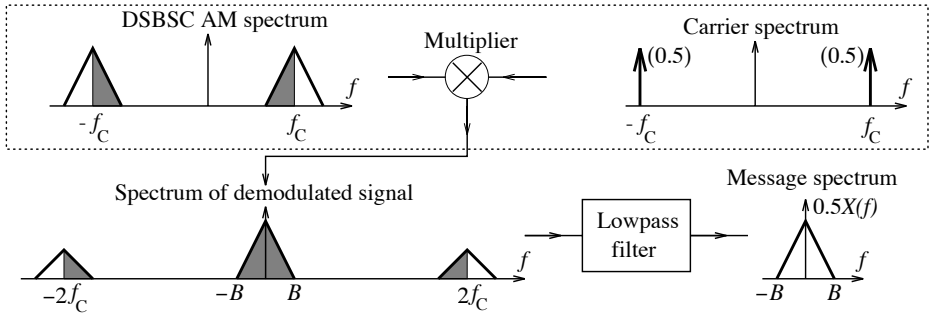


Figure 10.4 Synchronous demodulation of DSBSC AM

PGR2026-00033
Krisp Technologies EX1024 Page 317

With $x_M(t) = x_S(t)\cos(2\pi f_C t)$, the demodulated signal equals

$$x_D(t) = x_M(t)\cos(2\pi f_C t) = x_S(t)\cos(2\pi f_C t)\cos(2\pi f_C t) \tag{10.5}$$

Using the identity $\cos^2(\alpha) = 0.5(1 + \cos 2\alpha)$, we obtain

$$x_D(t) = 0.5x_S(t) + 0.5x_S(t)\cos(4\pi f_C t) \tag{10.6}$$

If $x_D(t)$ is passed through a lowpass filter whose bandwidth exceeds the message bandwidth B, the high-frequency component at $2f_C$ is blocked, and we recover $0.5x_S(t)$, which corresponds to the message signal.

The demodulation scheme for standard AM is shown in Figure 10.5.

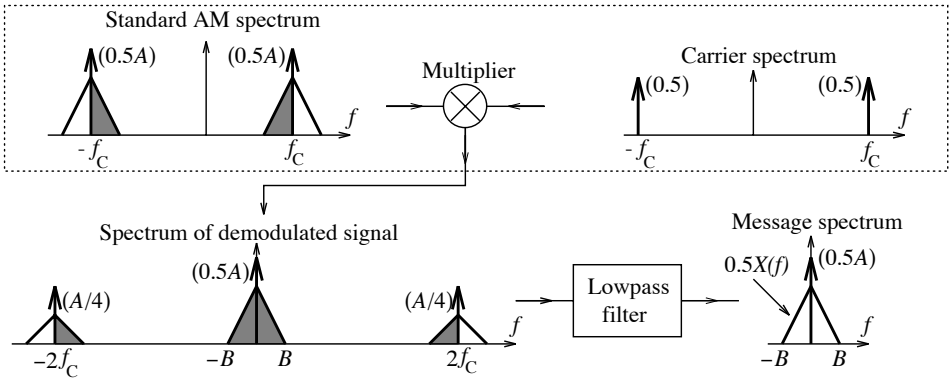


Figure 10.5 Synchronous demodulation of standard AM

With $x_M(t) = [A_C + x_S(t)]\cos(2\pi f_C t)$, the demodulated signal equals

$$x_D(t) = x_M(t)\cos(2\pi f_C t) = [A_C + x_S(t)]\cos(2\pi f_C t)\cos(2\pi f_C t) \tag{10.7}$$

Using the identity $\cos^2(\alpha) = 0.5(1 + \cos 2\alpha)$, we obtain

$$x_D(t) = 0.5[A_C + x_S(t)] + 0.5[A_C + x_S(t)]\cos(4\pi f_C t) \tag{10.8}$$

If $x_D(t)$ is passed through a lowpass filter whose bandwidth exceeds the message bandwidth B, the high-frequency component at $2f_C$ is blocked, and we recover $0.5[A_C + x_S(t)]$. If $x_S(t)$ contains no dc component, we can use ac coupling to eliminate the dc component $0.5A_C$ and recover the message component $0.5x_S(t)$.

This method is called **synchronous** or **coherent detection**, since it requires a signal whose frequency and phase (but not amplitude) must be synchronized or matched to the carrier signal at the transmitter. For DSBSC the coherent carrier may be obtained by transmitting a small fraction of the carrier (a *pilot signal*) along with $x_M(t)$. Synchronous detection is costly in practice.

REVIEW PANEL 10.4

Synchronous Detection of AM Signals

A carrier with the *original frequency and phase* modulates $x_M(t)$. A lowpass filter extracts $x_S(t)$.

PGR2026-00033
Krisp Technologies EX1024 Page 318

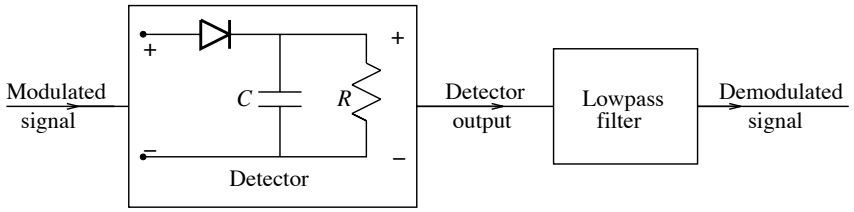


Figure 10.6 Envelope detection of AM signals

Asynchronous (Envelope) Detection

For standard AM (with $\beta \leq 1$) or DSBSC AM (with $x_S(t) \geq 0$), we can also recover the message $x_S(t)$ as the envelope of the demodulated signal using the circuit of Figure 10.6.

The diode turns on only when the input $x_D(t)$ exceeds the output. The detector output follows the message $x_S(t)$ (band-limited to B) only if the time constants are properly chosen. The detector output is smoothed by a lowpass filter to recover the message signal, as illustrated in Figure 10.7.

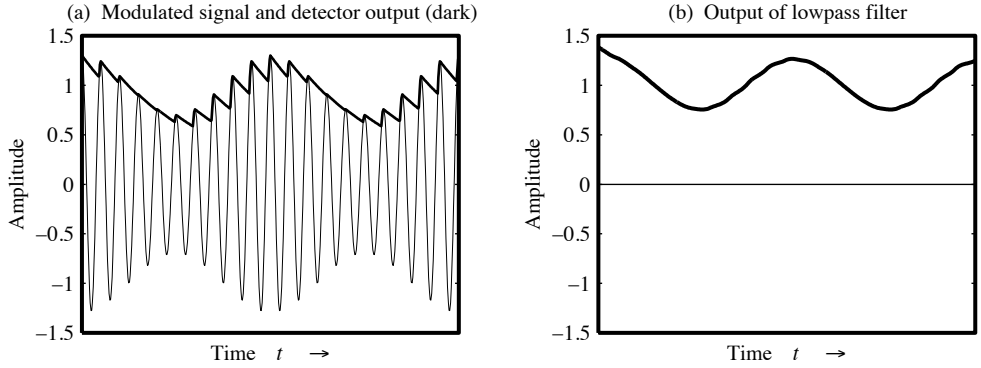


Figure 10.7 Various signals during envelope detection of AM signals

The charging constant (which equals R_DC, where R_D includes the source and diode resistance) is usually small enough (with $\tau_C \ll f_C$) to allow rapid charging and let the output follow the input (if $f_C \gg B$).

Figure 10.8 illustrates how the discharging time constant τ_D affects the detector output. The discharging time constant must be carefully chosen to satisfy $\tau_D > 1/f_C$, to ensure that the output is maintained between peaks; and $\tau_D < 1/B$, to ensure that the capacitor discharges rapidly enough to follow downward excursions of the message and does not miss any peaks. Envelope detection is an inexpensive scheme that finds widespread use in commercial AM receivers.

REVIEW PANEL 10.5

Envelope Detection of Standard AM Signals

We extract the *envelope* of $x_M(t)$ using a diode circuit and lowpass filter it to recover $x_S(t)$.
For proper recovery, $\tau_c \ll \frac{1}{f_C}$ (to follow the rising signal) and $\frac{1}{f_C} \ll \tau_d \ll \frac{1}{B}$ (to catch the peaks).

PGR2026-00033
Krisp Technologies EX1024 Page 319

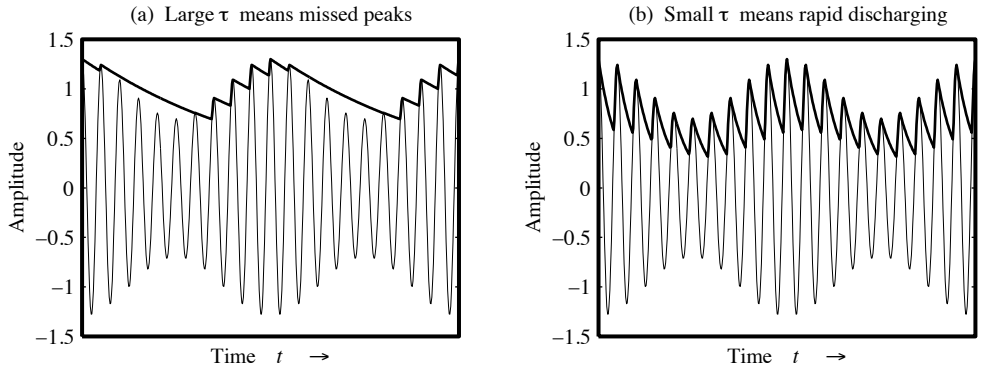


Figure 10.8 The effect of discharging time constant τ_D on the detected signal

10.1.5 Frequency Multiplexing and Demultiplexing

Since its inception, amplitude modulation has been widely used for commercial radio broadcasting. It also finds application in several other areas of communication. One such application is frequency multiplexing, which allows us to transmit many messages simultaneously, as illustrated in Figure 10.9.

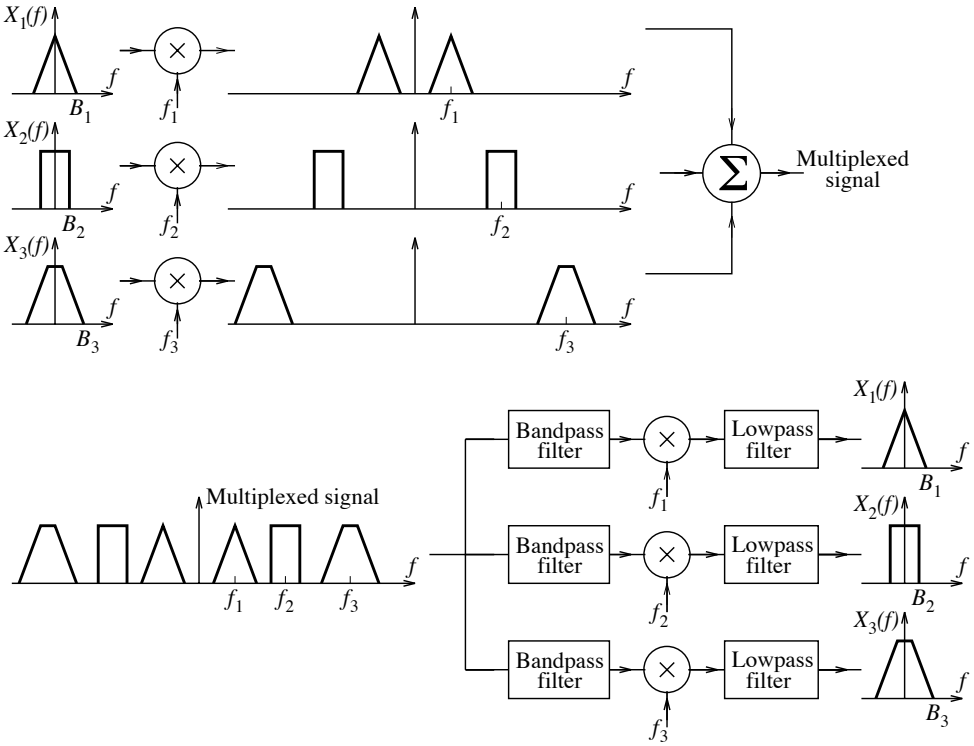


Figure 10.9 Frequency multiplexing

PGR2026-00033
Krisp Technologies EX1024 Page 320

Each message is modulated using a carrier at a different frequency in a way that the spectra of the modulated signals do not overlap. To recover the messages, the multiplexed signal is fed to bandpass filters with different center frequencies that extract the modulated signals. Each modulated signal is demodulated separately at the same carrier frequency as used in modulation. The problem is that this scheme is wasteful of equipment, requiring many local oscillators, and even more wasteful of transmission bandwidth and power.

EXAMPLE 10.2 (Frequency Multiplexing and Demultiplexing)
The spectra of four speech signals are shown in Figure E10.2. Sketch the spectrum of the frequency multiplexed signal if the carrier frequencies are chosen to provide a separation of 1.2 kHz and find the bandwidth of the multiplexed signal.

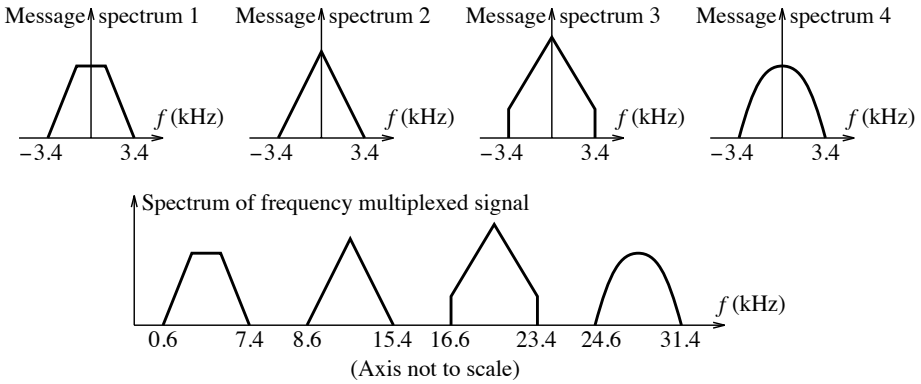


Figure E10.2 Signal spectra for Example 10.2 and spectrum of the multiplexed signal

The frequency multiplexed signal is shown in Figure E10.2. Since there are four messages, we must choose four carrier frequencies at 4, 12, 20, and 36 kHz, to provide a separation of 1.2 kHz between the messages. The spectrum of the multiplexed signal shows the message spectra in the bands 0.6–7.4 kHz, 8.6–15.4 kHz, 16.6–23.4 kHz, and 24.6–31.4 kHz. The bandwidth of the multiplexed signal is thus 31.4 kHz.

10.1.6 Heterodyning or Mixing

Heterodyning or **mixing** is a practical scheme for tuning in to messages modulated by carriers at different frequencies, or to recover a particular message from a frequency multiplexed signal. The idea is to first shift the message spectrum about a fixed *intermediate frequency* f_{IF} (for any carrier frequency) and then use a bandpass filter centered at f_{IF} and another carrier at f_{IF} to recover the message. The scheme is illustrated in Figure 10.10.

To recover the message $X_1(f)$ modulated by a carrier at f_1, it is mixed with a carrier generated by a tunable local oscillator whose frequency is adjusted to $f_1 + f_{\text{IF}}$ (or $f_1 - f_{\text{IF}}$). This produces a *heterodyned signal* with the message spectrum centered about f_{IF}. The heterodyned signal is passed through a bandpass filter (BPF) with a fixed center frequency of f_{IF} and finally demodulated by a carrier at f_{IF} to recover the message. By tuning the local oscillator, we can recover messages using carriers at different frequencies.

The main advantage of this method is that both the bandpass filter and the final demodulator operate at the fixed frequency f_{IF}, and this simplifies their design. Receivers using the principle of heterodyning are called **superhet** (or superheterodyne) receivers, even though the term *superheterodyne* originated with receivers that used a local oscillator frequency adjusted to $f_1 + f_{\text{IF}}$ (and not $f_1 - f_{\text{IF}}$).

PGR2026-00033
Krisp Technologies EX1024 Page 321

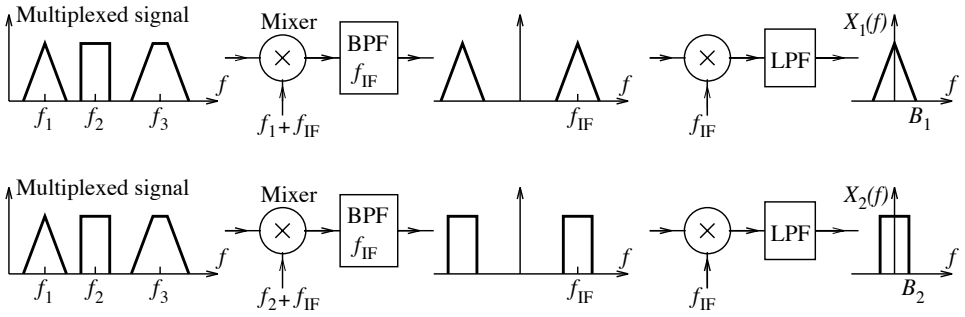


Figure 10.10 The principle of heterodyning

10.1.7 Superhet AM Receivers

In the United States of America, the carrier frequencies used for commercial AM radio stations range from 540 kHz to 1600 kHz with a separation of 10 kHz (or 30 kHz for stations in close proximity). The transmission bandwidth is 10 kHz (implying an audio message bandwidth of less than 5 kHz). The permissible carrier power ranges from 1 kW (local stations) to 50 kW, and a modulation index between 0.85 and 0.95 is required. Most low-cost commercial AM receivers have the architecture illustrated in Figure 10.11.

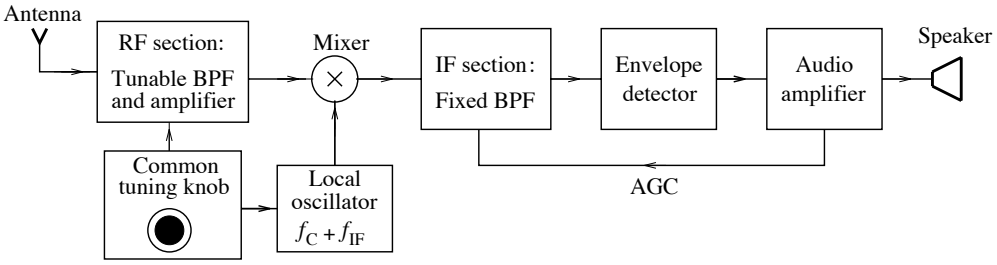


Figure 10.11 Block diagram of a superhet AM receiver

The incoming signal passes through an RF section that includes a tunable bandpass amplifier. The tuning knob serves two functions. It adjusts the center frequency of the RF amplifier to the required carrier frequency f_C. It also serves to adjust the local oscillator frequency to $f_C + f_{IF}$ in order to shift the incoming message to the fixed intermediate frequency $f_{IF} = 455$ kHz. The mixer produces the heterodyned signal, which is then filtered and amplified by an IF amplifier (with a bandwidth of 10 kHz) and demodulated using an envelope detector. The audio amplifier (a low pass filter with a cutoff frequency of about 5 kHz) raises the message power to a level suitable for the speakers. The automatic gain control (AGC) compensates for the variations in the RF signal level at the receiver input due to fading. It uses the dc offset of the detector output (which is proportional to the RF signal level) to control the gain of the IF amplifier.

For an RF tuning range of 540–1600 kHz, the frequency of the local oscillator must cover 85–1145 kHz if $f_{LO} = f_C - f_{IF}$, or 995–2055 kHz if $f_{LO} = f_C + f_{IF}$. The tuning range of 13:1 in the first case is harder to implement that the tuning range 2:1. For this reason, the local oscillator frequency is chosen as $f_{LO} = f_C + f_{IF}$ (hence the name *superheterodyne*). If the local oscillator is tuned to receive the frequency $f_C = f_{LO} - f_{IF}$, we will also pick up the *image frequency* $f_I = f_{LO} + f_{IF} = f_C + 2f_{IF}$. The bandwidth of the RF amplifier should thus be narrow enough to reject such *image signals*. Strong signals at subfrequencies (such as $f_C/2$) may also cause interference (due to nonlinearities in the IF amplifier).

PGR2026-00033
Krisp Technologies EX1024 Page 322

10.1.8 Spectrum Analyzers

Spectrum analyzers are devices that produce and display the magnitude spectrum or power spectral density (PSD) of signals. Conceptually, such a device consists of a narrow-band bandpass filter with adjustable center frequency, as shown in Figure 10.12.

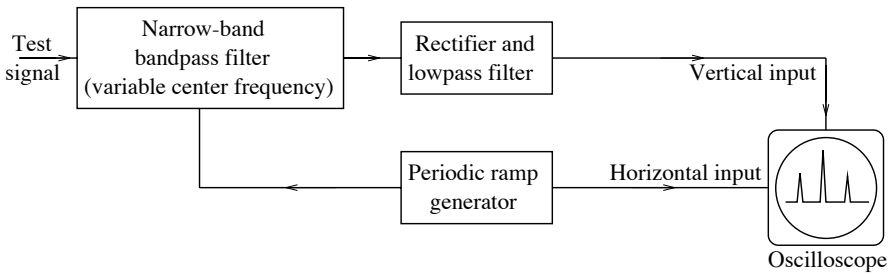


Figure 10.12 Conceptual block diagram of a spectrum analyzer

The center frequency of the bandpass filter can be varied in response to a ramp signal, which also serves as the horizontal channel input to an oscilloscope. The output of the bandpass filter is rectified to produce the rms value of the signal, filtered, and finally applied to the vertical channel input of the oscilloscope. The oscilloscope display approximates the magnitude spectrum of the input signal as the center frequency of the bandpass filter scans the frequency range of the input signal. If the rectifier is replaced by a square law detector, the display corresponds to the PSD. The accuracy of the display is governed by the sharpness of the bandpass filter and the sweep rate of the ramp generator. The narrower the passband of the bandpass filter, the better is the accuracy. The sweep rate of the ramp generator should be low enough to permit the bandpass filter to produce a steady-state response. If the filter bandwidth is Δf, its rise time will be around $1/\Delta f$. If the sweep rate (in Hz per second) is R, the filter will provide an adequate response only if $\Delta f/R > 1/\Delta f$ or $R < \Delta f^2$. The sweep rate is thus limited by the filter bandwidth.

In practice, it is difficult to design variable-frequency narrow-band bandpass amplifiers that can cover a wide range of frequencies. A practical alternative is to design a bandpass filter with a fixed center frequency f_0 and use a voltage-controlled oscillator (VCO) to tune in the signal frequencies, as illustrated in Figure 10.13.

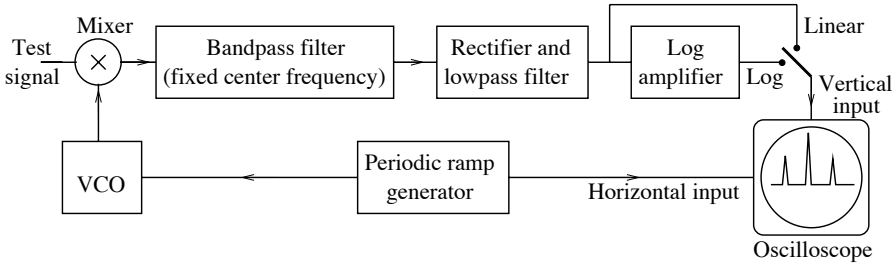


Figure 10.13 Block diagram of a practical spectrum analyzer

The VCO generates a frequency that sweeps from f_0 to $f_0 + B$, where B is the bandwidth of the test signal. At some input frequency f_{in}, the output of the mixer is at the frequencies $(f_0 + f_{\text{in}}) - f_{\text{in}} = f_0$ and $(f_0 + f_{\text{in}}) + f_{\text{in}} = f_0 + 2f_{\text{in}}$. Only the lower sideband component at f_0 is passed by the bandpass filter. The amplitude spectrum may also be displayed in decibels (dB) (using appropriate calibration) if the rectified and filtered signal is amplified by a logarithmic amplifier before being fed to the oscilloscope

PGR2026-00033
Krisp Technologies EX1024 Page 323

vertical channel. Logarithmic displays are useful in detecting small amplitude signals in the presence of large amplitude signals. Due to system imperfections and instrument noise, a spectrum analyzer does not display a decibel value of $-\infty$ if no input signal is present (as it ideally should), and its dynamic range is typically limited to between 60 and 90 dB.

10.2 Single-Sideband AM

Single-sideband (SSB) modulation makes use of the symmetry in the spectrum of an amplitude modulated signal to reduce the transmission bandwidth. Conceptually, we can transmit just the upper sideband (by using a bandpass or highpass filter) or the lower sideband (by using a bandpass or lowpass filter). This process requires filters with sharp cutoffs. A more practical method is based on the idea of the **Hilbert transform**, an operation that shifts the phase of $x(t)$ by $-\frac{\pi}{2}$. The phase shift can be achieved by passing $x(t)$ through a system whose transfer function $H(f)$ is

$$H(f) = -j\,\mathrm{sgn}(f) = \begin{cases} -j, & f > 0 \\ j, & f < 0 \end{cases} \tag{10.9}$$

In the time domain, the phase-shifted signal $\hat{x}(t)$ is given by the convolution

$$\hat{x}(t) = \frac{1}{\pi t} \star x(t) \tag{10.10}$$

The phase-shifted signal $\hat{x}(t)$ defines the Hilbert transform of $x(t)$. The spectrum $\hat{X}(f)$ of the Hilbert-transformed signal equals the product of $X(f)$ with the transform of $1/\pi t$. In other words,

$$\hat{X}(f) = -j\,\mathrm{sgn}(f)X(f) \tag{10.11}$$

A system that shifts the phase of a signal by $\pm\frac{\pi}{2}$ is called a **Hilbert transformer** or a **quadrature filter**. Such a system can be used to generate an SSB AM signal, as illustrated in Figure 10.14.

First, $x_S(t)$ is modulated by $x_C(t)$ to give the **in-phase component**:

$$x_{\mathrm{MI}}(t) = x_C(t)x_S(t) = \cos(2\pi f_C t)x_S(t) \tag{10.12}$$

Next, both $x_S(t)$ and $x_C(t)$ are shifted in phase by -90° to yield their Hilbert transforms $\hat{x}_S(t)$ and $\hat{x}_C(t)$. Note that shifting the carrier $x_C(t) = \cos(2\pi f_C t)$ by -90° simply yields $\hat{x}_C(t) = \sin(2\pi f_C t)$. Then, $\hat{x}_C(t)$ is modulated by $\hat{x}_S(t)$ to give the **quadrature component**:

$$x_{\mathrm{MQ}}(t) = \sin(2\pi f_C t)\hat{x}_S(t) \tag{10.13}$$

The lower sideband $x_{\mathrm{ML}}(t)$ and upper sideband $x_{\mathrm{MU}}(t)$ are obtained from $x_{\mathrm{MI}}(t)$ and $x_{\mathrm{MQ}}(t)$ as

$$x_{\mathrm{ML}}(t) = x_{\mathrm{MI}}(t) + x_{\mathrm{MQ}}(t) = x_S(t)\cos(2\pi f_C t) + \hat{x}_S(t)\sin(2\pi f_C t) \tag{10.14}$$

$$x_{\mathrm{MU}}(t) = x_{\mathrm{MI}}(t) - x_{\mathrm{MQ}}(t) = x_S(t)\cos(2\pi f_C t) - \hat{x}_S(t)\sin(2\pi f_C t) \tag{10.15}$$

Since the transmitted modulated signal contains only one sideband, its envelope does not correspond to the message signal.

REVIEW PANEL 10.6

The Envelope of SSB AM Does NOT Correspond to the Message Signal

Lower sideband SSB AM signal: $x_{ML}(t) = x_S(t)\cos(2\pi f_C t) + \hat{x}_S(t)\sin(2\pi f_C t)$

Upper sideband SSB AM signal: $x_{MU}(t) = x_S(t)\cos(2\pi f_C t) - \hat{x}_S(t)\sin(2\pi f_C t)$

PGR2026-00033
Krisp Technologies EX1024 Page 324

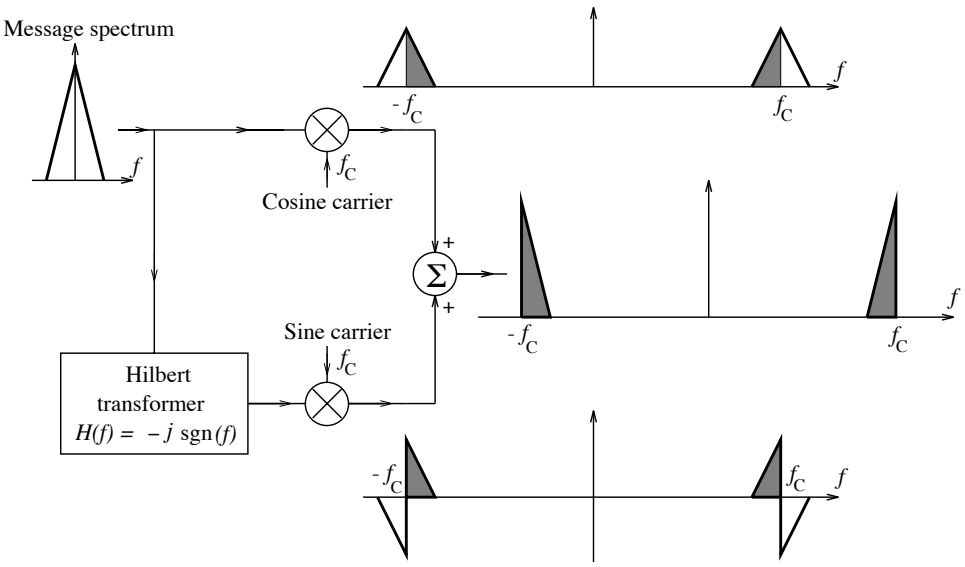


Figure 10.14 SSB amplitude modulation using the Hilbert transform

10.2.1 Detection of SSB AM Signals

The scheme for synchronous detection of SSB AM is shown in Figure 10.15.

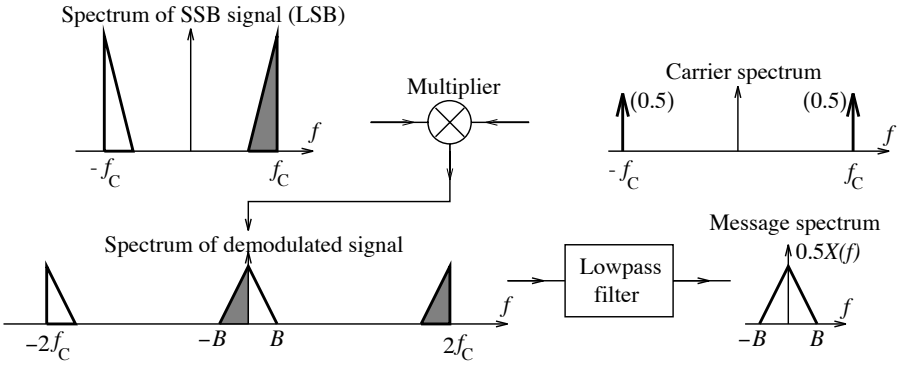


Figure 10.15 Synchronous detection of SSB AM

The demodulated signal is given by

$$x_D(t) = [x_S(t)\cos(2\pi f_C t) \mp \hat{x}_S(t)\sin(2\pi f_C t)]\cos(2\pi f_C t) \tag{10.16}$$

With $\cos^2(\alpha) = 0.5(1 + \cos 2\alpha)$ and $\sin(\alpha)\cos(\alpha) = 0.5\sin 2\alpha$, this simplifies to

$$x_D(t) = 0.5x_S(t)[1 + \cos(4\pi f_C t)] \mp 0.5\hat{x}_S(t)\sin(4\pi f_C t) \tag{10.17}$$

Once again, passing $x_D(t)$ through a lowpass filter removes the frequency components at $2f_C$ and yields the message signal $0.5x_S(t)$.

PGR2026-00033
Krisp Technologies EX1024 Page 325

EXAMPLE 10.3 (Single-Tone SSB Modulation)
Consider a single-tone message $x_S(t) = \cos(2\pi f_0 t)$, which modulates the carrier $\cos(2\pi f_C t)$, with $f_0 \ll f_C$. The in-phase component equals

$$x_{\rm MI}(t) = \cos(2\pi f_0 t)\cos(2\pi f_C t) = \tfrac{1}{2}\cos[2\pi t(f_0 - f_C)] + \tfrac{1}{2}\cos[2\pi t(f_0 + f_C)]$$

The phase-shifted (by $-90°$) signals $\sin(2\pi f_0 t)$ and $\sin(2\pi f_C t)$ yield the quadrature component

$$x_{\rm MQ}(t) = \sin(2\pi f_0 t)\sin(2\pi f_C t) = \tfrac{1}{2}\cos[2\pi t(f_0 - f_C)] - \tfrac{1}{2}\cos[2\pi t(f_0 + f_C)]$$

The sum and difference of $x_{\rm MI}(t)$ and $x_{\rm MQ}(t)$ generates the SSB signals:

$$\begin{aligned} x_{\rm ML}(t) = x_{\rm MI}(t) + x_{\rm MQ}(t) &= \cos[2\pi t(f_0 - f_C)] = \text{LSB signal} \\ x_{\rm MU}(t) = x_{\rm MI}(t) - x_{\rm MQ}(t) &= \cos[2\pi t(f_0 + f_C)] = \text{USB signal} \end{aligned}$$

The frequencies of the LSB and USB signals are $f_0 - f_C$ and $f_0 + f_C$, respectively.

10.3 Angle Modulation

Angle modulation collectively refers to schemes that vary the phase, or frequency, of a carrier signal in proportion to the message signal. Consider a sinusoidal carrier with variable phase

$$x(t) = A_C \cos[2\pi f_C t + \theta_C(t)] \tag{10.18}$$

The argument of the sinusoid may itself be regarded as an **instantaneous phase** angle $\phi_C(t)$, with

$$\phi_C(t) = 2\pi f_C t + \theta_C(t) \tag{10.19}$$

The carrier completes one full cycle as $\phi_C(t)$ changes by 2π. The **peak phase deviation** is defined as $|\theta(t)|_{\max}$.

The derivative $\phi_C'(t)$ describes the rate of change of $\phi_C(t)$ and yields the **instantaneous frequency** as

$$\omega_M(t) = \frac{d\phi_C}{dt} = 2\pi f_C + \theta_C'(t) \qquad f_M(t) = f_C + \frac{1}{2\pi}\theta_C'(t) \tag{10.20}$$

The **peak frequency deviation** equals $|\theta_C'(t)|_{\max}$ rad/s or $\frac{1}{2\pi}|\theta_C'(t)|_{\max}$ Hz.

Phase modulation (PM) refers to a scheme that varies $\phi_C(t)$ by varying the phase $\theta_C(t)$, whereas **frequency modulation** (FM) describes a scheme that varies $\phi_C(t)$ by varying the instantaneous frequency.

REVIEW PANEL 10.7

Instantaneous Phase $\phi_I(t)$ (radians) **and Instantaneous Frequency** $\omega_I(t)$ (radians/s)

For $x_M(t) = \cos[\phi_I(t)] = \cos[\omega_C t + \theta(t)]$, $\quad \phi_I(t) = \omega_C t + \theta(t) \quad \omega_I(t) = \frac{d\phi_I(t)}{dt} = \omega_C t + \frac{d\theta(t)}{dt}$

PGR2026-00033
Krisp Technologies EX1024 Page 326

10.3.1 Phase Modulation

In phase modulation, the phase $\theta_C(t)$ of the carrier is varied in proportion to $x_S(t)$. If $x_S(t)$ is constant, $\theta_C(t) = C$ and $f_M(t) = f_C$. If $x_S(t)$ shows a linear time-dependence, $\theta_C(t) = A + Bt$ and $f_M(t) = f_C + \frac{B}{2\pi}$. In either case, the instantaneous frequency $f_M(t)$ is constant. Formally, a phase-modulated signal $x_{\text{PM}}(t)$ may be written as

$$x_{\text{PM}}(t) = A_C \cos[2\pi f_C t + \theta_C(t)] = A_C \cos[2\pi f_C t + k_P x_S(t)] \tag{10.21}$$

where $\theta_C(t) = k_P x_S(t)$. The constant k_P (in radians/units of x_S) is called the **phase sensitivity**. Since $\theta_C(t)$ is cyclic, unique demodulation of a PM signal requires that $\theta_C(t)$ lie in the principal range $\pm\pi$.

The instantaneous phase and instantaneous frequency of a PM signal are given by

$$\phi_{\text{PM}}(t) = 2\pi f_C t + k_P x_S(t) \qquad f_{\text{PM}}(t) = f_C + \frac{k_P}{2\pi} x_S'(t) \tag{10.22}$$

The peak frequency deviation equals

$$\Delta f_P = \frac{k_P}{2\pi} |x_S'(t)|_{\max} \tag{10.23}$$

In analogy with AM, we also define a modulation index β_P, called the **deviation ratio**, in terms of the highest frequency B in the message as

$$\beta_P = \frac{\Delta f_P}{B} \tag{10.24}$$

10.3.2 Frequency Modulation

In frequency modulation, the instantaneous frequency of the carrier is varied linearly with $x_S(t)$ and may be written as

$$f_{\text{FM}}(t) = f_C + k_F x_S(t) \tag{10.25}$$

The constant k_F (in Hz/units of x_S) is called the **frequency sensitivity**.

Upon integrating $2\pi f_{\text{FM}}(t)$, we obtain the modulated angle $\phi_{\text{FM}}(t)$ and the frequency-modulated signal $x_{\text{FM}}(t)$ as

$$\phi_{\text{FM}}(t) = 2\pi f_C t + 2\pi k_F \int x_S(t)\, dt \qquad x_{\text{FM}}(t) = A_C \cos\Big[2\pi f_C t + 2\pi k_F \int x_S(t)\, dt\Big] \tag{10.26}$$

The peak frequency deviation Δf_F and deviation ratio β_F are given by

$$\Delta f_F = k_F |x_S(t)|_{\max} \qquad \beta_F = \frac{\Delta f_F}{f_B} \tag{10.27}$$

A PM signal may be viewed as a special case of an FM signal, with the modulating quantity as $\frac{d\,x_S(t)}{dt}$ (instead of $x_S(t)$). Similarly, an FM signal may be viewed as a special case of a PM signal, with the modulating quantity as $\int x_S(t)\, dt$ (instead of $x_S(t)$). This is illustrated in Figure 10.16.

REVIEW PANEL 10.8

Phase Modulation (PM) and Frequency Modulation (FM)

Modulated signal: $x_M(t) = \cos[\omega_C t + \theta(t)]$ **PM:** $\theta(t) = x_S(t)$ **FM:** $\theta(t) = \int x_S(t)\, dt$

PGR2026-00033
Krisp Technologies EX1024 Page 327

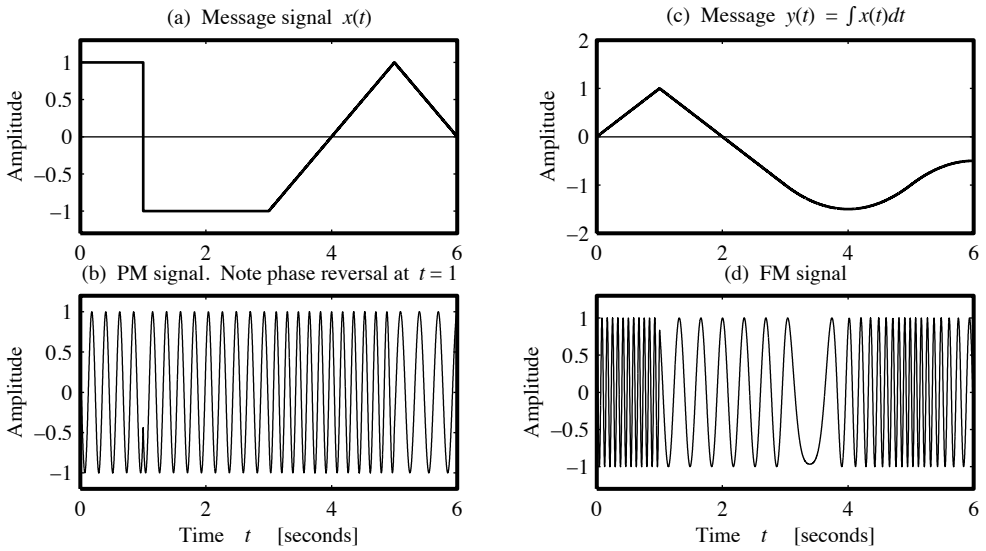


Figure 10.16 How phase modulation and frequency modulation are related

EXAMPLE 10.4 (Concepts in Angle Modulation)
Consider the angle-modulated signal $x_M(t) = \cos[200\pi t + 0.4\sin(10\pi t)]$.

(a) The carrier frequency and message frequency are $f_C = 100$ Hz and $f_0 = 5$ Hz, respectively.
Its instantaneous phase is $\phi_C(t) = 200\pi t + 0.4\sin(10\pi t)$.
The peak phase deviation thus equals 0.4 rad/s.
Its instantaneous frequency equals $f_M(t) = \frac{1}{2\pi}\phi_C'(t) = 100 + 2\cos(10\pi t)$.
The peak frequency deviation Δf thus equals 2 Hz.
The deviation ratio β then equals $\beta = \dfrac{\Delta f}{f_0} = \dfrac{2}{5} = 0.4$.

(b) If $x_M(t)$ is regarded as a PM signal, then $k_P x_S(t) = 0.4\sin(10\pi t)$.
Suppose $k_P = 5$ radians/unit. Then $x_S(t) = 0.08\sin(10\pi t)$. As a check, we evaluate
$\Delta f_P = \dfrac{k_P}{2\pi}|x_S'(t)|_{\max} = \dfrac{5(0.8\pi)}{2\pi} = 2$ Hz.

(c) If $x_M(t)$ is regarded as an FM signal, then $2\pi k_F \int x_S(t)\,dt = 0.4\sin(10\pi t)$.
Suppose $k_F = 4$ Hz/unit. Taking derivatives, we get
$2\pi k_F x_S(t) = 8\pi x_S(t) = 4\pi\cos(10\pi t)$ or $x_S(t) = 0.5\cos(10\pi t)$.
As a check, we evaluate $\Delta f_F = k_F|x_S(t)|_{\max} = 4(0.5) = 2$ Hz.

PGR2026-00033
Krisp Technologies EX1024 Page 328

10.3.3 Differences Between AM, FM, and PM

There are several major differences between amplitude-modulated signals and angle-modulated signals. In contrast to the varying envelope of AM signals, the envelope of FM or PM signals is always constant. The zero crossings of AM signals are always equispaced. Not so for PM and FM whose zero crossings are not equispaced because the angle varies with time. Unlike AM, an FM signal is a *nonlinear* function of $x_S(t)$. As a result, the spectrum of FM signals is governed by a much more complex form. As a consequence of this nonlinearity, the transmission bandwidth of FM signals is much larger than that of a corresponding AM signal.

10.3.4 Narrowband Angle Modulation

If $\theta_C(t) \ll 1$, we get a rather simple result for the modulated signal $x_M(t)$ and its spectrum $X_M(f)$. The signal $x_M(t) = A_C \cos[2\pi f_C t + \theta_C(t)]$ may be expanded as

$$x_M(t) = A_C \cos(2\pi f_C t)\cos[\theta_C(t)] - A_C \sin(2\pi f_C t)\sin[\theta_C(t)] \tag{10.28}$$

For $x \ll 1$ rad, $\cos(x) \approx 1$ and $\sin(x) \approx x$, and we can approximate $x_M(t)$ by

$$x_M(t) \approx A_C \cos(2\pi f_C t) - A_C \theta_C(t)\sin(2\pi f_C t) \tag{10.29}$$

If $\Theta_C(f)$ represents the Fourier transform of $\theta_C(t)$, the spectrum $X_M(f)$ of the modulated signal can be described by

$$X_M(f) = 0.5A_C[\delta(f + f_C) + \delta(f - f_C) - j\Theta_C(f + f_C) + j\Theta_C(f - f_C)] \tag{10.30}$$

For PM, $\theta_C(t)$ equals $k_P x_S(t)$ and $\Theta_C(f) = k_P X(f)$, and we get

$$X_{\text{PM}}(f) = 0.5A_C[\delta(f + f_C) + \delta(f - f_C) - jk_P X(f + f_C) + jk_P X(f - f_C)] \tag{10.31}$$

For FM, $\theta_C(t) = k_F \int x_S(t)\,dt$ and $\Theta_C(f) = k_F X(f)/j2\pi f$ (if we assume that $x_S(t)$ has no dc component, such that $X(0) = 0$), and we get

$$X_{\text{FM}}(f) = 0.5A_C \left[\delta(f + f_C) + \delta(f - f_C) - k_F \frac{X(f + f_C)}{(f + f_C)} + k_F \frac{X(f - f_C)}{(f - f_C)}\right] \tag{10.32}$$

Figure 10.17 compares the spectrum of standard AM, narrowband PM, and narrowband FM signals. The spectrum of both PM and FM signals contains two impulses at $f \pm f_C$, with strength $0.5A_C$, and two sidebands. The sidebands for PM retain the same shape as the message spectrum $X(f)$, but for FM, the factors $\frac{1}{f \pm f_C}$ make $X(f)$ droop as we move away from f_C. This makes FM more susceptible to noise. To counter the unwanted effects of noise, we use **pre-emphasis** to boost high frequencies in the message *before modulation* (using a differentiating filter with $|H(f)| \propto f$). At the receiver, demodulation is followed by **de-emphasis** (using a lowpass filter) to recover the original message.

REVIEW PANEL 10.9

Spectra of Narrowband (with $\theta(t) \ll 1$) PM and FM Signals

PM: The sidebands retain the shape of the message spectrum centered about the carrier frequency.
FM: The sidebands droop on either side of the carrier frequency.

PGR2026-00033
Krisp Technologies EX1024 Page 329

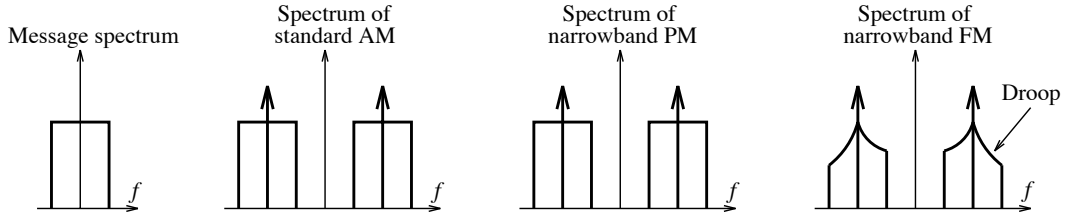


Figure 10.17 The spectra of AM, narrowband PM, and narrowband FM

EXAMPLE 10.5 (Single-Tone Angle Modulation)
Let us choose $\theta_C(t) = \beta \sin(2\pi f_0 t)$. Then $f_M(t) = f_C + \beta f_0 \cos(2\pi f_0 t)$.

Since $|\cos(2\pi f_0 t)| \leq 1,\ f_M(t)$ varies between $f_C + \beta f_0$ and $f_C - \beta f_0$, with a *peak* frequency excursion $\Delta f = \beta f_0$. For narrowband FM, $\theta_C(t) \ll 1$, and thus $\beta \ll 1$. If we substitute for $\theta_C(t)$, the modulated signal $x_M(t)$ equals

$$x_M(t) = A_C \cos[2\pi f_C t + \theta_C(t)] = A_C \cos[2\pi f_C t + \beta \sin(2\pi f_0 t)]$$

Using trigonometric identities, we expand this to

$$x_M(t) = A_C \cos(2\pi f_C t)\cos[\beta \sin(2\pi f_0 t)] - A_C \sin(2\pi f_C t)\sin[\beta \sin(2\pi f_0 t)]$$

For $x \ll 1$ rad, $\cos(x) \approx 1,\ \sin(x) \approx x$, and we have the narrowband approximation

$$x_M(t) \approx A_C \cos(2\pi f_C t) - A_C \beta \sin(2\pi f_C t)\sin(2\pi f_0 t)$$

Using trigonometric identities once again, this can be rewritten as

$$x_M(t) = A_C \cos(2\pi f_C t) + 0.5\beta A_C \cos[2\pi(f_C + f_0)t] - 0.5\beta A_C \cos[2\pi(f_C - f_0)t]$$

The spectrum of $x_M(t)$ consists of a carrier at f_C of strength A_C and two components at $f_C \pm f_0$ (both of strength $0.5\beta A_C$), whose phase is 0° at $f = f_C + f_0$ and 180° at $f = f_C - f_0$. Thus, a narrowband angle-modulated signal is similar to an AM signal except for a phase of 180° in the component at $f = f_C - f_0$.

Like AM, the bandwidth of the angle-modulated single-tone signal $x_M(t)$ also equals $2f_0$.

10.4 Wideband Angle Modulation

If $\theta_C(t)$ is not small enough, we have *wideband* angle modulation. The analysis of wideband angle modulation is often quite cumbersome. Some useful results obtain if $\theta_C(t)$ is assumed to be periodic. In such a case, we express $x_M(t)$, using Euler's identity, as

$$x_M(t) = A_C \cos[2\pi f_C t + \theta_C(t)] = \text{Re}\left\{A_C e^{j[2\pi f_C t + \theta_C(t)]}\right\} \tag{10.33}$$

Since $\theta_C(t)$ is assumed periodic, so is $e^{j\theta_C(t)}$, which can be described by its Fourier series:

$$e^{j\theta_C(t)} = \sum_{k=-\infty}^{\infty} X[k] e^{j2k\pi f_0 t} \qquad X[k] = \frac{1}{T}\int_T e^{j(\theta_C - 2k\pi f_0 t)} dt \tag{10.34}$$

PGR2026-00033
Krisp Technologies EX1024 Page 330

Substituting this into the expression for $x_M(t)$, we obtain

$$x_M(t) = \mathrm{Re}\left[A_C \sum_{k=-\infty}^{\infty} X[k] e^{j2\pi(f_C + kf_0)t}\right] \tag{10.35}$$

The spectrum of $x_M(t)$ shows components at the frequencies $f_C \pm kf_0$. Figure 10.18 shows some signals used for frequency modulation.

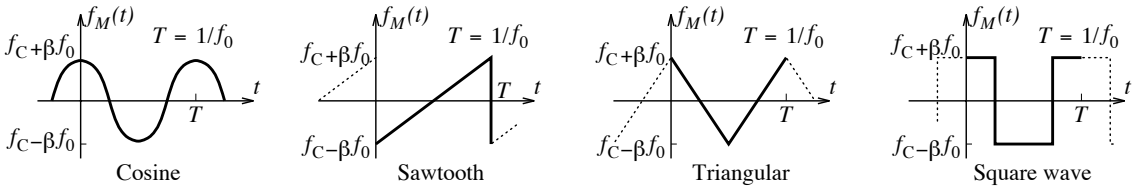


Figure 10.18 Some common modulating signals used in wideband FM

The coefficients $X[k]$ for each modulated signal are given by

Square wave: $X[k] = 0.5\,\mathrm{sinc}[0.5(\beta - k] + 0.5(-1)^{-k}\,\mathrm{sinc}[0.5(\beta + k)]$

Sinusoid: $X[k] = J_k(\beta) = \frac{1}{2\pi}\int_0^{2\pi} \exp[j(\beta \sin\tau - k\tau)]\,d\tau$. The $J_k(\beta)$ represent **Bessel functions** of integer order k and argument β.

Sawtooth wave: $X[k] = \frac{1}{\alpha}\exp(-j2\pi\lambda_k^2)\,[C(\lambda_k) - C(\gamma_k) + jS(\lambda_k) - jS(\gamma_k)]$. Here, we define $\alpha = \sqrt{2\beta}$, $\lambda = \frac{2k}{\alpha} + 0.5\alpha$, and $\gamma_k = \frac{2k}{\alpha} - 0.5\alpha$; and the quantities $C(x)$ and $S(x)$ are **Fresnel integrals** of the *first* and *second kind*, defined by

$$C(x) = \int_0^x \cos(0.5\pi\tau^2)\,d\tau \qquad S(x) = \int_0^x \sin(0.5\pi\tau^2)\,d\tau \tag{10.36}$$

Triangular wave: $X[k] = \frac{1}{\alpha}[C(\lambda_k) - C(\gamma_k)]\cos(0.5\pi\lambda_k^2) + \frac{1}{\alpha}[S(\lambda_k) - S(\gamma_k)]\sin(0.5\pi\lambda_k^2)$. Here, we define $\alpha = \sqrt{2\beta}$, $\lambda_k = \frac{k}{\alpha} + 0.5\alpha$, and $\gamma_k = \frac{k}{\alpha} - 0.5\alpha$.

Figure 10.19 shows the spectra, centered about f_C, of the resulting FM signals for each modulating signal (with $\beta = 10$). Except for square-wave modulation, the coefficients $|X[k]|$ are seen to be negligible for $|k| > \beta$, and most of the power is concentrated in the components in the range $f_C \pm \beta f_0$.

REVIEW PANEL 10.10

Spectral Coefficients $X[k]$ of Wideband FM for Periodic Modulating Signals

Sinusoid	Square Wave	Sawtooth or Triangular
$X[k]$ are Bessel functions.	$X[k]$ are sinc functions.	$X[k]$ are Fresnel integrals.

10.4.1 Bandwidth of Wideband FM

Theoretically, the bandwidth of a wideband FM signal is infinite. In practice, however, most of the power is concentrated in a few harmonics about f_C. As a result, the bandwidth is finite and still much smaller than the carrier frequency (even though it is much larger than the bandwidth for AM or narrowband FM). There are two ways to estimate the bandwidth B_{WB} of wideband FM. The first is **Carson's rule**, an empirical relation, which says

$$B_{\mathrm{WB}} = 2B(1 + \beta) = 2(\Delta f + B) \approx 2B\beta, \quad \beta > 100 \tag{10.37}$$

PGR2026-00033
Krisp Technologies EX1024 Page 331

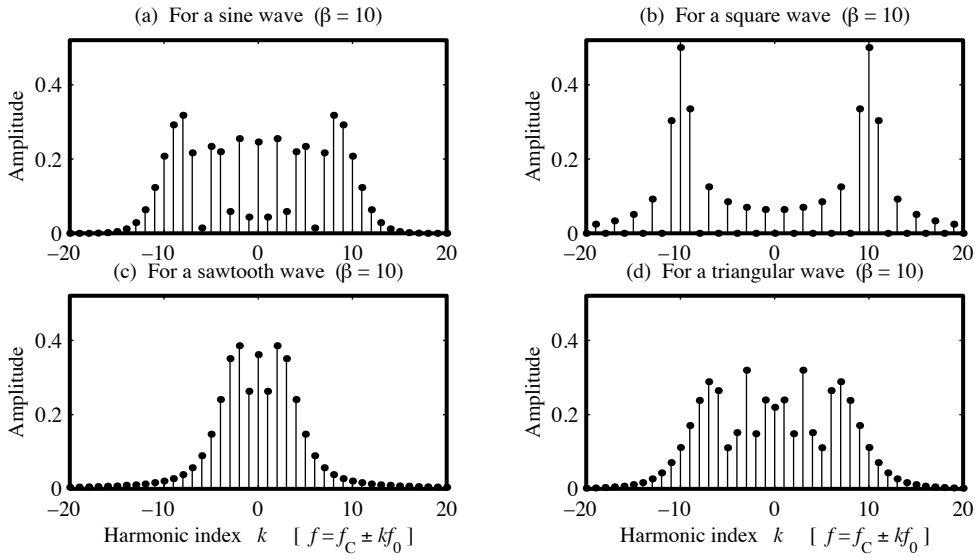


Figure 10.19 The spectrum of wideband FM using various modulating signals

For *narrowband* FM ($\beta < 0.25$ or so), the bandwidth is approximately $2B$, and more than 98% of the power lies in this frequency band for single-tone modulation. For larger values of β, Carson's rule underestimates the bandwidth, and for $\beta > 2$, it is sometimes modified to $B_{\text{WB}} = 2B(2+\beta)$. The second method is a numerical estimate based on applying Parseval's theorem to the signal $e^{j\theta_C(t)}$ and its spectral coefficients $X[k]$.

$$P = \frac{1}{T}\int_T \left|e^{j\theta_C(t)}\right|^2 dt = \sum_{k=-\infty}^{\infty} |X[k]|^2 = 1 \tag{10.38}$$

Since $|e^{j\theta_C(t)}| = 1$, the total power equals 1. The sum of $|X[k]|^2$ through a finite number of harmonics $k = N$ equals the fraction of the total power over the frequency band $f_C \pm Nf_0$. If we define B_{WB} as the range over which a large fraction—say, 98%—of the total power is concentrated, we can estimate N and compute the bandwidth as $B_{\text{WB}} \approx 2Nf_0$.

REVIEW PANEL 10.11

Bandwidth of Wideband FM Signals and Carson's Rule

Carson's rule: $B_{\text{WB}} \approx 2B(1+\beta)$ **Modified Carson's rule:** $B_{\text{WB}} \approx 2B(2+\beta)$

EXAMPLE 10.6 (Bandwidth of FM Signals)

(a) A 10-MHz carrier is modulated by a sinusoid at the frequency f_0 to yield a peak frequency deviation of 50 kHz. What is the *approximate* bandwidth of the modulated signal if $f_0 = 1$ kHz? 500 kHz?

For $f_0 = 1$ kHz, we find $\beta_F = \Delta f_F/f_0 = 50$. This value of β corresponds to wideband FM, and we may thus use Carson's rule to compute the bandwidth as $B_{\text{WB}} \approx= 2f_0(1+\beta) = 102$ kHz.

For $f_0 = 500$ kHz, we find $\beta_F = \Delta f_F/f_0 = 0.1$. This describes narrowband FM, and we may thus use the narrowband approximation to compute the bandwidth as $B_{\text{NB}} \approx 2f_0 = 1$ MHz.

PGR2026-00033
Krisp Technologies EX1024 Page 332

(b) An FM signal is generated using a single-tone sinusoidal modulating signal at a frequency f_0, with $\beta = 6$. What fraction of the total power is contained in the frequency range $f_C \pm kf_0,\ |k| \leq 4$?

For single-tone modulation, the Fourier series coefficients are $|X[k]| = |J_k(\beta)|$. With $\beta = 6$, we find $X[0] = |J_0(6)| = 0.1506$, $X[1] = |J_1(6)| = 0.2767$, $X[2] = |J_2(6)| = 0.2429$, $X[3] = |J_3(6)| = 0.1148$, and $X[0] = |J_4(6)| = 0.3576$. With $|X[-k]| = |X[k]|$, we use Parseval's relation to obtain the power as

$$P = \sum_{k=-4}^{4} |X[k]|^2 = 0.5759\,\mathrm{W}$$

This represents about 57.6% of the total power. In fact, it turns out that more than 99% of the total power is contained in the harmonics $|k| \leq 7$ and leads to a bandwidth of $B_{\mathrm{WB}} \approx 2Nf_0 = 14f_0$. This compares well with Carson's rule that says $B_{\mathrm{WB}} \approx 2f_0(1+\beta) = 14f_0$

(c) An FM signal is generated using a square-wave modulating signal whose fundamental frequency is f_0. What fraction of the total power is contained in the frequency range $f_C \pm kf_0,\ |k| \leq 4$, if $\beta = 6$?

For square-wave modulation, the Fourier series coefficients are

$$X[k] = 0.5\,\mathrm{sinc}[0.5(\beta - k)] + 0.5(-1)^k\,\mathrm{sinc}[0.5(\beta + k)]$$

With $\beta = 6$, we find $X[0] = 0$, $X[1] = 0.1091$, $X[2] = 0$, $X[3] = 0.1415$, and $X[4] = 0$.

With $|X[-k]| = |X[k]|$, we use Parseval's relation to obtain the power as

$$P = \sum_{k=-4}^{4} |X[k]|^2 = 0.0638\,\mathrm{W}$$

This represents only about 6.4% of the total power. Even though the first few harmonics contain a very small fraction of the total power, it turns out that 97.8% of the total power is contained in the harmonics $|k| \leq 7$ and leads to a bandwidth of $B_{\mathrm{WB}} \approx 2Nf_0 = 14f_0$. This also compares well with Carson's rule that says $B_{\mathrm{WB}} \approx 2f_0(1+\beta) = 14f_0$

10.5 Demodulation of FM Signals

Demodulation of FM signals is achieved using circuits called **discriminators**. Given the FM signal

$$x_{\mathrm{FM}}(t) = A_C \cos\left[2\pi f_C t + 2\pi k_F \int x_S(t)\,dt\right] \tag{10.39}$$

the idea is to generate a signal that is proportional to the modulating message signal $x_S(t)$. This may be accomplished in several ways. One commonly used method is differentiation of the FM signal followed by envelope detection. Assuming an ideal differentiator, its output will be

$$x_D(t) = x'_{\mathrm{FM}}(t) = -A_C\big[2\pi f_C + 2\pi k_F x_S(t)\big]\sin\left[2\pi f_C t + 2\pi k_F \int x_S(t)\,dt\right] \tag{10.40}$$

This describes an AM and FM signal. The AM portion contains the message that can be recovered by an envelope detector because the detector is insensitive to the FM characteristics of the signal. A practical scheme for demodulation based on this approach is illustrated in Figure 10.20.

PGR2026-00033
Krisp Technologies EX1024 Page 333

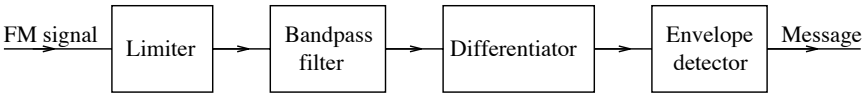


Figure 10.20 A practical scheme for demodulation of FM

The differentiator is preceded by a hard limiter and a bandpass filter. The limiter converts the FM signal to a square wave whose zero crossings correspond to those of the FM signal and thus contain information about the message signal. Its function is to remove any spurious amplitude variations that may have been introduced into the FM signal during generation or transmission. The bandpass filter is centered at the carrier frequency, and the filtered waveform is thus a sinusoidal FM signal with a constant peak amplitude (constant envelope). The differentiator introduces AM variations proportional to the message, as explained previously, and the envelope detector recovers the message.

Another method for demodulation of FM signals is based on the Hilbert transform (and described in the following section). Yet another demodulation method is based on the so-called phase-locked loop (PLL) described in Chapter 12.

10.5.1 FM Transmission

FM broadcasting was introduced in the 1930s and now includes the immensely popular stereo broadcasting (introduced in the United States of America in 1961) and (the much less frequently used) quadraphonic FM broadcasting (introduced in 1975). In the United States of America, commercial FM broadcast stations use carrier frequencies in the range 88–108 MHz, allocated by the FCC (Federal Communications Commission). The frequencies are separated by 200 kHz (to allow transmission of high-fidelity audio signals), and the frequency deviation is fixed at 75 kHz. During transmission the two stereo channels are combined to produce sum and difference signals. The difference signal amplitude-modulates a 38-kHz carrier to produce a DSB signal that is added to the sum signal. A 19-kHz pilot signal (obtained from the 38-kHz carrier by frequency division) is also added. Together, the three signals produce a stereo multiplexed signal whose spectrum $X(f)$ is shown in Figure 10.21. This multiplexed signal modulates the carrier to generate the FM signal for transmission.

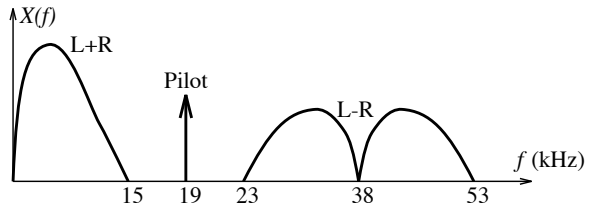


Figure 10.21 Spectrum of a stereo multiplexed signal for modulating the FM carrier

10.5.2 FM Receivers

The block diagram of a typical FM receiver is shown in Figure 10.22 and consists of an FM discriminator followed by circuits for generating either monophonic or stereo signals. For monophonic reception, the discriminator output is lowpass filtered by a filter whose bandwidth extends to 15 kHz to obtain the sum signal, which is fed to the speaker. In stereophonic reception, the discriminator output is also applied to a bandpass filter and then synchronously demodulated by a 38-kHz carrier (extracted from the pilot signal)

PGR2026-00033
Krisp Technologies EX1024 Page 334

and lowpass filtered to obtain the difference signal. Finally, the sum and difference signals are combined in a matrix to yield the left and right stereo-channel signals. In practice, pre-emphasis filters are used to combat the effect of noise during transmission (followed by de-emphasis filters at the receiver).

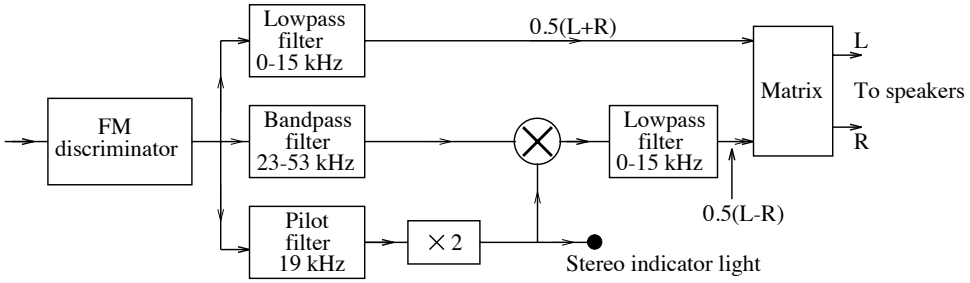


Figure 10.22 Block diagram of a stereo FM receiver

10.6 The Hilbert Transform

The Hilbert transform is an operation that shifts the phase of $x(t)$ by $-\pi/2$. The Hilbert transform of a signal $x(t)$ may be obtained by passing it through a system with a transfer function $H(f)$ given by

$$H(f) = -j\,\mathrm{sgn}(f) = \begin{cases} -j, & f > 0 \\ j, & f < 0 \end{cases} \tag{10.41}$$

This corresponds to a system whose impulse response is $h(t) = 1/\pi t$. In the time domain, the phase-shifted signal $\hat{x}(t)$ corresponds to the convolution

$$\hat{x}(t) = \frac{1}{\pi t} \star x(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\lambda)}{t-\lambda}\, d\lambda \tag{10.42}$$

The signal $\mathcal{H}[x(t)] = \hat{x}(t)$ is defined as the **Hilbert transform** of $x(t)$. Unlike most other transforms, the Hilbert transform belongs to the same domain as the signal transformed. Due to the singularity in $1/\pi t$ at $t = 0$, a formal evaluation of this relation requires complex variable theory. However, its phase-shifting property and the Fourier transform allow us to obtain the Hilbert transform of many signals without any formal integration. Some useful Hilbert transform pairs are listed in Table 10.1.

REVIEW PANEL 10.12

The Hilbert Transform Shifts the Phase Spectrum of a Signal by -90°

The Hilbert transform of a real signal is also real.

EXAMPLE 10.7 (Hilbert Transforms)

(a) The Hilbert transform of $x(t) = \cos(2\pi f t)$ is obtained by shifting the phase of $x(t)$ by $-\pi/2$ to give $\hat{x}(t) = \cos(2\pi f t - \frac{\pi}{2}) = \sin(2\pi f t)$.

(b) Likewise, the Hilbert transform of $x(t) = \sin(2\pi f t)$ is $\hat{x}(t) = \sin(2\pi f t - \frac{\pi}{2}) = -\cos(2\pi f t)$.

PGR2026-00033
Krisp Technologies EX1024 Page 335

(c) The Hilbert transform of $x(t) = e^{j2\pi ft}$ is $\hat{x}(t) = e^{j(2\pi ft - \pi/2)} = -je^{j2\pi ft} = \sin(2\pi ft) - j\cos(2\pi ft)$.

(d) The Hilbert transform of a periodic signal $x_p(t)$ with zero offset described by its Fourier series is

$$x_p(t) = \sum_{k=1}^{\infty} a_k \cos(k\omega_0 t) + \sum_{k=1}^{\infty} b_k \sin(k\omega_0 t) \qquad \hat{x}_p(t) = \sum_{k=1}^{\infty} a_k \sin(k\omega_0 t) - \sum_{k=1}^{\infty} b_k \cos(k\omega_0 t)$$

Table 10.1 Some Useful Hilbert Transform Pairs

Entry	**Signal** $x(t)$	**Hilbert Transform** $\hat{x}(t)$
1	$\cos(2\pi ft)$	$\sin(2\pi ft)$
2	$\sin(2\pi ft)$	$-\cos(2\pi ft)$
3	$x_B(t)\cos(2\pi f_C t)$ ($x_B(t)$ band-limited to $B < f_C$)	$x_B(t)\sin(2\pi f_C t)$
4	$x_B(t)\sin(2\pi f_C t)$ ($x_B(t)$ band-limited to $B < f_C$)	$-x_B(t)\cos(2\pi f_C t)$
5	$\text{rect}(t)$	$-\frac{1}{\pi}\ln\frac{\lvert t-0.5\rvert}{\lvert t+0.5\rvert}$
6	$\text{sinc}(t)$	$0.5\pi t\,\text{sinc}^2(t) = \frac{1-\cos(\pi t)}{\pi t}$
7	$\delta(t)$	$\frac{1}{\pi t}$
8	$\frac{1}{1+t^2}$	$\frac{t}{1+t^2}$
9	$\frac{1}{\pi t}$	$-\delta(t)$

10.6.1 Properties of the Hilbert Transform

Many of the properties of the Hilbert transform are based on its phase-shifting and convolutional nature. Since the convolution of real functions is real, the Hilbert transform of a real signal is also real. Since $\frac{1}{\pi t}$ is odd symmetric, the Hilbert transform (which describes a convolution) of an even symmetric signal is odd symmetric, and vice versa. Since the magnitude spectrum of $\frac{1}{\pi t}$ is unity, the magnitude spectra of $x(t)$ and its Hilbert transform $\hat{x}(t)$ are identical. The Hilbert transform of $x(t)$ taken twice in succession returns $-x(t)$ because each operation changes the phase of $x(t)$ by $-\frac{\pi}{2}$. This suggests that the **inverse Hilbert transform** of $x(t)$ equals the direct Hilbert transform with a sign change. The Hilbert transform of the scaled signal $y(t) = x(\alpha t)$ may be written as

$$\hat{y}(t) = x(\alpha t) \star \frac{1}{\pi t} = \alpha\left[x(\alpha t) \star \frac{1}{\pi \alpha t}\right] \tag{10.43}$$

The scaling property of convolution, $x(\alpha t) \star h(\alpha t) = \frac{y(\alpha t)}{|\alpha|}$, then gives

$$\hat{y}(t) = \frac{\alpha \hat{x}(\alpha t)}{|\alpha|} = \text{sgn}(\alpha)\hat{x}(\alpha t) \tag{10.44}$$

PGR2026-00033
Krisp Technologies EX1024 Page 336

This tells us, for example, that the Hilbert transform of the folded function $x(-t)$ is $-\hat{x}(-t)$.

If $x_B(t)$ is a band-limited signal and $x_C(t)$ is a signal whose spectrum does not overlap the spectrum of $x_B(t)$, the Hilbert transform of the product $x(t) = x_B(t)x_C(t)$ equals the product of $x_B(t)$ and the Hilbert transform of $x_C(t)$:

$$\hat{x}(t) = x_B(t)\hat{x}_C(t) \qquad \text{(for non-overlapping spectra)} \tag{10.45}$$

EXAMPLE 10.8 (Properties of the Hilbert Transform)

(a) The Hilbert transform of $x(t) = \cos(2\pi ft)$ is $\hat{x}(t) = \cos(2\pi ft - \frac{\pi}{2}) = \sin(2\pi ft)$.

(b) The Hilbert transform of $x(t) = \sin(2\pi ft)$ is $\hat{x}(t) = \sin(2\pi ft - \frac{\pi}{2}) = -\cos(2\pi ft) = -x(t)$.

(c) The Hilbert transform of $x(t) = \delta(t)$ is $\hat{x}(t) = \delta(t) \star \frac{1}{\pi t} = \frac{1}{\pi t}$.

(d) The Hilbert transform of $x(t) = \frac{1}{\pi t}$ is $\hat{x}(t) = \frac{1}{\pi t} \star \frac{1}{\pi t} = -\delta(t)$.

(e) The Hilbert transform of $x(t) = \text{sinc}(t)$ is $\hat{x}(t) = 0.5\pi t\,\text{sinc}^2(t)$.

Since $\text{sinc}(t)$ is even symmetric, the Hilbert transform of $y(t) = \text{sinc}(\alpha t)$ equals $\hat{y}(t) = 0.5\pi\alpha t\,\text{sinc}^2(\alpha t)$ for all $\alpha \neq 0$.

(f) Let $x_B(t)$ represent a signal band-limited to B that modulates a high-frequency carrier given by $x_C(t) = \cos(2\pi f_C t)$, with $f_C > B$, to yield the modulated signal $x_M(t) = x_B(t)\cos(2\pi f_C t)$. The Hilbert transform of $x_M(t)$ then equals $\hat{x}_M(t) = x_B(t)\sin(2\pi f_C t)$.

10.6.2 The Analytic Signal and Envelopes

Applications of the Hilbert transform to communications are based directly on the concept of the analytic signal and envelopes. The **analytic signal** $x_a(t)$ of a real signal $x(t)$ is defined by

$$x_a(t) = x(t) + j\hat{x}(t) \qquad \text{(analytic signal or pre-envelope)} \tag{10.46}$$

The real and imaginary parts of an analytic signal are Hilbert transforms. For example, the analytic signal of the real signal $x(t) = \cos(\omega t)$ is simply

$$x_a(t) = x(t) + j\hat{x}(t) = \cos(\omega t) + j\sin(\omega t) = e^{j\omega t} \tag{10.47}$$

The Fourier transform of the analytic signal $x_a(t) = x(t) + j\hat{x}(t)$ is always causal and given by

$$X_a(f) = X(f) + j[-j\,\text{sgn}(f)]X(f) = X(f)[1 + \text{sgn}(f)] = \begin{cases} 2X(f), & f > 0 \\ 0, & f < 0 \end{cases} \tag{10.48}$$

The phase of $x_a(t)$ is, however, identical to the phase of $x(t)$. Just as an analytic time signal corresponds to a causal transform, an analytic transform corresponds to a causal time signal. This is the basis for finding the causal impulse response corresponding to a given magnitude spectrum (or transfer function). We simply set up the analytic transform and find its inverse.

The analytic signal of $x(t)$ is also called the **pre-envelope** of $x(t)$. The **complex envelope** of $x(t)$ is defined, in analogy with phasors, at a given frequency f_C as

$$x_e(t) = x_a(t)e^{-j2\pi f_C t} \qquad \text{(complex envelope)} \tag{10.49}$$

PGR2026-00033
Krisp Technologies EX1024 Page 337

The **complex envelope** $x_e(t)$ is also an analytic signal in that its imaginary part equals the Hilbert transform of its real part:

$$x_e(t) = x_I(t) + jx_Q(t) = x_I(t) + j\hat{x}_I(t) \tag{10.50}$$

Note that $x_Q(t) = \hat{x}_I(t)$ is the Hilbert transform of the in-phase component. The quantities $x_I(t)$ and $x_Q(t)$ are called the **in-phase** and **quadrature** components of $x(t)$ at the frequency f_C. In analogy with phasors, we may also describe the real signal $x(t)$ in terms of its complex envelope at the frequency f_C as

$$x(t) = \text{Re}[x_a(t)] = \text{Re}\{x_e(t)e^{j2\pi f_C t}\} \tag{10.51}$$

The signal $x(t)$ may then be expressed in terms of its in-phase and quadrature components as

$$x(t) = \text{Re}\big[\{x_I(t) + jx_Q(t)\}e^{j2\pi f_C t}\big] = x_I(t)\cos(2\pi f_C t) - x_Q(t)\sin(2\pi f_C t) \tag{10.52}$$

This has the same form as a USB SSB AM signal. Since $x_e(t) = x_a(t)e^{-j2\pi f_C t}$, its spectrum can be found using the modulation property to give

$$X_e(f) = X_a(f + f_C) = 2X(f + f_C)u(f + f_C) \tag{10.53}$$

The spectrum of $x_e(t)$ is also one-sided. If $x(t)$ represents a signal band-limited to B such that $f_C \geq B$, the spectrum $X_e(f)$ is centered at f_C and confined to the frequency band $(f_C - B,\ f_C + B)$

The **natural envelope** of $x(t)$ is a real quantity that represents the magnitude of $x_a(t)$ or $x_e(t)$. If the spectrum $X(f)$ undergoes a constant phase change to $X(f)e^{j\phi}\text{sgn}(f)$, the analytic signal changes to $x_a(t)[\cos(\phi) + j\sin(\phi)]$, whose envelope is still $|x_a(t)|$ and thus invariant to the phase change. Since $a(t) \geq |x(t)|$, the natural envelope encloses the family of all possible signals we can generate from $x(t)$ by changing its phase spectrum by an arbitrary constant ϕ. At the instants where $a(t) = x(t)$, the Hilbert transform $\hat{x}(t)$ equals zero.

REVIEW PANEL 10.13

Envelopes of a Real Signal $x(t)$

Pre-Envelope (Analytic Signal)	**Complex Envelope**	**Natural Envelope**
$x_a(t) = x(t) + j\hat{x}(t)$	$x_e(t) = x_a(t)e^{-j2\pi f_C t}$	$a(t) = \lvert x_e(t)\rvert = \lvert x_a(t)\rvert$

EXAMPLE 10.9 (Envelopes)

(a) Let $x(t) = \cos(\alpha t)$. Its pre-envelope is $x_a(t) = \cos(\alpha t) + j\sin(\alpha t) = e^{j\alpha t}$. Its natural envelope is $a(t) = |x_a(t)| = 1$. This makes sense because the sinusoid $x(t)$ always has unit magnitude for any constant phase shift!

(b) Let $x(t) = \text{sinc}(t)$. Its pre-envelope is

$$x_a(t) = x(t) + j\hat{x}(t) = \frac{\sin(\pi t)}{\pi t} + j\frac{1 - \cos(\pi t)}{\pi t}$$

Its natural envelope $a(t) = |x_a(t)$ simplifies to

$$a(t) = \left|\frac{\sin(\pi t)}{\pi t} + j\frac{1 - \cos(\pi t)}{\pi t}\right| = \text{sinc}(0.5t)$$

This encloses all the signals we can generate from $\text{sinc}(t)$ by changing its phase spectrum by an arbitrary constant ϕ.

PGR2026-00033
Krisp Technologies EX1024 Page 338

(c) Consider the AM signal $x_M(t) = f(t)\cos(2\pi f_C t + \phi)$. Its pre-envelope is

$$x_a(t) = x_M(t) + j\hat{x}_M(t) = f(t)\cos(2\pi f_C t + \phi) + jf(t)\sin(2\pi f_C t + \phi) = f(t)e^{j(2\pi f_C t + \phi)}$$

Its complex envelope is $x_e(t) = x_a(t)e^{-j2\pi f_C t} = f(t)e^{j\phi}$. The natural envelope is $a(t) = |f(t)|$. The in-phase and quadrature components are

$$x_{MI}(t) = f(t)\cos(\phi) \qquad x_{MQ}(t) = f(t)\sin(\phi)$$

(d) Consider the PM signal $x_M(t) = \cos[2\pi f_C t + \phi(t)]$. Its pre-envelope is

$$x_a(t) = x_M(t) + j\hat{x}_M(t) = \cos[2\pi f_C t + \phi(t)] + j\sin[2\pi f_C t + \phi(t)] = e^{j[2\pi f_C t + \phi(t)]}$$

Its complex envelope is $x_e(t) = e^{j\phi(t)}$. Its natural envelope is $a(t) = 1$. Its in-phase and quadrature components are

$$x_{MI}(t) = \cos[(\phi(t)] \qquad x_{MQ}(t) = \sin[\phi(t)]$$

10.6.3 Detection of Modulated Signals by the Hilbert Transform

The Hilbert transform and the pre-envelope may be used for the detection of both amplitude-modulated signals and angle-modulated signals. Consider the modulated signal $x_M(t)$ given by

$$x_M(t) = a(t)\cos[2\pi f_C t + \theta(t)] \tag{10.54}$$

This represents the most general form of a modulated signal describing both amplitude and phase modulation of the carrier. In particular, $\theta(t) = 0$ for AM and $a(t) = 1$ for PM or FM. To demodulate $x_M(t)$, we first form its pre-envelope:

$$x_a(t) = x_M(t) + j\hat{x}_M(t) \tag{10.55}$$

If $a(t)$ is band-limited to a frequency $B < f_C$, then $\hat{x}_M(t) = a(t)\sin[2\pi f_C t + \theta(t)]$, and we get

$$x_a(t) = a(t)\cos[2\pi f_C t + \theta(t)] + ja(t)\sin[2\pi f_C t + \theta(t)] = a(t)e^{j[2\pi f_C t + \theta(t)]} \tag{10.56}$$

Finally, we multiply $x_a(t)$ by $e^{-j2\pi f_C t}$ to get

$$x_D(t) = x_a(t)e^{-j2\pi f_C t} = a(t)e^{j[2\pi f_C t + \theta(t)]}e^{-j2\pi f_C t} = a(t)e^{j\theta(t)} \tag{10.57}$$

The message signal corresponds to $\mathrm{Re}[x_D(t)$ for AM, to $\{\angle x_S(t)\}$ for PM, and $\frac{d\{\angle x_S(t)\}}{dt}$ for FM.

REVIEW PANEL 10.14

Detection of AM, FM, or PM Signals, Using Envelopes

Modulated signal: $x_M(t) = a(t)\cos[2\pi f_C t + \theta(t)]$

Pre-envelope: $x_a(t) = x_M(t) + j\hat{x}_M(t)$ **Demodulated signal:** $x_D(t) = x_a(t)e^{-j2\pi f_C t}$

Recovered message: For AM: $\mathrm{Re}\{x_D(t)\}$ For PM: $\{\angle x_D(t)\}$ For FM: $\frac{d\{\angle x_D(t)\}}{dt}$

PGR2026-00033
Krisp Technologies EX1024 Page 339

CHAPTER 10 PROBLEMS

DRILL AND REINFORCEMENT

10.1 (DSBSC AM) Consider the message $x(t) = 2\cos(2\pi f_1 t)$ used to modulate the carrier $\cos(2\pi f_C t)$, to generate a DSBSC AM signal $x_M(t)$.

(a) Write an expression for $x_M(t)$ and sketch its spectrum.

(b) Sketch the spectrum of the demodulated signal if demodulation is achieved by $\cos(2\pi f_C t)$.

10.2 (DSBSC AM) A 1.5-MHz carrier is modulated by a music signal whose frequencies range from 50 Hz to 15 kHz. What is the range of frequencies over which the upper and lower sidebands extend?

10.3 (Standard AM) Consider the message $x(t) = 2\cos(2\pi f_1 t) + \cos(2\pi f_2 t)$ used to modulate the carrier $\cos(2\pi f_C t)$ to generate $x_M(t) = [A_C + x(t)]\cos(2\pi f_C t)$.

(a) What value of A_C ensures a modulation index $\beta = 0.5$?

(b) What is the efficiency η when $\beta = 0.5$?

(c) Sketch the spectrum of $x(t)$ and $x_M(t)$ if f_1 =10 Hz, f_2 =20 Hz, and f_C =100 Hz.

10.4 (Standard AM) An AM station operates with a modulation index of 0.8 and transmits a total power of 50 kW.

(a) What is the power in the transmitted carrier?

(b) What fraction of the total power resides in the message?

10.5 (Standard AM) Find the efficiency of a single-tone AM modulator with a modulation index of 0.5 if the carrier power equals 32 W.

10.6 (Synchronous Demodulation) The signal $x(t) = 2\cos(2\pi f_1 t) + \cos(2\pi f_2 t)$ is used to modulate the carrier $\cos(2\pi f_C t)$, to generate $x_M(t) = [A_C + x(t)]\cos(2\pi f_C t)$. The modulated signal $x_M(t)$ is synchronously demodulated using the same carrier signal.

(a) Sketch the spectrum of the demodulated signal.

(b) Sketch the spectrum of an ideal lowpass filter that can be used to recover $x(t)$.

10.7 (Envelope Detection) The following signals are used to generate DSBSC AM signals. Which signals can be recovered using envelope detection?

(a) $x(t) = \cos(2\pi f_1 t)$ **(b)** $x(t) = 2\cos(2\pi f_1 t) + \cos(2\pi f_2 t)$

(c) $x(t) = 2 + \cos(2\pi f_1 t)$ **(d)** $x(t) = 2 + 2\cos(2\pi f_1 t) + \cos(2\pi f_2 t)$

10.8 (SSB AM) The message $x(t) = 2\cos(2\pi f_1 t) + \cos(2\pi f_2 t)$ modulates the carrier $\cos(2\pi f_C t)$, to generate $x_M(t) = x(t)\cos(2\pi f_C t)$. Find expressions for the upper and lower sideband components of $x_M(t)$.

10.9 (Instantaneous Frequency) Find the instantaneous frequency of the following signals.

(a) $x(t) = \cos(10\pi t + \frac{\pi}{4})$ **(b)** $x(t) = \cos(10\pi t + 2\pi t)$

(c) $x(t) = \cos(10\pi t + 2\pi t^2)$ **(d)** $x(t) = \cos[10\pi t + 2\sin(2\pi t)]$

PGR2026-00033
Krisp Technologies EX1024 Page 340

10.10 (FM) An FM signal is described by $x(t) = A\cos[2\pi 10^7 t + 50\sin(2\pi 10^4 t)]$.

(a) Identify the carrier frequency.
(b) Identify the frequency of the modulating signal.
(c) Find the peak frequency deviation and modulation index.

10.11 (FM) The peak deviation in commercial FM is 75 kHz. The frequency of the modulating signal varies between 50 Hz and 15 kHz. What is the permissible range for the modulation index?

10.12 (FM) A signal, band-limited to 15 kHz, is to be transmitted using FM. The peak deviation is 30 kHz. What bandwidth is required? Use Carson's rule or its modification as appropriate.

REVIEW AND EXPLORATION

10.13 (DSBSC AM) Consider the message $x(t) = 2\cos(2\pi f_1 t) + \cos(2\pi f_2 t)$ used to modulate the carrier $\cos(2\pi f_C t)$ to generate a DSBSC AM signal $x_M(t)$.

(a) Write an expression for $x_M(t)$ and sketch its spectrum.
(b) What fraction of the total signal power is contained in the sidebands?
(c) What is the modulation efficiency?

10.14 (Amplitude Modulation) A message signal $x(t)$ containing three unit-amplitude cosines at 100, 200, and 300 Hz is used to modulate a carrier at $f_C = 10$ kHz to generate the AM signals $x_1(t)$ and $x_2(t)$. Sketch the spectrum of each AM signal and compute the power in the sidebands as a fraction of the total power.

(a) $x_1(t) = [1 + 0.2x(t)]\cos(2\pi f_C t)$
(b) $x_2(t) = x(t)\cos(2\pi f_C t)$

10.15 (Standard AM) The periodic signal $x(t) = 4 + 4\cos(2\pi t) + 2\cos(4\pi t)$ modulates the carrier $\cos(40\pi t)$ to generate the standard AM signal $x_{\rm AM}(t) = [20 + x(t)]\cos(40\pi t)$. What is the modulation index? What is the power in the AM signal? What fraction of the total signal power resides in the sidebands?

10.16 (Standard AM) Consider the modulated signal

$$x_M(t) = 1000\cos(20000\pi t)\left[1 + \sum_k \frac{1}{2k}\cos(2000k\pi t)\right], \qquad k = 1, 2, 5$$

(a) What frequencies are present in the modulated signal?
(b) What are the magnitudes of the components at each frequency?
(c) What is the total transmitted power?
(d) What is the modulation index?
(e) What is the efficiency?

10.17 (Standard AM) Consider a periodic signal $x(t)$ with $|x(t)| < 1$, zero dc offset, band-limited to B Hz, and with power P. This signal modulates a high-frequency carrier $A\cos(2\pi f_C t)$, $f_C \gg B$, to generate the standard AM signal $x_{\rm AM}(t) = A[1 + x(t)]\cos(2\pi f_C t)$. What is the power in the AM signal as a function of P?

PGR2026-00033
Krisp Technologies EX1024 Page 341

10.18 (Synchronous Detection) Synchronous detection requires both phase and frequency coherence. Let the signal $x(t) = 2\cos(2\pi f_0 t)$ be used to modulate the carrier $\cos(2\pi f_C t)$ to generate a DSBSC AM signal $x_M(t)$.

(a) What is the spectrum of the demodulated signal if the demodulating carrier is $\cos(2\pi f_C t + \theta)$?
(b) Are there any values of θ for which it is impossible to recover the message?
(c) Let the demodulating carrier have a frequency offset to become $\cos[2\pi(f_C + \Delta f)t]$. Find the spectrum of the demodulated signal. Do we recover the message signal in this case?

10.19 (Synchronous Detection) Let $x(t) = \cos(100\pi t)$ be used to modulate the carrier $\cos(2000\pi t)$ to generate a DSBSC AM signal $x_M(t)$. The modulated signal $x_M(t)$ is demodulated using the signal $\cos[2\pi(1000+\Delta f)t+\theta]$. For each case, explain whether (and how) the original signal can be recovered from the demodulated signal.

(a) $\Delta f = 0$ $\quad \theta = 0$
(b) $\Delta f = 0$ $\quad \theta = 0.25\pi$
(c) $\Delta f = 0$ $\quad \theta = 0.5\pi$
(d) $\Delta f = 10$ Hz $\quad \theta = 0$
(e) $\Delta f = 10$ Hz $\quad \theta = 0.5\pi$

10.20 (Envelope Detection) A standard AM signal $x_M(t) = [A_C + x(t)]\cos(2\pi f t)$ is to be demodulated by an envelope detector. The message $x(t)$ is band-limited to 10 kHz, and the frequency of the carrier signal is 100 kHz. What are the allowable limits on the detector time constant to avoid distortion?

10.21 (SSB AM) The message $x(t) = \text{sinc}^2(t)$ modulates the carrier $\cos(10\pi t)$ to generate the modulated signal $x_M(t) = x(t)\cos(2\pi f_C t)$.

(a) Sketch the spectrum of $x(t)$, the carrier, and $x_M(t)$.
(b) Sketch the spectrum of the LSB SSB signal.
(c) The signal in part (b) is synchronously demodulated. Sketch the spectrum of the demodulated signal.

10.22 (Wideband FM) Consider an FM signal with $\beta = 5$. Find the power contained in the harmonics $f_C \pm kf_0$, $k = 0, 1, 2, 3$, for the following:

(a) Single-tone sinusoidal modulation **(b)** Square-wave modulation

10.23 (The Hilbert Transform) Unlike most other transforms, the Hilbert transform belongs to the same domain as the signal transformed. The Hilbert transform shifts the phase of $x(t)$ by $-\frac{\pi}{2}$. Find the Hilbert transforms of the following:

(a) $x(t) = \cos(2\pi f t)$ **(b)** $x(t) = \sin(2\pi f t)$ **(c)** $x(t) = \cos(2\pi f t) + \sin(2\pi f t)$
(d) $x(t) = e^{-j2\pi f t}$ **(e)** $x(t) = \delta(t)$ **(f)** $x(t) = \frac{1}{\pi t}$

10.24 (Properties of the Hilbert Transform) Using some of the results of the preceding problem as examples, or otherwise, verify the following properties of the Hilbert transform.

(a) The magnitude spectra of $x(t)$ and $\hat{x}(t)$ are identical.
(b) The Hilbert transform of $x(t)$ taken twice returns $-x(t)$.
(c) The Hilbert transform of an even function is odd, and vice versa.
(d) The Hilbert transform of $x(\alpha t)$ is $\text{sgn}(\alpha)\hat{x}(\alpha t)$.
(e) The Hilbert transform of $x(t) \star y(t)$ equals $x(t) \star \hat{y}(t)$ or $\hat{x}(t) \star y(t)$.
(f) The Hilbert transform of a real signal is also real.

PGR2026-00033
Krisp Technologies EX1024 Page 342

10.25 (Scaling) The Hilbert transform of $x(t) = \text{sinc}(t)$ is given by $\hat{x}(t) = 0.5\pi t\,\text{sinc}^2(t)$. Use the scaling property to find the Hilbert transform of $y(t) = \text{sinc}(\alpha t)$.

10.26 (Modulation) A signal $x_B(t)$ band-limited to B Hz modulates a carrier $x_C(t) = \cos(2\pi f_C t)$. If $f_C > B$, show that the Hilbert transform of the modulated signal $x_M(t) = x_B(t)\cos(2\pi f_C t)$ is given by the expression $\hat{x}_M(t) = x_B(t)\sin(2\pi f_C t)$.

COMPUTATION AND DESIGN

modgui A GUI for Visualization of Concepts in Modulation

The graphical user interface `modgui` allows you to visualize modulated signals.
You can select the modulation type and other parameters.
For AM, you can also display the demodulated signal using envelope detection.
To explore this routine, type `modgui` at the MATLAB prompt.

10.27 (Amplitude Modulation) The signal $x(t) = \cos(20\pi t)$ modulates a carrier at $f_C = 100$ Hz. Obtain and plot the modulated signal for each modulation scheme listed. For which cases does the message signal correspond to the envelope of the modulated signal? For which cases can the message signal be recovered by envelope detection? For which cases can the message signal be recovered by synchronous detection?

(a) DSBSC AM
(b) Standard AM with a modulation index of 0.5.
(c) Standard AM with a modulation index of 0.8.
(d) Standard AM with a modulation index of 1.2.
(e) SSB AM with only the upper sideband present.

10.28 (Demodulation of AM Signals) The signal $x(t) = \cos(20\pi t)$ modulates a carrier at $f_C = 100$ Hz. Generate the modulated signals as in Problem 10.27. To demodulate the signals, multiply each modulated signal by a carrier signal whose frequency and phase are identical to the modulated carrier and plot the resulting signal. How would you extract the message from the demodulated signal?

10.29 (Amplitude Modulation) The phase of the carrier is important in demodulating AM signals by synchronous detection. Generate the modulated signals as in Problem 10.27. Demodulate each signal, using the following carrier signals. For which cases can you recover the message exactly? Are there any cases for which the recovered message signal equals zero? Explain.

(a) Frequency f_0, same phase as the modulating carrier
(b) Frequency f_0, phase offset = 5°
(c) Frequency f_0, phase offset = 30°
(d) Frequency f_0, phase offset = 90°
(e) Frequency $1.1f_0$, phase offset = 0°

PGR2026-00033
Krisp Technologies EX1024 Page 343

Chapter 11

THE LAPLACE TRANSFORM

11.0 Scope and Objectives

Broadly speaking, the Laplace transform (LT) is a very useful tool for *system analysis*, much as the Fourier transform is a very useful tool for *signal analysis*. The genesis of the Laplace transform may be viewed as arising from the response (using convolution) of a linear system to the complex exponentials $\exp(st)$ that are its eigensignals, or as a modified version of the Fourier transform. To keep the discussion self-contained, we start with the Laplace transform as an independent transformation method and explore the connections between various transform methods toward the end of this chapter. This chapter begins with an introduction to the Laplace transform and its properties and the inverse Laplace transform. It then describes the Laplace transform as a tool for system analysis and concludes by explaining its connection with other transform methods.

11.1 The Laplace Transform

The **Laplace transform** $X(s)$ of a signal $x(t)$ is defined as

$$X(s) = \int_{0-}^{\infty} x(t)e^{-(\sigma+j\omega)t}\,dt = \int_{0-}^{\infty} x(t)e^{-st}\,dt \tag{11.1}$$

We also define the two-sided Laplace transform by allowing the lower limit to be $-\infty$. However, this form finds limited use and is not pursued here.

The complex quantity $s = \sigma + j\omega$ generalizes the concept of frequency to the complex domain. It is often referred to as the **complex frequency**, with ω measured in radians/second, and σ in nepers/second. The relation between $x(t)$ and $X(s)$ is denoted symbolically by the **transform pair** $x(t) \Leftrightarrow X(s)$. The double arrow suggests a one-to-one correspondence between the signal $x(t)$ and its Laplace transform $X(s)$.

The Laplace transform provides unique correspondence only for causal signals, since the Laplace transform of $x(t)$ and its causal version $x(t)u(t)$ are clearly identical due to the lower limit of 0. To include signals such as $\delta(t)$, which are discontinuous at the origin, the lower limit is chosen as $0-$ whenever appropriate.

REVIEW PANEL 11.1

Laplace Transform of Causal Signals

$X(s) = \int_0^{\infty} x(t)e^{st}dt$ (lower integration limit is $t = 0-$ if $x(t)$ contains impulses)

PGR2026-00033
Krisp Technologies EX1024 Page 344

11.1.1 Convergence

The Laplace transform is an *integral operation*. It exists if $x(t)e^{-\sigma t}$ is absolutely integrable. Clearly, only certain choices of σ will make this happen. The range of σ that ensures existence defines the **region of convergence** (ROC) of the Laplace transform. For example, the Laplace transform $X(s)$ of the signal $x(t) = e^{2t}u(t)$ exists only if $\sigma > 2$, which ensures that the product is a decaying exponential. Thus, $\sigma > 2$ defines the region of convergence of the transform $X(s)$. Since we shall deal only with causal signals (whose ROC is $\sigma > -\alpha,\ \alpha \geq 0$), we avoid explicit mention of the ROC (except in Example 11.1).

11.1.2 Some Laplace Transform Pairs

We use the defining relation to develop a stock list of Laplace transforms. These and other useful transform pairs are listed in Table 11.1.

Table 11.1 A Short Table of Laplace Transforms

Entry	$x(t)$	$X(s)$	**Entry**	$x(t)$	$X(s)$
1	$\delta(t)$	1	10	$t\cos(\beta t)u(t)$	$\dfrac{s^2-\beta^2}{(s^2+\beta^2)^2}$
2	$u(t)$	$\dfrac{1}{s}$	11	$t\sin(\beta t)u(t)$	$\dfrac{2s\beta}{(s^2+\beta^2)^2}$
3	$r(t)=tu(t)$	$\dfrac{1}{s^2}$	12	$\cos^2(\beta t)u(t)$	$\dfrac{s^2+2\beta^2}{s(s^2+4\beta^2)}$
4	$t^2u(t)$	$\dfrac{2}{s^3}$	13	$\sin^2(\beta t)u(t)$	$\dfrac{2\beta^2}{s(s^2+4\beta^2)}$
5	$t^nu(t)$	$\dfrac{n\,!}{s^{n+1}}$	14	$e^{-\alpha t}\cos(\beta t)u(t)$	$\dfrac{s+\alpha}{(s+\alpha)^2+\beta^2}$
6	$e^{-\alpha t}u(t)$	$\dfrac{1}{s+\alpha}$	15	$e^{-\alpha t}\sin(\beta t)u(t)$	$\dfrac{\beta}{(s+\alpha)^2+\beta^2}$
7	$te^{-\alpha t}u(t)$	$\dfrac{1}{(s+\alpha)^2}$	16	$te^{-\alpha t}\cos(\beta t)u(t)$	$\dfrac{(s+\alpha)^2-\beta^2}{[(s+\alpha)^2+\beta^2]^2}$
8	$t^ne^{-\alpha t}u(t)$	$\dfrac{n!}{(s+\alpha)^{n+1}}$	17	$te^{-\alpha t}\sin(\beta t)u(t)$	$\dfrac{2\beta(s+\alpha)}{[(s+\alpha)^2+\beta^2]^2}$
9	$\cos(\beta t)u(t)$	$\dfrac{s}{s^2+\beta^2}$	18	$\sin(\beta t)u(t)$	$\dfrac{\beta}{s^2+\beta^2}$

EXAMPLE 11.1 (Laplace Transforms from the Defining Relation)

(a) (Unit Impulse) Using the sifting property, the transform of $x(t) = \delta(t)$ is given by

$$X(s) = \int_{0-}^{\infty} \delta(t)e^{-st}\,dt = 1$$

PGR2026-00033
Krisp Technologies EX1024 Page 345

The choice 0− for the lower limit is due to the fact that $\delta(t)$ is discontinuous at $t = 0$. The ROC is the entire s plane.

(b) (Unit Step) From the defining relation, the transform of $x(t) = u(t)$ is given by

$$X(s) = \int_{0-}^{\infty} u(t)e^{-st}\,dt = \int_{0-}^{\infty} e^{-st}\,dt = \frac{1}{s}$$

Since $\exp(-st) = \exp[-(\sigma + j\omega)t]$, which equals zero at the upper limit only if $\sigma > 0$, the region of convergence of $X(s)$ is $\sigma > 0$.

(c) (Unit Ramp) The transform of $x(t) = r(t) = tu(t)$ is given by

$$X(s) = \int_{0}^{\infty} r(t)e^{-st}\,dt = \int_{0}^{\infty} te^{-st}\,dt = \frac{1}{s^2}$$

The region of convergence for this result is also $\sigma > 0$.

(d) (Decaying Exponential) The transform of $x(t) = e^{-\alpha t}u(t)$ is given by

$$X(s) = \int_{0}^{\infty} e^{-\alpha t}e^{-st}\,dt = \int_{0}^{\infty} e^{-(s+\alpha)t}\,dt = \frac{1}{s+\alpha}$$

The region of convergence of $X(s)$ is $\sigma > -\alpha$.

(e) (Switched Cosine) With $x(t) = \cos(\alpha t)u(t)$, we use Euler's relation to get

$$X(s) = \int_{0}^{\infty} \cos(\alpha t)e^{-st}\,dt = \frac{1}{2}\int_{0}^{\infty} \left[e^{-(s+j\alpha)t} + e^{-(s-j\alpha)t}\right] dt = \frac{s}{s^2+\alpha^2}$$

The region of convergence of $X(s)$ is $\sigma > 0$.

REVIEW PANEL 11.2

Four Basic Laplace Transform Pairs

$$\delta(t) \Longleftrightarrow 1 \qquad u(t) \Longleftrightarrow \frac{1}{s} \qquad r(t) \Longleftrightarrow \frac{1}{s^2} \qquad e^{-\alpha t}u(t) \Longleftrightarrow \frac{1}{s+\alpha}$$

11.2 Properties of the Laplace Transform

The Laplace transform is a linear operation, and its properties are summarized in Table 11.2. Most of these properties follow from the defining integral.

The **times-exp** property is obtained directly from the defining integral:

$$\int_{0}^{\infty} x(t)e^{-\alpha t}e^{-st}\,dt = \int_{0}^{\infty} x(t)e^{-(s+\alpha)t}\,dt = X(s+\alpha) \tag{11.2}$$

PGR2026-00033
Krisp Technologies EX1024 Page 346

Table 11.2 Operational Properties of the Laplace Transform

Note: $x(t)$ is to be regarded as the *causal* signal $x(t)u(t)$.			
Entry	**Property**	$x(t)$	$X(s)$
1	Superposition	$\alpha x_1(t) + \beta x_2(t)$	$\alpha X_1(s) + \beta X_2(s)$
2	Times-exp	$e^{-\alpha t}x(t)$	$X(s+\alpha)$
3	Times-cos	$\cos(\alpha t)x(t)$	$0.5[X(s+j\alpha) + X(s-j\alpha)]$
4	Times-sin	$\sin(\alpha t)x(t)$	$j0.5[X(s+j\alpha) - X(s-j\alpha)]$
5	Time Scaling	$x(\alpha t),\ \alpha > 0$	$\frac{1}{\alpha}X\left(\frac{s}{\alpha}\right)$
6	Time Shift	$x(t-\alpha)u(t-\alpha),\ \alpha > 0$	$e^{-\alpha s}X(s)$
7	Times-t	$tx(t)$	$-\frac{dX(s)}{ds}$
8		$t^n x(t)$	$(-1)^n \frac{d^n X(s)}{ds^n}$
9	Derivative	$x'(t)$	$sX(s) - x(0-)$
10		$x''(t)$	$s^2X(s) - sx(0-) - x'(0-)$
11		$x^{(n)}(t)$	$s^nX(s) - s^{n-1}x(0-) - \cdots - x^{n-1}(0-)$
12	Integral	$\int_{0-}^{t} x(t)\,dt$	$\frac{X(s)}{s}$
13	Convolution	$x(t) \star h(t)$	$X(s)H(s)$
14	Switched periodic, $x_1(t)$ = first period	$x_p(t)u(t)$	$\frac{X_1(s)}{1-e^{-sT}}$, $T =$ time period of $x(t)$
Laplace Transform Theorems			
15	Initial value	$x(0+) = \lim_{s\to\infty}[sX(s)]$ (if $X(s)$ is strictly proper)	
16	Final value	$x(t)\|_{t\to\infty} = \lim_{s\to 0}[sX(s)]$ (if poles of $X(s)$ lie in LHP)	

PGR2026-00033
Krisp Technologies EX1024 Page 347

The **times-sin** and **times-cos** properties form a direct extension if the sines and cosines are expressed as exponentials using Euler's relation:

$$\int_0^\infty x(t)\cos(\alpha t)e^{-st}dt = \int_0^\infty 0.5x(t)[e^{-j\alpha} + e^{j\alpha}]e^{-st}\,dt = 0.5[X(s+j\alpha) + X(s-j\alpha)] \tag{11.3}$$

The **time-scaling** property results from a change of variable:

$$\int_0^\infty x(\alpha t)e^{-st}\,dt \;\stackrel{\alpha t=\lambda}{\longrightarrow} = \frac{1}{\alpha}\int_0^\infty x(\lambda)e^{-s\lambda/\alpha}d\lambda = \frac{1}{\alpha}X\Big(\frac{s}{\alpha}\Big) \tag{11.4}$$

The **time-shift** property also results from a change of variable:

$$\int_\alpha^\infty x(t-\alpha)e^{-st}\,dt \;\stackrel{t-\alpha=\lambda}{\longrightarrow} = \int_0^\infty x(\lambda)e^{-s(\lambda+\alpha)}\,d\lambda = e^{-\alpha s}X(s) \tag{11.5}$$

The quantity e^{-s} may be regarded as a *unit-shift operator* that shifts or delays the time function by 1 unit. Similarly, $e^{-\alpha s}$ contributes a delay of α units to the time signal.

The **derivative** property follows from the defining relation:

$$\int_{0-}^\infty x'(t)e^{-st}\,dt = x(t)e^{-st}\Big|_{0-}^\infty + s\int_{0-}^\infty x(t)e^{-st}\,dt = -x(0-) + sX(s) \tag{11.6}$$

The **times-t** property involves expressing te^{-st} as $te^{-st} = -\frac{d}{ds}(e^{-st})$ and interchanging the order of differentiation and integration:

$$\int_0^\infty tx(t)e^{-st}\,dt = \int_0^\infty x(t)\left[-\frac{d}{ds}e^{-st}\right]dt = -\frac{d}{ds}\int_0^\infty x(t)e^{-st}\,dt = -\frac{dX(s)}{ds} \tag{11.7}$$

The times-t and derivative properties are duals. Multiplication by the variable (t or s) in one domain results in a derivative in the other.

REVIEW PANEL 11.3

Three Basic Laplace Transform Properties

$$e^{-\alpha t}x(t) \Longleftrightarrow X(s+\alpha) \qquad x(t-\alpha) \Longleftrightarrow X(s)e^{-\alpha s} \qquad tx(t) \Longleftrightarrow -\frac{dX(s)}{ds}$$

The **convolution** property involves changing the order of integration and a change of variable ($t-\lambda = \tau$):

$$\begin{aligned}\int_0^\infty \left[\int_0^\infty x(t-\lambda)u(t-\lambda)h(\lambda)\,d\lambda\right]e^{-st}\,dt &= \int_0^\infty \left[\int_\lambda^\infty x(t-\lambda)e^{-s(t-\lambda)}\,dt\right]h(\lambda)e^{-s\lambda}\,d\lambda \\ &= X(s)H(s)\end{aligned}$$

11.2.1 Using the Properties

To develop the Laplace transform of unfamiliar or complicated signals, it is often much easier to start with known transform pairs and then use properties. We illustrate this approach in the following examples.

PGR2026-00033
Krisp Technologies EX1024 Page 348

EXAMPLE 11.2 (Laplace Transforms Using Properties)

(a) (A Damped Exponential) Find the Laplace transform of $x(t) = te^{-3t}u(t)$.

- **(Easiest way)** Start with the known pair $r(t) = tu(t) \Leftrightarrow 1/s^2$ and use the times-exp property to give
$$e^{-3t}tu(t) \Longleftrightarrow \frac{1}{(s+3)^2}$$
- **(Still easy)** Start with the known pair $e^{-3t}u(t) \Leftrightarrow \frac{1}{s+3}$ and use the times-t property to give
$$te^{-3t}u(t) \Longleftrightarrow -\frac{d}{ds}\left[\frac{1}{s+3}\right] = \frac{1}{(s+3)^2}$$
- **(Another way)** Start with $te^{-t}u(t) \Leftrightarrow \dfrac{1}{(s+1)^2}$ as in part (a).

  Then, rewrite $x(t)$ as $\frac{1}{3}(3t)e^{-3t}u(t)$ and use the time-scaling property to get
$$X(s) = \frac{1}{3}\left[\frac{1/3}{(\frac{s}{3}+1)^2}\right] = \frac{1}{(s+3)^2}$$

(b) (A Sinusoid) Find the Laplace transform of $x(t) = \cos(3t + \frac{\pi}{4})u(t)$.

We use trigonometric identities to write $x(t) = \frac{1}{\sqrt{2}}\cos(3t)u(t) - \frac{1}{\sqrt{2}}\sin(3t)u(t)$.

From known transform pairs of the cosine and sine, and by superposition, we obtain
$$X(s) = \frac{s/\sqrt{2}}{s^2+9} - \frac{3/\sqrt{2}}{s^2+9} = \frac{(s-3)/\sqrt{2}}{s^2+9}$$

(c) Find the Laplace transform of $x(t) = t^2u(t-1)$.

- Start with $u(t) \Leftrightarrow \dfrac{1}{s}$ and shift to give $u(t-1) \Leftrightarrow \dfrac{e^{-s}}{s}$.

  Then, use the times-t property twice to give $t^2u(t-1) \Leftrightarrow \frac{d^2}{ds^2}[\frac{e^{-s}}{s}]$. This results in
$$X(s) = \frac{d^2}{ds^2}\left(\frac{e^{-s}}{s}\right) = \frac{d}{ds}\left(-\frac{e^{-s}}{s^2} - \frac{e^{-s}}{s}\right) = e^{-s}\left(\frac{2}{s^3} - \frac{2}{s^2} + \frac{1}{s}\right)$$
- **(Another way)** First, rewrite $x(t)$ as follows:
$$x(t) = (t-1+1)^2u(t-1) = (t-1)^2u(t-1) - 2(t-1)u(t-1) + u(t-1)$$
  Then, use the time-shift property to obtain
$$X(s) = e^{-s}\left(\frac{2}{s^3} - \frac{2}{s^2} + \frac{1}{s}\right)$$

  Comment: Note how we expressed $x(t)$ by terms containing the argument $t-1$ of the *shifted* step to ensure correct use of the time-shift property.

PGR2026-00033
Krisp Technologies EX1024 Page 349

(d) Find the transform of $x(t) = e^{-3t}\cos(2t)u(t)$.

- Start with the known pair $y(t) = \cos(2t)u(t) \Leftrightarrow Y(s) = \dfrac{s}{s^2+4}$.
  Then, use the times-exp property to give $x(t) = e^{-3t}\cos(2t)u(t) \Leftrightarrow Y(s+3)$ and
  $$e^{-3t}\cos(2t)u(t) \Longleftrightarrow \frac{s+3}{(s+3)^2+4}$$
- **(Another way)** Start with the known pair $y(t) = e^{-3t}u(t) \Leftrightarrow Y(s) = \dfrac{1}{s+3}$.
  Then, use the times-cos property $\cos(\alpha t)y(t) \Leftrightarrow 0.5[Y(s+j\alpha) + Y(s-j\alpha)]$ to give
  $$X(s) = 0.5\left(\frac{1}{s+j2+3} + \frac{1}{s-j2+3}\right) = \frac{s+3}{(s+3)^2+4}$$

(e) (Signals Described Graphically) Find the transform of the signals shown in Figure E11.2E.

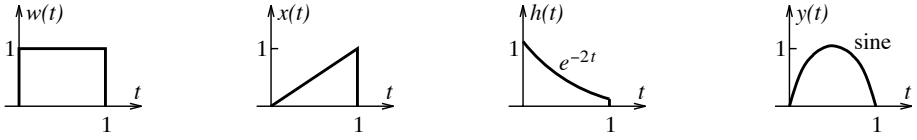


Figure E11.2E The signals for Example 11.2(e)

1. With $w(t) = u(t) - u(t-1)$, the time-shift property gives
   $$W(s) = \frac{1}{s} - \frac{1}{s}e^{-s} = \frac{1-e^{-s}}{s}$$
2. The signal $x(t)$ may be described as $x(t) = tw(t)$. The times-t property gives
   $$X(s) = -\frac{dW(s)}{ds} = -\frac{d}{ds}\left(\frac{1-e^{-s}}{s}\right) = \frac{1}{s^2} - \frac{e^{-s}}{s^2} - \frac{e^{-s}}{s}$$
   We can also express $x(t)$ as $x(t) = r(t) - r(t-1) - u(t-1)$ to get the same result.
3. The signal $y(t)$ may be described as $y(t) = e^{-2t}w(t)$. Using the times-exp property, we obtain
   $$Y(s) = W(s+2) = \frac{1-e^{-(s+2)}}{s+2}$$
4. The signal $h(t)$ may be described as $h(t) = \sin(\pi t)w(t)$. Using the times-sin property and the identity $e^{\pm j\pi} = -1$ (to simplify the result), we obtain
   $$H(s) = j0.5\left[\frac{1+e^{-(s+j\pi)}}{s+j\pi} - \frac{1+e^{-(s-j\pi)}}{s-j\pi}\right] = \frac{\pi(1+e^{-s})}{s^2+\pi^2}$$
   An easier way is to write $h(t) = \sin(\pi t)u(t) + \sin[\pi(t-1)]u(t-1)$ and use the time-shift property to get
   $$H(s) = \frac{\pi}{s^2+\pi^2} + \frac{\pi e^{-s}}{s^2+\pi^2} = \frac{\pi(1+e^{-s})}{s^2+\pi^2}$$

PGR2026-00033
Krisp Technologies EX1024 Page 350

11.2.2 Laplace Transform of Switched Periodic Signals

A periodic signal $x(t)$ turned on at $t = 0$ may be expressed as $x(t) = x_p(t)u(t)$. If the period of $x_p(t)$ equals T and $x_1(t)$ describes *the first period* of $x(t)$ over $(0,\ T)$, the signal $x(t)$ may be represented by $x_1(t)$ and its shifted versions as

$$x(t) = x_p(t)u(t) = x_1(t) + x_1(t-T) + x_1(t-2T) + \cdots \tag{11.8}$$

With $X_1(s)$ as the Laplace transform of $x_1(t)$, the time-shift property gives the Laplace transform $X(s)$ of $x(t)$ as

$$X(s) = X_1(s) + e^{-sT}X_1(s) + e^{-2sT}X_1(s) + \cdots = X_1(s)\Big(1 + e^{-sT} + e^{-2sT} + \cdots\Big) = \frac{X_1(s)}{1 - e^{-sT}} \tag{11.9}$$

REVIEW PANEL 11.4

The Laplace Transform of a Switched Periodic Signal $x(t)u(t)$ with Period T

$X(s) = \dfrac{X_1(s)}{1 - e^{-sT}}$ $\quad [X_1(s)$ is the Laplace transform of the first period of $x(t),\ 0 \le t \le T]$

EXAMPLE 11.3 (Laplace Transform of Switched Periodic Signals)

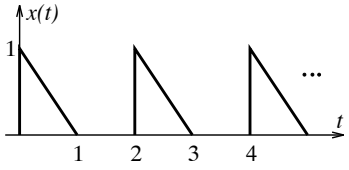
For the periodic signal $x(t)$ shown alongside, the period is $T = 2$.

Its first period is $x_1(t) = u(t) - r(t) + r(t-1)$.

Then, $X_1(s) = \dfrac{1}{s} - \dfrac{1}{s^2} + \dfrac{1}{s^2}e^{-s} = \dfrac{s - 1 + e^{-s}}{s^2}$.

The Laplace transform of $x(t)$ is thus

$X(s) = \dfrac{X_1(s)}{1 - e^{-sT}} = \dfrac{s - 1 + e^{-s}}{s^2(1 - e^{-2s})}$.



11.2.3 The Transfer Function

The concept of the transfer function lies at the heart of system analysis using transform methods. The response $y(t)$ of a relaxed LTI system with impulse response $h(t)$ to an input $x(t)$ is given by the convolution $y(t) = x(t) \star h(t)$. Since the convolution operation transforms to a product, we have the equivalent result

$$Y(s) = X(s)H(s) \qquad \text{or} \qquad H(s) = \frac{Y(s)}{X(s)} \tag{11.10}$$

The time-domain and s-domain descriptions are illustrated in Figure 11.1. We emphasize that the transfer function is defined only for *relaxed* LTI systems, either as the ratio of the transformed output $Y(s)$ and transformed input $X(s)$ or as the Laplace transform of the system impulse response $h(t)$.

An LTI system may also be described by the differential equation

$$y^{(N)}(t) + a_1 y^{(N-1)}(t) + \cdots + a_{N-1}y^{(1)}(t) + a_N y(t) = b_0 x^{(M)}(t) + b_1 x^{(M-1)}(t) + \cdots + b_{M-1}x^{(1)}(t) + b_M x(t)$$

Its Laplace transform results in the transfer function $H(s) = Y(s)/X(s)$ as

$$H(s) = \frac{Y(s)}{X(s)} = \frac{b_0 s^M + b_1 s^{M-1} + \cdots + b_M}{s^N + a_1 s^{N-1} + \cdots + a_N} \tag{11.11}$$

The transfer function is thus a ratio of polynomials in s. It allows access to both the impulse response and the differential equation of the relaxed system in the time domain.

PGR2026-00033
Krisp Technologies EX1024 Page 351

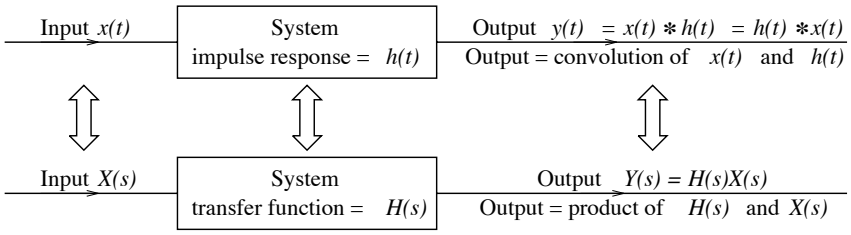


Figure 11.1 The concept of the transfer function

REVIEW PANEL 11.5

The Transfer Function of Relaxed LTI Systems Is a Rational Function of s

$$H(s) = \frac{Y(s) \text{ (transformed output)}}{X(s) \text{ (transformed input)}} \qquad H(s) = \mathcal{L}[h(t)] \text{ (Laplace transform of impulse response)}$$

11.2.4 System Representation in Various Forms

A relaxed LTI system may be described in various forms: by its differential equation, by its impulse response, or by its transfer function. Given one form, it is fairly easy to obtain the others using time-domain methods and/or frequency-domain transformations. For example, given the impulse response $h(t)$, we can find the transfer function from its Laplace transform $H(s)$. If we express the transfer function as $H(s) = Y(s)/X(s)$, cross-multiplication and inverse transformation leads to the system differential equation.

REVIEW PANEL 11.6

Three Representations of a Relaxed LTI System

Differential equation: From $H(s) = \frac{Y(s)}{X(s)}$ by cross-multiplication and inverse Laplace transform (ILT)

Transfer function: $H(s) = \frac{Y(s)}{X(s)}$ **Impulse response:** $h(t) = \text{ILT}\{H(s)\}$

EXAMPLE 11.4 (System Representation in Various Forms)

(a) Consider a system governed by $y''(t) + 3y'(t) + 2y(t) = 2x'(t) + 3x(t)$.

Using the derivative property, we obtain $s^2Y(s) + 3sY(s) + 2Y(s) = 2sX(s) + 3X(s)$.

Thus, $H(s) = \dfrac{Y(s)}{X(s)} = \dfrac{2s+3}{s^2+3s+2}$.

(b) We can write $H(s) = \dfrac{1}{s+1} + \dfrac{1}{s+2}$.

The impulse response thus equals $h(t) = e^{-t}u(t) + e^{-2t}u(t)$.

(c) Since $H(s) = Y(s)/X(s)$, we can also write $(s^2 + 3s + 2)Y(s) = (2s + 3)X(s)$.

This leads to $y''(t) + 3y'(t) + 2y(t) = 2x'(t) + 3x(t)$.

Note that $H(s), h(t)$, and the differential equation are three ways of describing the same system.

PGR2026-00033
Krisp Technologies EX1024 Page 352

11.3 Poles and Zeros of the Transfer Function

The transfer function of LTI systems is a rational function (a ratio of polynomials) whose general form can be expressed as

$$H(s) = \frac{P(s)}{Q(s)} = \frac{b_0 s^M + b_1 s^{M-1} + \cdots + b_M}{s^N + a_1 s^{N-1} + \cdots + a_N} \tag{11.12}$$

It is customary to normalize the coefficient of s^N in the denominator polynomial to unity. If the degree N of the denominator $Q(s)$ *exceeds* the degree M of the denominator $P(s)$, (i.e, $N > M$), $H(s)$ describes a **strictly proper rational function**. If N *exceeds or equals* M (i.e, $N \geq M$), $H(s)$ is called a **proper rational function**.

REVIEW PANEL 11.7

When Is a Transfer Function $H(s) = N(s)/D(s)$ Proper and Strictly Proper?

Proper: If degree of $N(s) \leq D(s)$ **Strictly Proper:** If degree of $N(s) < D(s)$

The transfer function may also be expressed in factored form as

$$H(s) = \frac{P(s)}{Q(s)} = K\frac{(s - z_1)(s - z_2)\cdots(s - z_M)}{(s - p_1)(s - p_2)\cdots(s - p_N)} \tag{11.13}$$

The transfer function $H(s)$ equals zero if $s = z_k$ and infinity if $s = p_k$. The roots z_k of the numerator polynomial $P(s)$ are termed the **zeros** of $H(s)$, and the roots p_k of the denominator polynomial $Q(s)$ are called the **poles** of $H(s)$. A plot of the poles (shown by $\times$) and zeros (shown by circles) of $H(s)$ in the s-plane constitutes the **pole-zero plot** and provides a visual picture of the root locations. For roots that are repeated m times, we show the multiplicity m alongside the roots. Clearly, we can find $H(s)$ directly from a pole-zero plot of the root locations but only to within the multiplicative constant K. If the value of K is also shown on the plot, $H(s)$ is then known in its entirety. The poles of $H(s)$ are also called **natural modes**, or **natural frequencies**. The poles of $Y(s) = H(s)X(s)$ determine the form of the system response. The natural frequencies of $H(s)$ always appear in the system response (unless they are canceled by any corresponding zeros in $X(s)$). The zeros of $H(s)$ are often called **transmission zeros**. Since $H(s) = 0$ at its zeros, the response $Y(s) = X(s)H(s)$ is also zero. The zeros of $H(s)$ may thus be regarded as the frequencies that are blocked by the system.

REVIEW PANEL 11.8

Poles and Zeros of a Rational Transfer Function $H(s) = N(s)/D(s)$

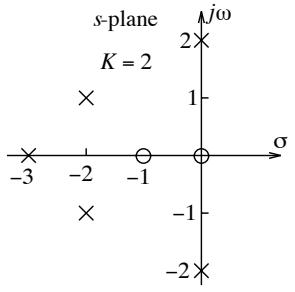
Zeros: Roots of the numerator $N(s)$ **Poles:** Roots of the denominator $D(s)$

EXAMPLE 11.5 (Pole-Zero Plots)

Let $H(s) = \dfrac{2s(s+1)}{(s+3)(s^2+4)(s^2+4s+5)}$.

The constant is $K = 2$. The numerator degree is 2.
The two zeros are $s = 0$ and $s = -1$.
The denominator degree is 5. The pole locations are
$s = -3$, $s = \pm j2$, and $s = -2 \pm j$.

The pole-zero plot is shown in the figure.



PGR2026-00033
Krisp Technologies EX1024 Page 353

11.4 The Inverse Laplace Transform

The inverse Laplace transform is also governed by an integral operation but with respect to the variable s. For a particular value of σ in the region of convergence, the signal $x(t)$ is given by

$$x(t) = \int_{-\infty}^{\infty} X(\sigma + j2\pi f) e^{(\sigma + j2\pi f)t} df = e^{\sigma t} \int_{-\infty}^{\infty} X(\sigma + j2\pi f) e^{j2\pi f t} df \tag{11.14}$$

The change of variable $s = \sigma + j2\pi f$ gives $df = ds/2\pi$, and we obtain

$$x(t) = \frac{1}{2\pi j} \int_{\sigma - j\infty}^{\sigma + j\infty} X(s) e^{st}\, ds \tag{11.15}$$

Since s is complex, the solution requires a knowledge of complex variables. A simpler alternative, and the only one we pursue in this text, is to express $X(s)$ as a sum of terms whose inverse transforms can be readily recognized from a stock list of transform pairs. We form such a sum by resorting to a *partial fraction expansion* of $X(s)$.

11.4.1 Partial Fraction Expansion

The denominator $Q(s)$ of an Nth-order rational function can be factored into linear (perhaps repeated) factors. A **partial fraction expansion** (PFE) allows a *strictly proper* rational fraction $P(s)/Q(s)$ to be expressed as a sum of terms whose numerators are constant and whose denominators correspond to the linear and/or repeated factors of $Q(s)$. This in turn allows us to relate such terms to their corresponding inverse transform.

The partial fraction method requires not only a strictly proper $X(s)$ but also a denominator $Q(s)$ in factored form. The form of the expansion depends on the nature of the factors in $Q(s)$. The constants in the numerator of the partial fraction expansion terms are often called **residues**.

11.4.2 Distinct Linear Factors

If $Q(s)$ contains only distinct linear factors, we write $X(s)$ in the form

$$X(s) = \frac{P(s)}{(s+p_1)(s+p_2)\cdots(s+p_N)} = \frac{K_1}{s+p_1} + \frac{K_2}{s+p_2} + \cdots + \frac{K_N}{s+p_N} \tag{11.16}$$

To find the mth coefficient (or residue) K_m, we multiply both sides by $(s+p_m)$ to get

$$(s+p_m)X(s) = K_1 \frac{(s+p_m)}{(s+p_1)} + \cdots + K_m + \cdots + K_N \frac{(s+p_m)}{(s+p_N)} \tag{11.17}$$

With both sides evaluated at $s = -p_m$, we obtain K_m as

$$K_m = (s+p_m)X(s)\Big|_{s=-p_m} \tag{11.18}$$

Note that $X(s)$ will in general contain terms with real constants and pairs of terms with complex conjugate residues and can be written as

$$X(s) = \frac{K_1}{s+p_1} + \frac{K_2}{s+p_2} + \cdots + \frac{A_1}{s+r_1} + \frac{A_1^*}{s+r_1^*} + \frac{A_2}{s+r_2} + \frac{A_2^*}{s+r_2^*} + \cdots \tag{11.19}$$

For a real root, the residue (coefficient) will also be real. For each pair of complex conjugate roots, the residues will also be complex conjugates, and we thus need compute only one of these.

PGR2026-00033
Krisp Technologies EX1024 Page 354

11.4.3 Repeated Factors

The preceding results require modification if $Q(s)$ contains repeated factors. If a factor is repeated k times, it is said to have a *multiplicity* of k. If $Q(s)$ contains the repeated term $(s+r)^k$ and one other linear factor, the PFE of $X(s) = P(s)/Q(s)$ may be set up as

$$X(s) = \frac{P(s)}{(s+p_1)(s+r)^k} = \frac{K_1}{s+p_1} + \frac{A_0}{(s+r)^k} + \frac{A_1}{(s+r)^{k-1}} + \cdots + \frac{A_{k-1}}{s+r} \tag{11.20}$$

Observe that the constants A_j ascend in index j from 0 to $k-1$, whereas the denominators $(s+r)^m$ descend in power m from k to 1. The coefficient K_1 for the non-repeated term $\frac{K_1}{s+p_1}$ is found by evaluating $(s+p_1)X(s)$ at $s = -p_1$:

$$K_1 = (s+p_1)X(s)\Big|_{s=-p_1} \tag{11.21}$$

The coefficients A_j for the repeated roots require $(s+r)^k X(s)$ (and its derivatives) for their evaluation. We successively find

$$A_0 = (s+r)^k X(s)\Big|_{s=-r} \qquad A_2 = \frac{1}{2!}\frac{d^2}{ds^2}\left[(s+r)^k X(s)\right]\Big|_{s=-r}$$

$$A_1 = \frac{d}{ds}\left[(s+r)^k X(s)\right]\Big|_{s=-r} \qquad A_n = \frac{1}{n!}\frac{d^n}{ds^n}\left[(s+r)^k X(s)\right]\Big|_{s=-r}$$

Even though this process allows us to find the coefficients independently of each other, the algebra in finding the derivatives can become tedious if the multiplicity k of the roots exceeds 2 or 3.

REVIEW PANEL 11.9

Inverse Transformation of $X(s)$ Relies on Partial Fractions and Table Look-Up

Distinct roots: $X(s) = \displaystyle\prod_{m=1}^{N} \frac{K_m}{s+p_m} = \sum_{m=1}^{N} \frac{P(s)}{s+p_m}$, where $K_m = (s+p_m)X(s)|_{s=-p_m}$

Repeated: $\dfrac{1}{(s+r)^k}\displaystyle\prod_{m=1}^{N} \frac{P(s)}{s+p_m} = \sum_{m=1}^{N} \frac{K_m}{s+p_m} + \sum_{n=0}^{k-1} \frac{A_n}{(s+r)^{k-n}}$ $\qquad A_n = \dfrac{1}{n!}\dfrac{d^n}{ds^n}\left[(s+r)^k X(s)\right]\Big|_{s=-r}$

11.4.4 The Inverse Transform

Once the partial fraction expansion is established, the inverse transform for each term can be found with the help of a transform-pair table. A summary of the various forms appears in Table 11.3. The entries in this table are based on the following observations:

1. All terms show an exponential decay or damped exponential form.

2. Terms corresponding to real factors will have the form $Ke^{-pt}u(t)$.

3. Terms corresponding to each complex conjugate pair of roots will be of the form $Ae^{-rt} + A^*e^{-r^*t}$. Using Euler's relation, this can be reduced to a real term in one of the forms listed in Table 11.3.

PGR2026-00033
Krisp Technologies EX1024 Page 355

Table 11.3 Inverse Laplace Transforms of Partial Fraction Expansion Terms

Entry	Partial Fraction Expansion Term	Inverse Transform
1	$\frac{K}{s+\alpha}$	$Ke^{-\alpha t}u(t)$
2	$\frac{K}{(s+\alpha)^n}$	$\frac{K}{(n-1)!}t^{n-1}e^{-\alpha t}u(t)$
3	$\frac{Cs+D}{(s+\alpha)^2+\beta^2}$	$e^{-\alpha t}\left[C\cos(\beta t)+\frac{D-\alpha C}{\beta}\sin(\beta t)\right]u(t)$
4	$\frac{A+jB}{s+\alpha+j\beta}+\frac{A-jB}{s+\alpha-j\beta}$	$2e^{-\alpha t}[A\cos(\beta t)+B\sin(\beta t)]u(t)$
5	$\frac{A+jB}{(s+\alpha+j\beta)^n}+\frac{A-jB}{(s+\alpha-j\beta)^n}$	$\frac{2}{(n-1)!}t^{n-1}e^{-\alpha t}[A\cos(\beta t)+B\sin(\beta t)]u(t)$
6	$\frac{M\angle\theta}{s+\alpha+j\beta}+\frac{M\angle-\theta}{s+\alpha-j\beta}$	$2Me^{-\alpha t}\cos(\beta t-\theta)u(t)$
7	$\frac{M\angle\theta}{(s+\alpha+j\beta)^n}+\frac{M\angle-\theta}{(s+\alpha-j\beta)^n}$	$\frac{2M}{(n-1)!}t^{n-1}e^{-\alpha t}\cos(\beta t-\theta)u(t)$

EXAMPLE 11.6 (Partial Fraction Expansion)

(a) (Non-Repeated Poles) Let $X(s)=\dfrac{2s^3+8s^2+4s+8}{s(s+1)(s^2+4s+8)}$. This can be factored as

$$X(s)=\frac{K_1}{s}+\frac{K_2}{s+1}+\frac{A}{s+2+j2}+\frac{A^*}{s+2-j2}$$

We successively evaluate

$$\begin{aligned}
K_1 &= sX(s)\Big|_{s=0}=\frac{2s^3+8s^2+4s+8}{(s+1)(s^2+4s+8)}\Big|_{s=0}=\frac{8}{8}=1\\
K_2 &= (s+1)X(s)\Big|_{s=-1}=\frac{2s^3+8s^2+4s+8}{s(s^2+4s+8)}\Big|_{s=-1}=\frac{10}{-5}=-2\\
A &= (s+2+j2)X(s)\Big|_{s=-2-j2}=\frac{2s^3+8s^2+4s+8}{s(s+1)(s+2-j2)}\Big|_{s=-2-j2}=1.5+j0.5
\end{aligned}$$

The partial fraction expansion thus becomes

$$X(s)=\frac{1}{s}-\frac{2}{s+1}+\frac{1.5+j0.5}{s+2+j2}+\frac{1.5-j0.5}{s+2-j2}$$

Table 11.3 thus yields $x(t)$ as

$$x(t)=u(t)-2e^{-t}u(t)+2e^{-2t}[1.5\cos(2t)+0.5\sin(2t)]u(t)$$

PGR2026-00033
Krisp Technologies EX1024 Page 356

With $1.5 + j0.5 = 1.581\angle 18.4^\circ = 1.581\angle 0.1024\pi$, Table 11.3 also yields $x(t)$ as

$$x(t) = u(t) - 2e^{-t}u(t) + 3.162e^{-2t}\cos(2t - 0.1024\pi)u(t)$$

(b) (Repeated Poles) Let $X(s) = \dfrac{4}{(s+1)(s+2)^3}$. Its partial fraction expansion is

$$X(s) = \frac{K_1}{s+1} + \frac{A_0}{(s+2)^3} + \frac{A_1}{(s+2)^2} + \frac{A_2}{(s+2)}$$

We compute $K_1 = (s+1)X(s)\Big|_{s=-1} = \dfrac{4}{(s+2)^3}\Big|_{s=-1} = 4.$

Since $(s+2)^3X(s) = \dfrac{4}{s+1}$, we also successively compute

$$\begin{aligned} A_0 &= \frac{4}{s+1}\Big|_{s=-2} = -4 \\ A_1 &= \frac{d}{ds}\left[\frac{4}{s+1}\right]\Big|_{s=-2} = -\frac{4}{(s+1)^2}\Big|_{s=-2} = -4 \\ A_2 &= \tfrac{1}{2}\frac{d^2}{ds^2}\left[\frac{4}{s+1}\right]\Big|_{s=-2} = \frac{4}{(s+1)^3}\Big|_{s=-2} = -4 \end{aligned}$$

This gives the result

$$X(s) = \frac{4}{s+1} - \frac{4}{(s+2)^3} - \frac{4}{(s+2)^2} - \frac{4}{s+2}$$

From Table 11.3, $x(t) = 4e^{-t}u(t) - 2t^2e^{-2t}u(t) - 4te^{-2t}u(t) - 4e^{-2t}u(t)$.

11.4.5 Some Nonstandard Forms

Certain exceptions can also be handled by using the partial fraction method.

1. If $X(s)$ is a rational function but not strictly proper, we use long division to express $X(s)$ as the sum of a quotient $G(s)$ and a strictly proper function $X_1(s)$. The inverse transform then equals $x_1(t) + g(t)$.
2. If the numerator of $X(s)$ contains exponentials of the form $\exp(-\alpha s)$, this reflects a time delay. First, we find the partial fractions and the inverse without this factor. Then, we replace t by $t - \alpha$ in this inverse to account for the time delay it produces.

EXAMPLE 11.7 (Inverse Transforms of Nonstandard Forms)

(a) (An Improper Rational Function) Let $X(s) = \dfrac{s^2 + 4s + 2}{(s+1)(s+2)}$. Since $X(s)$ is not strictly proper, we use long division followed by partial fractions to obtain

$$X(s) = \frac{s^2 + 4s + 2}{(s+1)(s+2)} = 1 + \frac{s}{(s+1)(s+2)} = 1 + \frac{2}{s+2} - \frac{1}{s+1}$$

Thus, $x(t) = \delta(t) + 2e^{-2t}u(t) - e^{-t}u(t)$.

PGR2026-00033
Krisp Technologies EX1024 Page 357

(b) (The Effect of Delay) Let $X(s) = \dfrac{se^{-2s}+1}{(s+1)(s+2)}$. We split this into two terms:

$$X(s) = \frac{se^{-2s}}{(s+1)(s+2)} + \frac{1}{(s+1)(s+2)} = e^{-2s}X_1(s) + X_2(s)$$

We find the partial fractions of $X_1(s)$ and $X_2(s)$ to get

$$X_1(s) = \frac{s}{(s+1)(s+2)} = \frac{2}{s+2} - \frac{1}{s+1} \qquad X_2(s) = \frac{1}{(s+1)(s+2)} = \frac{1}{s+1} - \frac{1}{s+2}$$

Thus, $x_1(t) = 2e^{-2t}u(t) - e^{-t}u(t)$ and $x_2(t) = e^{-t)}(t) - e^{-2t}u(t)$.

From the time-shift property, the inverse transform of $e^{-2s}X_1(s)$ equals $x_1(t-2)$, and we get

$$x(t) = x_1(t-2) + x_2(t) = [2e^{-2(t-2)} - e^{-(t-2)}]u(t-2) + (e^{-t} - e^{-2t})u(t)$$

11.5 The *s*-Plane and BIBO Stability

In the time domain, BIBO (bounded-input, bounded-output) stability of an LTI system requires a differential equation in which the highest derivative of the input never exceeds the highest derivative of the output and a characteristic equation whose roots have negative real parts. Equivalently, we require the impulse response $h(t)$ to be absolutely integrable. In the s-domain, this translates into requiring a *proper* transfer function $H(s)$ (with common factors canceled) whose poles have negative real parts or lie in the left half of the s-plane (excluding the $j\omega$-axis). Here is why:

1. A pole in the right half-plane (RHP) leads to exponential growth in $h(t)$. We thus have an unbounded response for any input, bounded or otherwise. For example, $H(s) = 1/(s-2)$ results in the growing exponential $\exp(2t)$.
2. Multiple poles on the $j\omega$-axis lead to terms of the form $1/(s^2+\alpha^2)^2$ or $1/s^2$ in $H(s)$ and result in polynomial growth of the form $t\cos(\alpha t+\beta)$. For example, $H(s) = 1/s^2$ produces a growing ramp, and $H(s) = 1/(s^2+1)^2$ produces the growing term $t\cos(t)$.
3. Non-repeated (simple) poles on the $j\omega$-axis lead to factors of the form $1/(s^2+\alpha^2)$ or $1/s$ in $H(s)$. If the applied input also has the same form, the response will have factors of the form $1/s^2$ or $1/(s^2+\alpha^2)^2$ and show polynomial growth. This excludes the $j\omega$-axis from the region of stability.

None of the terms corresponding to these poles is absolutely integrable. *For BIBO stability, $H(s)$ (with common factors canceled) must be proper, and its poles must lie entirely in the left half-plane (LHP) and exclude the $j\omega$-axis.* This is both a necessary and sufficient condition for stability. We make the following remarks:

- BIBO stability requires a system differential equation in which the highest derivative of the input never exceeds the highest derivative of the output. This condition is equivalent to a *proper* transfer function $H(s) = P(s)/Q(s)$, where the degree M of $P(s)$ never exceeds the degree N of $Q(s)$. If $M > N$, the system will be unstable due to $j\omega$-axis poles at $s = j\omega = \infty$. Alternatively, $h(t)$ will contain derivatives of $\delta(t)$ that are not absolutely integrable.
- If a linear system has no right half-plane zeros or poles, the system is called a **minimum-phase system**. The transfer function of such a system may be obtained from a knowledge of its magnitude spectrum alone. We discuss minimum-phase systems in later chapters.

PGR2026-00033
Krisp Technologies EX1024 Page 358

- A system with poles in the left half-plane and simple, finite poles on the $j\omega$-axis is also called **wide sense stable**, or **marginally stable**.
- A polynomial with real coefficients whose roots lie entirely in the left half of the s-plane is called **strictly Hurwitz**. If it also has simple $j\omega$-axis roots, it is called **Hurwitz**. The coefficients of Hurwitz polynomials are all nonzero and of the same sign. A formal method called the **Routh test** allows us to check if a polynomial is Hurwitz (without having to find its roots) and also gives us the number (but not the location) of right half-plane roots.

REVIEW PANEL 11.10

BIBO Stability from the Transfer Function or Impulse Response

From transfer function: $H(s)$ must be *strictly proper* and its poles must lie *inside* the LHP.
From impulse response: $h(t)$ must be *absolutely integrable* ($\int |h(\lambda)|\, d\lambda < \infty$).

EXAMPLE 11.8 (System Stability)
Investigate the stability of the following systems.

(a) $H(s) = \dfrac{s+2}{s^2-2s-3} = \dfrac{s+2}{(s+1)(s-3)}$; unstable due to the RHP pole at $s = 3$.

(b) $H(s) = \dfrac{s+1}{s^2+4}$; unstable (marginally stable) due to the simple poles at $s = \pm j2$.

(c) $H(s) = \dfrac{s^2+2}{s^2+3s+2}$; stable because $H(s)$ is proper and there are no RHP poles.

(d) $H(s) = \dfrac{s^3+2}{s^2+3s+2}$; unstable because $H(s)$ is not proper.

11.5.1 Other Forms of Stability

BIBO stability relates to external inputs. Other definitions of stability rely only on the specifics of a system, such as its structure or state.

Asymptotic Stability

If the zero-input response of a system decays to zero with time, we classify the system as being **asymptotically stable**. Such a system has a transfer function whose poles lie entirely within the LHP and this also implies BIBO stability. BIBO stability and asymptotic stability are often used interchangeably.

Any passive linear system that includes one or more resistive elements is always asymptotically stable. The resistive elements dissipate energy and allow the system to relax to zero state no matter what the initial state.

Liapunov Stability

If the zero-input response always remains bounded (it may or may not approach zero), we classify the system as being **stable in the sense of Liapunov**. In addition to poles in the LHP, the transfer function of such a system may also contain *simple* (not repeated) poles on the $j\omega$-axis. Asymptotic stability is a special case of Liapunov stability.

PGR2026-00033
Krisp Technologies EX1024 Page 359

A passive system with resistors is both asymptotically and Liapunov stable. A passive system with only lossless elements (e.g., L and C) is Liapunov stable but not asymptotically stable. The energy in such a system due to a nonzero initial state remains constant. It cannot increase because the system is passive. It cannot be dissipated because the system is lossless.

11.5.2 The Initial Value Theorem and Final Value Theorem

The **initial value theorem** predicts the initial value $x(0+)$ of the time signal $x(t)$ *from its strictly proper transform* $X(s)$ as

$$x(0+) = \lim_{s\to\infty}[sX(s)] \qquad \text{(initial value theorem)} \tag{11.22}$$

Note that this theorem predicts the initial value at $t = 0+$. If $X(s)$ is not strictly proper, we get correct results by using the *strictly proper* part of $X(s)$ obtained after long division since the remaining part corresponds to impulses and their derivatives (which occur at $t = 0$ and are zero at $t = 0+$). If $X(s)$ or its strictly proper part equals $P(s)/Q(s)$ and $P(s)$ is of degree M and $Q(s)$ is of degree N, $x(0+) = 0$ if $N - M > 1$. If $N - M = 1$, $x(0+)$ equals the ratio of the leading coefficients of $P(s)$ and $Q(s)$. In other words, $x(0+)$ is nonzero only if $N - M = 1$.

The **final value theorem** predicts the value of the time signal $x(t)$ as $t \to \infty$ from its transform $X(s)$, and reads

$$x(\infty) = \lim_{s\to 0}[sX(s)] \qquad \text{(final value theorem)} \tag{11.23}$$

Note that this theorem applies only if *the poles of* $X(s) = P(s)/Q(s)$*, with common factors in* $P(s)$ *and* $Q(s)$ *canceled, lie in the left half of the s-plane*. The only $j\omega$-axis pole permitted is a simple pole at $s = 0$. We should expect $x(\infty) = 0$ if all poles of $X(s)$ lie to the left of the $j\omega$-axis (because $x(t)$ contains only damped exponentials). We should also expect $x(\infty)$ to be constant if there is a single pole at $s = 0$ (which corresponds to a step function). Finally, $x(\infty)$ will be indeterminate if there are conjugate pole pairs on the $j\omega$-axis (because $x(t)$ includes sinusoids whose final value is indeterminate). *The final value theorem yields erroneous results if used in this case.*

REVIEW PANEL 11.11

Finding the Initial Value $x(0+)$ **and Final Value** $x(\infty)$ **of** $x(t)$ **from a Rational** $X(s)$

$$x(0+) = \lim_{s\to\infty} sX_P(s) \quad [X_P(s) = \textit{strictly proper} \text{ part of } X(s)] \qquad x(\infty) = \lim_{s\to 0} sX(s)$$

Initial value: If $X_P(s) = N(s)/D(s)$ and $sN(s)$ and $D(s)$ are not of the same order, $x(0+) = 0$.
Final value: Meaningful only if poles of $X(s)$ lie *inside* the LHP (with one pole allowed at $s = 0$).

EXAMPLE 11.9 (Initial and Final Value Theorems)
Find the initial and final value of $x(t)$ for the following:

(a) $X(s) = \dfrac{12(s+1)}{s(s^2+4)}$.

We find

Initial value: $x(0+) = \lim_{s\to\infty} sX(s) = 0$ since $N - M = 2$ for $X(s)$.

Final value: Indeterminate ($X(s)$ has conjugate poles $s = \pm j2$ on $j\omega$-axis).

PGR2026-00033
Krisp Technologies EX1024 Page 360

(b) $X(s) = \dfrac{2s+6}{s(4s+2)}$.

We find

$$\begin{aligned} x(0+) &= \lim_{s\to\infty} sX(s) = \lim_{s\to\infty} \frac{2s+6}{4s+2} = \lim_{s\to\infty} \frac{2+6/s}{4+2/s} = 0.5 \\ x(\infty) &= \lim_{s\to 0} sX(s) = \lim_{s\to 0} \frac{s+6}{s+2} = 3 \end{aligned}$$

Both the initial and final value are nonzero.

(c) $X(s) = \dfrac{4s+5}{2s+1}$.

Since this is not strictly proper, we use long division to rewrite it as

$$X(s) = 2 + \frac{3}{2s+1} = 2 + Y(s)$$

The strictly proper part $Y(s)$ gives the initial value of $x(t)$ as

$$x(0+) = \lim_{s\to\infty} sY(s) = \lim_{s\to\infty} \frac{3s}{2s+1} = 1.5$$

For the final value we use $X(s)$ directly, to obtain

$$x(\infty) = \lim_{s\to 0} sX(s) = \lim_{s\to 0} \frac{4s^2+5s}{2s+1} = 0$$

11.6 The Laplace Transform and System Analysis

The Laplace transform is a useful tool for the analysis of LTI systems. The convolution, derivative, and integration properties provide the key to simplifying analysis using Laplace transforms because

1. Convolution transforms to the much simpler multiplication operation.
2. Transformed differential equations assume an algebraic form.

For a relaxed LTI system, the response $Y(s)$ to an input $X(s)$ is $H(s)X(s)$, where the transfer function $H(s)$ may be found in several ways:

1. From the Laplace transform of the impulse response $h(t)$.
2. From the transformed differential equation.
3. From an s-domain model of an electric circuit.

If the system is not relaxed, the effect of initial conditions is easy to include. In any case, the response is obtained in the transformed domain. A time-domain solution requires inverse transformation. This additional step is a penalty exacted by all transformed-domain methods.

11.6.1 Systems Described by Differential Equations

For a system governed by a differential equation and subject to arbitrary initial conditions, we transform the differential equation (using the derivative property), incorporate the effect of initial conditions, and invoke partial fractions to find the time-domain result. Recall that for an LTI system with nonzero initial

PGR2026-00033
Krisp Technologies EX1024 Page 361

conditions the total response equals the sum of the zero-state response (due only to the input) and the zero-input response (due only to the initial conditions). These components can be easily identified using Laplace transforms.

REVIEW PANEL 11.12

The Solution of Differential Equations Requires the Derivative Property

$x'(t) \Leftrightarrow sX(s) - x(0)$ $\quad x''(t) \Leftrightarrow s^2X(s) - sx'(0) - x''(0)$ $\quad$ **For zero IC:** $x^{(n)}(t) \Leftrightarrow s^nX(s)$

EXAMPLE 11.10 (Solving Differential Equations)
Let $y''(t) + 3y'(t) + 2y(t) = 4e^{-2t}$, with $y(0) = 3$ and $y'(0) = 4$.

Transformation to the s-domain, using the derivative property, yields

$$s^2Y(s) - sy(0) - y'(0) + 3[sY(s) - y(0)] + 2Y(s) = \frac{4}{s+2}$$

(a) (Total Response) Substitute for the initial conditions and rearrange:

$$(s^2 + 3s + 2)Y(s) = 3s + 4 + 9 + \frac{4}{s+2} = \frac{3s^2 + 19s + 30}{s+2}$$

Upon simplification, we obtain $Y(s)$ and its partial fraction form as

$$Y(s) = \frac{3s^2 + 19s + 30}{(s+1)(s+2)^2} = \frac{K_1}{s+1} + \frac{A_0}{(s+2)^2} + \frac{A_1}{s+2}$$

Solving for the constants, we obtain

$$K_1 = \frac{3s^2 + 19s + 30}{(s+2)^2}\Big|_{s=-1} = 14$$

$$A_0 = \frac{3s^2 + 19s + 30}{s+1}\Big|_{s=-2} = -4$$

$$A_1 = \frac{d}{ds}\left[\frac{3s^2 + 19s + 30}{s+1}\right]\Big|_{s=-2} = -11$$

Upon inverse transformation, $y(t) = (14e^{-t} - 4te^{-2t} - 11e^{-2t})u(t)$.

As a check, we confirm that $y(0) = 3$ and $y'(0) = -14 + 22 - 4 = 4$.

(b) (Zero-State Response) For the zero-state response, we assume zero initial conditions to obtain

$$(s^2 + 3s + 2)Y_{zs}(s) = \frac{4}{s+2}$$

This gives

$$Y_{zs}(s) = \frac{4}{(s+2)(s^2 + 3s + 2)} = \frac{4}{s+1} - \frac{4}{(s+2)^2} - \frac{4}{s+2}$$

Inverse transformation gives $y_{zs}(t) = (4e^{-t} - 4te^{-2t} - 4e^{-2t})u(t)$.

PGR2026-00033
Krisp Technologies EX1024 Page 362

(c) (Zero-Input Response) For the zero-input response, we assume zero input to obtain

$$(s^2+3s+2)Y_{\text{zi}}(s) = 3s+13 \qquad Y_{\text{zi}}(s) = \frac{3s+13}{s^2+3s+2} = \frac{10}{s+1} - \frac{7}{s+2}$$

Upon inverse transformation, $y_{\text{zi}}(t) = (10e^{-t} - 7e^{-2t})u(t)$. The total response equals

$$y(t) = y_{\text{zs}}(t) + y_{\text{zi}}(t) = (14e^{-t} - 4te^{-2t} - 11e^{-2t})u(t)$$

This matches the result found from the direct solution.

11.6.2 System Response from the Transfer Function

For an LTI system with nonzero initial conditions, the response $Y(s)$ in the s-domain is the sum of the zero-state response (ZSR) due only to the *transformed input* and the zero-input response (ZIR) due to the *transformed initial conditions*. With $H(s) = N(s)/D(s)$, the response $Y(s)$ of a relaxed system to an input $X(s)$ may be expressed as $Y(s) = X(s)H(s) = X(s)N(s)/D(s)$. This corresponds to the zero-state response. If the system is not relaxed, the initial conditions result in an additional contribution, the zero-input response $Y_{\text{zi}}(s)$. To evaluate $Y_{\text{zi}}(s)$, we must find and transform the system differential equation using the initial conditions.

EXAMPLE 11.11 (System Response from the Transfer Function)

Consider a system whose transfer function is $H(s) = \dfrac{1}{s^2+3s+2}$. We find its zero-state, zero-input, and total response, assuming the input $x(t) = 4e^{-2t}$ and the initial conditions $y(0) = 3$ and $y'(0) = 4$.

(a) (Zero-State Response) For the zero-state response, we transform $x(t)$ to $X(s) = \dfrac{4}{s+2}$ and obtain the response as $Y_{\text{zs}}(s) = H(s)X(s)$ to give

$$Y_{\text{zs}}(s) = \frac{4}{(s+2)(s^2+3s+2)} = \frac{4}{s+1} - \frac{4}{(s+2)^2} - \frac{4}{s+2}$$

Its inverse transform gives $y_{\text{zs}}(t) = (4e^{-t} - 4te^{-2t} - 4e^{-2t})u(t)$.

(b) (Zero-Input Response) For the zero-input response, write the transfer function $H(s) = \frac{Y(s)}{X(s)}$ as $(s^2+3s+2)Y(s) = X(s)$ to obtain the system differential equation $y''(t) + 3y'(t) + 2y(t) = x(t)$.

Assuming zero input and nonzero initial conditions, this transforms to

$$s^2Y_{\text{zi}}(s) - sy(0) - y'(0) + 3[sY_{\text{zi}}(s) - y(0)] + 2Y_{\text{zi}}(s) = 0$$

With the given initial conditions, $y(0) = 3$ and $y'(0) = 4$, this results in

$$(s^2+3s+2)Y_{\text{zi}}(s) = 3s+13 \qquad Y_{\text{zi}}(s) = \frac{3s+13}{s^2+3s+2} = \frac{10}{s+1} - \frac{7}{s+2}$$

Its inverse transform gives $y_{\text{zi}}(t) = (10e^{-t} - 7e^{-2t})u(t)$.

(c) (Total Response) The total response is the sum of the zero-state response and the zero-input response. Thus,

$$y(t) = y_{\text{zs}}(t) + y_{\text{zi}}(t) = (14e^{-t} - 4te^{-2t} - 11e^{-2t})u(t)$$

PGR2026-00033
Krisp Technologies EX1024 Page 363

11.6.3 The Transfer Function and Circuit Analysis

For circuit analysis, we write Kirchhoff's laws and transform these to establish $H(s)$, or the system response. This process is entirely equivalent to transforming the circuit itself to the s-domain using the constitutive relations for circuit elements and then writing the (much simpler) algebraic equations for the transformed circuit. For a relaxed RLC circuit, we obtain the transformations

$$v_R(t) = Ri_R(t) \Leftrightarrow V_R(s) = RI_R(s) \tag{11.24}$$

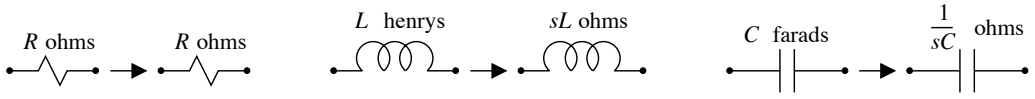
$$v_L(t) = L\frac{di_L(t)}{dt} \Leftrightarrow V_L(s) = sLI_L(s) \tag{11.25}$$

$$i_C(t) = C\frac{dv_C(t)}{dt} \Leftrightarrow I_C(s) = sCV_C(s) \tag{11.26}$$

The quantity $Z(s) = V(s)/I(s)$ defines an **impedance** with units of ohms. These relations describe a general form of Ohm's law $V(s) = Z(s)I(s)$. We use these relations to transform a circuit to the s-domain by replacing the elements R, L, and C by their impedances Z_R, Z_L, and Z_C and replacing sources by their Laplace transforms. Node or mesh equations for this **transformed circuit** now assume a much simpler algebraic form. It is important to keep in mind that the definitions of both impedance and the transfer function are based on *relaxed* LTI systems.

REVIEW PANEL 11.13

Transformation of Relaxed Circuits to the s-Domain



11.6.4 Circuits with Nonzero Initial Conditions

For a circuit with nonzero initial conditions, we can include the effect of the initial conditions $i_L(0-)$ and $v_C(0-)$ in the constitutive equations for L and C to give

$$v_L(t) = L\frac{di_L(t)}{dt} \Leftrightarrow V_L(s) = L[sI_L(s) - i_L(0-)] = Z_L I_L(s) - Li_L(0-) \tag{11.27}$$

$$i_C(t) = C\frac{dv_C(t)}{dt} \Leftrightarrow I_C(s) = C[sV_C(s) - v_C(0-)] = \frac{V_C(s)}{Z_C} - Cv_C(0-) \tag{11.28}$$

This allows the initial conditions to be included by modeling

1. A capacitor C by the impedance $Z_C = 1/sC$ *in parallel with a current source* $-Cv_C(0-)$.
2. An inductor L by the impedance $Z_L = sL$ *in series with a voltage source* $-Li_L(0-)$.

The *constants* $Li_L(0-)$ and $Cv_C(0-)$ represent the Laplace transform of the *impulses* $Li_L(0-)\delta(t)$ and $Cv_C(0-)\delta(t)$, which are responsible for the sudden changes in state leading to the nonzero initial conditions $i_L(0-)$ and $v_C(0-)$. The Laplace transform thus automatically accounts for any initial conditions. A great advantage indeed!

REVIEW PANEL 11.14

Transformation of Electrical Circuits to the s-Domain

$L \Rightarrow sL$ in *parallel* with a current source $i_L(0)/s$ having the same direction as $i_L(t)$.
$C \Rightarrow 1/sC$ in *series* with a voltage source $v_C(0)/s$ having the same polarity as $v_C(t)$.

PGR2026-00033
Krisp Technologies EX1024 Page 364

The inductor current and capacitor voltage may also be written as

$$I_L(s) = \frac{V_L(s)}{Z_L} + \frac{i_L(0-)}{s} \qquad V_C(s) = I_L(s)Z_C + \frac{v(0-)}{s} \tag{11.29}$$

This provides an alternative means of including the effect of initial conditions by modeling

1. A capacitor by $Z_C = 1/sC$ *in series with a voltage source* $v_C(0-)/s$.
2. An inductor by $Z_L = sL$ *in parallel with a current source* $i_L(0-)/s$.

These models are illustrated in Figure 11.2. The two forms for each element are equivalent (and related by the well-known Thevenin's and Norton's theorems of basic electric circuit analysis).

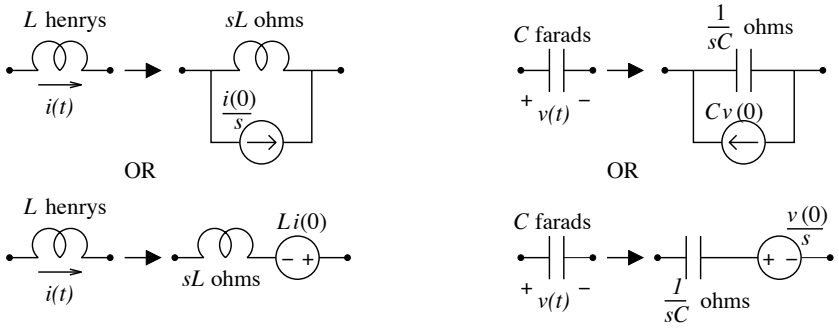


Figure 11.2 Modeling inductors and capacitors with nonzero initial conditions

EXAMPLE 11.12 (Response of Electric Circuits)

(a) (A Relaxed Circuit) Consider the relaxed RL circuit and its transformed s-domain version shown in Figure E11.12A.

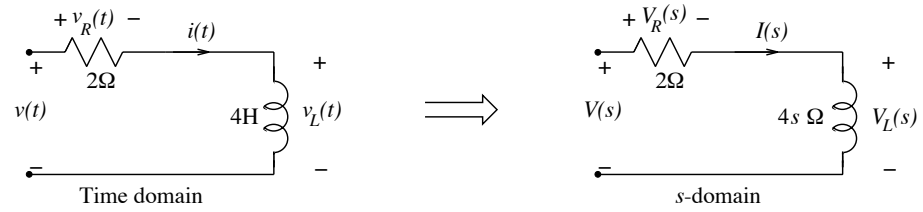


Figure E11.12A Circuit for Example 11.12(a)

The transfer function and impulse response for each of the outputs labeled is found as

$$H_R(s) = \frac{V_R(s)}{V(s)} = \frac{2}{4s+2} = \frac{0.5}{s+0.5} \qquad h_R(t) = 0.5e^{-0.5t}u(t)$$

$$H_L(s) = \frac{V_L(s)}{V(s)} = \frac{4s}{4s+2} = 1 - \frac{0.5}{s+0.5} \qquad h_L(t) = \delta(t) - 0.5e^{-0.5t}u(t)$$

$$H_I(s) = \frac{I(s)}{V(s)} = \frac{1}{4s+2} = \frac{0.25}{s+0.5} \qquad h_I(t) = 0.25e^{-0.5t}u(t)$$

The system transfer function depends on what is specified as the output.

PGR2026-00033
Krisp Technologies EX1024 Page 365

(b) (The Effect of Initial Conditions) Consider the RL circuit and its s-domain versions in the absence of initial conditions and with the initial condition $i(0) = 1$ A, as shown in Figure E11.12B.

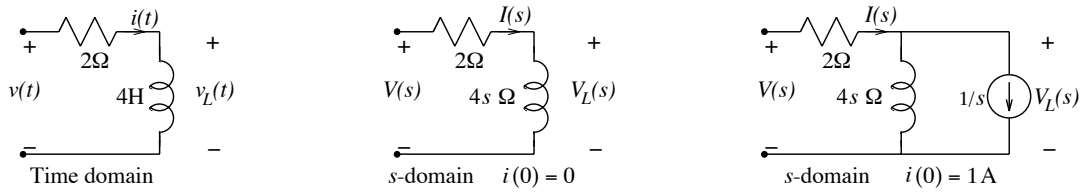


Figure E11.12B Circuit for Example 11.12(b)

The output is specified as $v_L(t)$. Let the input be $v(t) = e^{-t}u(t)$.

1. If the initial conditions are zero, we find the zero-state response as

$$V_{\text{zs}}(s) = V(s)H_L(s) = \frac{s}{(s+1)(s+0.5)} = \frac{2}{s+1} - \frac{1}{s+0.5}$$

   The zero-state response then equals $v_{\text{zs}}(t) = 2e^{-t}u(t) - e^{-0.5t}u(t)$.

2. With the initial condition $i(0) = 1$ A, we must include the effect of the initial conditions to transform the circuit as shown. A node equation gives

$$\frac{V_L(s) - V(s)}{2} + \frac{V_L(s)}{4s} + \frac{1}{s} = 0 \quad \text{or} \quad \left[\frac{1}{2} + \frac{1}{4s}\right] V_L(s) = \frac{1}{2}V(s) - \frac{1}{s}$$

   Substituting $V(s) = \dfrac{1}{s+1}$, and solving for $V_L(s)$, we get

$$V_L(s) = -\frac{s+2}{(s+1)(s+0.5)} = \frac{2}{s+1} - \frac{3}{s+0.5}$$

   Taking inverse transforms, we finally obtain the total response as

$$v_L(t) = 2e^{-t}u(t) - 3e^{-0.5t}u(t)$$

3. Since the zero-state response is $v_{\text{zs}}(t) = 2e^{-t}u(t) - e^{-0.5t}u(t)$, the initial condition is responsible for the additional term $-2e^{-0.5t}u(t)$, which corresponds to the zero-input response $v_{\text{zi}}(t)$.

   We could also compute the zero-input response $v_{\text{zi}}(t)$ separately by short circuiting the input and finding $V_{\text{zi}}(s)$ from the resulting node equation:

$$\frac{V_{\text{zi}}(s)}{2} + \frac{V_{\text{zi}}(s)}{4s} + \frac{1}{s} = 0$$

   Upon solving this, we get

$$V_{\text{zi}}(s) = -\frac{2}{s+0.2} \qquad v_{\text{zi}}(t) = -2e^{-0.5t}u(t)$$

PGR2026-00033
Krisp Technologies EX1024 Page 366

11.6.5 The Steady-State Response to Harmonic Inputs

Everlasting harmonic inputs are the eigensignals of LTI systems. Only their magnitude and phase is modified by the magnitude and phase of the system function $H(\omega)$ at the frequency of the input. As a result, the steady-state response of an LTI system to a sinusoid or harmonic input is also a harmonic at the input frequency. To find the steady-state response $y_{ss}(t)$ to the sinusoidal input $x(t) = A\cos(\omega_0 t + \theta)$, we first evaluate the transfer function at the input frequency ω_0 to obtain $H(\omega_0) = K\angle\phi$. The output is then given by $y_{ss}(t) = KA\cos(\omega_0 t + \theta + \phi)$. In effect, we multiply the input magnitude A by the transfer function magnitude K, and add the transfer function phase ϕ to the input phase θ. For multiple inputs, we simply use superposition *of the time-domain response* due to each input.

REVIEW PANEL 11.15

Finding the Steady-State Response of an LTI System to a Sinusoidal Input

Input: $x(t) = A\cos(\omega_0 t + \theta)$ **Transfer function:** Evaluate $H(s)$ at $s = j\omega_0$ as $K\angle\phi$.
Steady-state output: $y_{ss}(t) = KA\cos(\omega_0 t + \theta + \phi)$

EXAMPLE 11.13 (Steady-State Response to Harmonic Inputs)

Consider a system with $H(s) = \dfrac{s-2}{s^2+4s+4}$.

Find the steady-state response $y_{ss}(t)$ due to each of the following inputs:

1. $x(t) = 8\cos(2t)$: Since $\omega = 2$, we have $H(2) = \dfrac{2j-2}{-4+j8+4} = 0.3536\angle 45^\circ$.

   This changes the input $x(t) = 8\cos(2t)$ to yield the steady-state response

   $$y_{ss}(t) = 8(0.3536)\cos(2t + 45^\circ) = 2.8284\cos(2t + 45^\circ)$$

---

2. $x(t) = 8\cos(2t) + 16\sin(2t)$: Both terms are at the same frequency, $\omega = 2$ rad/s.

   Since $H(2) = 0.3536\angle 45^\circ$, each input term is affected by this transfer function to give

   $$y_{ss}(t) = 8(0.3536)\cos(2t+45^\circ) + 16(0.3536)\sin(2t+45^\circ) = 2.8284\cos(2t+45^\circ) + 5.6569\sin(2t+45^\circ)$$

---

3. $x(t) = 4u(t)$: We evaluate $H(0) = -0.5$ and $y_{ss}(t) = (4)(-0.5) = -2$.

---

4. $x(t) = 4u(t) + 8\cos(2t + 15^\circ)$: Each term has a different frequency.

   With $H(0) = -0.5$, the steady-state response to $4u(t)$ is -2.

   With $H(2) = 0.3536\angle 45^\circ$, the steady-state response to $8\cos(2t)$ is $2.8284\cos(2t + 60^\circ)$.

   By superposition $y_{ss}(t) = -2 + 2.8284\cos(2t + 60^\circ)$.

Comment: Note how we computed the time-domain response *before using superposition*. Of course, we can find $y_{ss}(t)$ by first finding the *total response* $y(t)$ using Laplace transforms, partial fractions, and inverse transforms, and then discarding the natural component to obtain $y_{ss}(t)$. This roundabout method is hardly worth the effort if all we want is the steady-state component $y_{ss}(t)$!

PGR2026-00033
Krisp Technologies EX1024 Page 367

11.6.6 Forced Response from the Transfer Function

The method of finding the steady-state response of an LTI system to a harmonic input generalizes to complex exponential inputs of the form $x(t) = e^{s_0 t}$ because such signals are also the eigenvalues of LTI systems. The main difference is that the system function is now evaluated at a complex frequency s_0. For example, the input $Ae^{(-2+j3)t}$ corresponds to the complex frequency $s_0 = -2 + j3$. To find the forced response to a damped sinusoid of the form $x(t) = Ae^{-\alpha t}\cos(\omega_0 t + \theta)$, we recognize it as the real part of the complex exponential $Ae^{j\theta}e^{(-\alpha+j\omega_0)t}$ at the frequency $s_0 = -\alpha + j\omega_0$ and evaluate $H(s)$ at $s = s_0$. Once the transfer function $H(s_0) = H_0\angle\phi$ is obtained, it changes only the magnitude A and phase θ of the input, and we obtain the forced response as steady-state response $y_F(t) = AH_0e^{-\alpha t}\cos(\omega_0 t + \theta + \phi)$. For multiple inputs, we simply use superposition *of the time-domain response* due to each input.

EXAMPLE 11.14 (Finding the Forced Response)

(a) Find the forced response of the system $H(s) = \dfrac{s+8}{(s+1)(s+2)}$ to the input $x(t) = 4e^{-3t}u(t)$.

First, we recognize the complex input frequency as $s_0 = -3$ and evaluate $H(s)$ at $s = s_0 = -3$ to get

$$H(s)\Big|_{s=-3} = \frac{5}{(-2)(-1)} = 2.5$$

The forced response is thus $y_F(t) = (2.5)(4)e^{-3t}u(t) = 10e^{-3t}u(t)$.

Note that if s_0 corresponds to a pole frequency, this method breaks down and must be modified.

(b) Find the forced response of the system $H(s) = \dfrac{s^2}{s^2+3s+2}$ to the input $x(t) = 3e^{-t}\sin(t - 75^\circ)u(t)$.

We recognize the complex input frequency as $s_0 = -1 + j$ and evaluate $H(s)$ at $s = s_0 = -1 + j$:

$$H(s)\Big|_{s=-1+j} = \frac{(-1+j)^2}{(-1+j)^2 + 3(-1+j) + 2} = 1.4142\angle 135^\circ$$

The forced response thus equals

$$y_F(t) = 3(1.4142)e^{-t}\sin(t - 75^\circ + 135^\circ)u(t) = 4.2426e^{-t}\sin(t + 60^\circ)u(t)$$

11.6.7 The Response to Switched Periodic Inputs

The response of a system to a switched periodic input $x_p(t)u(t)$ consists of the natural response that decays with time and a periodic steady-state component. We can use the Laplace transform to find the response $y_1(t)$ to the first period of the input and then invoke time invariance and superposition to find the total response as the sum of $y_1(t)$ and its successively shifted versions. However, it is no easy task to obtain the results in closed form. The steady-state portion of the response can be found by subtracting the natural response from the total response $y_1(t)$ for one period of the input, but even this requires some effort. Formally, the form of the natural response may be expressed as

$$y_N(t) = \sum_k A_k e^{p_k t} \tag{11.30}$$

Here, p_k are the poles of the transfer function $H(s)$ and A_k are constants. The steady-state response $y_{ss}(t)$ for the first period $(0, T)$ is then

$$y_{ss}(t) = y_1(t) - y_N(t), \qquad 0 \le t \le T \tag{11.31}$$

PGR2026-00033
Krisp Technologies EX1024 Page 368

The constants A_k in $y_{ss}(t)$ may be evaluated by letting $y_{ss}(0) = y_{ss}(T)$ (which ensures periodicity of $y_{ss}(t)$) or by evaluating $A_k = (s-p_k)Y(s)|_{s=p_k}$, where $Y(s) = X(S)H(S)$ and $X(s)$ is the Laplace transform of the switched periodic signal $x_p(t)u(t)$. Both approaches are illustrated in the following example.

EXAMPLE 11.15 (Steady-State Response to Switched Periodic Inputs)
Consider an RC circuit excited by a rectangular pulse train, as shown in Figure E11.15. Find the steady-state response of this circuit.

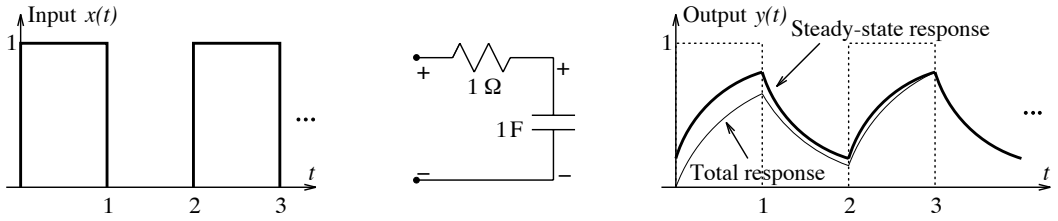


Figure E11.15 Figure for Example 11.15

The transfer function of the circuit is $H(s) = \dfrac{1}{s+1}$. The Laplace transform of $x(t)$ is $X(s) = \dfrac{X_1(s)}{1-e^{-2s}}$, where $X_1(s) = \dfrac{1}{s} - \dfrac{e^{-s}}{s}$ is the Laplace transform of one period $x_1(t), (0,2)$ of the input. The response equals

$$Y(s) = \frac{1-e^{-s}}{s(s+1)(1-e^{-2s})} = \left(\frac{1-e^{-s}}{1-e^{-2s}}\right)\left[\frac{1}{s} - \frac{1}{s+1}\right]$$

If the factor $\dfrac{1}{1-e^{-2s}}$ were not present, the response $y_1(t)$ would equal

$$y_1(t) = [1-e^{-t}]u(t) - [1-e^{-(t-1)}]u(t-1)$$

Including the factor $\dfrac{1}{1-e^{-2s}}$ results in the superposition of the shifted versions of $y_1(t)$ and leads to the total response $y(t)$ as

$$y(t) = y_1(t) + y_1(t-2)u(t-2) + y_2(t-4)u(t-4) + \cdots$$

It is not easy to simplify this result.

To find the steady-state response $y_{ss}(t)$, we subtract the natural response $y_N(t) = Ke^{-t}u(t)$ from $y_1(t)$ (the response to the first period of the input) to give

$$y_{ss}(t) = y_1(t) - y_N(t) = [1-e^{-t}]u(t) - [1-e^{-(t-1)}]u(t-1) - Ke^{-t}u(t)$$

We can find K using one of two ways:

1. We equate $y_{ss}(0)$ and $y_{ss}(T) = y_{ss}(2)$ to give

$$0 - K = (1-e^{-2}) - (1-e^{-1}) - Ke^{-2} \qquad \text{or} \qquad K = -\frac{1}{1+e}$$

2. We can also find K by evaluating

$$K = (s+1)Y(s)\Big|_{s=-1} = (s+1)H(s)X(s)\Big|_{s=-1} = \frac{1-e^{-s}}{s(1-e^{-2s})}\Big|_{s=-1} = -\frac{1}{1+e}$$

PGR2026-00033
Krisp Technologies EX1024 Page 369

The steady-state response for the first period (0, 2) then equals

$$y_{ss}(t) = y_1(t) - Ke^{-t}u(t) = (1 - e^{-t})u(t) - [1 - e^{-(t-1)}]u(t-1) + \frac{e^{-t}}{1+e}u(t)$$

We can also express $y_{ss}(t)$ by intervals as

$$y_{ss}(t) = \begin{cases} 1 - \dfrac{e^{-(t-1)}}{1+e}, & 0 \le t \le 1 \\ \dfrac{e^{-(t-2)}}{1+e}, & 1 \le t \le 2 \end{cases}$$

By the way, we can also find $y_{ss}(t)$ using periodic convolution (see Example 6.12).

11.6.8 An Application-Oriented Example: Drug Metabolism

The drug response of the body is important in determining the appropriate dose of a given drug and the proper time interval between doses in order to provide maximum effectiveness and reduce harmful side effects. In its simplest form, a first-order model for drug metabolism assumes that a dose is delivered and mixes with the bloodstream instantaneously and the rate at which the drug is metabolized in the bloodstream is directly proportional to its concentration (or the amount still unmetabolized). The drug concentration $c(t)$ at any time may then be described by

$$c'(t) + \frac{1}{\tau}c(t) = x(t) \tag{11.32}$$

where $x(t)$ is the input (the dose) and τ is the *clearance rate*, a measure of how fast the drug is metabolized (does its work) and cleared from the bloodstream. A drug regimen often requires repeated doses. If the dose c_0 is administered instantaneously at intervals of T, the input $x(t)$ may be modeled as an impulse train (a switched periodic signal):

$$x(t) = c_0\delta(t) + c_0\delta(t-T) + c_0\delta(t-2T) + \cdots \tag{11.33}$$

Since the impulse response of the system is $h(t) = e^{-t/\tau}u(t)$, the response to repeated doses is

$$c(t) = \left[c_0e^{-t/\tau} + c_0e^{-(t-T)/\tau} + c_0e^{-(t-2T)/\tau} + \cdots\right]u(t) \tag{11.34}$$

The concentration following the Nth dose and prior to the next dose is

$$c(t) = \left[c_0e^{-t/\tau} + c_0e^{-(t-T)/\tau} + c_0e^{-(t-2T)/\tau} + \cdots + c_0e^{-(t-NT)/\tau}\right], \qquad NT \le t \le (N+1)T \tag{11.35}$$

Reversing the terms, this may be simplified as

$$c(t) = c_0e^{-(t-NT)/\tau}\left[1 + e^{-T/\tau} + e^{-2T/\tau} + \cdots + e^{-NT/\tau}\right], \qquad NT \le t \le (N+1)T \tag{11.36}$$

As N becomes large, the infinite series sums to $1/(1 - e^{-T/\tau})$, and the concentration reaches a steady state described by

$$c(t) = \frac{c_0}{1 - e^{-T/\tau}}e^{-(t-NT)/\tau}, \qquad NT \le t \le (N+1)T \tag{11.37}$$

Immediately after the Nth dose ($t = NT$) and just prior to the $(N+1)$th dose ($t = (N+1)T$), the concentrations are

$$c_{sat} = \frac{c_0}{1 - e^{-T/\tau}} \qquad c_{min} = \frac{c_0e^{-T/\tau}}{1 - e^{-T/\tau}} = c_{sat} - c_0 \tag{11.38}$$

PGR2026-00033
Krisp Technologies EX1024 Page 370

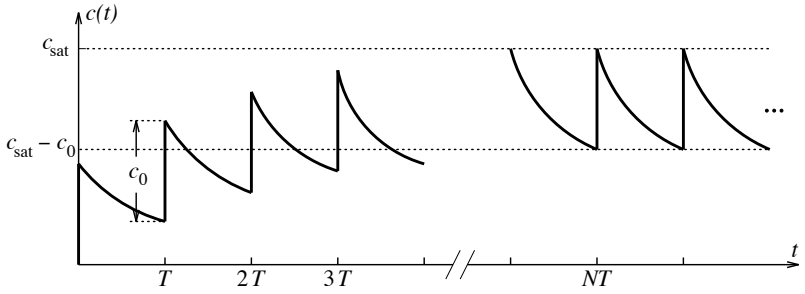


Figure 11.3 The response to repeated drug doses

The response $c(t)$ is sketched in Figure 11.3. In the steady state, the drug concentration hovers between a saturation level of c_{sat} and a minimum level of $c_{\text{sat}} - c_0$. Too low a saturation level may be ineffective, while too high a saturation level may cause harmful side effects. We must thus choose the interval T and the appropriate dose c_0 carefully in order to arrive at an optimum saturation level c_{sat} for a given drug.

EXAMPLE 11.16 (Drug Metabolism)
A drug requires concentrations between 5 mgL^{-1} and 20 mgL^{-1} for effective treatment. What is the shortest safe interval for repeated doses of 10 mgL^{-1} if the clearance rate of the drug is 7 hours? What is the maximum and minimum concentration level in the steady state?

The maximum safe level is $c_{\max} = c_{\text{sat}} = 20$ mgL^{-1}.
The minimum concentration level is $c_{\min} = c_{\text{sat}} - c_0 = 20 - 10 = 10$ mgL^{-1}. For repeated doses, we require

$$c_{\text{sat}} = 20 \approx \frac{c_0}{1 - e^{-T/\tau}} = \frac{10}{1 - e^{-T/7}}$$

This gives $T = 4.83$ hours. So, the safe interval between doses of 10 mgL^{-1} is about 5 hours.

11.6.9 Interconnected Systems

The Laplace transform is well suited to the study of interconnected LTI systems. Figure 11.4 shows the transfer function of two *relaxed* systems in cascade and in parallel.

For N systems *in cascade*, the overall impulse response $h_C(t)$ is the convolution of the individual impulse responses with $h_C(t) = h_1(t) \star h_2(t) \star \cdots \star h_N(t)$. Since the convolution operation transforms to a product, it follows that

$$H_C(s) = H_1(s) \times H_2(s) \times \cdots \times H_N(s) \qquad \text{(for } N \text{ systems in cascade)} \tag{11.39}$$

In words, the overall transfer function of a cascaded system is the product of the individual transfer functions (assuming ideal cascading with no loading effects).

Similarly, for N systems *in parallel*, the overall transfer function is the algebraic sum of the N individual transfer functions:

$$H_P(s) = H_1(s) + H_2(s) + \cdots + H_N(s) \qquad \text{(for } N \text{ systems in parallel)} \tag{11.40}$$

PGR2026-00033
Krisp Technologies EX1024 Page 371

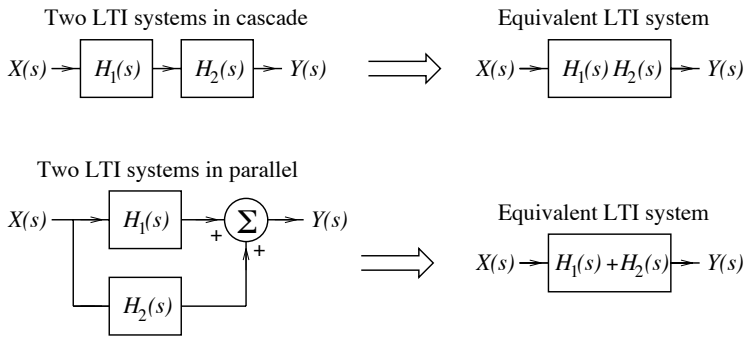


Figure 11.4 Cascaded and parallel systems and their equivalents

REVIEW PANEL 11.16

Overall Impulse Response and Transfer Function of Systems in Cascade and Parallel

Cascade: Convolve individual impulse responses. Multiply individual transfer functions.
Parallel: Add individual impulse responses. Add individual transfer functions.

11.7 Connections

The Fourier transform describes a signal as a sum (integral) of weighted harmonics, or complex exponentials. However, it cannot handle exponentially growing signals, and it cannot handle initial conditions in system analysis. In addition, the Fourier transform of signals that are not absolutely integrable usually includes impulses. The Laplace transform overcomes these shortcomings by redefining the Fourier transform as the sum of exponentially weighted harmonics:

$$X(s) = \int_{-\infty}^{\infty} x(t)e^{-(\sigma+j2\pi f)t}\,dt = \int_{-\infty}^{\infty} x(t)e^{-st}\,dt \tag{11.41}$$

This is the *two-sided Laplace transform*. If we change the lower limit in the integration to zero, we obtain the *one-sided Laplace transform*:

$$X(s) = \int_{0}^{\infty} x(t)e^{-st}\,dt \tag{11.42}$$

It applies only to causal signals but permits the analysis of systems with arbitrary initial conditions to arbitrary inputs.

11.7.1 The Convergence Factor

The exponential weighting $e^{-\sigma t}$ serves as a **convergence factor** and allows transformation of signals that are not absolutely integrable. An appropriate choice of σ can make $x(t)e^{-\sigma t}$ absolutely integrable and ensure a Laplace transform that requires no impulses. The choice of σ that makes convergence possible defines the region of convergence. We point out that even though the convergence factor $e^{-\sigma t}$ can handle exponentially growing signals of the form $e^{\alpha t}u(t)$, neither the Laplace transform nor the Fourier transform can handle signals of faster growth, such as $e^{\alpha t^2}u(t)$. Due to the presence of the convergence factor $e^{-\sigma t}$, the Laplace transform $X(s)$ of a signal $x(t)u(t)$ no longer displays the kind of symmetry present in its Fourier transform $X(f)$.

PGR2026-00033
Krisp Technologies EX1024 Page 372

11.7.2 The Laplace Transform and Fourier Transform

The Fourier transform $X(f)$ of a signal $x(t)$ corresponds to the Laplace transform $X(s)$ evaluated on the $j\omega$-axis ($\sigma = 0$), with s replaced by $j2\pi f$, but only if both of the following conditions are met:

1. The signal $x(t)$ is causal (zero for $t < 0$).
2. The signal $x(t)$ is absolutely integrable.

This connection allows us to relate the Laplace transform and Fourier transform of absolutely integrable causal signals such as the decaying exponential $e^{\alpha t}u(t)$, the exponentially damped ramp $t^n e^{\alpha t}u(t)$, and the exponentially damped sinusoid $e^{-\alpha t}\cos(\beta t + \theta)u(t)$.

Transform of Signals That Are Not Absolutely Integrable

For causal signals that are not absolutely integrable and whose Fourier transform includes impulses, such as the step $u(t)$ and the switched sinusoid $\cos(\alpha t + \theta)u(t)$, the Laplace transform still corresponds to the non-impulsive portion of the Fourier transform $X(f)$ with $j2\pi f \to s$.

We can thus always find the Laplace transform of causal signals from their Fourier transform, but not the other way around.

REVIEW PANEL 11.17

Relating the Laplace Transform and the Fourier Transform

From $X(s)$ **to** $X(f)$: If $x(t)$ is *causal* and *absolutely integrable*, simply replace s by $j2\pi f$.
From $X(f)$ **to** $X(s)$: If $x(t)$ is *causal*, delete impulsive terms in $X(f)$ and replace $j2\pi f$ by s.

11.7.3 Properties of the Laplace and Fourier Transforms

Most operational properties of the Laplace and Fourier transform also show correspondence. These include superposition, shifting, scaling, convolution, and products. The derivative property must be modified to account for initial values. Other properties, such as symmetry, do not translate to the Laplace transform. Since $s = \sigma + j\omega$ is complex, the Laplace transform may be plotted as a surface in the s-plane. The Fourier transform (with $s \to j\omega = j2\pi f$) is just the cross section of this plot along the $j\omega$-axis.

EXAMPLE 11.17 (Connecting the Laplace and Fourier Transforms)

(a) The signal $x(t) = e^{-\alpha t}u(t)$ is absolutely integrable. Its Fourier transform is $X(f) = \dfrac{1}{j2\pi f + \alpha}$.

We can thus find the Laplace transform from $X(f)$ as $X(s) = \dfrac{1}{s + \alpha}$.

The term $\epsilon^{-\sigma t}e^{-\alpha t}u(t)$ is absolutely integrable if $\sigma > -\alpha$. For positive α, this includes $\sigma = 0$ (or the $j\omega$-axis).

(b) The signal $x(t) = u(t)$ is not absolutely integrable, but $e^{-\sigma t}x(t)$ is absolutely integrable for $\sigma > 0$. Since this excludes $\sigma = 0$ (the $j\omega$-axis), we find the Laplace of $u(t)$ by dropping the impulsive part of $X(f) = 0.5\delta(f) + \frac{1}{j2\pi f}$ and replacing the quantity $j2\pi f$ by s, to give $X(s) = \frac{1}{s}$.

PGR2026-00033
Krisp Technologies EX1024 Page 373

CHAPTER 11 PROBLEMS

DRILL AND REINFORCEMENT

11.1 (The Laplace Transform and its ROC) Use the defining relation to find the Laplace transform and its region of convergence for the following:

(a) $x(t) = e^{-3t}u(t)$
(b) $x(t) = e^{3t}u(t)$

11.2 (Laplace Transforms) Find the Laplace transforms of the following:

(a) $x(t) = e^{-2t+4}u(t)$ **(b)** $x(t) = te^{-2t+4}u(t)$
(c) $x(t) = e^{-2t+4}u(t-1)$ **(d)** $x(t) = (1-e^{-t})u(t)$
(e) $x(t) = (t-2)u(t-1)$ **(f)** $x(t) = (t-2)^2u(t-1)$

11.3 (Properties) The Laplace transform of a signal $x(t)$ is $X(s) = \dfrac{4}{(s+2)^2}$. Find the Laplace transform of the following without computing $x(t)$.

(a) $y(t) = x(t-2)$ **(b)** $f(t) = x(2t)$
(c) $g(t) = x(2t-2)$ **(d)** $h(t) = x'(t)$
(e) $r(t) = x'(t-2)$ **(f)** $v(t) = x'(2t)$

11.4 (Properties) The Laplace transform of $x(t) = e^{-2t}u(t)$ is $X(s)$. Find the time signal corresponding to the following transforms without computing $X(s)$.

(a) $Y(s) = X(2s)$ **(b)** $F(s) = X'(s)$
(c) $G(s) = sX(s)$ **(d)** $H(s) = sX'(s)$

11.5 (Pole-Zero Patterns) Sketch the pole-zero patterns of the following systems. Which of these describe stable systems?

(a) $H(s) = \dfrac{(s+1)^2}{s^2+1}$ **(b)** $H(s) = \dfrac{s^2}{(s+2)(s^2+2s-3)}$
(c) $H(s) = \dfrac{s-2}{s(s+1)}$ **(d)** $H(s) = \dfrac{2(s^2+4)}{s(s^2+1)(s+2)}$
(e) $H(s) = \dfrac{16}{s^2(s+4)}$ **(f)** $H(s) = \dfrac{2(s+1)}{s(s^2+1)^2}$

11.6 (Initial and Final Value Theorems) Find the initial and final values for each $X(s)$ for which the theorems apply.

(a) $X(s) = \dfrac{s}{s+1}$ **(b)** $X(s) = \dfrac{(s+2)^2}{s+1}$
(c) $X(s) = \dfrac{(s+1)^2}{s^2+1}$ **(d)** $X(s) = \dfrac{s^2}{(s+2)(s^2+2s+2)}$
(e) $X(s) = \dfrac{2}{s(s+1)}$ **(f)** $X(s) = \dfrac{2(s^2+1)}{s(s+2)(s+5)}$
(g) $X(s) = \dfrac{16}{(s^2+4)^2}$ **(h)** $X(s) = \dfrac{2(s+1)}{s(s^2+1)^2}$

PGR2026-00033
Krisp Technologies EX1024 Page 374

11.7 (Partial Fractions and Inverse Transforms) Find the inverse transforms of the following:

(a) $H(s) = \dfrac{2}{s(s+1)}$ **(b)** $H(s) = \dfrac{2s}{(s+1)(s+2)(s+3)}$

(c) $H(s) = \dfrac{4s}{(s+3)(s+1)^2}$ **(d)** $H(s) = \dfrac{4(s+2)}{(s+3)(s+1)^2}$

(e) $H(s) = \dfrac{2}{(s+2)(s^2+4s+5)}$ **(f)** $H(s) = \dfrac{4(s+1)}{(s+2)(s^2+2s+2)}$

(g) $H(s) = \dfrac{2(s^2+2)}{(s+2)(s^2+4s+5)}$ **(h)** $H(s) = \dfrac{2}{(s+2)(s+1)^3}$

11.8 (Inverse Transforms for Nonstandard Forms) Find the time signal $h(t)$ corresponding to each Laplace transform $H(s)$.

(a) $H(s) = \dfrac{s}{s+1}$ **(b)** $H(s) = \dfrac{(s+2)^2}{s+1}$

(c) $H(s) = \dfrac{4(s-2e^{-s})}{(s+1)(s+2)}$ **(d)** $H(s) = \dfrac{4(s^2-e^{-s})}{(s+1)(s+2)}$

11.9 (Transfer Function) Find the transfer function, differential equation, and order of the following systems and determine if the system is stable.

(a) $h(t) = e^{-2t}u(t)$ **(b)** $h(t) = (1-e^{-2t})u(t)$
(c) $h(t) = te^{-t}u(t)$ **(d)** $h(t) = 0.5\delta(t)$
(e) $h(t) = \delta(t) - e^{-t}u(t)$ **(f)** $h(t) = (e^{-t} + e^{-2t})u(t)$

11.10 (Transfer Function) Find the transfer function and impulse response of the following systems.

(a) $y''(t) + 3y'(t) + 2y(t) = 2x'(t) + x(t)$
(b) $y''(t) + 4y'(t) + 4y(t) = 2x'(t) + x(t)$
(c) $y(t) = 0.2x(t)$

11.11 (System Formulation) Set up the impulse response and the system differential equation from each transfer function.

(a) $H(s) = \dfrac{3}{s+2}$ **(b)** $H(s) = \dfrac{1+2s+s^2}{(1+s^2)(4+s^2)}$

(c) $H(s) = \dfrac{2}{1+s} - \dfrac{1}{2+s}$ **(d)** $H(s) = \dfrac{2s}{1+s} - \dfrac{s}{2+s}$

11.12 (System Response) The transfer function of a system is $H(s) = \dfrac{2+2s}{4+4s+s^2}$. Find its response $y(t)$ for each input $x(t)$.

(a) $x(t) = \delta(t)$ **(b)** $x(t) = 2\delta(t) + \delta'(t)$
(c) $x(t) = e^{-t}u(t)$ **(d)** $x(t) = te^{-t}u(t)$
(e) $x(t) = 4\cos(2t)u(t)$ **(f)** $x(t) = [4\cos(2t) + 4\sin(2t)]u(t)$

11.13 (System Analysis) Find the zero-state, zero-input, and total response of the following systems, assuming $x(t) = e^{-2t}u(t), y(0) = 1$ and $y'(0) = 2$.

(a) $y''(t) + 4y'(t) + 3y(t) = 2x'(t) + x(t)$
(b) $y''(t) + 4y'(t) + 4y(t) = 2x'(t) + x(t)$
(c) $y''(t) + 4y'(t) + 5y(t) = 2x'(t) + x(t)$

PGR2026-00033
Krisp Technologies EX1024 Page 375

11.14 (Transfer Function) Find the transfer function $H(s)$ and impulse response $h(t)$ of each circuit shown in Figure P11.14. Assume that $R = 1\ \Omega$, $C = 1$ F, and $L = 1$ H where required.

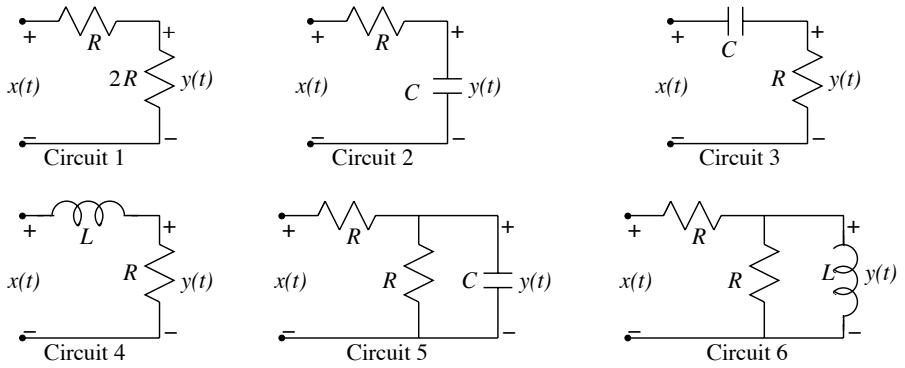


Figure P11.14 Circuits for Problem 11.14

11.15 (Steady-State Response) The transfer function of a system is $H(s) = \dfrac{2+2s}{4+4s+s^2}$. Find the *steady-state response* of this system for the following inputs.

(a) $x(t) = 4u(t)$ **(b)** $x(t) = 4\cos(2t)u(t)$
(c) $x(t) = [\cos(2t) + \sin(2t)]u(t)$ **(d)** $x(t) = [4\cos(t) + 4\sin(2t)]u(t)$

11.16 (Response to Periodic Inputs) Find the steady-state response and the total response of an RC lowpass filter with time constant $\tau = 2$ for the switched periodic inputs $x(t)$ whose one period $x_1(t)$ and time period T are defined as follows:

(a) $x_1(t) = u(t) - u(t-1)$, with $T = 2$.
(b) $x_1(t) = t[u(t) - u(t-1)]$, with $T = 1$.

REVIEW AND EXPLORATION

11.17 (The Laplace Transform) Sketch the following signals and find their Laplace transform. Are any of these signal forms or their transforms identical?

(a) $x(t) = r(t-2)$
(b) $x(t) = tu(t) - 2u(t-2)$
(c) $x(t) = tu(t-2) - 2u(t)$
(d) $x(t) = tu(t-2) - 2u(t-2)$
(e) $x(t) = tu(t-2) - 2tu(t)$

11.18 (The Laplace Transform) Find the Laplace transform of the following signals.

(a) $x(t) = \cos(t - \frac{\pi}{4})u(t)$ **(b)** $x(t) = \cos(t - \frac{\pi}{4})u(t - \frac{\pi}{4})$
(c) $x(t) = \cos(t)u(t - \frac{\pi}{4})$ **(d)** $x(t) = |\sin(\pi t)|u(t)$
(e) $x(t) = u[\sin(\pi t)]u(t)$ **(f)** $x(t) = \delta[\sin(\pi t)]u(t)$

11.19 (Properties) The Laplace transform of a signal $x(t)$ is $X(s) = \dfrac{4}{(s+2)^2}$. Find the Laplace transform of the following without computing $x(t)$.

PGR2026-00033
Krisp Technologies EX1024 Page 376

(a) $y(t) = x'(2t-2)$ **(b)** $f(t) = e^{-2t}x(t)$
(c) $g(t) = e^{-2t}x(2t)$ **(d)** $h(t) = e^{-2t-2}x(2t-2)$
(e) $r(t) = tx'(t)$ **(f)** $v(t) = 2te^{-2t}x'(2t-2)$

11.20 (Properties) The Laplace transform of $x(t) = e^{-2t}u(t)$ is $X(s)$. Find the time signal corresponding to the following transforms without computing $X(s)$.

(a) $Y(s) = sX'(2s)$
(b) $G(s) = e^{-2s}X(s)$
(c) $H(s) = se^{-2s}X(2s)$

11.21 (Partial Fractions and Inverse Transforms) Find the partial fraction expansion for each $H(s)$. Then compute the time signal $x(t)$.

(a) $H(s) = \dfrac{4}{(s^2+2s+2)^2}$ **(b)** $H(s) = \dfrac{4s}{(s^2+2s+2)^2}$
(c) $H(s) = \dfrac{32}{(s^2+4)^2}$ **(d)** $H(s) = \dfrac{4}{s(s^2+1)^2}$

11.22 (Inverse Transforms for Nonstandard Forms) Find the time signal corresponding to the following Laplace transforms.

(a) $H(s) = \dfrac{(s+1)^2}{s^2+1}$ **(b)** $H(s) = \dfrac{s^3}{(s+2)(s^2+2s+2)}$
(c) $H(s) = \dfrac{1-e^{-s}}{s(1-e^{-3s})}$ **(d)** $H(s) = \dfrac{1-e^{-s}}{s^2(1+e^{-s})}$

11.23 (Initial Value Theorem) If $H(s) = N(s)/D(s)$ is a ratio of polynomials with $N(s)$ of degree N and $D(s)$ of degree D, in which of the following cases does the initial value theorem apply, and if it does, what can we say about the initial value $h(0+)$?

(a) $D \leq N$
(b) $D > N+1$
(c) $D = N+1$

11.24 (Inversion and Partial Fractions) There are several ways to set up a partial fraction expansion and find the inverse transform. For a quadratic denominator $[(s+\alpha)^2+\beta^2]$ with complex roots, we use a linear term $As+B$ in the numerator. For $H(s) = \dfrac{As+B}{(s+\alpha)^2+\beta^2}$, we find $h(t) = e^{-\alpha t}[K_1\cos(\beta t) + K_2\sin(\beta t)]$.

(a) Express the constants K_1 and K_2 in terms of A, B, α, and β.
(b) Let $H(s) = \dfrac{2(s^2+2)}{(s+2)(s^2+2s+2)} = \dfrac{C}{s+2} + \dfrac{As+B}{(s+\alpha)^2+\beta^2}$. Find the constants A, B, and C by comparing the numerator of $H(s)$ with the assumed form and then find $h(t)$.
(c) Extend these results to find $x(t)$ if $X(s) = \dfrac{40}{(s^2+4)(s^2+2s+2)}$.

11.25 (Inversion and Partial Fractions) Let $H(s) = \dfrac{s+2}{(s+3)(s+4)^3}$. Its partial fraction expansion (PFE) has the form

$$H(s) = \frac{K}{s+3} + \frac{A_0}{(s+4)^3} + \frac{A_1}{(s+4)^2} + \frac{A_2}{s+4}$$

PGR2026-00033
Krisp Technologies EX1024 Page 377

Finding the two constants K and A_0 is easy, but A_1 and A_2 require derivatives of $(s+1)^3H(s)$. An alternative is to recognize that $H(s)$ and its PFE are valid for any value of s, excluding the poles. All we need is to evaluate the PFE at two values of s to yield two equations in two unknowns, assuming K and A_0 are already known.

(a) Try this approach using $s = -2$ and -5 to find the PFE constants A_1 and A_2, assuming you have already evaluated K and A_0.

(b) Repeat part (a), choosing $s = 0$ and $s = -6$. Is there a "best choice"?

11.26 (Integral Equations) Integral equations arise in many contexts. One approach to solving integral equations is by using the convolution property of the Laplace transform. Find the transfer function and impulse response of a filter whose input-output relation is described by

(a) $y(t) = x(t) - 2\displaystyle\int_{-\infty}^{t} y(\lambda)e^{-(t-\lambda)}u(t-\lambda)\,d\lambda$

(b) $y(t) = x(t) + \displaystyle\int_{-\infty}^{t} y(\lambda)e^{-3(t-\lambda)}u(t-\lambda)\,d\lambda$

11.27 (System Analysis) Consider a system whose impulse response is $h(t) = 2e^{-2t}u(t)$. Find its response to the following inputs.

(a) $x(t) = \delta(t)$ **(b)** $x(t) = u(t)$
(c) $x(t) = e^{-t}u(t)$ **(d)** $x(t) = e^{-2t}(t)$
(e) $x(t) = \cos(t)$ **(f)** $x(t) = \cos(t)u(t)$
(g) $x(t) = \cos(2t)$ **(h)** $x(t) = \cos(2t)u(t)$

11.28 (System Analysis) For each circuit shown in Figure P11.28, assume that $R = 1\ \Omega$, $C = 1$ F, and $L = 1$ H where required and

(a) Find the transfer function $H(s)$ and the impulse response $h(t)$.

(b) Find the response to $x(t) = e^{-t}u(t)$, assuming $v_C(0) = 0$ and $i_L(0) = 0$.

(c) Find the response to $x(t) = u(t)$, assuming $v_C(0) = 1$ V and $i_L(0) = 2$ A (directed down).

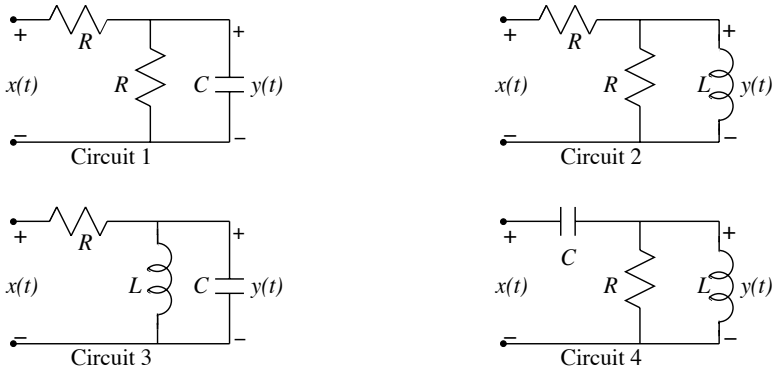


Figure P11.28 Circuits for Problem 11.28

11.29 (System Analysis) Consider a system whose impulse response is $h(t) = 2e^{-2t}\cos(t)u(t)$. Let an input $x(t)$ produce the output $y(t)$. Find $x(t)$ for the following outputs.

(a) $y(t) = \cos(2t)$ **(b)** $y(t) = 2 + \cos(2t)$ **(c)** $y(t) = \cos^2(t)$

PGR2026-00033
Krisp Technologies EX1024 Page 378

11.30 (System Analysis and Convolution) Consider the following system

$$x(t) \longrightarrow \boxed{h(t) = 2e^{-t}u(t) - \delta(t)} \longrightarrow y(t)$$

(a) Let $x(t) = e^{-t}u(t)$. Find the output $y(t)$, using Laplace transforms. Verify your results, using time-domain convolution.

(b) Let $x(t) = u(t)$. Find the output $y(t)$, using Laplace transforms. Verify your results, using time-domain convolution.

(c) Let $x(t) = \cos(t)$. Find the output $y(t)$, using Laplace transforms. Verify your results, using time-domain convolution.

11.31 (Response to Periodic Inputs) Find the steady-state response and the total response of the following systems to the switched periodic inputs $x(t)$ whose one period equals $x_1(t)$.

(a) $H(s) = \dfrac{s+3}{s^2+3s+2}$ $\qquad x_1(t) = [u(t) - u(t-1)]$ $\qquad T = 2$

(b) $H(s) = \dfrac{s+2)}{s^2+4s+3}$ $\qquad x_1(t) = \text{tri}(t-1)$ $\qquad T = 2$

11.32 (Stability) A perfect differentiator is described by $y(t) = \dfrac{d\,x(t)}{dt}$.

(a) Find its transfer function $H(s)$ and use the condition for BIBO stability to show that the system is unstable.

(b) Verify your conclusion by finding the impulse response $h(t)$ and applying the condition for BIBO stability in the time domain.

11.33 (Stability) A perfect integrator is described by $y(t) = \displaystyle\int_{-\infty}^{t} x(t)\,dt$.

(a) Find its transfer function $H(s)$ and use the condition for BIBO stability to show that the system is unstable.

(b) Verify your conclusion by finding the impulse response $h(t)$ and applying the condition for BIBO stability in the time domain.

11.34 (Model-Order Reduction) For a stable system, the effect of poles much farther from the $j\omega$-axis than the rest is negligible after some time, and the behavior of the system may be approximated by a lower-order model from the remaining or **dominant poles**.

(a) Let $H(s) = \dfrac{100}{(s+1)(s+20)}$. Find its impulse response $h(t)$, discard the term in $h(t)$ that makes the smallest contribution to obtain the reduced impulse response $h_R(t)$, and establish the transfer function of the reduced model by computing $H_R(s)$.

(b) If the poles $(s + \alpha_k)$ of $H(s)$ to be neglected are written in the form $\alpha(1 + \frac{s}{\alpha_k})$, $H_R(s)$ may also be computed directly from $H(s)$ if we discard just the factors $(1 + \frac{s}{\alpha_k})$ from $H(s)$. Obtain $H_R(s)$ from $H(s)$ using this method and explain any differences from the results of part (a).

(c) As a rule of thumb, poles with magnitudes ten times larger than the rest may be neglected. Use this idea to find the reduced model $H_R(s)$ and its order if $H(s) = \dfrac{400}{(s^2+2s+200)(s+20)(s+2)}$.

PGR2026-00033
Krisp Technologies EX1024 Page 379

COMPUTATION AND DESIGN

ctfgui A GUI for Analog System Visualization

The graphical user interface `ctfgui` allows you to visualize the impulse response, the pole-zero plot, and the frequency response of an analog system.
You can select the transfer function and plot results using various options.
To explore this routine, type `ctfgui` at the MATLAB prompt.

11.35 (System Response in Symbolic Form) The ADSP routine `sysresp2` yields a symbolic expression for the system response (see Chapter 21 for examples of its usage). Consider a system described by the differential equation $y'(t) + 2y(t) = 2x(t)$. Use `sysresp2` to obtain

(a) Its step response.
(b) Its impulse response.
(c) Its zero-state response to $x(t) = 4e^{-3t}u(t)$.
(d) Its complete response to $x(t) = 4e^{-3t}u(t)$, with $y(0) = 5$.

11.36 (System Response in Symbolic Form) Consider the system $y''(t) + 4y(t) + Cy(t) = x(t)$.

(a) Use `sysresp2` to obtain its step response and impulse response for $C = 3,\ 4,\ 5$ and plot each response.
(b) How does the step response differ for each value of C? For what value of C would you expect the smallest rise time? For what value of C would you expect the smallest 3% settling time?
(c) Confirm your predictions in part (b) by numerically estimating the rise time and settling time, using the ADSP routine `trbw`.

11.37 (Steady-State Response in Symbolic Form) The ADSP routine `ssresp` yields a symbolic expression for the steady-state response to sinusoidal inputs (see Chapter 21 for examples of its usage). Find the steady-state response to the input $x(t) = 2\cos(3t - \frac{\pi}{3})$ for each of the following systems and plot the results over $0 \le t \le 3$.

(a) $y'(t) + \alpha y(t) = 2x(t)$, for $\alpha = 1,\ 2$
(b) $y''(t) + 4y(t) + Cy(t) = x(t)$, for $C = 3,\ 4,\ 5$

PGR2026-00033
Krisp Technologies EX1024 Page 380

Chapter 12

APPLICATIONS OF THE LAPLACE TRANSFORM

12.0 Scope and Objectives

Applications of the Laplace transform cut across many disciplines. This chapter formalizes the concept of frequency response and introduces Bode plots as a useful means of describing the frequency response. It quantifies measures related to system performance in both the time-domain and frequency-domain, and concludes with an introduction to feedback and its application to signal processing.

12.1 Frequency Response

The frequency-domain performance of systems is based on the concept of the frequency response. The frequency response $H(\omega)$ of a system represents the transfer function for harmonic (and dc) inputs and is found from $H(s)$ by the substitution $s \to j\omega$ (with $\omega = 0$ for dc). It is useful primarily for stable systems whose natural (transient) response does indeed decay with time. Since the transfer function $H(s)$ of LTI systems is a ratio of polynomials in s, the frequency response $H(\omega)$ is a ratio of polynomials in $j\omega$. It is characterized by *conjugate symmetry* with $H(-\omega) = H^*(\omega)$. In other words, if we express the frequency response in magnitude-phase form as $H(\omega) = |H(\omega)|\angle\theta(\omega)$, its magnitude $|H(\omega)|$ is an even symmetric function of ω, and its phase $\theta(\omega)$ is an odd symmetric function of ω.

12.1.1 Phase Delay and Group Delay

The phase and group delay of a system whose transfer function is $H(\omega) = |H(\omega)|\angle\theta(\omega)$ is defined as

$$t_p = -\frac{\theta(\omega)}{\omega} \quad \text{(phase delay)} \qquad t_g = -\frac{d\,\theta(\omega)}{d\omega} \quad \text{(group delay)} \tag{12.1}$$

If $\theta(\omega)$ varies linearly with frequency, t_p and t_g are not only constant but are also equal. For LTI systems (with rational transfer functions), the phase $\theta(\omega)$ is a transcendental function, but *the group delay is always a rational function of* ω^2 and is much easier to work with in many filter applications.

REVIEW PANEL 12.1

The Group Delay of LTI Systems Is Always an Even Rational Function of ω^2

Phase delay: $t_p(\omega) = -\dfrac{\theta(\omega)}{\omega}$ **Group delay:** $t_g(\omega) = -\dfrac{d\,\theta(\omega)}{d\omega} = f(\omega^2)$

PGR2026-00033
Krisp Technologies EX1024 Page 381

EXAMPLE 12.1 (Frequency Response)

(a) Discuss the frequency response of the RC circuit shown in Figure E12.1A(1).

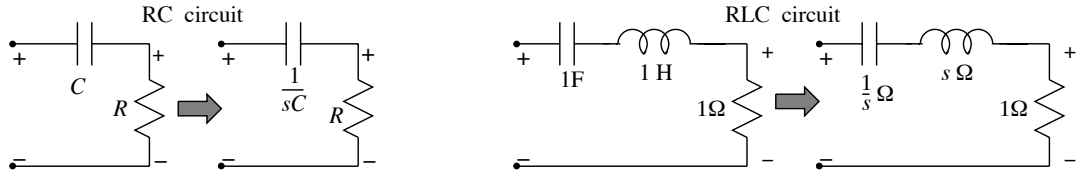


Figure E12.1A(1) The RC circuit (left) and RLC circuit (right) for Example 12.1(a and b)

The output is the resistor voltage. The transfer function may be written as

$$H(s) = \frac{Y(s)}{X(s)} = \frac{R}{R + 1/sC} = \frac{sRC}{1 + sRC} = \frac{s\tau}{1 + s\tau}$$

Its frequency response is given by

$$H(\omega) = \frac{j\omega\tau}{1 + j\omega\tau} \qquad |H(\omega)| = \frac{\omega\tau}{\sqrt{1 + \omega^2\tau^2}} \qquad \phi(\omega) = \tfrac{\pi}{2} - \tan^{-1}(\omega\tau)$$

The magnitude and phase of the transfer function is sketched in Figure E12.1A(2).

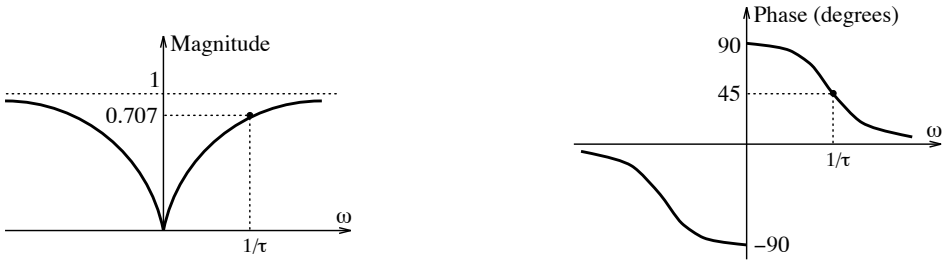


Figure E12.1A(2) Frequency response of the RC circuit for Example 12.1(a)

The system describes a *highpass filter* because $|H(\omega)|$ increases monotonically from $|H(0)| = 0$ to a maximum of $H_{\max} = |H(\infty)| = 1$. Its half-power frequency (where $|H(\omega)| = H_{\max}/\sqrt{2}$) is $\omega = 1/\tau$, but its half-power bandwidth is infinite. The phase decreases from a maximum of 90° at $\omega = 0$ to 0° as $\omega \to \infty$. The phase at $\omega = 1/\tau$ is 45°.

(b) Discuss the frequency response of the RLC filter shown in Figure E12.1A(1).

We find the transfer function $H(s)$ and frequency response $H(\omega)$ of this filter as

$$H(s) = \frac{s}{s^2 + s + 1} \qquad H(\omega) = \frac{j\omega}{(1 - \omega^2) + j\omega}$$

This describes a bandpass filter because its magnitude response is zero at $\omega = 0$ and very small at high frequencies, with a peak in between. Frequency-domain measures for bandpass filters include the half-power frequencies ω_1, ω_2 at which the gain is $1/\sqrt{2}$ times the peak gain, the center frequency ω_0 (defined as $\omega_0 = \sqrt{\omega_1\omega_2}$, the geometric mean of the half-power frequencies), and the half-power bandwidth (the frequency band $B = \omega_2 - \omega_1$ covered by the half-power frequencies). The peak gain

PGR2026-00033
Krisp Technologies EX1024 Page 382

of this bandpass filter is unity and occurs $\omega = 1$ rad/s. If we let $H(\omega)|^2 = 0.5$, we find its half-power frequencies as $\omega_1 = 0.618$ rad/s and $\omega_2 = 1.618$ rad/s or $\frac{1}{2}(\sqrt{5}\mp 1)$. The two frequencies are reciprocals of each other. Thus, its center frequency equals $\omega_0 = \sqrt{\omega_1\omega_2} = 1$ rad/s. Its half-power bandwidth is $B = \omega_2 - \omega_1 = 1$ rad/s. Its phase equals zero at $\omega = \omega_0$, 45° at $\omega = 0.618$ rad/s, and −45° at $\omega = 1.618$ rad/s. The phase response and group delay of this filter can be computed as

$$\phi(\omega) = 0.5\pi - \tan^{-1}\left[\frac{\omega}{1-\omega^2}\right] \qquad t_g = \frac{\omega^2+1}{\omega^4-\omega^2+1}$$

The step response of this filter is given by

$$S(s) = \frac{H(s)}{s} = \frac{1}{s^2+s+1} = \frac{1}{(s+\frac{1}{2})^2+(\frac{1}{2}\sqrt{3})^2}$$

Inverse transformation yields $s(t) = \dfrac{\sqrt{3}}{2}e^{-t/2}\sin(\frac{\sqrt{3}}{2}t)$. This is a damped sinusoid that is typical of the impulse response of bandpass filters. The impulse response nearly dies out in about 5 time constants (or 10 s), or less than three half-cycles.

12.2 Minimum-Phase Filters

A minimum-phase system has all its poles and zeros in the left half-plane of the s-plane. Its transfer function $H(s)$ has the factored form

$$H(s) = K\frac{(s-z_1)(s-z_2)\cdots(s-z_M)}{(s-p_1)(s-p_2)\cdots(s-p_N)}, \qquad N > M,\ K > 0,\ \text{Re}[z_k] < 0,\ \text{Re}[p_k] < 0 \tag{12.2}$$

It has the smallest group delay and smallest deviation from zero phase, at every frequency, among all transfer functions with the same magnitude spectrum $|H(\omega)|$. A *stable* system is called **mixed phase** if some of its zeros lie in the right half-plane (RHP) and **maximum phase** if all its zeros lie in the RHP. Of all possible stable systems with the same magnitude response, there is only one minimum-phase system.

REVIEW PANEL 12.2

Minimum-Phase Systems Are Always Stable and Have the Smallest Group Delay

Stable minimum-phase system: All poles and zeros in the LHP; smallest group delay
Mixed phase: Some zeros in the RHP **Maximum phase:** All zeros in the RHP

EXAMPLE 12.2 (The Minimum-Phase Concept)
Consider the transfer functions of the following stable systems:

$$H_1(s) = \frac{(s+1)(s+2)}{(s+3)(s+4)(s+5)} \qquad H_2(s) = \frac{(s-1)(s+2)}{(s+3)(s+4)(s+5)} \qquad H_3(s) = \frac{(s-1)(s-2)}{(s+3)(s+4)(s+5)}$$

All have the same magnitude but different phase and delay. $H_1(s)$ is minimum phase (no zeros in the RHP), $H_2(s)$ is mixed phase (one zero outside the RHP), and $H_3(s)$ is maximum phase (all zeros in the RHP).

PGR2026-00033
Krisp Technologies EX1024 Page 383

12.2.1 Minimum-Phase Transfer Function from the Magnitude Spectrum

A minimum-phase transfer function $H(s)$ may be found from its magnitude spectrum $|H(\omega)|$ alone. We start by writing $|H(\omega)|^2 = H(\omega)H(-\omega)$. Since $H(\omega)$ possesses conjugate symmetry, $|H(\omega)|^2$ describes a non-negative, even polynomial in ω^2. Now, if we create $H(s)H(-s)|_{s=j\omega} = |H(\omega)|^2$, its poles and zeros will show **quadrantal symmetry** (which describes both conjugate symmetry and mirror symmetry about the $j\omega$-axis). This is illustrated in Figure 12.1.

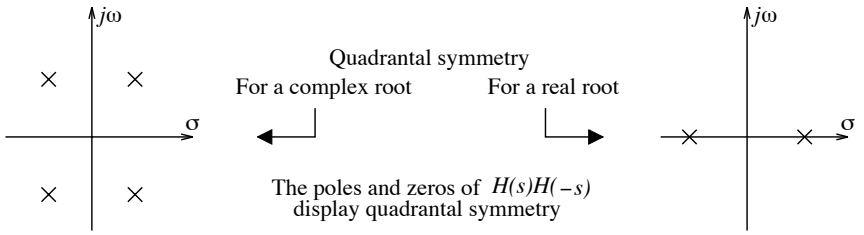


Figure 12.1 Illustrating quadrantal symmetry

The system transfer function $H(s) = KP(s)/Q(s)$ is obtained from $H(s)H(-s)$ by selecting only the left half-plane poles (for stability) and left half-plane zeros (to ensure minimum phase) and a value of K that matches the magnitudes of $H(s)$ and $H(\omega)$ at a convenient frequency (such as dc).

REVIEW PANEL 12.3

Minimum-Phase Transfer Function $H(s)$ from the Magnitude Spectrum $|H(\omega)|$

Generate $Y(s) = |H(\omega)|^2|_{j\omega\to s}$. Pick only the LHP poles and zeros of $Y(s)$ to generate $H(s)$.

EXAMPLE 12.3 (Minimum-Phase Transfer Functions)

Find the minimum-phase transfer function corresponding to $|H(\omega)|^2 = \dfrac{4(9+\omega^2)}{4+5\omega^2+\omega^4}$.

First, we replace $j\omega$ by s (or ω^2 by $-s^2$) to give $|H(s)|^2 = H(s)H(-s) = \dfrac{4(9-s^2)}{4-5s^2+s^4}$.

Next, we factor this to get $H(s)H(-s) = \dfrac{4(3-s)(3+s)}{(2-s)(2+s)(1-s)(1+s)}$.

The minimum-phase transfer function $H(s)$ is found by choosing only the LHP poles and zeros as

$$H(s) = \frac{K(s+3)}{(s+1)(s+2)}$$

Finally, $|H(\omega)| = \sqrt{9} = 3$ at $\omega = 0$, and $H(s) = 1.5K$ at $s = 0$. To match the dc gain, $1.5K = 3$ or $K = 2$.

Thus, $K = 2$ and $H(s) = \dfrac{2(s+3)}{(s+1)(s+2)}$.

12.3 Bode Plots

Bode plots allow us to plot the frequency response over a wide frequency and amplitude range by using logarithmic *scale compression*. The magnitude $H(\omega)$ is plotted in **decibels** as $H_{\text{dB}} = 20\log(|H(\omega)|)$ against

PGR2026-00033
Krisp Technologies EX1024 Page 384

$\log(\omega)$. Since any phase variation can be brought into the range $\pm\pi$, the phase $\phi(\omega)$ is typically plotted on a *linear* scale versus $\log(\omega)$. Since $\log(0) = -\infty$ and the logarithm of a negative number is not real, log and Bode plots use only positive frequencies and exclude dc. For LTI systems whose transfer function is a ratio of polynomials in $j\omega$, a rough sketch can be quickly generated using linear approximations called **asymptotes** over different frequency ranges to obtain **asymptotic Bode plots**.

The log operator provides nonlinear compression. A **decade** (tenfold) change in $|H(\omega)|$ results in exactly a 20-dB change in its decibel magnitude, because

$$20\log|10H(\omega)| = 20\log|H(\omega)| + 20\log 10 = H_{\text{dB}} + 20 \tag{12.3}$$

Similarly, a twofold (an octave) change in $|H(\omega)|$ results in approximately a 6-dB change in its decibel magnitude, because

$$20\log|2H(\omega)| = 20\log|H(\omega)| + 20\log 2 \approx H_{\text{dB}} + 6 \tag{12.4}$$

The slope of the magnitude is measured in decibels per decade (dB/dec) or decibels per octave (dB/oct). For example, if $|H(\omega)| = A\omega^k$, the slope of H_{dB} versus ω is $20k$ dB/dec or $6k$ dB/oct.

REVIEW PANEL 12.4

Bode Plots Use Magnitude and Frequency Compression

Magnitude: In decibels ($20\log|H(f)|$) versus $\log(f)$ **Phase:** In degrees or radians against $\log(f)$
Asymptotic plots: Linear approximations in the low- and high-frequency regions

The Transfer Function

The steady-state transfer function of an LTI system may be expressed as

$$H(\omega) = \frac{B_m(j\omega)^m + \cdots + B_2(j\omega)^2 + B_1(j\omega) + B_0}{A_n(j\omega)^n + \cdots + A_2(j\omega)^2 + A_1(j\omega) + A_0} \tag{12.5}$$

The numerator and denominator can be factored into linear and quadratic factors in $j\omega$ with real coefficients. A **standard form** is obtained by setting the real part of each factored term to unity to obtain

$$H(\omega) = \frac{K_1(j\omega)^k(1 + j\omega/a_1)(1 + j\omega/a_2)\cdots[1 + C_1(j\omega/c_1) + (j\omega/c_1^2)]\cdots}{(1 + j\omega/d_1)(1 + j\omega/d_2)\cdots[1 + E_1(j\omega/e_1) + (j\omega/e_1^2)]\cdots} \tag{12.6}$$

The standard factored form offers two advantages. Upon taking logarithms, products transform to a sum. This allows us to sketch the decibel magnitude of the simpler individual factors separately and then use superposition to plot the composite response. Similarly, the phase is the algebraic sum of the phase contributions due to each factor, and we can thus plot the composite phase by superposition. Note that $H(\omega)$ is a constant or proportional to some power of $j\omega$ as $\omega \to \infty$ or $\omega \to 0$. There are only four types of terms that we must consider in order to plot the spectrum for any rational transfer function: a constant, the term $j\omega$, the term $(1 + j\omega/\alpha)$, and a quadratic term.

12.3.1 Bode Magnitude Plots

The asymptotic Bode magnitude of various terms is illustrated in Figure 12.2. The decibel magnitude of $H(\omega) = j\omega$ is simply $H_{\text{dB}} = 20\log(\omega)$ dB. This describes a straight line with a slope of 20 dB/dec for all ω. A convenient reference for plotting this term is $\omega = 1$ rad/s when $H_{\text{dB}} = 0$ dB. Its phase is 90° for all ω.

PGR2026-00033
Krisp Technologies EX1024 Page 385

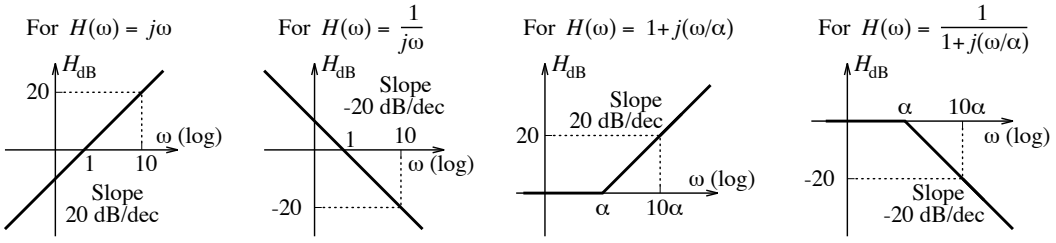


Figure 12.2 Asymptotic Bode magnitude plots for for some standard forms

The decibel magnitude of $H(\omega) = 1 + j\frac{\omega}{\alpha}$ is $H_{\text{dB}} = 20\log|1 + j\frac{\omega}{\alpha}|$ dB, and yields separate linear approximations for low and high frequencies

$$H_{\text{dB}} = 20\log\left|1 + j\frac{\omega}{\alpha}\right| \approx \begin{cases} 0 \text{ (slope = 0 dB/dec)}, & \omega \ll \alpha \\ 20\log\left|\frac{\omega}{\alpha}\right| \text{ (slope = 20 dB/dec)}, & \omega \gg \alpha \end{cases} \tag{12.7}$$

The two straight lines intersect at $\omega = \alpha$ where $H_{\text{dB}} \approx 3$ dB. The frequency α is called the **break frequency**, or **corner frequency**. For a linear factor, it is also called the **half-power frequency**, or the **3-dB frequency**. The difference between the asymptotic and true magnitude equals 3 dB at $\omega = \alpha$, 1 dB at frequencies an octave above and below α, and nearly 0 dB at frequencies a decade above and below α. All the slopes and magnitudes change sign if a term is in the denominator and increase k-fold if a term is repeated k times.
The decibel value of a constant $H(\omega) = K$ is $H_{\text{dB}} = 20\log|K|$ dB, a constant for all ω. If K is negative, the negative sign contributes a phase of 180° to the phase plot.

REVIEW PANEL 12.5

Asymptotic Magnitude Plot of Linear and Repeated Linear Terms

Term $= j\omega$: Straight line with slope = 20 dB/dec for *all* ω with $H_{\text{dB}} = 0$ at $\omega = 1$ rad/s.
Term $= 1 + j\frac{\omega}{\alpha}$: For $\omega \le \alpha$, $H_{\text{dB}} = 0$. For $\omega \ge \alpha$, straight line with slope = 20 dB/dec.
If repeated k **times:** Multiply slopes by k. **If in denominator:** Slopes are negative.

The Bode magnitude plot for a transfer function with several terms is the sum of similar plots for each of its individual terms. Since the asymptotic plots are linear, so is the composite plot. It can actually be sketched directly if we keep track of the following in the transfer function:

1. The initial slope is zero if terms of the form $(j\omega)^{\pm k}$ are absent. Otherwise, the initial slope equals $\pm 20k$ dB and the asymptote always passes through 0 dB at $\omega = 1$ *if we ignore the constant term.*

2. If the break frequencies are arranged in ascending order, the slope changes at each successive break frequency. The change in slope equals $\pm 20k$ dB/dec for $(1 + j\omega/\omega_B)^{\pm k}$. The slope increases by $20k$ dB/dec if ω_B corresponds to a numerator term, and decreases by $20k$ dB/dec otherwise.

3. A constant term contributes a constant decibel value and shifts the entire plot. We include this last.

As a consistency check, the final slope should equal $20n$ dB/dec, where n is the difference between the order of the numerator and denominator polynomials in $H(\omega)$.

PGR2026-00033
Krisp Technologies EX1024 Page 386

REVIEW PANEL 12.6

Guidelines for Sketching a Composite Asymptotic Magnitude Plot for $H(\omega)$

Initial slope: 20 dB/dec (with $H_{\text{dB}} = 0$ at $\omega = 1$ rad/s) if a term $j\omega$ is present; 0 dB/dec if absent.
At each break frequency: The slope *changes* by 20 dB/dec due to a linear term.
If repeated k times: Multiply slopes by k. **If in denominator:** Slopes are negative.

A rough sketch of the true plot may also be generated by adding approximate **correction factors** at the break frequencies. The correction factor at the break frequency α, corresponding to a linear factor $1 + j\frac{\omega}{\alpha}$ is 3 dB if the surrounding break frequencies are at least a decade apart or 2 dB if an adjacent break frequency is an octave away. For repeated factors with multiplicity k, these values are multiplied by k. At all other frequencies, the true value must be computed directly from $H(\omega)$.

REVIEW PANEL 12.7

Magnitude Correction at a Break Frequency in Composite Asymptotic Plots

Correction = 3 dB: If surrounding break frequencies are at least a *decade* away.
Correction = 2 dB: If one of the surrounding break frequencies is an *octave* away.
If repeated k times: Multiply correction by k. **If in denominator:** Change sign.

EXAMPLE 12.4 (Bode Magnitude Plots)

(a) (Linear Factors) Let $H(\omega) = \dfrac{40(0.25 + j\omega)(10 + j\omega)}{j\omega(20 + j\omega)}$.

We write this in the standard form

$$H(\omega) = \frac{5(1 + j\omega/0.25)(1 + j\omega/10)}{j\omega(1 + j\omega/20)}$$

The break frequencies, in ascending order, are:

$$\omega_1 = 0.25 \text{ rad/s (numerator)} \quad \omega_2 = 10 \text{ rad/s (numerator)} \quad \omega_3 = 20 \text{ rad/s (denominator)}$$

The term $1/j\omega$ provides a starting asymptote -20 dB/dec whose value is 0 dB at $\omega = 1$ rad/s. We can now sketch a composite plot by including the other terms:

At $\omega_1 = 0.25$ rad/s (numerator), the slope changes by +20 dB/dec to 0 dB/dec.

At $\omega_2 = 10$ rad/s (numerator), the slope changes by +20 dB/dec to 20 dB/dec.

At $\omega_3 = 20$ rad/s (denominator), the slope changes by -20 dB/dec to 0 dB/dec.

We find the asymptotic magnitudes at the break frequencies:

$\omega_1 = 0.25$ rad/s (numerator): 12 dB (2 octaves below $\omega = 1$ rad/s).

$\omega_2 = 10$ rad/s (numerator): 12 dB (zero slope from $\omega = 0.5$ rad/s to $\omega = 10$ rad/s).

$\omega_3 = 20$ rad/s (denominator): 18 dB (1 octave above $\omega = 10$ rad/s, with slope 6 dB/oct).

Finally, the constant 5 shifts the plot by $20 \log 5 = 14$ dB. Its Bode plot is shown in Figure E12.4(a).

PGR2026-00033
Krisp Technologies EX1024 Page 387

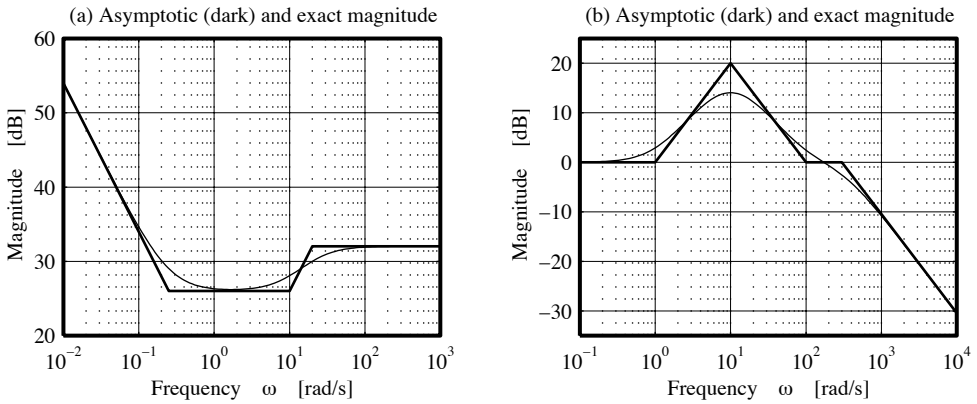


Figure E12.4 Bode magnitude plots for Example 12.4(a and b)

We apply the following correction factors at the break frequencies:

ω_B **(rad/s)**	**Surrounding Frequency**	**Correction**	**Corrected dB Magnitude**
0.25 (numerator)	10 (>1 decade away)	+3 dB	$14+12+3=29$ dB
10 (numerator)	20 (1 octave away)	+2 dB	$14+12+2=28$ dB
20 (denominator)	10 (1 octave away)	−2 dB	$14+18-2=30$ dB

For the exact decibel magnitude—say, at $\omega = 20$ rad/s—we use $H(\omega)$ to compute

$$H_{\text{dB}} = 20\log\left|5\frac{(1+j\omega/0.25)(1+j\omega/10)}{j\omega(1+j\omega/20)}\right|_{\omega=20} = 30.0007 \text{ dB}$$

(b) (Repeated Factors) Let $H(\omega) = \dfrac{(1+j\omega)(1+j\omega/100)}{(1+j\omega/10)^2(1+j\omega/300)}$.

Its Bode plot is sketched in Figure E12.4(b). We make the following remarks:

The starting slope is 0 dB/dec since a term of the form $(j\omega)^k$ is absent.

The slope changes by -40 dB/dec at $\omega_B = 10$ rad/s due to the repeated factor.

For a true plot, the correction factor at $\omega_B = 10$ rad/s (denominator) is -6 dB.

The true magnitudes at $\omega = 100$ rad/s and $\omega = 300$ rad/s, which are neither an octave apart nor a decade apart, must be computed directly from $H(\omega)$.

(c) For the Bode magnitude plot shown in Figure E12.4C, find α, β, and $H(s)$.

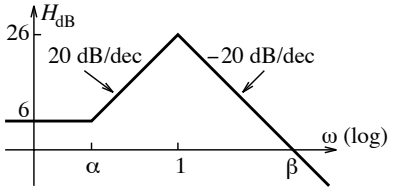


Figure E12.4C Bode plot for Example 12.4(c)

PGR2026-00033
Krisp Technologies EX1024 Page 388

The 20-dB/dec slope and a 20-dB rise from α to 1 rad/s places α a decade below 1 rad/s. Thus, $\alpha = 0.1$ rad/s. The -20-dB/dec (or -6 dB/oct) slope and a 26-dB drop from 1 rad/s to β places β a decade (a 20-dB drop) plus an octave (a 6-dB drop) away, at $\beta = 20$ rad/s. We may also find β from the "slope = rise over run" rule provided we compute the run from logarithms of the frequencies:

$$\text{slope} = -20 \text{ dB/dec} = \frac{0 - 26}{\log \beta - \log 1} \qquad \log \beta = 1.3 \qquad \beta \approx 20 \text{ rad/s}$$

The 20-dB/dec slope at 0.1 rad/s corresponds to a numerator factor $(1 + j\omega/0.1)$, and the slope change of -40 dB/dec at 1 rad/s corresponds to the quadratic denominator factor $(1 + j\omega)^2$. The decibel shift of 6 dB corresponds to a linear gain of $K = 2$ (because $20 \log K = 6$ dB). Thus,

$$H(\omega) = \frac{2(1 + j\omega/0.1)}{(1 + j\omega)^2} \qquad H(s) = \frac{2(1 + s/0.1)}{(1 + s)^2} = \frac{0.2(s + 0.1)}{(s + 1)^2}$$

12.3.2 Bode Phase Plots

The Bode phase plots of various terms are illustrated in Figure 12.3.

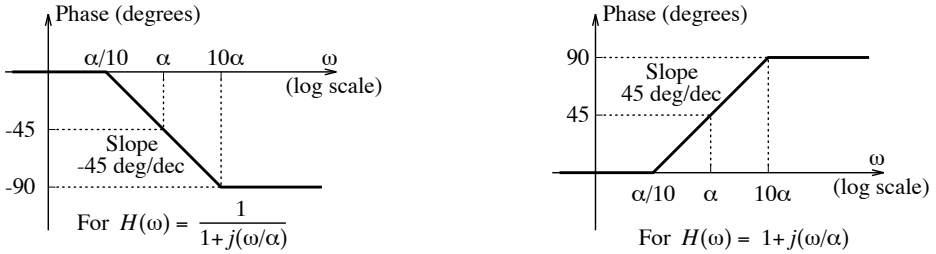


Figure 12.3 Asymptotic Bode phase plots of various terms

The phase of $j\omega/\alpha$ is 90° for all ω. The phase of $(1 + j\omega/\alpha)$ is $\phi(\omega) = \tan^{-1}(\omega/\alpha)$. At $\omega = \alpha$, we have the *exact* result $\phi(\omega) = 45°$. For low and high frequencies, we find that $\phi(\omega) \approx 0$, $\omega \ll \alpha$ and $\phi(\omega) \approx 90°$, $\omega \gg \alpha$. To generate an asymptotic phase plot, we also need a linear approximation in the intermediate frequency range. The break frequency $\omega = \alpha$ provides a convenient reference. Since the difference between the true and asymptotic magnitude almost disappears within a decade of the break frequency, it is customary to start the high-frequency and low-frequency approximations a decade above and a decade below $\omega = \alpha$. Over the two-decade range $(0.1\alpha,\ 10\alpha)$, we approximate the phase by a straight line with a slope of 45°/dec (which corresponds to 13.5°/oct). As with Bode magnitudes, all the slopes and phase values change sign if a term is in the denominator and increase k-fold if a term is repeated k times. The constant term contributes zero phase if positive and ±180° if negative.

The Bode phase plot for a transfer function with several terms is just the sum of similar plots for each of its individual terms. Since asymptotic plots are linear, the composite plot can be sketched directly without sketching the individual terms, if we note the following:

1. A negative constant contributes 180°, while the term $(j\omega)^{\pm k}$ also contributes a constant value $\pm 90k°$. Both may be used as the last step to shift the phase plot by an appropriate amount.

2. If the break frequencies are arranged in ascending order, the slope changes at frequencies one decade on either side of each successive break frequency. The slope increases by $\pm 45k°$/dec for a factor of the form $(1 + j\omega/\omega_B)^{\pm k}$.

PGR2026-00033
Krisp Technologies EX1024 Page 389

Here are some consistency checks. The initial and final asymptotes have zero slope. The initial phase is nonzero only if a term of the form $(j\omega)^{\pm k}$ is present. The final phase equals $90n^\circ$, where n is the difference between the order of the numerator and denominator polynomials in $H(\omega)$.

REVIEW PANEL 12.8

Asymptotic Phase Plot of Linear and Repeated Linear Terms

Term $= j\omega$: Phase $= 90^\circ$ for *all* ω.
Term $= 1 + j\frac{\omega}{\alpha}$: 0° for $\omega \leq 0.1\alpha$, 90° for $\omega \geq 10\alpha$, linear in between (slope $= 45^\circ$/dec $= 13.5^\circ$/oct).
If repeated k **times:** Multiply phase by k. **If in denominator:** Negative phase and slope.

EXAMPLE 12.5 (Bode Phase Plots)

Sketch the Bode phase for $H(\omega) = \dfrac{(0.5 + j\omega)(10 + j\omega)}{j\omega(20 + j\omega)^2}$.

First, check the initial and final phase: $\phi_i = -90^\circ$ and $\phi_f = -90^\circ$.
Next, arrange break frequencies in increasing order and list their multiplicity:

$\omega_1 = 0.5$ rad/s (numerator) $\omega_2 = 10$ rad/s (numerator) $\omega_3 = 20$ rad/s (denominator, multiplicity $k = 2$)

Next, we find the slopes of the asymptotes for the individual terms.

1. $\omega_1 = 0.5$ rad/s (numerator): $+45^\circ$/dec between 0.05 rad/s and 5 rad/s

2. $\omega_2 = 10$ rad/s (numerator): $+45^\circ$/dec between 1 rad/s and 100 rad/s

3. $\omega_3 = 20$ rad/s (denominator, $k = 2$): -90°/dec between 2 rad/s and 200 rad/s

The frequencies at which the slopes change are [0.05, 1, 2, 5, 100, 200] rad/s.
The slopes for the composite plot are found as

1. 0°/dec before 0.05 rad/s and past 200 rad/s.

2. 45°/dec between 0.05 rad/s and 1 rad/s (due to the first term only).

3. 90°/dec between 1 rad/s and 2 rad/s (due to the first two terms).

4. 0°/dec between 2 rad/s and 5 rad/s (due to all three terms).

5. -45°/dec between 5 rad/s and 100 rad/s (due to the last two terms).

6. -90°/dec between 100 rad/s and 200 rad/s (due to the last term only).

We sketch the asymptotic plot and shift it by -90° due to the term $1/j\omega$. As a check, both the initial and final values of the phase equal -90°. The phase plot is shown in Figure E12.5.

PGR2026-00033
Krisp Technologies EX1024 Page 390

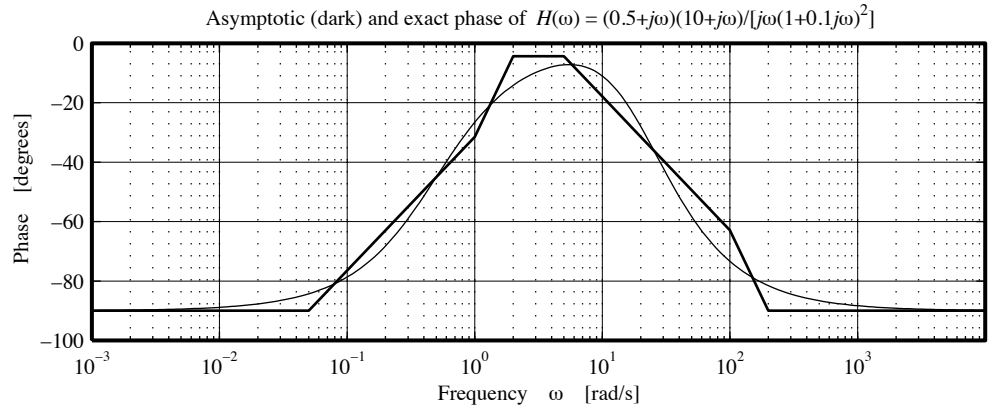


Figure E12.5 Bode phase plot for Example 12.5

12.3.3 The Quadratic Term

A quadratic term may be written in the form

$$H(\omega) = 1 + j2\zeta(\omega/\alpha) + (j\omega/\alpha)^2 = 1 + j\,\tfrac{1}{Q}(\omega/\alpha) + (j\omega/\alpha)^2 \tag{12.8}$$

We consider this term only for $\zeta < 1$ (or $Q > 0.5$), which gives complex roots (otherwise it is the product of linear or repeated factors). The quantity ζ is called the **damping factor**, and $Q = 1/2\zeta$ is called the **quality factor**. For low and high frequencies, it behaves as a squared linear factor with

$$H_{\text{dB}} \approx \begin{cases} 0 \text{ (slope = 0 dB/dec)}, & \omega \ll \alpha \\ 40 \log \left|\dfrac{\omega}{\alpha}\right| \text{ (slope = 40 dB/dec)}, & \omega \gg \alpha \end{cases} \tag{12.9}$$

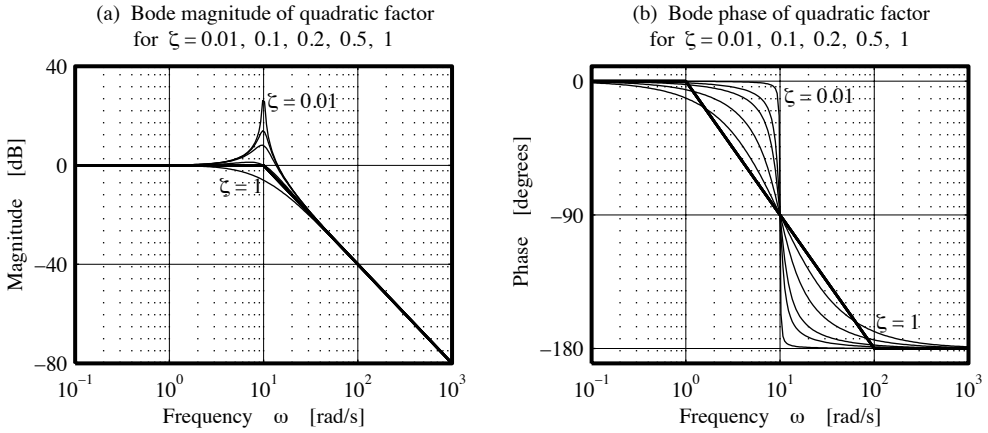


Figure 12.4 Magnitude and phase of a quadratic numerator at the break frequency

Figure 12.4 shows the magnitude and phase plots for various values of ζ. At the break frequency $\omega = \alpha$, the true magnitude may differ significantly from the asymptotic, and depends on ζ (or Q) with

$$H_{\text{dB}} = 20 \log(2\zeta) = -20 \log Q \tag{12.10}$$

PGR2026-00033
Krisp Technologies EX1024 Page 391

For $\zeta = 0.5$ (or $Q = 1$), the true value matches the asymptotic value but for $\zeta < 0.5$ (or $Q > 1$), the true value exceeds the asymptotic value.

The phase plot reveals even greater differences. For $\zeta = 1$ (or $Q = 0.5$), the asymptotic phase plot is identical to that for a squared linear factor. For $\zeta < 1$ (or $Q > 0.5$), the transition between 0° and 180° occurs in less than two decades. For $\zeta < 0.1$ (or $Q > 5$), the phase change occurs in an almost step-like fashion, and the asymptotes extend almost up to the break frequency itself.

EXAMPLE 12.6 (Composite Plot for Quadratic Terms)

Sketch the Bode plots for the transfer function $H(\omega) = \dfrac{100(1 + j2\omega)}{-\omega^2 + 2j\omega + 100}$.

With $-\omega^2 = (j\omega)^2$, we write this in the standard form

$$H(\omega) = \frac{(1 + j\omega/0.5)}{1 + j2(0.1)(\omega/10) + (j\omega/10)^2}$$

The Bode magnitude and phase plots are sketched in Figure E12.6. The Bode magnitude is 0 dB up to $\omega = 0.5$ rad/s with a 20-dB/dec asymptote starting at $\omega = 0.5$ rad/s and another of -20 dB/dec starting at $\omega = 10$ rad/s. The asymptotic magnitude at $\omega = 10$ rad/s is 14 dB. The true magnitude differs significantly. With $\zeta = 0.1$, we compute $-20\log(2\zeta) = -20\log(0.2) = 14$ dB. Thus, the true value at $\omega = 10$ rad/s differs from the asymptotic value by 14 dB and equals 28 dB. The phase plot shows zero phase until $\omega = 0.05$ rad/s, followed by asymptotes with slopes of 45°/dec until $\omega = 1$ rad/s, -45°/dec until $\omega = 5$ rad/s, and -90°/dec until $\omega = 100$ rad/s. After $\omega = 100$ rad/s, the phase stays constant at -90°. Note how different the asymptotic phase plot is compared with the exact phase plot, especially in the mid-frequency range.

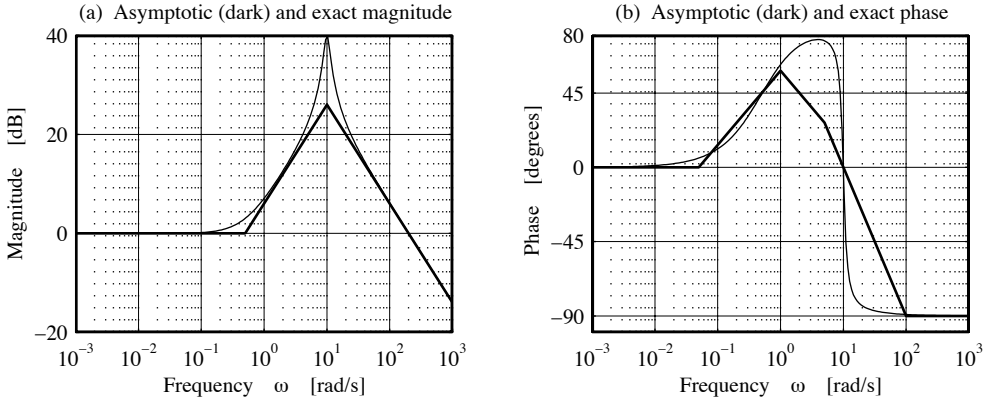


Figure E12.6 Bode magnitude and phase plot for Example 12.6

12.4 Performance Measures

The performance of a filter is measured at two levels. Its steady-state performance is based on the response to sinusoids and measured in the frequency domain by its steady-state transfer function or frequency response $H(\omega)$. For lowpass filters, transient performance (speed of response) is measured in the time domain using the step response. Typical time measures based on the step response are summarized in Table 12.1. A sharper frequency response $|H(\omega)|$ can be achieved only at the expense of a slower time response.

PGR2026-00033
Krisp Technologies EX1024 Page 392

Table 12.1 Time-Domain Performance Measures for Real Filters

Measure	Explanation
Time delay	Time between application of input and appearance of response Typical measure: Time to reach 50% of final value
Rise time	Measure of the steepness of initial slope of response Typical measure: Time to rise from 10% to 90% of final value
Overshoot	Deviation (if any) beyond the final value Typical measure: Peak overshoot
Settling time	Time for oscillations to settle to within a specified value Typical measure: Time to settle to within 5% or 2% of final value
Damping	Rate of change toward final value Typical measure: Damping factor ζ or quality factor Q

12.4.1 Performance of Second-Order Lowpass Filters

A second-order lowpass filter with unit dc gain may be described by

$$H(s) = \frac{\omega_p^2}{s^2 + \frac{\omega_p}{Q}s + \omega_p^2} \tag{12.11}$$

Here, ω_p is called the **undamped natural frequency**, or **pole frequency**, and Q represents the **quality factor** and is a measure of the losses in the circuit. The quantity $\zeta = 1/2Q$ is called the **damping factor**. For second-order filters, Q and ω_p control the performance in both the time domain and the frequency domain. The two poles of the second-order filter are

$$p_{1,2} = -\sigma_k \pm j\omega_k = -\frac{\omega_p}{2Q} \pm j\omega_p\sqrt{1 - (1/4Q^2)} \tag{12.12}$$

For $Q > 0.5$, the poles are complex conjugates, and lie on a circle of radius ω_p in the s-plane, as shown in Figure 12.5. Their angular orientation θ_k is related to Q. The larger the Q, the closer the poles are to the $j\omega$-axis, the less stable is the system, and the more sensitive it is to component variations.

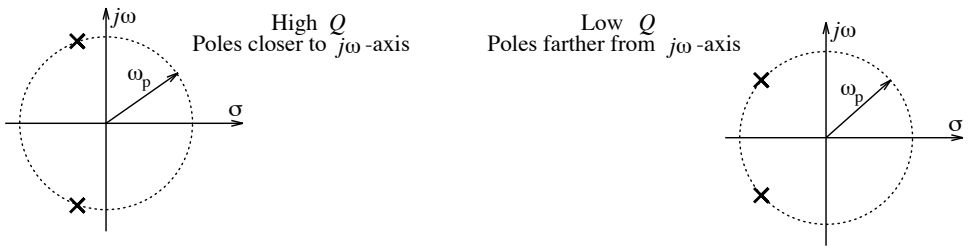


Figure 12.5 Pole locations of a second-order filter

Frequency-Domain Measures

For $Q < 1/\sqrt{2}$, the magnitude $|H(\omega)|$ is monotonic, but for $Q > 1/\sqrt{2}$ it shows overshoot and a peak near ω_p, and the peaking increases with Q, as illustrated in Figure 12.6.

PGR2026-00033
Krisp Technologies EX1024 Page 393

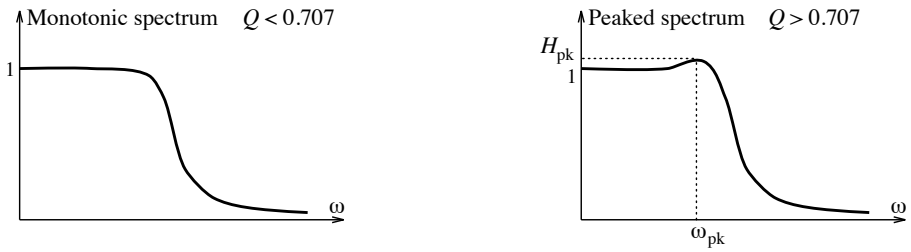


Figure 12.6 Frequency response of a second-order system

For $Q > 1/\sqrt{2}$, the frequency ω_{pk} of the peak and the peak magnitude H_{pk} are found by setting $d|H(\omega)|/d\omega$ to zero to give

$$\omega_{\text{pk}} = \omega_p[1 - (1/2Q^2)]^{1/2} \qquad H_{\text{pk}} = \frac{Q}{\omega_p[1 - (1/4Q^2)]^{1/2}}, \qquad Q > 1/\sqrt{2} \tag{12.13}$$

For $Q \gg 1, \omega_{\text{pk}}$ and H_{pk} can be approximated by

$$\omega_{\text{pk}} \approx \omega_p, \quad Q \gg 1 \qquad H_{\text{pk}} = Q/\omega_p, \quad Q \gg 1 \tag{12.14}$$

The half-power frequency where $|H(\omega)| = H_{\text{pk}}/\sqrt{2}$ must usually be computed numerically. The phase response and group delay of second-order filters are given by

$$\theta(\omega) = -\tan^{-1}\left[\frac{(\omega/\omega_p)/Q}{1 - (\omega/\omega_p)^2}\right] \qquad t_g(\omega) = \frac{2(\omega_p/Q)(\omega_p^2 + \omega^2)}{(\omega_p^2 - \omega^2)^2 + (\omega_p^2/Q^2)\omega^2} \tag{12.15}$$

Time-Domain Measures

The nature of the step response depends on the poles of $H(s)$, as illustrated in Figure 12.7.

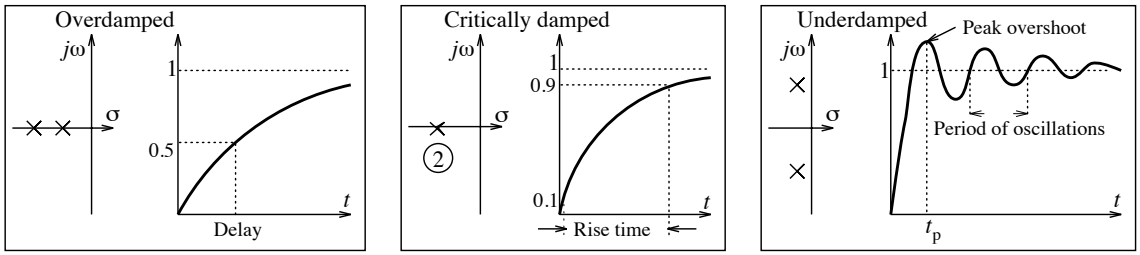


Figure 12.7 Step response of a second-order system

For $Q < 0.5$, the poles are real and distinct, and the step response shows a smooth, monotonic rise to the final value (**overdamped**) with a large time constant. For $Q > 0.5$, the poles are complex conjugates, and the step response is **underdamped** with overshoot and decaying oscillations (ringing) about the final value. This results in a small rise time but a large settling time. The frequency of oscillations increases with Q. For $Q = 0.5$, the poles are real and equal, and the response is **critically damped** and yields the fastest monotonic approach to the steady-state value with no overshoot. The 5% settling time and 2% settling time

PGR2026-00033
Krisp Technologies EX1024 Page 394

are reasonably well approximated by

$$t_{5\%} \approx \frac{6Q}{\omega_p} \qquad t_{2\%} \approx \frac{8Q}{\omega_p}, \quad Q > 0.6 \tag{12.16}$$

An exact expression for the peak overshoot P_{OS} and its time of occurrence is given by

$$P_{\text{OS}} = e^{-\pi/\sqrt{4Q^2-1}} \qquad t_{\text{OS}} = \frac{2\pi Q}{\omega_p \sqrt{4Q_2 - 1}} \tag{12.17}$$

It is difficult to obtain explicit expressions for the delay time and rise time, which must be computed numerically. For $Q > 0.6$, the rise time decreases almost inversely with ω_p.

Some General Observations

In general, for a monotonic $|H(\omega)|$ as well as a small rise time, we must tolerate some overshoot in the step response. The overshoot depends largely on how sharp the transition between the passband and stopband is. A sharper transition requires a higher filter order, resulting in a larger group delay and more nonlinearity in the phase. A flatter delay requires a more gradual transition band in order to achieve a smoother step response that is largely free of overshoot. These observations are typical of the compromises that must be made when designing practical filters.

EXAMPLE 12.7 (Performance Measures of Real Filters)
We analyze the two filters shown in Figure E12.7.



Figure E12.7 The filters for Example 12.7

(a) The second-order Bessel filter yields $H(s)$ and $H(\omega)$ as

$$H(s) = \frac{3}{s^2 + 3s + 3} \qquad H(\omega) = \frac{3}{3 - \omega^2 + j3\omega}$$

Comparison of the denominator with $s^2 + s\omega_p/Q + \omega_p^2$ gives $\omega_p = \sqrt{3}$ rad/s and $Q = 1/\sqrt{3} \approx 0.577$. Since $Q < 1/\sqrt{2}$, we should expect a monotonic frequency response. We find the half-power frequency by setting $|H(\omega)| = \sqrt{\frac{1}{2}}$ or $|H(\omega)|^2 = 0.5$ to get $\omega^4 + 3\omega^2 - 9 = 0$. Solving for ω^2 and taking its square root gives the half-power frequency $\omega = 1.3617$ rad/s.

Since $Q > 0.5$ we should expect an underdamped step response. The peak overshoot is

$$P_{\text{OS}} = \exp(-\pi/\sqrt{4Q^2 - 1}) = e^{-\pi/\sqrt{3}} = 0.0043$$

The peak overshoot occurs at $t_{\text{OS}} = 3.63$ s. The approximate 5% and 2% settling times and their exact values (in parentheses) are

$$t_{5\%} \approx \frac{6Q}{\omega_p} = 2 \ (2.19) \text{ s} \qquad t_{2\%} \approx \frac{8Q}{\omega_p} = 2.66 \ (2.5) \text{ s}$$

PGR2026-00033
Krisp Technologies EX1024 Page 395

The rise time is numerically estimated as $t_r = 1.573$ s, and the delay time as $t_d = 0.9$ s.

The phase response $\theta(\omega)$ and group delay t_g are given by

$$\theta(\omega) = -\tan^{-1}\left[\frac{3\omega}{3-\omega^2}\right] \qquad t_g = -\frac{d\,\theta(\omega)}{d\omega} = \frac{3\omega^2+9}{\omega^4+3\omega^2+9}$$

The group delay is almost linear in the passband.

(b) The second-order lowpass Butterworth filter is described by $H(s) = \dfrac{1}{s^2 + s\sqrt{2} + 1}$.

Comparison of the denominator with $s^2 + s\omega_p/Q + \omega_p^2$ gives $\omega_p = 1$ rad/s and $Q = 1/\sqrt{2}$ (the highest value of Q that still ensures a monotonic $|H(\omega)|$). The half-power frequency is $\omega = 1$ rad/s.

Since $Q > 0.5$, the step response will be underdamped and show overshoot. In fact, the step response $s(t)$ corresponds to the inverse transform of

$$S(s) = \frac{H(s)}{s} = \frac{1}{s(s^2 + s\sqrt{2} + 1)} = \frac{1}{s} + \frac{-0.5 - j0.5}{s + 1/\sqrt{2} - j/\sqrt{2}} + \frac{-0.5 + j0.5}{s + 1/\sqrt{2} + j/\sqrt{2}}$$

This gives

$$s(t) = [1 + \sqrt{2}e^{-t/\sqrt{2}}\cos(t/\sqrt{2} + 135^\circ)]u(t)$$

The peak overshoot is 0.0432 and occurs at 4.44 s. Compared to the Bessel filter, this overshoot is larger, but it is offset by the flatter magnitude response of the Butterworth filter. The settling time approximations and true values (in parentheses) are

$$t_{5\%} \approx \frac{6Q}{\omega_p} = 4.24\ (2.93)\ \text{s} \qquad t_{2\%} \approx \frac{8Q}{\omega_p} = 5.66\ (5.97)\ \text{s}$$

The rise time is numerically estimated as $t_r = 2.15$ s and the delay time as $t_d = 1.43$ s. The settling time, rise time, and delay time are much larger than those for the Bessel filter.

12.5 Feedback

Feedback plays an extremely important role in regulating the performance of systems. In essence, we compare the output of a system with the desired input and correct for any deficiencies to make improvements. Feedback systems abound in nature. Temperature regulation in humans is an example. When we are cold, the body responds by shivering to increase the temperature. The concept of feedback has been effectively used to both study and design systems in a wide variety of disciplines ranging from biomedical engineering to space exploration and has spawned the fast growing discipline of automatic control systems. A typical feedback system involves the structure shown in Figure 12.8.

The system is typically analyzed under normal operating conditions or about a **set point** that describes a steady state. Each variable then describes its deviation from its corresponding steady-state value. The **plant** describes the system to be controlled. The input $R(s)$ is a **reference signal**. The output $C(s)$ is fed back to generate an **error signal** $E(s)$, which controls the output in a manner that can be adjusted by **compensators** in either the forward or return path or both. Any disturbances (unwanted signals) that cause a change in the steady state $C(s)$ are modeled as a separate signal $D(s)$.

PGR2026-00033
Krisp Technologies EX1024 Page 396

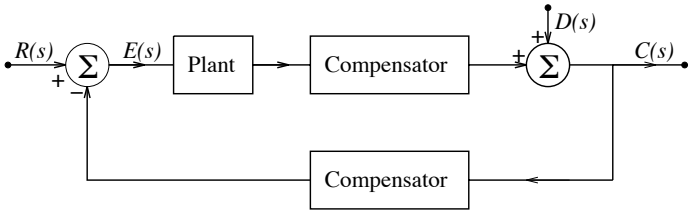


Figure 12.8 A typical feedback system

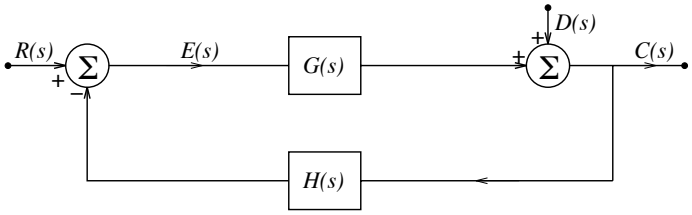


Figure 12.9 A modified feedback system

If we combine the forward compensator and the plant, we obtain the feedback system of Figure 12.9. Depending on the requirements, such a feedback system can be operated for **tracking** or for **regulation**. In the tracking mode, the system tracks (or responds to) the reference signal such that $C(s) \approx R(s)$. This requires minimization of the error signal $E(s)$. During regulation, the system minimizes changes in $C(s)$ (in the presence of a disturbance), assuming a constant reference. Regulation is important in situations where we must maintain the response in the face of changing system parameters. How feedback affects system stability is also an important aspect.

12.5.1 Analysis of a Feedback System

The response $C(s)$ of the feedback system of Figure 12.9 may be written as

$$C(s) = D(s) + G(s)[R(s) - C(s)H(s)] \qquad C(s) = R(s)\frac{G(s)}{1 + G(s)H(s)} + D(s)\frac{1}{1 + G(s)H(s)} \tag{12.18}$$

In the absence of a disturbance, we obtain

$$T(s) = \frac{C(s)}{R(s)} = \frac{G(s)}{1 + G(s)H(s)} \tag{12.19}$$

This describes the **closed-loop transfer function** of the feedback system. Note that feedback reduces the overall gain. If a disturbance is present but there is no feedback ($H(s) = 0$), we have

$$C(s) = R(s)G(s) + D(s) \tag{12.20}$$

Tracking
Since good tracking requires $C(s) \approx R(s)$, this implies

$$\frac{G(s)}{1 + G(s)H(s)} \approx 1 \tag{12.21}$$

To ensure this we could, for example, choose $H(s) = 1$ (unity feedback) and $G(s) \gg 1$.

PGR2026-00033
Krisp Technologies EX1024 Page 397

Regulation
The regulating effect of feedback is based on choosing a large gain for the term $G(s)H(s)$, such that $|G(s)H(s)| \gg 1$. Then

$$T(s) = \frac{G(s)}{1 + G(s)H(s)} \approx \frac{1}{H(s)} \tag{12.22}$$

The system transfer function depends only on the feedback path. Since it is also independent of $G(s)$, the system is insensitive to variations in the plant dynamics.

The Open-Loop Gain
Feedback is hardly useful when the open-loop gain is small ($|GH| \ll 1$), since $T(s) \approx G(s)$. Clearly, the influence of feedback (closing the loop) depends on the term $G(s)H(s)$, regardless of whether we want to study tracking, regulation, or stability. Curiously, $G(s)H(s)$ (with its sign reversed) is called the **open-loop transfer function**. If we consider only the parts in the feedback loop, as shown in Figure 12.10, and open the loop at any point, an input $I(s)$ inserted at one end returns as $-G(s)H(s)I(s)$ at the other end (to act as the new input at its starting point). Thus, the open-loop transfer function equals

$$\text{OLT} = \frac{O(s)}{I(s)} = -G(s)H(s) \tag{12.23}$$

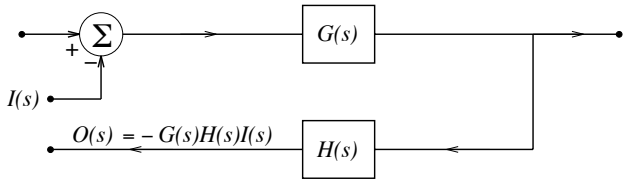


Figure 12.10 The concept of open-loop gain

Advantages of Feedback
Even though feedback reduces the overall gain, it can lead to a stable overall system, regardless of the complexity, variations, or even lack of stability of the plant itself. Since $H(s)$ can usually be realized using inexpensive, highly reliable devices and elements, feedback affords a simple, inexpensive means of regulating stable systems or stabilizing unstable ones. The major advantages of feedback include:

1. Insensitivity to plant variations
2. Reduced effects of distortion
3. Improvement in the system bandwidth
4. Improvement in the system linearity and stability

For a large open-loop gain, the overall transfer function of a feedback system is $T(s) \approx 1/H(s)$, and its stability (and linearity) is dictated primarily by $H(s)$. The plant $G(s)$ itself could be unstable or highly nonlinear, but if $H(s)$ is stable (or linear), so is the overall system.

EXAMPLE 12.8 (Advantages of Feedback)

(a) (Insensitivity to Plant Variations) Consider an amplifier with gain $G = 10$. To minimize the variation in its gain due to factors such as aging, power supply fluctuations, and temperature changes, we can put this amplifier in the forward path of a feedback system that also includes two compensating blocks A_1 and A_2, as illustrated in Figure E12.8A.

PGR2026-00033
Krisp Technologies EX1024 Page 398

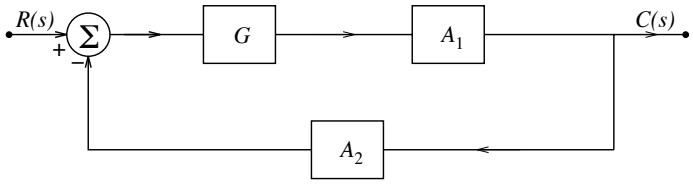


Figure E12.8A The feedback system for Example 12.8(a)

The overall gain T of the feedback system is

$$T = \frac{GA_1}{1 + GA_1A_2}$$

With $A_1 = 100$ and $A_2 = 0.099$, we see that $T \approx 10$. Even if the amplifier gain G reduces by half to $G = 5$, the overall system gain is $T = 9.9$, a reduction of only 1%! Other choices for A_1 and A_2 that ensure $T \approx 10$ but increase the open-loop gain GA_1A_2 will yield even more impressive results.

(b) (Disturbance Rejection) Amplifiers are prone to distortion at full-power output. If the distortion is small, they may be modeled by the linear system of Figure E12.8B. The structure is similar to that of part (a) but the output $C(s)$ now equals the sum of an undistorted output and a term $D(s)$ that accounts for distortion.

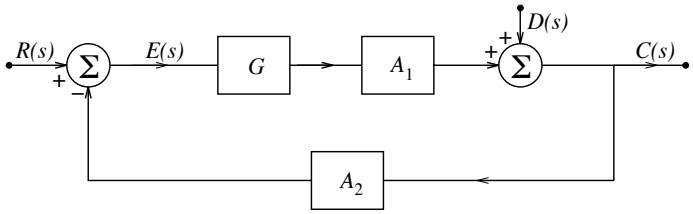


Figure E12.8B The feedback system for Example 12.8(b)

We obtain the output $C(s)$ as

$$C(s) = \frac{GA_1}{1 + GA_1A_2}R(s) + \frac{1}{1 + GA_1A_2}D(s)$$

If $A_1 = 1 + GA_1A_2$, the input $R(s)$ is amplified by G, but the distortion $D(s)$ is reduced by the factor $(1 + GA_1A_2)$. To minimize the distortion, we require a large open-loop gain (i.e., $GA_1A_2 \gg 1$).

(c) (Noise Reduction) Feedback is effective in combating noise only if it appears in the output, not if it adds to the input. The error signal $E(s)$ is generated by comparing the input with the measured output. Measurement error results in a noisy signal $C'(s) = C(s) + N(s)$ and an error that equals $E(s) = R(s) - C'(s) = R(s) - N(s) - C(s)$. This is equivalent to a system with the noisy input $R'(s) = R(s) - N(s)$. In the absence of any other disturbances (if $D(s) = 0$), we get

$$C(s) = \frac{GA_1}{1 + GA_1A_2}R'(s) = \frac{GA_1}{1 + GA_1A_2}R(s) - \frac{GA_1}{1 + GA_1A_2}N(s)$$

It should be obvious that both the input $R(s)$ and the noise $N(s)$ are multiplied by the same factor, and no noise reduction is thus possible! The noise must be minimized by other means. For example, the effect of measurement noise may be minimized by using more accurate sensors.

PGR2026-00033
Krisp Technologies EX1024 Page 399

(d) (Bandwidth Improvement) Consider an amplifier with transfer function $G(s) = K/(1 + s\tau)$. Its half-power bandwidth is $\omega_B = 1/\tau$. If this amplifier is made part of a feedback system as illustrated in Figure E12.8D, the overall transfer function may be written as

$$T(s) = \frac{G(s)}{1 + AG(s)} = \frac{K_1}{1 + s\tau_1}, \quad \text{where} \quad K_1 = \frac{K}{1 + KA} \quad \text{and} \quad \tau_1 = \frac{\tau}{1 + KA}$$

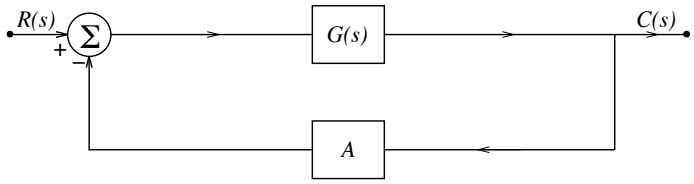


Figure E12.8D The feedback system for Example 12.8(d)

For a large loop gain ($KA \gg 1$), we get $\tau_1 \approx \frac{\tau}{KA}$ and $K_1 \approx \frac{K}{KA}$. The bandwidth ($1/\tau_1$) of the overall system is KA times the amplifier bandwidth ($1/\tau$). However, this increase in bandwidth is made possible only at the expense of a reduction in the overall gain (by the same factor KA).

12.5.2 Stability of Feedback Systems

If we design a stable open-loop system, it will remain stable under any conditions. A feedback system, on the other hand, may not always be stable. Stability is an important factor in the design of feedback systems. One way to study stability is through the closed-loop transfer function $T(s)$. To ensure a stable system, the poles of $T(s)$ must lie in the left half of the s-plane. In other words, the roots of $1 + G(s)H(s) = 0$ must have negative real parts. A large loop gain implies that $T(s) \approx 1/H(s)$, and the poles of the overall system are very nearly those of $H(s)$ itself.

For an intuitive way to see why feedback systems may suffer from instability, we return to Figure 12.10 and consider its behavior in the frequency domain. Since the return signal is 180° out of phase, it opposes the insert signal as often as it reinforces it (in alternate round-trips). The system can thus stay stable even if the open-loop gain is very large. However, if the phase of $G(f)H(f)$ equals 180° at some frequency f_C, the return signal is *in phase* with the insert signal and always reinforces it. And if the loop gain exceeds unity ($|G(f_C)H(f_C)| > 1$), the output continues to increase, even in the absence of an applied input, leading to system instability. The apparent solution is to avoid operating at the frequency f_C. In practice, however, any random noise with even a small component at f_C will in time become large enough to produce instability. The real solution is to ensure $|G(f)H(f)| < 1$ at the frequencies where its phase equals −180°. This leads to the definitions of the gain margin and phase margin of feedback systems as follows:

Gain margin G_M: The gain at the frequency where the phase equals 180°
Phase margin ϕ_M: The *phase deviation* from 180° when the gain is unity

The larger the phase margin, the better off we are in terms of system stability.

12.5.3 Feedback and Inverse Systems

Feedback is an effective way of designing inverse systems. As shown in Figure 12.11, for a large loop gain ($|G| \gg 1$) the transfer function $T(s)$ of a feedback system is approximately equal to $1/H(s)$, where $H(s)$ is the block in the feedback path.

$$T(s) = \frac{C(s)}{R(s)} = \frac{G}{1 + GH(s)} \approx \frac{1}{H(s)}, \quad |G| \gg 1 \tag{12.24}$$

PGR2026-00033
Krisp Technologies EX1024 Page 400

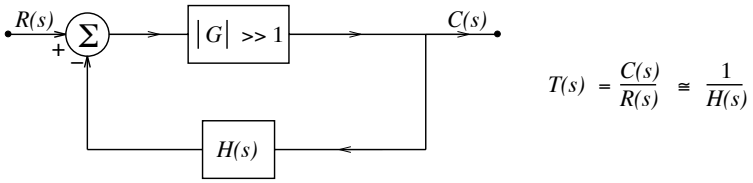


Figure 12.11 Realizing an inverse system

It is thus possible to realize the inverse system corresponding to some $H(s)$ simply by designing a feedback system with $H(s)$ in the feedback path, and a large open-loop gain. And as an added advantage, the exact details of $H(s)$ need not even be known! This is a useful way to undo the damaging influence of some system $H(s)$, such as a sensor or measurement device, on a signal. We feed the signal to the device and feed its output to a feedback system with the same device in the feedback path. The overall transfer function of the cascaded system is unity, and we recover the input signal!

12.6 Application of Feedback: The Phase-Locked Loop

The **phase-locked loop** (PLL) is a versatile phase-tracking feedback system used in the detection of angle-modulated and DSBSC AM signals. A block diagram for the detection of FM signals is shown in Figure 12.12. The PLL tracks the phase variation of the input signal. The PLL output depends on the rate of change of the input phase.

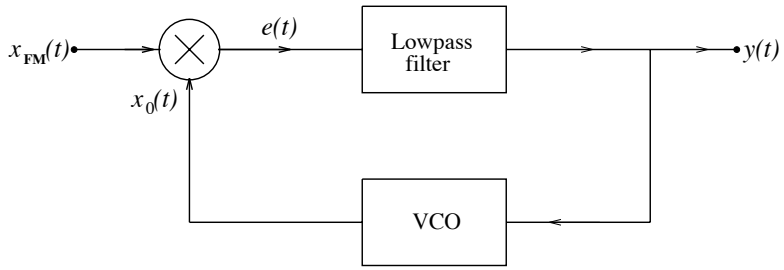


Figure 12.12 Detection of FM signals using the phase-locked loop

12.6.1 PLL Operation

The input to the PLL is the FM signal

$$x_{\mathrm{FM}}(t) = A\cos[2\pi f_C t + \theta_i(t)] \tag{12.25}$$

Here, the phase $\theta_i(t)$ is related to the message signal $m(t)$ by

$$\theta_i(t) = 2\pi k_F \int_0^t m(t)\,dt \qquad\qquad m(t) = \frac{1}{2\pi k_F}\,\frac{d\theta_i}{dt} \tag{12.26}$$

The heart of the PLL is a **voltage-controlled oscillator** (VCO) or *frequency modulator* whose frequency is proportional to the input signal. The output of the VCO is a sinusoid $x_0(t)$ of the form

$$x_0(t) = B\sin[2\pi f_C t + \theta_0(t)] \tag{12.27}$$

PGR2026-00033
Krisp Technologies EX1024 Page 401

Under ideal conditions, the rate of change of $\theta_0(t)$ is also proportional to the PLL output $y(t)$ such that

$$\theta_0(t) = 2\pi k_0 \int_0^t y(t)\,dt \qquad\qquad y(t) = \frac{1}{2\pi k_F}\,\frac{d\theta_i}{dt} \tag{12.28}$$

If $\theta_0(t) \approx \theta_i(t)$, it follows that $y(t) \approx m(t)(k_F/k_0)$ describes a scaled version of the demodulated message. In other words, for demodulation, the PLL behaves as an ideal differentiator.

12.6.2 Analysis of the PLL

The phase detector constitutes a multiplier and an ideal filter that eliminates frequencies outside the band $f \le |2f_C|$. The error signal $e(t)$ at the output of the multiplier is simply

$$e(t) = x_0(t)x_M(t) = AB\sin(2\pi f_C t + \theta_0)\cos(2\pi f_C t + \theta_i) = 0.5AB[\sin(4\pi f_C t + \theta_0 + \theta_i) + \sin(\theta_0 - \theta_i)]$$

If we assume that both $\theta_i(t)$ and $\theta_0(t)$ vary slowly compared to $2\pi f_C t$, and the filter rejects frequencies outside the range $f \le 2f_C$, the filter output may be written as

$$y(t) = 0.5AB\sin(\theta_0 - \theta_i) = -0.5AB\sin(\theta_i - \theta_0) \tag{12.29}$$

Since $d\theta_0/dt = 2\pi k_0 y(t)$, we obtain the *nonlinear* differential equation

$$\frac{d\theta_0}{dt} = -0.5AB(2\pi k_0)\sin(\theta_i - \theta_0) = K\sin(\theta_i - \theta_0) \tag{12.30}$$

Here, $K = -0.5AB(2\pi k_0)$. The system is said to be **synchronized** or in **phase-lock** if $\theta_0 \approx \theta_i$. When this happens, we obtain the *linear approximation*

$$\frac{d\theta_0}{dt} \approx K[\theta_i(t) - \theta_0(t)] \tag{12.31}$$

This linear model is amenable to analysis using Laplace transforms. The Laplace transform of both sides gives

$$s\Theta_0(s) = K[\Theta_i(s) - \Theta_0(s)] \qquad\qquad \Theta_0(s) = \frac{K}{s+K}\Theta_i(s) \tag{12.32}$$

Since $d\theta_0/dt = 2\pi k_0 y(t)$ and $d\theta_i/dt = 2\pi k_F m(t)$, we have $Y(s) = s\Theta_0(s)/(2\pi k_0)$ and $\Theta_i(s) = 2\pi k_F M(s)/s$. We then obtain the following:

$$Y(s) = \frac{s\Theta_0(s)}{2\pi k_0} = \frac{Ks}{2\pi k_0(s+K)}\Theta_i(s) \qquad\qquad Y(s) = \frac{(k_F/k_0)K}{s+K}M(s) \tag{12.33}$$

For demodulation, we would like $y(t) \approx m(t)$ or $Y(s) \approx M(s)$. If $K \gg 1$ (implying a very large loop gain), we obtain $Y(s) \approx M(s)(k_F/k_0)$ and do in fact recover the message signal (scaled by the factor k_F/k_0, of course). It is important to realize that this result holds only when the PLL has locked on to the phase of the FM signal and when the phase error $\theta_i - \theta_0$ is small. If the phase error is not restricted to small values, the output will suffer from distortion. Naturally, during the locking process, the phase error is much too large to warrant the use of a linear model.

12.6.3 A Feedback Model of the PLL

Under phase lock, a general feedback model of the PLL can be developed, as shown in Figure 12.13. The objective is to track the input phase $\theta_i(t)$. The feedback block K/s describes the action of the VCO and

PGR2026-00033
Krisp Technologies EX1024 Page 402

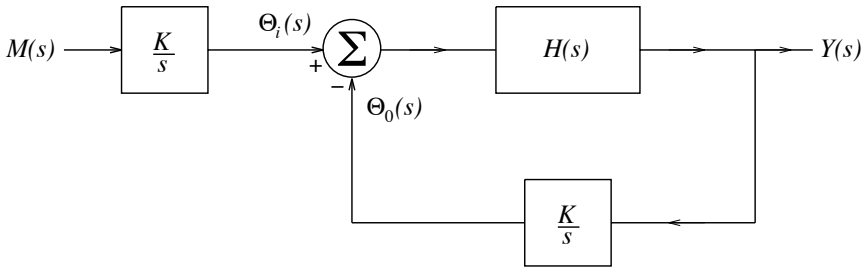


Figure 12.13 A feedback model of the PLL

produces a phase that varies with $y(t)$. Under linear operation, the output of the detector (multiplier and lowpass filter) is approximated by $\theta_i - \theta_0$, and its action is thus modeled by a summing junction.

The various transfer functions are readily obtained as

$$T(s) = \frac{Y(s)}{M(s)} = \frac{KH(s)}{s + KH(s)} \qquad \frac{\Theta_0(s)}{\Theta_i(s)} = \frac{KH(s)}{s + KH(s)} \qquad \frac{Y(s)}{\Theta_i(s)} = \frac{sH(s)}{s + KH(s)} \tag{12.34}$$

The phase-error transfer function, defined as $[\Theta_i(s) - \Theta_0(s)]/\Theta_i(s)$, yields

$$\frac{\Theta_i(s) - \Theta_0(s)}{\Theta_i(s)} = 1 - \frac{\Theta_0(s)}{\Theta_i(s)} = \frac{s}{s + KH(s)} \tag{12.35}$$

For a large open-loop gain $|KH(s)| \gg 1$, we have $T(s) = Y(s)/M(s) \approx 1$, and the output $y(t)$ equals the message signal $m(t)$ as required. If the input phase does not vary, the error signal $\theta_i(s) - \theta_0(s)$ equals zero and so does the steady-state response.

12.6.4 Implementation

The block $H(s)$ may be realized in passive or active configurations. If $H(s) = 1$, we obtain a **first-order loop** with $T(s) = K/(s + K)$. Its frequency response and phase error may be written as

$$\frac{Y(f)}{M(f)} = \frac{\Theta_0(f)}{\Theta_i(f)} = \frac{K}{K + j2\pi f} \qquad \frac{\Theta_i(f) - \Theta_0(f)}{\Theta_i(f)} = \frac{j2\pi f}{K + j2\pi f} \tag{12.36}$$

The loop gain K completely characterizes the closed-loop response. If $H(s)$ is a first-order system of the form $H(s) = (s\tau_2 + 1)/(s\tau_1 + \alpha)$, we obtain a **second-order loop** whose frequency response and phase error can be described by

$$\frac{Y(s)}{M(s)} = \frac{\Theta_0(s)}{\Theta_i(s)} = \frac{K(s\tau_2 + 1)}{\tau_1 s^2 + (\alpha + K\tau_2)s + K} \tag{12.37}$$

$$\frac{\Theta_i(s) - \Theta_0(s)}{\Theta_i(s)} = \frac{s}{s + KH(s)} = \frac{s(s\tau_1 + \alpha)}{\tau_1 s^2 + (\alpha + K\tau_2)s + K} \tag{12.38}$$

The parameters K, α, τ_1, and τ_2 are chosen to confine the phase error to small values while minimizing the distortion in the output.

PGR2026-00033
Krisp Technologies EX1024 Page 403

CHAPTER 12 PROBLEMS

DRILL AND REINFORCEMENT

12.1 (Filter Response) Consider a filter described by $H(s) = \dfrac{0.2s}{s^2 + 0.2s + 16}$.

(a) What traditional filter type does it describe?

(b) Estimate its response to the input $x(t) = \cos(0.2t) + \sin(4t) + \cos(50t)$.

12.2 (Filter Analysis) Find the impulse response $h(t)$ and step response $s(t)$ of each filter in Figure P12.2.

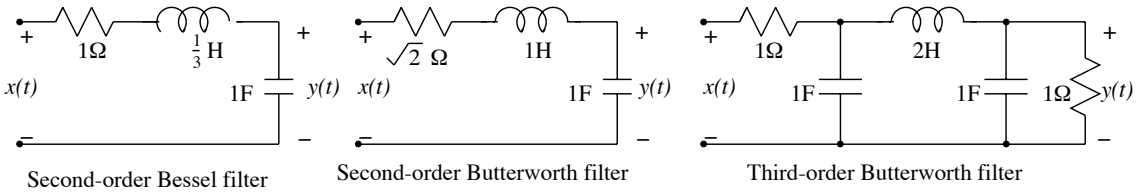


Figure P12.2 Filters for Problem 12.2

12.3 (Phase Delay and Group Delay) For each filter, find the phase delay and group delay.

(a) $H(s) = \dfrac{s-1}{s+1}$ (an allpass filter)

(b) $H(s) = \dfrac{1}{s^2 + \sqrt{2}s + 1}$ (a second-order Butterworth lowpass filter)

(c) $H(s) = \dfrac{3}{s^2 + 3s + 3}$ (a second-order Bessel lowpass filter)

12.4 (Minimum-Phase Systems) Find the stable minimum phase transfer function $H(s)$ from the following $|H(\omega)|^2$ and classify each filter by type. Also find a stable transfer function that is not minimum phase, if possible.

(a) $|H(\omega)|^2 = \dfrac{\omega^2}{\omega^2 + 4}$

(b) $|H(\omega)|^2 = \dfrac{4\omega^2}{\omega^4 + 17\omega^2 + 16}$

(c) $|H(\omega)|^2 = \dfrac{\omega^4 + 8\omega^2 + 16}{\omega^4 + 17\omega^2 + 16}$

12.5 (Bode Plots) Sketch the Bode magnitude and phase plots for each transfer function.

(a) $H(\omega) = \dfrac{10 + j\omega}{(1 + j\omega)(100 + j\omega)}$ **(b)** $H(\omega) = \dfrac{10j\omega(10 + j\omega)}{(1 + j\omega)(100 + j\omega)}$

(c) $H(\omega) = \dfrac{100 + j\omega}{j\omega(10 + j\omega)^2}$ **(d)** $H(\omega) = \dfrac{1 + j\omega}{(10 + j\omega)(5 + j\omega)}$

(e) $H(s) = \dfrac{10s(s+10)}{(s+1)(s+5)^2}$ **(f)** $H(s) = \dfrac{10(s+1)}{s(s^2 + 10s + 100)}$

PGR2026-00033
Krisp Technologies EX1024 Page 404

12.6 (Bode Plots) The transfer function of a system is $H(\omega) = \dfrac{10 + 10j\omega}{(10 + j\omega)(2 + j\omega)}$. Compute the following quantities in decibels (dB) at $\omega = 1$, 2, and 10 rad/s.

(a) The asymptotic, corrected and exact magnitude

(b) The asymptotic and exact phase

12.7 (Bode Plots) The transfer function of a system is $H(\omega) = \dfrac{1 + j\omega}{(0.1 + j\omega)^2}$.

(a) Find the asymptotic phase in degrees at ω =0.01, 0.1, 1, and 10 rad/s.

(b) What is the exact phase in degrees at $\omega = 0.01,\ 0.1,\ 1$, and 10 rad/s?

12.8 (Minimum-Phase Systems) For each Bode magnitude plot shown in Figure P12.8,

(a) Find the minimum-phase transfer function $H(s)$.

(b) Plot the asymptotic Bode phase plot from $H(s)$.

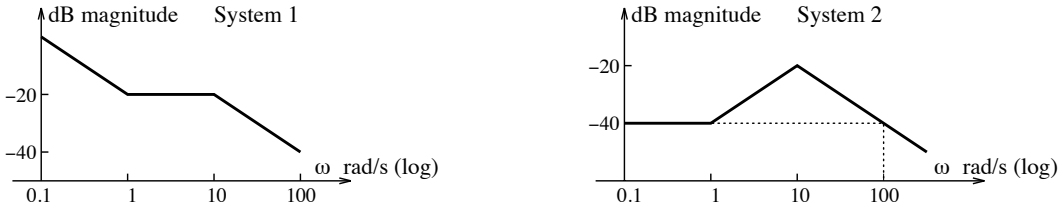


Figure P12.8 Bode magnitude plots for Problem 12.8

12.9 (Frequency Response and Bode Plots) Find and sketch the frequency response magnitude of $H(\omega)$ and the Bode magnitude plots for each of the following filters. What is the expected and exact value of the decibel magnitude of each filter at $\omega = 1$ rad/s?

(a) $H(s) = \dfrac{1}{s+1}$ (a first-order Butterworth lowpass filter)

(b) $H(s) = \dfrac{1}{s^2 + s\sqrt{2} + 1}$ (a second-order Butterworth lowpass filter)

(c) $H(s) = \dfrac{1}{(s+1)(s^2 + s + 1)}$ (a third-order Butterworth lowpass filter)

12.10 (Bode Plots of Nonminimum-Phase Filters) Even though it is usually customary to use Bode plots for minimum-phase filters, they can also be used for nonminimum-phase filters. Consider the filters described by

(a) $H(s) = \dfrac{10(s+1)}{s(s+10)}$ **(b)** $H(s) = \dfrac{10(s-1)}{s(s+10)}$ **(c)** $H(s) = \dfrac{10(s+1)e^{-s}}{s(s+10)}$

Which of these describe minimum-phase filters? Sketch the Bode magnitude and phase plots for each filter and compare. What are the differences (if any)? Which transfer function do you obtain (working backward) from the Bode magnitude plot of each filter alone? Why?

12.11 (Lead-Lag Compensators) Lead-lag compensators are often used in control systems and have the generic form described by $H(s) = \dfrac{1 + s\tau_1}{1 + s\tau_2}$. Sketch the Bode magnitude and phase plots of this compensator for the following values of τ_1 and τ_2.

(a) $\tau_1 = 0.1$ s, $\tau_2 = 10$ s **(b)** $\tau_1 = 10$ s, $\tau_2 = 0.1$ s

PGR2026-00033
Krisp Technologies EX1024 Page 405

REVIEW AND EXPLORATION

12.12 (Minimum-Phase Filters) The gain of an nth-order Butterworth lowpass filter is described by

$$|H(\omega)| = \frac{1}{\sqrt{1 + (\frac{\omega}{\alpha})^{2n}}}$$

(a) Argue that the Butterworth filter is a minimum-phase filter.

(b) Show that the half-power frequency of this filter is $\omega = \alpha$ for any order n.

(c) Find the minimum-phase transfer function $H(s)$ for $\alpha = 1$ and $n = 1$.

(d) Find the minimum-phase transfer function $H(s)$ for $\alpha = 1$ and $n = 2$.

(e) Find the minimum-phase transfer function $H(s)$ for $\alpha = 1$ and $n = 3$.

12.13 (Subsonic Filters) Some audio amplifiers include a subsonic filter to remove or reduce unwanted low-frequency noise (due, for example, to warped records) that might otherwise modulate the audible frequencies, causing intermodulation distortion. One manufacturer states that their subsonic filter provides a 12-dB/oct cut for frequencies below 15 Hz. Sketch the Bode plot of such a filter and find its transfer function $H(s)$.

12.14 (Audio Equalizers) Many audio amplifiers have bass, midrange, and treble controls to tailor the frequency response. The controls adjust the cutoff frequencies of lowpass, bandpass, and highpass filters connected in cascade (not parallel). The bass control provides a low-frequency boost (or cut) without affecting the high-frequency response. Examples of passive bass and treble control circuits are shown in Figure P12.14. The potentiometer setting controls the boost or cut.

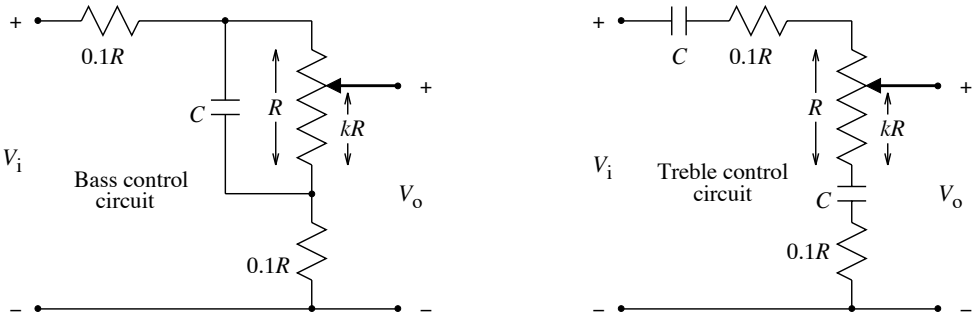


Figure P12.14 Figure for Problem 12.14

(a) Find the transfer function of the bass-control filter. What is its gain for $k = 0$, $k = 0.5$, and $k = 1$? What is the high-frequency gain? What are the minimum and maximum values of the low-frequency boost and cut in decibels provided by this filter?

(b) Find the transfer function of the treble-control filter. What is its gain for $k = 0$, $k = 0.5$, and $k = 1$? What is the low-frequency gain? What are the minimum and maximum values of the high-frequency boost and cut in decibels provided by this filter?

(c) The Sony TA-AX 500 audio amplifier lists values of the half-power frequency for its bass control filter as 300 Hz and for its treble control filter as 5 kHz. Select the circuit element values for the two filters to satisfy these half-power frequency requirements.

PGR2026-00033
Krisp Technologies EX1024 Page 406

12.15 (Allpass Filters) Allpass filters exhibit a constant gain and have a transfer function of the form $H(s) = P(-s)/P(s)$. Such filters are used as *delay equalizers* in cascade with other filters to adjust the phase or to compensate for the effects of phase distortion. Two examples are shown in Figure P12.15.

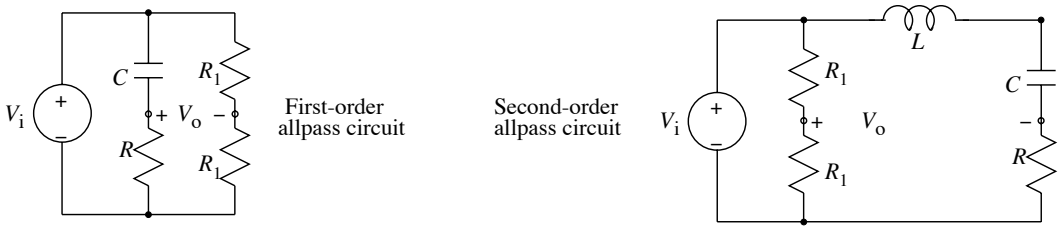


Figure P12.15 Allpass filters for Problem 12.15

(a) Show that the transfer functions of the two filters are given by

$$H_1(s) = 0.5\frac{s - 1/RC}{s + 1/RC} \qquad H_2(s) = 0.5\frac{s^2 - s(R/L) + (1/LC)}{s^2 + s(R/L) + (1/LC)}$$

(b) Assume $R = 1\ \Omega$, $L = 1$ H, and $C = 1$ F and find expressions for the gain, phase delay, and group delay of each filter as a function of frequency.

12.16 (Allpass Filters) Consider a lowpass filter with impulse response $h(t) = e^{-t}u(t)$. The input to this filter is $x(t) = \cos(t)$. We expect the output to be of the form $y(t) = A\cos(t+\theta)$. Find the values of A and θ. What should be the transfer function $H_1(s)$ of a first-order allpass filter that can be cascaded with the lowpass filter to correct for the phase distortion and produce the signal $z(t) = B\cos(t)$ at its output? If we use the first-order allpass filter of Problem 12.15, what will be the value of B?

12.17 (RIAA Equalization) Audio signals usually undergo a high-frequency boost (and low-frequency cut) before being used to make the master for commercial production of phonograph records. During playback, the signal from the phono cartridge is fed to a preamplifier (equalizer) that restores the original signal. The frequency response of the preamplifier is based on the RIAA (Recording Industry Association of America) equalization curve whose Bode plot is shown in Figure P12.17, with break frequencies at 50, 500, and 2122 Hz.

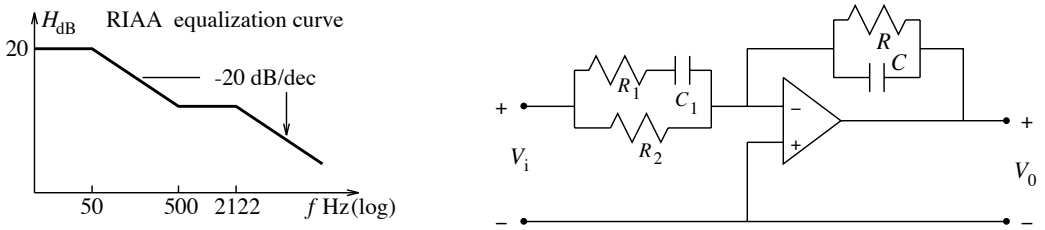


Figure P12.17 RIAA equalization and op-amp circuit for Problem 12.17

(a) What is the transfer function $H(s)$ of the RIAA equalizer? From your expression, obtain the exact decibel magnitude at 1 kHz.

(b) It is claimed that the op-amp circuit shown in Figure P12.17 can realize the transfer function $H(s)$ (except for a sign inversion). Justify this claim and determine the values for the circuit elements if $R_2 = 10\ \text{k}\Omega$.

12.18 (Pre-Emphasis and De-Emphasis Filters for FM) To improve efficiency and signal-to-noise ratio, commercial FM stations have been using pre-emphasis (a high-frequency boost) for input signals

PGR2026-00033
Krisp Technologies EX1024 Page 407

prior to modulation and transmission. At the receiver, after demodulation, the signal is restored using a de-emphasis circuit that reverses this process. Figure P12.18 shows the circuits and their Bode plots.

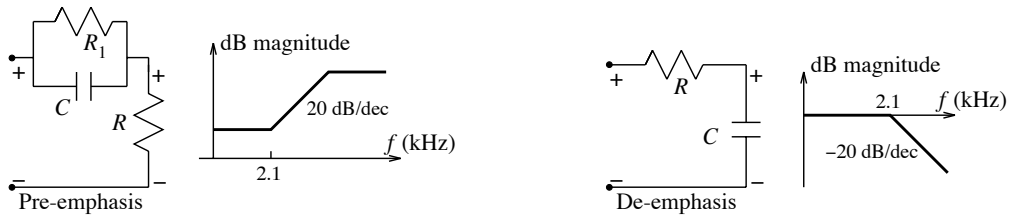


Figure P12.18 Pre-emphasis and de-emphasis circuits for Problem 12.18

The lower break-frequency corresponds to $f = 2122$ Hz (or $\tau = 75\mu$s), and the upper break-frequency should exceed the highest frequency transmitted (say, 15 kHz). Find the transfer function of each circuit and compute the element values required. While pre-emphasis and de-emphasis schemes have been used for a long time, FM stations are turning to alternative compression schemes (such as Dolby B) to transmit high-fidelity signals without compromising their frequency response or dynamic range.

12.19 (Allpass Filters) Argue that for a stable allpass filter, the location of each left half-plane pole must be matched by a right half-plane zero. Does an allpass filter represent a minimum-phase, mixed-phase, or maximum-phase system?

12.20 (Allpass Filters) Argue that the group delay t_g of a stable allpass filter is always greater than or equal to zero.

12.21 (Allpass Filters) The overall delay of cascaded systems is simply the sum of the individual delays. By cascading a minimum-phase filter with an allpass filter, argue that a minimum-phase filter has the smallest group delay from among all filters with the same magnitude spectrum.

12.22 (Design of a Notch Filter) To design a second-order notch filter with a notch frequency of ω_0, we select conjugate zeros at $s = \pm j\omega_0$ where the response goes to zero. The poles are then placed close to the zeros but to the left of the $j\omega$-axis to ensure a stable filter. We thus locate the poles at $s = -\alpha\omega_0 \pm j\omega_0$, where the parameter α determines the sharpness of the notch.

(a) Find the transfer function $H(s)$ of a notch filter with $\omega_0 = 10$ rad/s and α as a variable.
(b) Find the bandwidth of the notch filter for $\alpha = 0.1$.
(c) Will smaller values of α (i.e., $\alpha \ll 1$) result in a sharper notch?

12.23 (Design of a Bandpass Filter) Let us design a bandpass filter with lower and upper cutoff frequencies ω_1 and ω_2 by locating conjugate poles at $s = -\alpha\omega_1 \pm j\omega_1$ and $s = -\alpha\omega_2 \pm j\omega_2$ (where $\alpha < 1$) and a pair of zeros at $s = 0$.

(a) Find the transfer function $H(s)$ of a bandpass filter with $\omega_1 = 40$ rad/s, $\omega_2 = 50$ rad/s, and $\alpha = 0.1$.
(b) What is the filter order?
(c) Estimate the center frequency and bandwidth.
(d) What is the effect of changing α on the frequency response?

12.24 (Frequency Response of Filters) The transfer function of a second-order Butterworth filter is $H(s) = \dfrac{1}{s^2 + \sqrt{2}s + 1}$.

(a) Sketch its magnitude and phase spectrum and its Bode magnitude plot.

PGR2026-00033
Krisp Technologies EX1024 Page 408

(b) What is the dc gain A_0 of this filter?
(c) At what frequency does the gain equal $A_0/\sqrt{2}$?
(d) At what frequency (approximately) does the gain equal $0.01A_0$?

12.25 (Frequency Response of Filters) The transfer function of a second-order Butterworth lowpass filter is $H(s) = \dfrac{1}{(0.1s)^2 + 0.1\sqrt{2}s + 1}$.

(a) What is the half-power frequency of this filter?
(b) At what frequency (approximately) does the gain equal 1% of the peak gain?
(c) Sketch the response $y(t)$ of this filter if $x(t) = u(t)$. Does the response show overshoot, or is it monotonic?

12.26 (Frequency Response of Filters) A bandpass filter is described by $H(s) = \dfrac{0.1s}{(0.1s)^2 + 0.1s + 1}$.

(a) Find the center frequency ω_0, the half-power frequencies ω_1 and ω_2, and the half-power bandwidth $\Delta\omega$ of this filter.
(b) Compute and sketch its response $y(t)$ if $x(t) = \cos(\omega_0 t)$.
(c) Compute and sketch its response $y(t)$ if $x(t) = 1 + \cos(0.1\omega_1 t) + \cos(\omega_0 t) + \cos(10\omega_2 t)$.
(d) Regard the frequencies outside the passband of the filter as noise and define the signal-to-noise ratio (SNR) in dB as

$$\text{SNR } = 10\log\left(\frac{\text{signal power}}{\text{noise power}}\right) \text{ dB}$$

What is the SNR at the input of the filter and the SNR at the output of the filter if the input is $x(t) = \cos(0.2\omega_1 t) + \cos(\omega_0 t)$?
(e) What is the SNR at the input of the filter and the SNR at the output of the filter if the input is $x(t) = \cos(\omega_0 t) + \cos(5\omega_2 t)$?
(f) Compute the difference in the decibel gain of $H(\omega)$ at ω_0 and $0.2\omega_1$. Also compute the difference in the decibel gain of $H(\omega)$ at ω_0 and $5\omega_2$. Is there a connection between these numbers and the SNR figures computed in parts (d) and (e)?

12.27 (Compensating Probes for Oscilloscopes) Probes are often used to attenuate large amplitude signals before displaying them on an oscilloscope. Ideally, the probe should provide a constant attenuation for all frequencies. However, the capacitive input impedance of the oscilloscope and other stray capacitance effects cause loading and distort the displayed signal (especially if it contains step changes). Compensating probes employ an adjustable compensating capacitor to minimize the effects of loading. This allows us, for example, to display a square wave as a square wave (with no undershoot or ringing). An equivalent circuit using a compensating probe is shown in Figure P12.27.

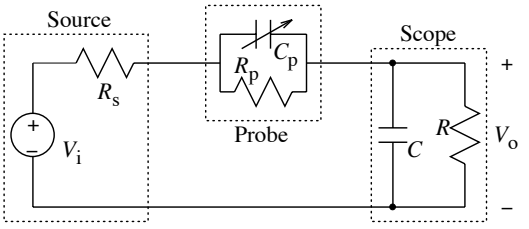


Figure P12.27 Figure for Problem 12.27

(a) Assuming $R_s = 0.1R$, show that $R_p = 8.9R$ provides a 10:1 attenuation at dc.

PGR2026-00033
Krisp Technologies EX1024 Page 409

(b) In the absence of C_p (no compensation) but with $R_s = 0.1R$ and $R_p = 8.9R$ as before, find the transfer function $H(s) = V_o(s)/V_i(s)$. Compute and sketch its step response. What are the time constant and rise time?

(c) In the presence of C_p, and with $R_s = 0.1R$ and $R_p = 8.9R$ as before, compute the transfer function $H_p(s) = V_o(s)/V_i(s)$. Show that if $R_pC_p = RC$, then $H_p(s) = 0.1/(1 + 0.01sRC)$. Compute and sketch its step response. How does the presence and choice of the compensating capacitor affect the composite time constant and rise time? Does compensation work?

12.28 (Design of Allpass Filters) Consider a lowpass filter with impulse response $h(t) = e^{-t}u(t)$. The input to this filter is $x(t) = \cos(t) + \cos(t/3)$. Design an allpass filter $H_2(s)$ that can be cascaded with the lowpass filter to correct for the phase distortion and produce a signal that has the form $z(t) = B\cos(t) + C\cos(t/3)$ at its output. If the allpass filter has a gain of 0.5, what will be the values of B and C?

COMPUTATION AND DESIGN

bodegui A GUI for Visualization of Bode Plots

The graphical user interface `bodegui` allows you to visualize bode plots of a transfer function. Both exact and asymptotic plots are displayed against f or ω.
To explore this routine, type `bodegui` at the MATLAB prompt.

12.29 (The Minimum-Phase Concept) Consider three filters described by

$$H_1(s) = \frac{10(s+1)}{s(s+10)} \qquad H_2(s) = \frac{10(s-1)}{s(s+10)} \qquad H_3(s) = \frac{10(s+1)e^{-s}}{s(s+10)}$$

(a) Use MATLAB to plot the magnitude spectrum of all filters on the same graph. Confirm that all filters have identical magnitude spectra.

(b) Use MATLAB to plot the phase spectrum of all filters on the same graph. How does their phase differ?

(c) Use MATLAB to plot the group delay of all filters on the same graph. How does their group delay differ?

(d) From which plot can you identify the minimum-phase filter, and why?

12.30 (Bode Plots) The gain of an nth-order Butterworth lowpass filter is described by

$$|H(\omega)| = \frac{1}{\sqrt{1 + (\frac{\omega}{\alpha})^{2n}}}$$

(a) Show that the half-power bandwidth of this filter is $\omega = \alpha$ for any n.

(b) Use paper and pencil to make a sketch of the asymptotic Bode plot of a Butterworth filter for $\alpha = 1$ and $n = 2,\ 3,\ 4$. What is the asymptotic and true magnitude of each filter at the half-power frequency?

(c) Determine the minimum-phase transfer function $H(s)$ for $n = 2,\ 3,\ 4,\ 5$.

(d) Use MATLAB to plot the Bode magnitude plot for $\alpha = 1$ and $n = 2,\ 3,\ 4,\ 5$ on the same graph. Do these plots confirm your results from the previous part?

PGR2026-00033
Krisp Technologies EX1024 Page 410

(e) Use MATLAB to plot the phase for $\alpha = 1$ and $n = 2,\ 3,\ 4,\ 5$ on the same graph. Describe how the phase spectrum changes as the order n is increased. Which filter has the most nearly linear phase in the passband?

(f) Use MATLAB to plot the group delay for $\alpha = 1$ and $n = 2,\ 3,\ 4,\ 5$ on the same plot. Describe how the group delay changes as the order n is increased. Which filter has the most nearly constant group delay in the passband?

12.31 (Bode Plots) The ADSP routine `bodelin` plots the asymptotic Bode magnitude and phase plots of a minimum-phase transfer function $H(s)$. Plot the asymptotic Bode magnitude and phase plots of the following filters.

(a) $H(s) = \dfrac{1}{s^2 + \sqrt{2}s + 1}$ (a second-order Butterworth lowpass filter)

(b) $H(s) = \dfrac{s}{s^2 + s + 1}$ (a second-order bandpass filter)

(c) $y'''(t) + 2y''(t) + 2y'(t) + y(t) = x(t)$ (a third-order Butterworth lowpass filter)

12.32 (System Response) Use the ADSP routine `sysresp2` to find the step response and impulse response of the following filters and plot each result. Compare the features of the step response of each filter. Compare the features of the impulse response of each filter.

(a) $y'(t) + y(t) = x(t)$ (a first-order lowpass filter)

(b) $y''(t) + \sqrt{2}y'(t) + y(t) = x(t)$ (a second-order Butterworth lowpass filter)

(c) $y''(t) + y'(t) + y(t) = x'(t)$ (a bandpass filter)

(d) $y'''(t) + 2y''(t) + 2y'(t) + y(t) = x(t)$ (a third-order Butterworth lowpass filter)

12.33 (Frequency Response) Use the ADSP routine `bodelin` to obtain Bode plots of the following filters. Also plot the frequency response on a linear scale. Compare the features of the frequency response of each filter.

(a) $y'(t) + y(t) = x(t)$ (a first-order lowpass filter)

(b) $y''(t) + \sqrt{2}y'(t) + y(t) = x(t)$ (a second-order Butterworth lowpass filter)

(c) $y''(t) + y'(t) + y(t) = x'(t)$ (a bandpass filter)

(d) $y'''(t) + 2y''(t) + 2y'(t) + y(t) = x(t)$ (a third-order Butterworth lowpass filter)

12.34 (Performance Measures) Performance measures for lowpass filters include the rise time, the 5% settling time, the half-power bandwidth, and the time-bandwidth product. Plot the step response, magnitude spectrum, and phase spectrum of the following systems. Then, use the ADSP routine `trbw` to numerically estimate each performance measure.

(a) A first-order lowpass filter with $\tau = 1$

(b) The cascade of two first-order filters, each with $\tau = 1$

(c) $H(s) = \dfrac{1}{s^2 + \sqrt{2}s + 1}$ (a second-order Butterworth lowpass filter)

(d) $y'''(t) + 2y''(t) + 2y'(t) + y(t) = x(t)$ (a third-order Butterworth lowpass filter)

12.35 (Steady-State Response in Symbolic Form) The ADSP routine `ssresp` yields a symbolic expression for the steady-state response to sinusoidal inputs (see Chapter 21 for examples of its usage). Find the steady-state response to the input $x(t) = 2\cos(3t - \frac{\pi}{3})$ for each of the following systems.

(a) $H(s) = \dfrac{2}{s + \alpha}$, for $\alpha = 1,\ 2$

(b) $y''(t) + 4y(t) + Cy(t) = x(t)$, for $C = 3,\ 4,\ 5$

PGR2026-00033
Krisp Technologies EX1024 Page 411

Chapter 13

ANALOG FILTERS

13.0 Scope and Objectives

Analog filters are indispensable in many situations. The front end of most digital signal-processing devices is an **anti-aliasing** analog filter that limits the input signal to a range of frequencies that the digital filter can handle. The design of classical analog filters is based on approximating the magnitude or phase specifications by polynomials or rational functions. This chapter describes the synthesis and design of practical analog filters and develops design methods for Butterworth, Chebyshev I, Chebyshev II (inverse Chebyshev), and elliptic filters whose design calls for magnitude specifications, and Bessel filters with nearly linear phase whose design calls for phase specifications.

13.1 Introduction

A filter may be regarded as a *frequency-selective device* that allows us to shape the magnitude or phase response in a prescribed manner. The terminology of filters based on magnitude specifications is illustrated in Figure 13.1. The passband ripple (deviation from maximum gain) and stopband ripple (deviation from zero gain) are usually described by their attenuation (reciprocal of the linear gain).

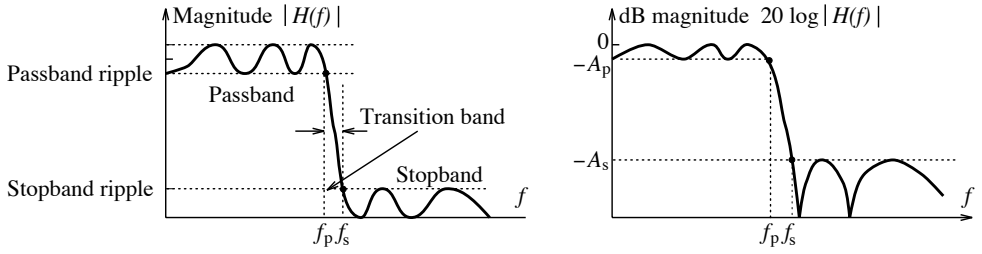


Figure 13.1 Filter terminology

REVIEW PANEL 13.1

How Are the Gain and Attenuation Related on a Linear or Decibel Scale?

Linear: Attenuation equals the *reciprocal* of gain: $G = |H(f)|, \quad A = 1/|H(f)|$.
Decibel: Attenuation equals the *negative* of dB gain: $G_{\text{dB}} = 20\log|H(f)|, \quad A_{\text{dB}} = -20\log|H(f)|$.

PGR2026-00033
Krisp Technologies EX1024 Page 412

Of the four classical filter types based on magnitude specifications, the Butterworth filter is monotonic in the passband and stopband, the Chebyshev I filter displays ripples in the passband but is monotonic in the stopband, the Chebyshev II filter is monotonic in the passband but has ripples in the stopband, and the elliptic filter has ripples in both bands.

REVIEW PANEL 13.2

Magnitude Characteristics of Four Classical Filters

Butterworth: Monotonic in both bands **Chebyshev I:** Monotonic passband, rippled stopband
Elliptic: Rippled in both bands **Chebyshev II:** Rippled passband, monotonic stopband

The design of analog filters typically relies on frequency specifications (passband and stopband edge(s)) and magnitude specifications (maximum passband attenuation and minimum stopband attenuation) to generate a minimum-phase filter transfer function with the smallest order that meets or exceeds specifications. Most design strategies are based on converting the given frequency specifications to those applicable to a **lowpass prototype** (LPP) with a cutoff frequency of 1 rad/s (typically the passband edge), designing the lowpass prototype, and converting to the required filter type using frequency transformations.

13.1.1 Prototype Transformations

The lowpass-to-lowpass (LP2LP) transformation converts a lowpass prototype $H_P(s)$ with a cutoff frequency of 1 rad/s to a lowpass filter $H(s)$ with a cutoff frequency of ω_x rad/s using the transformation $s \Rightarrow s/\omega_x$, as shown in Figure 13.2. This is just linear frequency scaling.

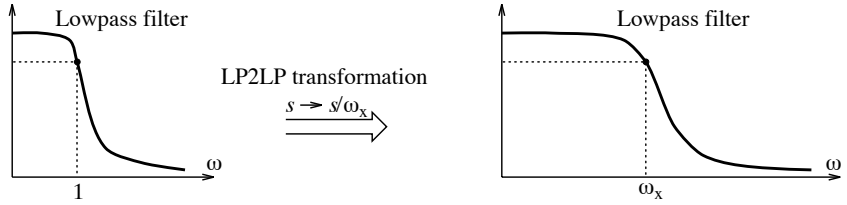


Figure 13.2 The lowpass-to-lowpass transformation

The lowpass-to-highpass (LP2HP) transformation converts a lowpass prototype $H_P(s)$ with a cutoff frequency of 1 rad/s to a highpass filter $H(s)$ with a cutoff frequency of ω_x rad/s, using the *nonlinear* transformation $s \Rightarrow \omega_x/s$. This is illustrated in Figure 13.3.

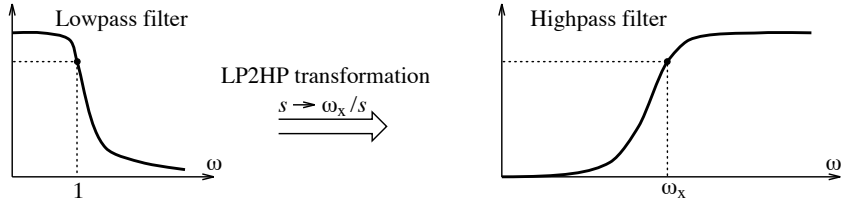


Figure 13.3 The lowpass-to-highpass transformation

The lowpass-to-bandpass (LP2BP) transformation is illustrated in Figure 13.4. It converts a lowpass prototype $H_P(s)$ with a cutoff frequency of 1 rad/s to a bandpass filter $H(s)$ with a center frequency of

PGR2026-00033
Krisp Technologies EX1024 Page 413

ω_0 rad/s and a passband of B rad/s, using the nonlinear, quadratic transformation

$$s \Rightarrow \frac{s^2 + \omega_0^2}{sB} \qquad \omega_{\text{LP}} \longrightarrow \frac{\omega_{\text{BP}}^2 - \omega_0^2}{\omega_{\text{BP}} B} \tag{13.1}$$

Here, ω_0 is the *geometric* mean of the band edges ω_L and ω_H, with $\omega_L \omega_H = \omega_0^2$, and the bandwidth is given by $B = \omega_H - \omega_L$. Any pair of geometrically symmetric bandpass frequencies ω_a and ω_b, with $\omega_a \omega_b = \omega_0^2$, corresponds to the lowpass prototype frequency $(\omega_b - \omega_a)/B$. The lowpass prototype frequency at infinity is mapped to the bandpass origin. This quadratic transformation yields a transfer function with *twice* the order of the lowpass filter.

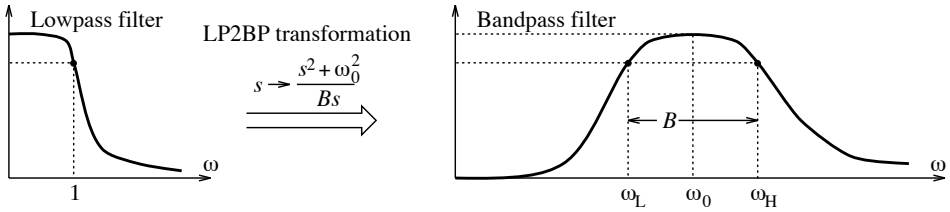


Figure 13.4 The lowpass-to-bandpass transformation

The lowpass-to-bandstop (LP2BS) transformation is illustrated in Figure 13.5. It converts a lowpass prototype $H_P(s)$ with a cutoff frequency of 1 rad/s to a bandstop filter $H(s)$ with a center frequency of ω_0 rad/s and a stopband of B rad/s, using the nonlinear, quadratic transformation

$$s \longrightarrow \frac{sB}{s^2 + \omega_0^2} \qquad \omega_{\text{LP}} \longrightarrow \frac{\omega_{\text{BS}} B}{\omega_0^2 - \omega_{\text{BS}}^2} \tag{13.2}$$

Here $B = \omega_H - \omega_L$ and $\omega_0^2 = \omega_H \omega_L$. The lowpass origin maps to the bandstop frequency ω_0. Since the roles of the passband and the stopband are now reversed, a pair of geometrically symmetric bandstop frequencies ω_a and ω_b, with $\omega_a \omega_b = \omega_x^2$, maps to the lowpass prototype frequency $\omega_{\text{LP}} = B/(\omega_b - \omega_a)$. This quadratic transformation also yields a transfer function with *twice* the order of the lowpass filter.

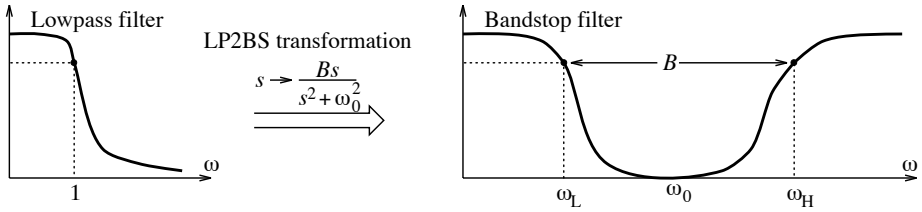


Figure 13.5 The lowpass-to-bandstop transformation

13.1.2 Lowpass Prototype Specifications

Given the frequency specifications of a lowpass filter with band edges ω_p and ω_s, the specifications for a lowpass prototype with a passband edge of 1 rad/s are $\nu_p = 1$ rad/s and $\nu_s = \omega_s/\omega_p$ rad/s. The LP2LP transformation is $s \rightarrow s/\omega_p$. For a lowpass prototype with a stopband edge of 1 rad/s, we would use $\nu_p = \omega_s/\omega_p$ rad/s and $\nu_s = 1$ rad/s instead.

PGR2026-00033
Krisp Technologies EX1024 Page 414

For a highpass filter with band edges ω_p and ω_s, the specifications for a lowpass prototype with a passband edge of 1 rad/s are $\nu_p = 1$ rad/s and $\nu_s = \omega_p/\omega_s$ rad/s. The LP2HP transformation is $s \to \omega_p/s$.

For bandpass or bandstop filters, we start with the band-edge frequencies arranged in increasing order as $[\omega_1,\ \omega_2,\ \omega_3,\ \omega_4]$. If these frequencies show *geometric symmetry*, the center frequency ω_0 is $\omega_0^2 = \omega_2\omega_3 = \omega_1\omega_4$. The band edges of the lowpass prototype are $\nu_p = 1$ rad/s and $\nu_s = (\omega_4 - \omega_1)/(\omega_3 - \omega_2)$ rad/s for both the bandpass and bandstop case. The bandwidth is $B = \omega_4 - \omega_1$ for bandstop filters and $B = \omega_4 - \omega_1$ for bandpass filters. These values of ω_0 and B are used in subsequent transformation of the lowpass prototype to the required filter type.

REVIEW PANEL 13.3

Obtaining Frequency Specifications for the Lowpass Prototype

From lowpass and highpass (given $[f_1,\ f_2]$): $\nu_p = 1$ rad/s $\quad \nu_s = f_2/f_1$ rad/s

From bandpass and bandstop (given $[f_1,\ f_2,\ f_3,\ f_4]$): $\nu_p = 1$ rad/s $\quad \nu_s = \dfrac{f_4 - f_1}{f_3 - f_2}$ rad/s

Note: $[f_1,\ f_2,\ f_3,\ f_4]$ are in *increasing order* with $f_1 f_4 = f_2 f_3 = f_0^2$ (*geometric symmetry*).

REVIEW PANEL 13.4

Transforming the Lowpass Prototype to the Required Form

To LP: $s \Rightarrow \dfrac{s}{2\pi f_p}$ **To HP:** $s \Rightarrow \dfrac{2\pi f_p}{s}$ **To BP:** $s \Rightarrow \dfrac{s^2 + \omega_0^2}{sB_{\text{BP}}}$ **To BS:** $s \Rightarrow \dfrac{sB_{\text{BS}}}{s^2 + \omega_0^2}$

Notation: $\omega_0^2 = 4\pi^2 f_0^2 \qquad f_0^2 = f_1 f_4 = f_2 f_3 \qquad B_{\text{BP}} = 2\pi(f_3 - f_2) \qquad B_{\text{BS}} = 2\pi(f_4 - f_1).$

If the given specifications $[\omega_1, \omega_2, \omega_3, \omega_4]$ are not geometrically symmetric, some of the band edges must be relocated (increased or decreased) in a way that the new transition widths do not exceed the original. This is illustrated in Figure 13.6. It is common practice to hold the passband edges fixed and relocate the other frequencies. For example, for the bandpass case, the passband edges $[\omega_2,\ \omega_3]$ are held fixed, and $\omega_0^2 = \omega_2\omega_3$. To ensure that the new transition widths do not exceed the original, we must *increase* ω_1 or *decrease* ω_4 (or do both) such that $\omega_0^2 = \omega_1\omega_4$ (for geometric symmetry).

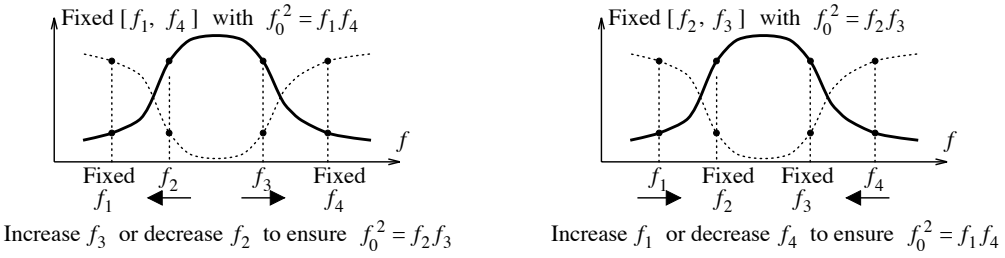


Figure 13.6 The band edges for bandpass and bandstop filters

REVIEW PANEL 13.5

Making the Band Edges Geometrically Symmetric for Bandpass and Bandstop Filters

Bandpass (fixed passband edges $[f_2,\ f_3]$): *Increase* f_1 or *decrease* f_4 such that $f_1 f_4 = f_2 f_3$.

Bandstop (fixed passband edges $[f_1,\ f_4]$): *Decrease* f_2 or *increase* f_3 such that $f_2 f_3 = f_1 f_4$.

Rationale: We make sure that the new transition widths do not exceed the given ones.

PGR2026-00033
Krisp Technologies EX1024 Page 415

EXAMPLE 13.1 (Prototype Specifications and Transformations)

(a) For a lowpass filter with $f_p = 200$ Hz and $f_s = 500$ Hz, normalization by f_p gives the lowpass prototype band edges $\nu_p = 1$ rad/s and $\nu_s = 2.5$ rad/s. The actual filter is designed from the lowpass prototype using the LP2LP transformation $s \to s/\omega_p$, with $\omega_p = 2\pi f_p = 400\pi$ rad/s.

(b) For a highpass filter with $\omega_p = 500$ rad/s and $\omega_s = 200$ rad/s, $[\omega_1,\ w_2] = [200,\ 500]$ rad/s. The lowpass prototype band edges are $\nu_p = 1$ rad/s and $\nu_s = 2.5$ rad/s. The highpass filter is designed from the lowpass prototype, using the LP2HP transformation $s \to \omega_p/s$, with $\omega_p = 500$ rad/s.

(c) Consider a bandpass filter with band edges $[\omega_1,\ \omega_2,\ \omega_3,\ \omega_4] = [16,\ 18,\ 32,\ 48]$ rad/s.
For a fixed passband, we choose $B = \omega_3 - \omega_2$ and $\omega_0^2 = \omega_2\omega_3 = 576$.
Since $\omega_1\omega_4 > \omega_0^2$, we recompute $\omega_4 = \omega_0^2/\omega_1 = 36$ rad/s, which ensures both geometric symmetry and transition widths not exceeding the original.
Thus, $[\omega_1,\ \omega_2,\ \omega_3,\ \omega_4] = [16,\ 18,\ 32,\ 36]$ rad/s and $B = \omega_3 - \omega_2 = 32 - 18 = 14$ rad/s.
With $\omega_4 - \omega_1 = 36 - 16 = 20$ rad/s, the lowpass prototype band edges are

$$\nu_p = 1 \text{ rad/s} \qquad \nu_s = \frac{\omega_4 - \omega_1}{B} = \frac{20}{14} = 1.4286 \text{ rad/s}$$

The bandpass filter is designed from the LPP using the LP2BP transformation

$$s \to \frac{s^2 + \omega_0^2}{sB} = \frac{s^2 + 576}{14s}$$

(d) Consider a bandstop filter with band edges $[\omega_1,\ \omega_2,\ \omega_3,\ \omega_4] = [16,\ 18,\ 32,\ 48]$ rad/s.
For a fixed passband, we choose $B = \omega_4 - \omega_1$, and $\omega_0^2 = (16)(48) = 768$.

Since $\omega_2\omega_3 < \omega_0^2$, we recompute $\omega_3 = \omega_0^2\omega_2 = 42.667$ to give

$$[\omega_1,\ \omega_2,\ \omega_3,\ \omega_4] = [16,\ 18,\ 42.6667,\ 48] \text{ rad/s} \qquad B = \omega_4 - \omega_1 = 48 - 16 = 32 \text{ rad/s}$$

With $\omega_3 - \omega_2 = 24.6667$ rad/s, the lowpass prototype band edges are

$$\nu_p = 1 \text{ rad/s} \qquad \nu_s = \frac{B}{\omega_3 - \omega_2} = \frac{32}{24.6667} = 1.2973 \text{ rad/s}$$

The bandstop filter is designed from the LPP using the LP2BS transformation

$$s \to \frac{sB}{s^2 + \omega_0^2} = \frac{32s}{s^2 + 768}$$

13.2 The Design Process

The filter magnitude specifications are usually given by the passband *attenuation* A_p and stopband *attenuation* A_s, expressed in decibels. The filter gain and attenuation are reciprocals of each other on a linear scale and differ only in sign on a decibel scale. Classical filters approximate the magnitude squared function $|H(\nu)|^2$ of a lowpass prototype by a polynomial or ratio of polynomials in ν^2 and derive a minimum-phase

PGR2026-00033
Krisp Technologies EX1024 Page 416

filter transfer function from this approximation. The magnitude squared function and attenuation for an nth-order lowpass prototype are of the form

$$|H(\nu)|^2 = \frac{1}{1+\alpha^2 L_n^2(\nu)} \qquad A_{\text{dB}}(\nu) = 10\log[1+\alpha^2 L_n^2(\nu)] \ \ (\text{dB}) \tag{13.3}$$

Here, $L_n(\nu)$ is an nth-order polynomial or rational function, and α controls the passband ripple (deviation from maximum gain). The various classical filters differ primarily in the choice for $L_n^2(\nu)$ to best meet desired specifications. We require $L_n^2(\nu) \approx 0$ in the passband (to ensure a gain of nearly unity) and $L_n^2(\nu) \gg 0$ in the stopband (to ensure a gain of nearly zero). The attenuation equation is the only design equation we need. We find the filter order n and the parameter α by evaluating the attenuation relation at the passband and stopband edges. This establishes the exact form of $|H(\nu)|^2$ from which the prototype transfer function $H_P(s)$ can be readily obtained. If we design the prototype to *exactly meet* attenuation requirements at the cutoff frequency (usually the passband edge), the attenuation requirements will be *exceeded* at all the other frequencies. The final step is to use frequency transformations to obtain the required filter.

13.3 The Butterworth Filter

The Butterworth lowpass prototype is based on the choice $L(\nu) = \nu^n$ to give

$$|H(\nu)|^2 = \frac{1}{1+\epsilon^2\nu^{2n}} \qquad A_{\text{dB}}(\nu) = 10\log(1+\epsilon^2\nu^{2n}) \ \ (\text{dB}) \tag{13.4}$$

The Butterworth filter is also called a **maximally flat filter** because the choice $L(\nu) = \nu^n$ (with all its lower-order coefficients missing) forces all its derivatives to zero at $\nu = 0$ and ensures not only the flattest response at $\nu = 0$ but also the least deviation from unit gain in the passband. All the degrees of freedom (the free coefficients in the nth-order polynomial $L(\nu)$) are spent in making the Butterworth magnitude response maximally flat at the origin (by setting them to zero). This yields a monotonic frequency response free of overshoot but also results in a slow transition from the passband to the stopband.

The quantity ϵ^2 is a measure of the deviation from unit passband magnitude. The magnitude is *monotonic* and decreases with frequency. The filter gain at the passband edge $\nu = 1$ equals

$$|H(1)| = \frac{1}{\sqrt{1+\epsilon^2}} \tag{13.5}$$

A typical magnitude spectrum is illustrated in Figure 13.7.

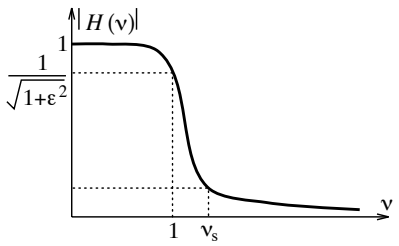


Figure 13.7 A Butterworth lowpass filter

For an nth-order Butterworth filter, the attenuation rate at high frequencies is $20n$ dB/dec and the attenuation may be approximated by

$$A(\nu) = 10\log(1+\epsilon^2\nu^{2n}) \approx 10\log(\epsilon^2\nu^{2n}) = 20\log\epsilon + 20n\log\nu \tag{13.6}$$

PGR2026-00033
Krisp Technologies EX1024 Page 417

REVIEW PANEL 13.6

The Butterworth Filter Has a Maximally Flat Magnitude Squared Spectrum

$$|H(\nu)|^2 = \frac{1}{1+\epsilon^2\nu^{2n}} \qquad |H(0)| = 1 \qquad |H(1)| = \frac{1}{\sqrt{1+\epsilon^2}} \qquad H_{\text{dB}} = -10\log(1+\epsilon^2\nu^{2n})$$

13.3.1 Butterworth Filter Design

The form of $|H(\nu)|^2$ requires only the ripple parameter ϵ and the filter order n. We find ϵ^2 by evaluating the attenuation $A_{\text{dB}}(\nu)$ at the passband edge ($\nu = \nu_p = 1$ rad/s):

$$A_p = A_{\text{dB}}(1) = 10\log(1+\epsilon^2) \qquad \epsilon^2 = 10^{0.1A_p} - 1 \tag{13.7}$$

We find n by evaluating $A_{\text{dB}}(\nu)$ at the stopband edge $\nu = \nu_s$ rad/s to give

$$A_s = A_{\text{dB}}(\nu_s) = 10\log(1+\epsilon^2\nu_s^{2n}) \qquad n = \frac{\log[(10^{0.1A_s}-1)/\epsilon^2]^{1/2}}{\log \nu_s} \tag{13.8}$$

The filter order must be *rounded up* to an integer value. If the attenuation is specified at several frequencies, the order must be found for each frequency, and the *largest* value chosen for n. The smaller the passband attenuation A_p (meaning a smaller ϵ^2 or a flatter passband) or the larger the stopband attenuation A_s (meaning a sharper transition), the higher is the filter order n.

The half-power frequency ν_3 is found by setting $|H(\nu_3)|^2 = 0.5$ to give

$$\epsilon^2\nu_3^{2n} = 1 \qquad \text{or} \qquad \nu_3 = (1/\epsilon)^{1/n} \tag{13.9}$$

REVIEW PANEL 13.7

Ripple (ϵ), Order (n), and 3-dB Frequency (ν_3) of a Butterworth Prototype

$$\epsilon^2 = 10^{0.1A_p} - 1 \qquad n = \frac{\log[(10^{0.1A_s}-1)/\epsilon^2]^{1/2}}{\log \nu_s} \qquad \nu_3 = (1/\epsilon)^{1/n}$$

13.3.2 The Butterworth Prototype Transfer Function

To find the Butterworth lowpass prototype transfer function $H_P(s)$ from $|H(\nu)|^2$, we replace ν^2 by $-s^2$ and pick only the left half-plane (LHP) roots to generate the required minimum-phase transfer function. With $\nu^2 = -s^2$, we obtain

$$H_P(s)H_P(-s) = \frac{1}{1+\epsilon^2(-s^2)^n} = \frac{1}{1+(-1)^n(s/R)^{2n}} \qquad R = (1/\epsilon)^{1/n} \tag{13.10}$$

This has $2n$ poles, only half of which are in the left half-plane. The pole locations are given by

$$(-1)^n(s/R)^{2n} = -1 \qquad (-js/R)^{2n} = -1 \tag{13.11}$$

This equation suggests that we find the $2n$ roots of -1. With $e^{j(2k-1)\pi} = -1$, we get

$$(-js/R)^{2n} = e^{j(2k-1)\pi}, \qquad k = 1, 2, \ldots, 2n \tag{13.12}$$

This yields the $2n$ poles p_k as

$$p_k = Re^{j(\theta_k+\pi/2)} \qquad \theta_k = \frac{(2k-1)\pi}{2n}, \quad k = 1, 2, \ldots, 2n \tag{13.13}$$

PGR2026-00033
Krisp Technologies EX1024 Page 418

The poles lie on a circle of radius $R = (1/\epsilon)^{1/n}$ in the s-plane. The poles are equally spaced π/n radians apart. Their angular orientation (with respect to the positive $j\omega$-axis) is given by $\theta_k = (2k-1)\pi/2n$. There can never be a pole on the $j\omega$-axis because θ_k can never be zero. The pole locations are illustrated for $n = 2$ and $n = 3$ in Figure 13.8.

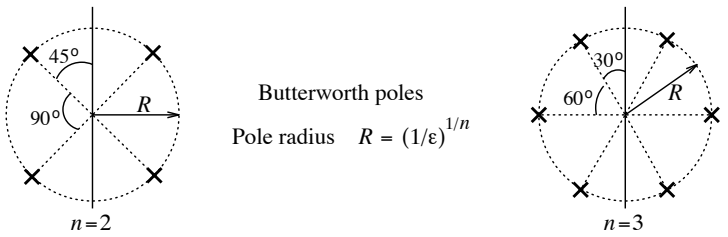


Figure 13.8 The poles of a Butterworth lowpass filter lie on a circle.

The real and imaginary parts of the left half-plane poles are given by

$$p_k = -R\sin\theta_k + jR\cos\theta_k, \qquad k = 1, 2, \ldots, n \tag{13.14}$$

Each conjugate pole pair yields a real quadratic factor of the form

$$(s - p_k)(s - p_k^*) = s^2 + 2sR\,\sin\theta_k + R^2 \tag{13.15}$$

For odd n, there is always one real pole located at $s = -R$. The factored form for $H_P(s)$ (which is useful for realizing higher-order filters by cascading individual sections) may then be written

$$H_P(s) = \frac{K}{Q_P(s)} \qquad Q_P(s) = (s+R)^M \prod_{k=1}^{\text{int}(n/2)} (s^2 + 2sR\,\sin\theta_k + R^2), \quad \begin{cases} M = 0, & n \text{ even} \\ M = 1, & n \text{ odd} \end{cases} \tag{13.16}$$

For unit peak gain, we choose $K = Q_P(0)$. We can also express the denominator $Q_P(s)$ in polynomial (unfactored) form as

$$Q_P(s) = q_0 s^n + q_1 s^{n-1} + q_2 s^{n-2} + \cdots + q_{n-1}s + q_n \tag{13.17}$$

To find q_k, we use the recursion relation

$$q_0 = 1 \qquad\qquad q_k = \frac{R\cos[(k-1)\pi/2n]}{\sin(k\pi/2n)} q_{k-1} \tag{13.18}$$

REVIEW PANEL 13.8

Poles p_k and Transfer Function $H_P(s)$ of a Butterworth Lowpass Prototype

The LHP poles p_k lie on a *circle* of radius $R = \nu_3 = (1/n)^{1/n}$ at angles θ_k with respect to the $j\omega$-axis. No poles lie on the $j\omega$-axis. For odd n, one pole is on the real axis at $s = -R$.

$p_k = -R\sin\theta_k + jR\cos\theta_k$ where $\theta_k = \dfrac{(2k-1)\pi}{2n}$, $k = 1, 2, \ldots, n$ (with respect to the $j\omega$-axis)

$H_P(s) = \dfrac{K}{Q_P(s)}$ $\qquad Q_P(s) = \displaystyle\prod_{k=1}^{n}(s - p_k)$ $\qquad K = Q_P(0) = \displaystyle\prod_{k=1}^{n} p_k$ (for unit peak or dc gain)

PGR2026-00033
Krisp Technologies EX1024 Page 419

EXAMPLE 13.2 (Butterworth Filter Computations)

(a) What is the minimum order of a Butterworth lowpass filter that shows an attenuation of no more than 2 dB at 5 rad/s and at least 10 dB at 8 rad/s, 20 dB at 10 rad/s, and 40 dB at 30 rad/s?

Assuming a passband edge of 5 rad/s, we find $\epsilon^2 = 10^{0.1A_p} - 1 = 0.5849$.

Next, we compute the order from $n = \dfrac{\log[(10^{0.1A_x} - 1)/\epsilon^2]^{1/2}}{\log \nu_x}$ for each attenuation to get

$$A_1 = 10 \text{ dB} \qquad \nu = 8/5 = 1.6 \qquad n = 2.91$$
$$A_2 = 20 \text{ dB} \qquad \nu = 10/5 = 2 \qquad n = 3.7$$
$$A_3 = 40 \text{ dB} \qquad \nu = 30/5 = 6 \qquad n = 2.72$$

The largest value of n is rounded up to give the filter order as $n = 4$.

(b) Find the attenuation at $f = 7.5$ kHz of a fourth-order Butterworth filter whose 1 dB passband edge is located at 3 kHz.

We find $\epsilon^2 = 10^{0.1A_p} - 1 = 0.2589$. Since $f = 7.5$ kHz corresponds to a lowpass prototype (normalized) frequency of $\nu = \frac{7.5}{3} = 2.5$ rad/s, and $n = 4$, we find

$$A = 10 \log(1 + \epsilon^2 \nu^{2n}) = 10 \log(1 + 0.2589\nu^{10}) = 25.98 \text{ dB}$$

(c) Find the pole radius and orientations of a third-order Butterworth lowpass filter with a passband edge of 100 rad/s and $\epsilon = 0.7$.

The lowpass prototype poles lie on a circle of radius $R = \nu_3 = (1/\epsilon)^{1/n} = (1/0.7)^{1/3} = 1.1262$. The pole radius of the actual filter is $100R = 112.62$ rad/s. The left half-plane pole orientations are $\theta_k = \frac{(2k-1)\pi}{2n} = \frac{(2k-1)\pi}{6}$, $k = 1, 2, 3$, or $30°$, $90°$, and $150°$, with respect to the $j\omega$-axis.

(d) (Design from a Specified Order) What is the transfer function $H(s)$ of a third-order Butterworth lowpass prototype with $\epsilon = 0.7$ and a peak gain of 10?

The lowpass prototype poles lie on a circle of radius $R = \nu_3 = (1/\epsilon)^{1/n} = (1/0.7)^{1/3} = 1.1262$. The left half-plane pole orientations are $[\frac{\pi}{6}, \frac{\pi}{2}, \frac{5\pi}{6}]$ rad. The pole locations are thus $s = -1.1262$ and $s = -1.1262 \sin(\pi/6) \pm 1.1262 \cos(\pi/6) = -0.5651 \pm j0.9754$. The denominator of $H(s)$ is thus

$$Q(s) = (s + 0.5651 + j0.9754)(s + 0.5651 - j0.9754)(s + 1.1262) = s^3 + 2.2525s^2 + 2.5369s + 1.4286$$

For a peak gain of 10, the numerator K of $H(s) = K/Q(s)$ is $K = 10Q(0) = 14.286$, and thus

$$H(s) = \frac{K}{Q(s)} = \frac{14.286}{s^3 + 2.2525s^2 + 2.5369s + 1.4286}$$

(e) (Design from High-Frequency Decay Rate) Find the transfer function of a 3-dB Butterworth lowpass filter with a passband edge of 50 Hz and a high-frequency decay rate of 40 dB/dec.

For an nth-order filter, the high-frequency decay rate is $20n$ dB/dec. Thus, $n = 2$. The pole radius of a 3-dB lowpass prototype is $R = 1$. With $\theta_k = [\pi/4, 3\pi/4]$ rad, the pole locations are described by

PGR2026-00033
Krisp Technologies EX1024 Page 420

$s = -\sin(\pi/4) \pm j\cos(\pi/4) = -0.707 \pm j0.707$. This gives $Q_P(s) = s^2 + \sqrt{2} + 1$ and $H_P(s) = 1/Q_P(s)$. With $\omega_p = 2\pi f_p = 100\pi$ rad/s, the LP2LP transformation gives the required transfer function

$$H(s) = H_P(s/100\pi) = \frac{9.8696(10)^4}{s^2 + 444.29 + 9.8696(10)^4}$$

(f) (Design from a Specified Order) Design a fourth-order Butterworth bandpass filter with a 2-dB passband of 200 Hz and a center frequency of $f_0 = 1$ kHz.

To design the filter, we start with a second-order lowpass prototype. We find $\epsilon^2 = 10^{0.1A_p} - 1 = 0.5849$. With $n = 2$, the pole radius is $R = (1/\epsilon)^{1/n} = 1.1435$.

The pole locations are $s = -R\sin(\pi/4) \pm jR\cos(\pi/4) = -0.8086 \pm j0.8086$.

The prototype denominator is

$$Q_P(s) = (s + 0.8086 + j0.8086)(s + 0.8086 + j0.8086) = s^2 + 1.6171s + 1.3076$$

The prototype transfer function is

$$H_P(s) = \frac{1.3076}{s^2 + 1.6171s + 1.3076}$$

We now use the LP2BP transformation $s \to (s^2 + \omega_0^2)/Bs$, with $\omega_0 = 2\pi f_0 = 2000\pi$ rad/s and $B = 2\pi(200) = 400\pi$ rad/s, to obtain the required transfer function:

$$H_{\text{BP}}(s) = \frac{2.0648(10)^6 s^2}{s^4 + 2.0321(10)^3 s^3 + 8.1022(10)^7 s^2 + 8.0226(10)^{10} s + 1.5585(10)^{15}}$$

Exactly Meeting the Stopband Attenuation

The lowpass prototype $H_P(s)$ shows an attenuation of A_p at the passband edge $\nu_p = 1$ rad/s. Since the filter exceeds specifications everywhere else, the exact value of A_s occurs at a frequency $\tilde{\nu}_s$ that is smaller than the design frequency ν_s. To *exactly* match A_s to the stopband edge ν_s, we must frequency scale (stretch) $H_P(s)$ to $H(s/\beta)$, where the factor β is given by

$$\beta = \frac{\text{specified stopband edge}}{\text{actual stopband edge}} = \frac{\nu_s}{\tilde{\nu}_s} \tag{13.19}$$

The frequency $\tilde{\nu}_s$ is found from the attenuation expression for $A(\nu)$, with $\nu = \tilde{\nu}_s$, as

$$\tilde{\nu}_s = \left(\frac{10^{0.1A_s} - 1}{\epsilon^2}\right)^{1/2n} \tag{13.20}$$

The new filter $H_S(s) = H_P(s/\beta)$ will show an exact match to A_s dB at ν_s. At the passband edge, however, it will exceed specifications and its attenuation will be *less than* A_p.

13.3.3 Butterworth Prototypes from 3-dB Filters

If the passband attenuation is $A_p = 3$ dB, then $\epsilon^2 = 1$ and $R = 1$. The poles of a Butterworth 3-dB filter thus lie on a unit circle. The coefficients of the denominator polynomial $Q_3(s)$ in the filter transfer function $H_3(s) = K/Q_3(s)$ show symmetry about their midpoint with $q_k = q_{n-k},\ k = 0, 1, 2, \ldots, n$. Consequently, only about half the coefficients need to be computed to establish $Q_3(s)$. Because of these simplifications, it is common practice to start with a 3-dB Butterworth filter $H_3(s)$ and use frequency scaling ($s \to s/R$) to design the lowpass prototype $H_P(s) = H_3(s/R)$. In fact, this scaling can also be applied as the final step, in combination with any other frequency transformations, to obtain the final filter. Table 13.1 lists both the factored and polynomial forms of $Q_3(s)$ for 3-dB Butterworth lowpass prototypes.

PGR2026-00033
Krisp Technologies EX1024 Page 421

Table 13.1 3-dB Butterworth Lowpass Prototype Transfer Functions

Order n	**Denominator $Q(s)$ in Polynomial Form**
1	$1 + s$
2	$1 + \sqrt{2}s + s^2$
3	$1 + 2s + 2s^2 + s^3$
4	$1 + 2.613s + 3.414s^2 + 2.613s^3 + s^4$
5	$1 + 3.236s + 5.236s^2 + 5.236s^3 + 3.236s^4 + s^5$
6	$1 + 3.864s + 7.464s^2 + 9.141s^3 + 7.464s^4 + 3.864s^5 + s^6$

Order n	**Denominator $Q(s)$ in Factored Form**
1	$1 + s$
2	$1 + \sqrt{2}s + s^2$
3	$(1 + s)(1 + s + s^2)$
4	$(1 + 0.76536s + s^2)(1 + 1.84776s + s^2)$
5	$(1 + s)(1 + 0.6180s + s^2)(1 + 1.6180s + s^2)$
6	$(1 + 0.5176s + s^2)(1 + \sqrt{2}s + s^2)(1 + 1.9318s + s^2)$

EXAMPLE 13.3 (Butterworth Filter Design)

(a) A Butterworth lowpass filter is to meet the following specifications: $A_p = 1$ dB for $\omega \leq 4$ rad/s and $A_s \geq 20$ dB for $\omega \geq 8$ rad/s. Find its transfer function $H(s)$.

The lowpass prototype specifications are $\nu_p = 1$ rad/s and $\nu_s = \omega_s/\omega_p = 2$ rad/s. We compute

$$\epsilon^2 = 10^{0.1A_p} - 1 = 0.2589 \qquad \frac{\log[(10^{0.1A_s} - 1)/\epsilon^2]^{1/2}}{\log \nu_s} = 4.289 \Rightarrow n = 5$$

The pole radius is given by $R = (1/\epsilon)^{1/n} = 1.1447$.

The pole orientations are $\theta_k = (2k - 1)\pi/2n = (2k - 1)18°, k = 1, 2, \ldots, 5$. With $H_P(s) = K/Q_P(s)$, where $Q_P(s) = q_0 s^5 + q_1 s^4 + q_2 s^3 + q_3 s^2 + q_4 s + q_5$, we use the recursion

$$q_0 = 1 \qquad q_k = \frac{R\cos[(k-1)\pi/2n]}{\sin(k\pi/2n)} = \frac{R\cos[(k-1)\pi/10]}{\sin(k\pi/10)} q_{k-1}, \quad k = 1, 2, 3, 4, 5$$

to find the coefficients of $Q_P(s)$. This gives

$$Q_P(s) = s^5 + 3.704s^4 + 6.861s^3 + 7.853s^2 + 5.556s + 1.965$$

PGR2026-00033
Krisp Technologies EX1024 Page 422

Choosing $K = 1.965$ for unit gain, the prototype transfer function becomes

$$H_P(s) = \frac{1.965}{s^5 + 3.704s^4 + 6.861s^3 + 7.853s^2 + 5.556s + 1.965}$$

We obtain the required filter as $H(s) = H_P(s/4)$ using the LP2LP transformation $s \to s/4$ to get

$$H(s) = H_P(s/4) = \frac{2012.4}{s^5 + 14.82s^4 + 109.8s^3 + 502.6s^2 + 1422.3s + 2012.4}$$

Comment: Had we been asked to exactly meet the attenuation at the stopband edge, we would first compute $\tilde{\nu}_s$ and the stretching factor $\beta = \nu_s/\tilde{\nu}_s$ as

$$\tilde{\nu}_s = \left(\frac{10^{0.1A_s} - 1}{\epsilon^2}\right)^{1/2n} = 1.5833 \qquad \beta = \nu_s/\tilde{\nu}_s = 2/1.8124 = 1.1035$$

We would need to frequency scale $H_P(s)$ using $s \to s/\beta$ and then denormalize using $s \to s/4$. We can combine these operations and scale $H_P(s)$ to $H_2(s) = H_P(s/4\beta) = H_P(s/4.4141)$ to match the attenuation at the stopband edge. We get

$$H_2(s) = H_P(s/4.4141) = \frac{3293.31}{s^5 + 16.35s^4 + 133.68s^3 + 675.44.05s^2 + 2109.23s + 3293.31}$$

The linear and decibel magnitude of the two filters is sketched in Figure E13.3. The passband gain is 0.89 (corresponding to $A_p = 1$ dB). The stopband gain is 0.1 (corresponding to $A_s = 20$ dB). Note how the design attenuation is exactly met at the passband edge for the original filter $H(s)$ (shown light), and at the stopband edge for the redesigned filter $H_2(s)$ (shown dark).

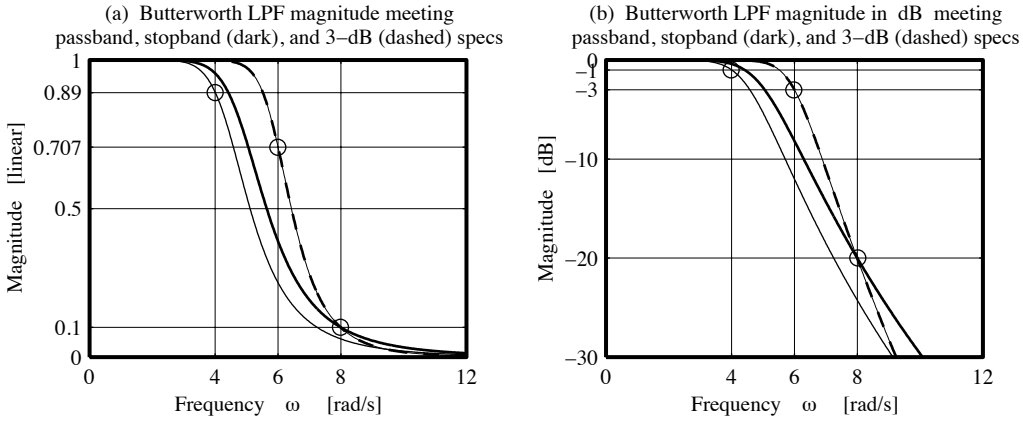


Figure E13.3 Butterworth lowpass filters for Example 13.3(a and b)

(b) (A Design with Additional Constraints) Consider the design of a Butterworth lowpass filter to meet the following specifications: $A_p \leq 1$ dB for $\omega \leq 4$ rad/s, $A_s \geq 20$ dB for $\omega \geq 8$ rad/s, and a half-power frequency of $\omega_3 = 6$ rad/s.

Since we must exactly meet specifications at the half-power frequency ω_3, it is convenient to select the cutoff frequency as $\nu_3 = 1$ rad/s and the prototype band edges as $\nu_p = \omega_p/\omega_3 = 4/6$ rad/s and

PGR2026-00033
Krisp Technologies EX1024 Page 423

$\nu_s = \omega_s/\omega_3 = 8/6$ rad/s. Since $\nu_3 = 1$, we have $\epsilon^2 = 1$. To find the filter order, we solve the attenuation relation for n at both $\nu_p = \omega_p/\omega_3$ and $\nu_s = \omega_s/\omega_3$ and choose n as the larger value. Thus,

$$\begin{aligned} A_p &= 10\log(1+\nu_p^{2n}) \qquad n = 0.5\log(10^{0.1A_p}-1)/\log(\nu_p) = 1.6663 \\ A_s &= 10\log(1+\nu_s^{2n}) \qquad n = 0.5\log(10^{0.1A_s}-1)/\log(\nu_s) = 7.9865 \end{aligned}$$

Thus, $n = 8$. The prototype transfer function of the 3-dB filter is found to be

$$H_P(s) = \frac{1}{s^8 + 5.126s^7 + 13.138s^6 + 21.848s^5 + 25.691s^4 + 21.848s^3 + 13.138s^2 + 5.126s + 1}$$

The actual filter transfer function is $H(s) = H_P(s/\omega_3) = H_P(s/6)$. We find

$$H(s) = \frac{1679616}{s^8 + 30.75s^7 + 473s^6 + 4718s^5 + 33290s^4 + 169876s^3 + 612923s^2 + 1434905s + 1679616}$$

Its linear and decibel magnitude, shown dashed in Figure E13.3, exactly meets the half-power (3-dB) frequency requirement, and exceeds the attenuation specifications at the other band edges.

(c) (A Bandstop Filter) Design a Butterworth bandstop filter with 2-dB passband edges of 30 Hz and 100 Hz, and 40-dB stopband edges of 50 Hz and 70 Hz.

The band edges are $[f_1,\ f_2,\ f_3,\ f_4] = [30,\ 50,\ 70,\ 100]$ Hz. Since $f_1 f_4 = 3000$ and $f_2 f_3 = 3500$, the specifications are not geometrically symmetric. Assuming a fixed passband, we relocate the upper stopband edge f_3 to ensure geometric symmetry $f_2 f_3 = f_1 f_4$. This gives $f_3 = (30)(100)/50 = 60$ Hz.

The lowpass prototype band edges are $\nu_p = 1$ rad/s and $\nu_s = \dfrac{f_4 - f_1}{f_3 - f_2} = 7$ rad/s.

We compute $\epsilon^2 = 10^{0.1A_p} - 1 = 0.5849$ and the lowpass prototype order as

$$n = \frac{\log[(10^{0.1A_s} - 1)/\epsilon^2]^{1/2}}{\log \nu_s} \Rightarrow n = 3$$

The pole radius is $R = (1/\epsilon)^{1/n} = 1.0935$. The pole angles are $\theta_k = [\frac{\pi}{6},\ \frac{\pi}{2},\ \frac{5\pi}{6}]$ rad.

The pole locations are $s_k = -R\sin\theta_k + jR\cos\theta_k$, and lead to the lowpass prototype

$$H_P(s) = \frac{1.3076}{s^3 + 2.1870s^2 + 2.3915s + 1.3076}$$

With $\omega_0^2 = \omega_1\omega_4 = 4\pi^2(3000)$ and $B = 2\pi(f_4 - f_1) = 2\pi(70)$ rad/s, the LP2BS transformation $s \to sB/(s^2 + \omega_0^2)$ gives

$$H(s) = \frac{s^6 + 3.55(10)^5 s^4 + 4.21(10)^{10} s^2 + 1.66(10)^{15}}{s^6 + 8.04(10)^2 s^5 + 6.79(10)^5 s^4 + 2.56(10)^8 s^3 + 8.04(10)^{10} s^2 + 1.13(10)^{13} s + 1.66(10)^{15}}$$

The linear and decibel magnitude of this filter is sketched in Figure E13.3C.

PGR2026-00033
Krisp Technologies EX1024 Page 424

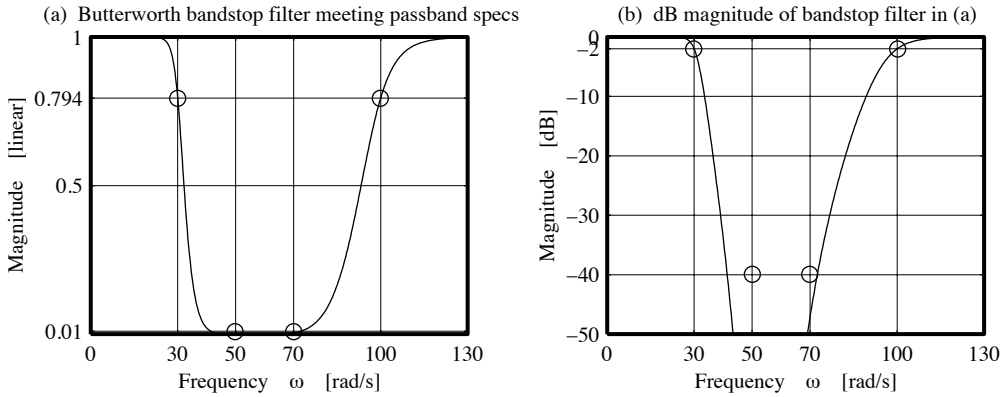


Figure E13.3C Butterworth bandstop filter of Example 13.3(c)

13.3.4 Phase and Group Delay of Butterworth Filters

The phase of Butterworth filters is best evaluated from $H_P(s)$ in factored form. The contribution due to a first-order section (linear factor) is simply

$$H_P(s) = \frac{1}{s+R} \qquad H_P(\nu) = \frac{1}{R+j\nu} \qquad \phi(\nu) = -\tan^{-1}(\nu/R) \tag{13.21}$$

For a second-order section (quadratic factor), conversion from $H_P(s)$ to $H_P(\nu)$ leads to

$$H_P(s) = \frac{1}{s^2 + 2sR\sin\theta_k + R^2} \tag{13.22}$$

$$H_P(\nu) = \frac{1}{[1-(\nu/R)^2) + j2(\nu/R)\sin\theta_k]} \qquad \phi_P(\nu) = -\tan^{-1}\left[\frac{2(\nu/R)\sin\theta_k}{1-(\nu/R)^2}\right] \tag{13.23}$$

The total phase $\phi(\nu)$ equals the sum of the contributions due to each of the individual sections. The group delay $t_g = -d\phi/d\nu$ may be found by differentiating the phase and leads to the result

$$t_g(\nu) = \frac{1}{1+(\nu/R)^{2n}} \sum_{k=1}^{n} \frac{(\nu/R)^{2(k-1)}}{\sin\theta_k} \tag{13.24}$$

The phase approaches $-n\pi/2$ radians at high frequencies. The group delay is fairly constant for small frequencies but peaks near the passband edge. As the filter order n increases, the delay becomes less flat, and the peaking is more pronounced. The phase or group delay of other filter types may be found using the appropriate frequency transformations.

REVIEW PANEL 13.9

The Group Delay of a Butterworth Prototype Peaks Near the Passband Edge

$$t_g(\nu) = \frac{1}{1+(\nu/R)^{2n}} \sum_{k=1}^{n} \frac{(\nu/R)^{2(k-1)}}{\sin\theta_k} \qquad \text{where} \qquad \theta_k = \frac{(2k-1)\pi}{2n} \quad \text{and} \quad R = \nu_3 = (1/\epsilon)^{1/n}$$

PGR2026-00033
Krisp Technologies EX1024 Page 425

13.4 The Chebyshev Approximation

The Chebyshev approximation for lowpass prototypes is also based on a polynomial fit. Instead of selecting the coefficients (degrees of freedom) of the polynomial $L_n(\nu)$ to make the response maximally flat at the origin alone, we now try to make the response uniformly good by ensuring an exact match at several frequencies spread out over the passband and minimizing the error at all other frequencies in the passband. Such an approximation is called a **uniform approximation**. A uniform approximation requires the maxima and minima of the error to have equal magnitudes or be **equiripple**. The number of points where the error is zero corresponds to the order of the approximating polynomial (or filter order). Note that least squares methods, which only minimize the mean squared error, provide no guarantee that a small mean squared error will also result in a small absolute error *throughout the approximation interval*. The equiripple polynomial with the smallest maximum absolute error is called a **minmax polynomial**. Although minmax approximations usually require cumbersome iterative methods, the good news is that Chebyshev polynomials are themselves minmax, and approximations based on Chebyshev polynomials are quite easily obtained.

13.4.1 The Chebyshev Polynomials

The nth-order **Chebyshev polynomial** (*of the first kind*) is described by a relation that looks trigonometric in nature

$$T_n(x) = \cos(n \cos^{-1} x) \tag{13.25}$$

Here, $T_n(x)$ is actually a *polynomial* of degree n. The polynomial form shows up if we use the transformation $x = \cos\theta$ to give $T_n(x) = T_n(\cos\theta) = \cos(n\theta)$. Using trigonometric identities for $\cos(n\theta)$, we easily obtain the first few Chebyshev polynomials as

$$\begin{array}{llll} \cos 0 = 1 & \cos\theta = \cos\theta & \cos 2\theta = 2\cos^2\theta - 1 & \cos 3\theta = 4\cos^3\theta - 3\cos\theta \\ T_0(x) = 1 & T_1(x) = x & T_2(x) = 2x^2 - 1 & T_3(x) = 4x^3 - 3x \end{array}$$

The Chebyshev polynomials can also be obtained from the recursion relation

$$T_0(x) = 1 \qquad T_1(x) \qquad\qquad T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x), \quad n > 1 \tag{13.26}$$

For example, $T_4(x) = 2xT_3(x) - T_2(x) = 8x^4 - 8x^2 + 1$.

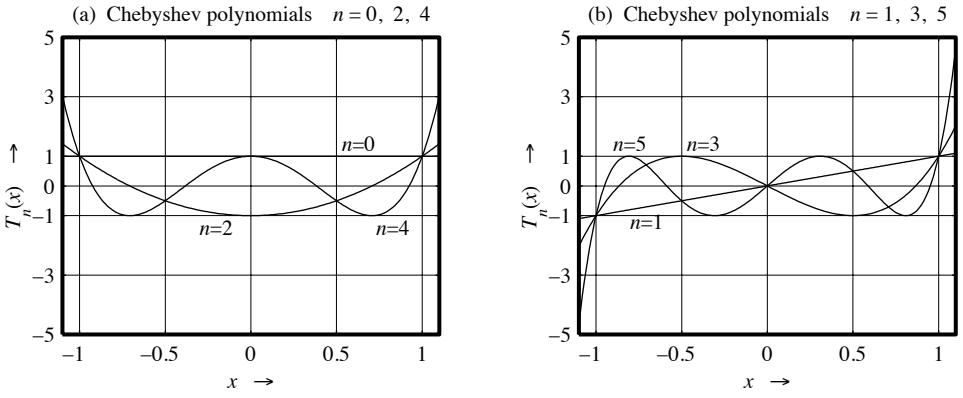


Figure 13.9 Chebyshev polynomials

PGR2026-00033
Krisp Technologies EX1024 Page 426

Figure 13.9 illustrates the characteristics of Chebyshev polynomials. The leading coefficient of an nth-order Chebyshev polynomial always equals 2^{n-1}. The constant coefficient is always ± 1 for even n and 0 for odd n. The nonzero coefficients alternate in sign and point to the oscillatory nature of the polynomial. Chebyshev polynomials possess two remarkable characteristics in the context of both polynomial approximation and filter design. First, $T_n(x)$ oscillates about zero in the interval $(-1,\ 1)$, with $n+1$ maxima and minima (that equal ± 1). Outside the interval $(-1,\ 1)$, $|T_n(x)|$ shows a very steep increase. Over $(-1,\ 1)$, the *normalized Chebyshev polynomial* $\tilde{T}_n(x) = T_n(x)/2^{n-1}$ has the smallest *maximum* absolute value (which equals $1/2^{n-1}$) among all normalized nth-degree polynomials. This is the celebrated **Chebyshev theorem**.

13.4.2 The Chebyshev Lowpass Prototype

The nth-order Chebyshev lowpass prototype is described by

$$|H(\nu)|^2 = \frac{1}{1+\epsilon^2 T_n^2(\nu)} \qquad A_{\text{dB}}(\nu) = 10\log[1+\epsilon^2 T_n^2(\nu)] \quad \text{dB} \tag{13.27}$$

Here, ϵ is a measure of the passband **ripple**. The magnitude of the passband maxima and minima is governed only by ϵ^2, but a larger order n yields more of them and results in a sharper transition due to the higher polynomial order. Since $T_n(1) = 1$, the magnitude at the passband edge ($\nu_p = 1$ rad/s) is

$$|H(\nu_p)| = \frac{1}{\sqrt{1+\epsilon^2}} \tag{13.28}$$

Since $T_n(0) = 1$ for even n and $T_n(0) = 0$ for odd n, the dc gain $|H(0)|$ is

$$|H(0)| = \begin{cases} \dfrac{1}{\sqrt{1+\epsilon^2}}, & n \text{ even} \\ 1, & n \text{ odd} \end{cases} \tag{13.29}$$

The *peak* gain, however, equals 1 for both odd n and even n. The magnitude spectrum for $n = 2$ and $n = 3$ is illustrated in Figure 13.10.

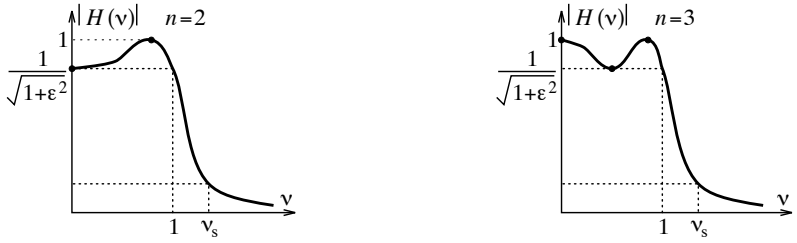


Figure 13.10 Gain of Chebyshev filters of odd and even order

The filter order can be found by counting the number of maxima and minima in the passband (excluding the passband edge). Since $|T_n(\nu)|$ increases monotonically (and rapidly) for $\nu > 1$, the gain decreases rapidly outside the passband. The attenuation rate at high frequencies is $20n$ dB/dec. Since $T_n(\nu) \approx 2^{n-1}\nu^n$ for large ν, the attenuation may be approximated by

$$A(\nu) = 10\log[1+\epsilon^2 T_n^2(\nu)] \approx 20\log \epsilon T_n(\nu) \approx 20\log \epsilon 2^{n-1}\nu^n = 20\log\epsilon + 20(n-1)\log 2 + 20n\log\nu \tag{13.30}$$

The Chebyshev filter provides an additional $6(n-1)$ dB of attenuation in the stopband compared with the Butterworth filter, and this results in a much sharper transition for the same filter order. This improvement is made at the expense of introducing ripples in the passband, however.

PGR2026-00033
Krisp Technologies EX1024 Page 427

REVIEW PANEL 13.10

The Magnitude Spectrum of a Chebyshev I Filter Has an Equiripple Passband

The number of passband peaks in the *two-sided* spectrum equals the filter order n.

$$|H(\nu)|^2 = \frac{1}{1+\epsilon^2 T_n^2(\nu)} \qquad H_{\text{dB}} = -10\log[1+\epsilon^2 T_n^2(\nu)]$$

$$|H(0)| = \begin{cases} 1, & n \text{ odd} \\ \dfrac{1}{\sqrt{1+\epsilon^2}}, & n \text{ even} \end{cases} \qquad |H(1)| = \frac{1}{\sqrt{1+\epsilon^2}}$$

13.4.3 Chebyshev Prototype Design

Chebyshev prototype design starts with the attenuation relation and requires only ϵ and the filter order n. We find ϵ^2 by evaluating the attenuation $A_{\text{dB}}(\nu)$ at the passband edge ($\nu = \nu_p = 1$ rad/s), where $T_n(\nu) = 1$:

$$A_p = A_{\text{dB}}(1) = 10\log(1+\epsilon^2) \qquad \epsilon^2 = 10^{0.1A_p} - 1 \tag{13.31}$$

This result is identical to the corresponding Butterworth result. We find the filter order n by evaluating $A_{\text{dB}}(\nu)$ at the stopband edge ($\nu = \nu_s$ rad/s) to give

$$A_s = A_{\text{dB}}(\nu_s) = 10\log(1+\epsilon^2 T_n^2(\nu)) \qquad n = \frac{\cosh^{-1}[(10^{0.1A_s}-1)/\epsilon^2]^{1/2}}{\cosh^{-1}\nu_s} \tag{13.32}$$

The filter order must be *rounded up* to an integer value. Note that for $\nu_s > 1$, the computation of $T_n(\nu_s)$ requires the hyperbolic form, $T_n(\nu_s) = \cosh(n\cosh^{-1}\nu_s)$, as used in the above relation.

The half-power frequency ν_3 at which $|H(\nu)| = 0.707$ or $\epsilon^2 T_n^2(\nu) = 1$ can be evaluated in terms of the filter order n, using the hyperbolic form for $T_n(\nu)$:

$$\nu_3 = \cosh\left[\frac{\cosh^{-1}(1/\epsilon)}{n}\right], \qquad \epsilon \le 1 \tag{13.33}$$

We can also express the filter order n in terms of the half-power frequency ν_3 as

$$n = \cosh^{-1}\left[\frac{1}{\epsilon\cosh^{-1}\nu_3}\right] \tag{13.34}$$

REVIEW PANEL 13.11

Ripple (ϵ), Order (n), and 3-dB Frequency (ν_3) of the Chebyshev I Prototype

$$\epsilon^2 = 10^{0.1A_p} - 1 \qquad n = \frac{\cosh^{-1}[(10^{0.1A_s}-1)/\epsilon^2]^{1/2}}{\cosh^{-1}\nu_s} \qquad \nu_3 = \cosh\left[\frac{1}{n}\cosh^{-1}\left(\frac{1}{\epsilon}\right)\right]$$

13.4.4 The Chebyshev Prototype Transfer Function

To find the prototype transfer function $H_P(s)$, we substitute $\nu^2 = -s^2$ into $|H(\nu)|^2$ and obtain

$$|H(\nu)|^2_{\nu^2=-s^2} = \frac{1}{1+\epsilon^2 T_n^2(s/j)} \tag{13.35}$$

Its poles are found from

$$1+\epsilon^2 T_n^2(s/j) = 0 \quad \text{or} \quad T_n(s/j) = \cos[n\cos^{-1}(s/j)] = \pm j/\epsilon \tag{13.36}$$

It is convenient to define a new complex variable z in the form

$$z = \theta + j\alpha = \cos^{-1}(s/j) \quad \text{or} \quad s = j\cos z \tag{13.37}$$

PGR2026-00033
Krisp Technologies EX1024 Page 428

This results in

$$\cos(nz) = \cos(n\theta + jn\alpha) = \cos(n\theta)\cosh(n\alpha) - j\sin(n\theta)\sinh(n\alpha) = \pm j/\epsilon \tag{13.38}$$

Equating the real and imaginary parts, we obtain the two relations

$$\cos(n\theta)\cosh(n\alpha) = 0 \qquad \sin(n\theta)\sinh(n\alpha) = \pm\frac{1}{\epsilon} \tag{13.39}$$

Since $\cosh(n\alpha) \geq 1$ for all $n\alpha$, the first relation gives

$$\cos(n\theta) = 0 \qquad \text{or} \qquad \theta_k = \frac{(2k-1)\pi}{2n}, \qquad k = 1, 2, \ldots, 2n \tag{13.40}$$

Note that θ_k (*in radians*) has the same expression as the Butterworth poles. For $\theta > 0,\ \sin(n\theta) = 1$, we get

$$\sinh(n\alpha) = \frac{1}{\epsilon} \qquad \text{or} \qquad \alpha = \frac{1}{n\,\sinh^{-1}(1/\epsilon)} \tag{13.41}$$

Since $s = j\cos z$, we have

$$s = j\cos z = j\cos(\theta_k + j\alpha) = \sin\theta_k \sinh\alpha + j\cos\theta_k \cosh\alpha \tag{13.42}$$

The poles of $H_P(s)$ are just the left half-plane roots (with negative real parts):

$$p_k = \sigma_k + j\omega_k = -\sin\theta_k \sinh\alpha + j\cos\theta_k \cosh\alpha, \qquad k = 1, 2, \ldots, n \tag{13.43}$$

The real and imaginary parts also satisfy

$$\frac{\sigma_k^2}{\sinh^2\alpha} + \frac{\omega_k^2}{\cosh^2\alpha} = 1 \tag{13.44}$$

This suggests that *the Chebyshev poles lie on an ellipse* in the s-plane, with a major semi-axis (along the $j\omega$-axis) that equals $\cosh\alpha$ and a minor semi-axis (along the σ-axis) that equals $\sinh\alpha$. The pole locations for $n = 2$ and $n = 3$ are illustrated in Figure 13.11.

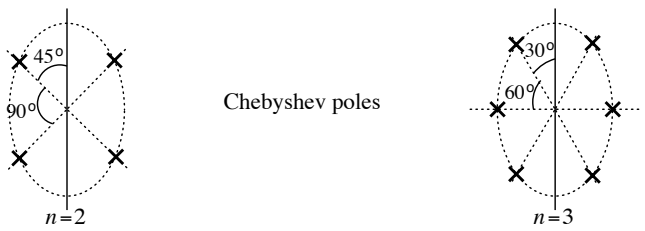


Figure 13.11 Pole locations of Chebyshev lowpass prototypes

Comparison with the 3-dB Butterworth prototype poles $p_B = -\sin\theta_k + j\cos\theta_k$ reveals that the real part of the Chebyshev pole is scaled by $\sinh\alpha$, and the imaginary part by $\cosh\alpha$. The Chebyshev poles may thus be found directly from the orientations of the Butterworth poles (for the same order) that lie on a circle and contracting the circle to an ellipse. This leads to the geometric evaluation illustrated in Figure 13.12.

We draw concentric circles of radii $\cosh\alpha$ and $\sinh\alpha$ with radial lines along θ_k, the angular orientation of the Butterworth poles. On each radial line, label the point of intersection with the larger circle (of radius

PGR2026-00033
Krisp Technologies EX1024 Page 429

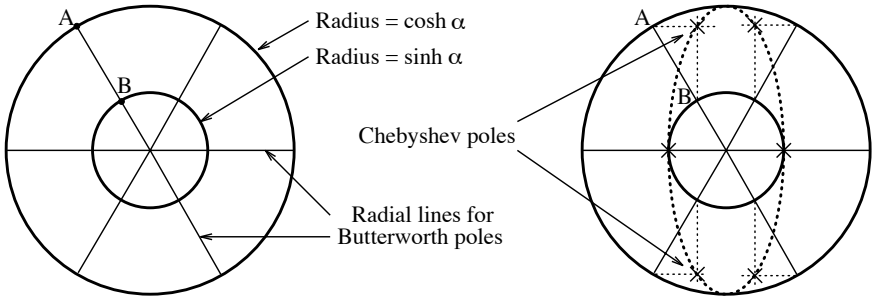


Figure 13.12 Locating Chebyshev poles from Butterworth poles

$\cosh\alpha$) as A, and with the smaller circle as B. Draw a horizontal line through A and a vertical line through B. Their intersection represents a Chebyshev pole location.

The lowpass prototype transfer function $H_P(s) = K/Q_P(s)$ may be readily obtained in factored form by expressing the denominator $Q_P(s)$ as

$$Q_P(s) = (s - p_1)(s - p_2)\cdots(s - p_k) \tag{13.45}$$

Each conjugate pole pair will yield a quadratic factor. For odd n, $Q_P(s)$ will also contain the linear factor $s + \sinh\alpha$.

For unit dc gain, we choose $K = Q_P(0)$ for any filter order. For unit peak, however, we choose

$$K = \begin{cases} Q_P(0), & n \text{ odd} \\ \dfrac{Q_P(0)}{\sqrt{1+\epsilon^2}}, & n \text{ even} \end{cases} \tag{13.46}$$

REVIEW PANEL 13.12

Poles and Transfer Function of the Chebyshev I Prototype

The poles p_k lie on an *ellipse*.

No poles lie on the $j\omega$-axis. For odd n, one pole is on the real axis at $s = -\sinh\alpha$.

$$p_k = -\sin\theta_k \sinh\alpha + j\cos\theta_k \cosh\alpha \qquad \theta_k = \frac{(2k-1)\pi}{2n}, \quad k = 1, 2, \ldots, n \qquad \alpha = \frac{1}{n}\sinh^{-1}\left(\frac{1}{\epsilon}\right)$$

$$H(s) = \frac{K}{Q_P(s)} \qquad Q_P(s) = \prod_{k=1}^{n}(s - p_k) \qquad K = \begin{cases} Q_P(0)/\sqrt{1+\epsilon^2}, & n \text{ even} \\ Q_P(0), & n \text{ odd} \end{cases} \quad \text{(for unit } \textit{peak} \text{ gain)}$$

EXAMPLE 13.4 (Chebyshev Filter Computations)

(a) The magnitude of a Chebyshev lowpass prototype is shown in Figure E13.4A. What is $|H(\nu)|^2$?

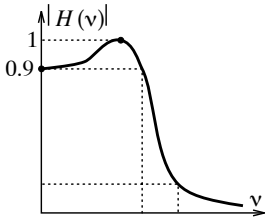


Figure E13.4A Chebyshev magnitude spectrum for Example 13.4(a)

PGR2026-00033
Krisp Technologies EX1024 Page 430

The filter order is $n = 2$ because we see a total of two maxima and minima in the passband. Since $T_2(0) = 0$ and $|H(0)| = 0.9$, we find $|H(0)|^2 = 1/(1+\epsilon^2) = (0.9)^2$. This gives $\epsilon^2 = 0.2346$. Since $T_2(\nu) = 2\nu^2 - 1$, we have $T_2^2(\nu) = 4\nu^4 - 4\nu^2 + 1$. Thus,

$$|H(\nu)|^2 = \frac{1}{1+\epsilon^2 T_2^2(\nu)} = \frac{1}{1+\epsilon^2(4\nu^4 - 4\nu^2 + 1)} = \frac{1}{1.2346 - 0.9383\nu^2 + 0.9383\nu^4}$$

(b) What is the attenuation at $f = 1.5$ kHz of a second-order Chebyshev lowpass filter with a 1-dB passband of 500 Hz?

With $A_p = 1$ dB, we find $\epsilon^2 = 10^{0.1A_p} - 1 = 0.2589$. The prototype (normalized) frequency corresponding to $f = 1.5$ kHz is $\nu = 3$ rad/s, and the attenuation at $\nu = 3$ rad/s is found from

$$A = 10\log[1 + \epsilon^2 T_n^2(\nu)] = 10\log[1 + 0.2589T_2^2(\nu)] \text{ dB}$$

With $T_2(\nu) = 2\nu^2 - 1$, we compute $A = 18.8$ dB.

(c) Find the half-power frequency of a fifth-order Chebyshev lowpass filter with a 2-dB passband edge at 1 kHz.

With $A_p = 2$ dB, we find $\epsilon^2 = 10^{0.1A_p} - 1 = 0.5849,\ \epsilon = 0.7648$.

The prototype half-power frequency is $\nu_3 = \cosh[(1/n)\cosh^{-1}(1/\epsilon)] = 1.01174$.

The half-power frequency of the actual filter is thus $f_3 = 1000\nu_3 = 1011.74$ Hz.

(d) (Design from a Specified Order) Find the transfer function of a third-order Chebyshev lowpass prototype with a passband ripple of 1.5 dB.

We find $\epsilon^2 = 10^{0.1A_p} - 1 = 0.4125$. With $n = 3$, we find

$$\theta_k = [\pi/6,\ \pi/2,\ 5\pi/6] \qquad \alpha = (1/n)\sinh^{-1}(1/\epsilon) = 0.4086 \qquad \sinh\alpha = 0.4201 \qquad \cosh\alpha = 1.0847$$

The Chebyshev poles are $p_k = -\sin\theta_k \sinh\alpha + j\cos\theta_k \cosh\alpha = [-0.2101 \pm j0.9393,\ -0.4201]$.

The denominator of $H_P(s)$ is thus $Q_P(s) = s^3 + 0.8402s^2 + 1.1030s + 0.3892$.

Since n is odd, the prototype transfer function $H_P(s)$ with unit (dc or peak) gain is

$$H_P(s) = \frac{0.3892}{s^3 + 0.8402s^2 + 1.1030s + 0.3892}$$

(e) (A Bandpass Filter) Design a fourth-order Chebyshev bandpass filter with a peak gain of unity, a 2-dB passband of 200 Hz, and a center frequency of $f_0 = 1$ kHz.

To design the filter, we start with a second-order lowpass prototype. We find $\epsilon^2 = 10^{0.1A_p} - 1 = 0.5849$. With $n = 2,\ \theta_k = [\pi/4,\ 3\pi/4]$, $\alpha = (1/n)\sinh^{-1}(1/\epsilon) = 0.5415,\ \sinh\alpha = 0.5684$, and $\cosh\alpha = 1.1502$. The Chebyshev poles are $p_k = -\sin\theta_k \sinh\alpha + j\cos\theta_k \cosh\alpha = -0.4019 \pm j0.8133$. The denominator of the prototype transfer function $H_P(s)$ is thus $Q_P(s) = s^2 + 0.8038s + 0.8231$. Since n is even, the numerator of $H_P(s)$ for peak unit gain is $K = Q_P(0)/\sqrt{1+\epsilon^2} = 0.8231/\sqrt{1.5849} = 0.6538$, and the prototype transfer function with unit peak gain is

$$H_P(s) = \frac{K}{Q_P(s)} = \frac{0.6538}{s^2 + 0.8038s + 0.8231}$$

PGR2026-00033
Krisp Technologies EX1024 Page 431

With $\omega_0 = 2\pi f_0 = 2000\pi$ rad/s and $B = 400\pi$ rad/s, the LP2BP transformation $s \to (s^2 + \omega_0^2)/Bs$ yields the required transfer function $H(s)$ as

$$H(s) = \frac{1.0234(10)^6 s^2}{s^4 + 1.0101(10)^3 s^3 + 8.0257(10)^7 s^2 + 3.9877(10)^{10} s + 1.5585(10)^{15}}$$

Exactly Meeting the Stopband Attenuation

The lowpass prototype filter $H_P(s)$ shows an attenuation of A_p at the passband edge $\nu_p = 1$ rad/s. Since the filter exceeds specifications everywhere else, the exact value of A_s occurs at a frequency $\tilde{\nu}_s$ that is smaller than the design frequency ν_s. To *exactly* match A_s to the stopband edge ν_s, we must frequency scale (stretch) $H_P(s)$ by the factor $\beta = \nu_s/\tilde{\nu}_s$, using $s \to s/\beta$. The frequency $\tilde{\nu}_s$ is found from the attenuation expression for $A(\nu)$ with $\nu = \tilde{\nu}_s$ as

$$\tilde{\nu}_s = \cosh\left\{\frac{1}{n}\cosh^{-1}\left[\frac{1}{\epsilon}(10^{0.1A_s} - 1)^{1/2}\right]\right\} \tag{13.47}$$

The new filter $H_S(s) = H_P(s/\beta)$ will show an exact match to A_s dB at ν_s rad/s. At the passband edge, however, it will exceed specifications and its attenuation will be less than A_p.

EXAMPLE 13.5 (Chebyshev Filter Design)

(a) Design of a Chebyshev lowpass filter to meet the following specifications: $A_p = 1$ dB at $\omega = 4$ rad/s and $A_s \geq 20$ dB for $\omega \geq 8$ rad/s.

The prototype band edges are $\nu_p = \omega_p/\omega_p = 1$ rad/s and $\nu_s = \omega_s/\omega_p = 2$ rad/s. We find

$$\epsilon^2 = 10^{0.1A_p} - 1 = 0.2589 \qquad n = \frac{\cosh^{-1}[(10^{0.1A_s} - 1)/\epsilon^2]^{1/2}}{\cosh^{-1}\nu_s} = 2.783 \Rightarrow n = 3$$

The half-power frequency is $\nu_3 = \cosh[\frac{1}{3}\cosh^{-1}(1/\epsilon)] = 1.0949$. We also compute

$$\alpha = (1/n)\sinh^{-1}(1/\epsilon) = 0.4760 \qquad \theta_k(\text{rad}) = \frac{(2k-1)\pi}{2n} = \frac{(2k-1)\pi}{6}^{\circ}, \quad k = 1, 2, 3$$

The LHP poles $p_k = -\sinh\alpha\sin\theta_k + j\cosh\alpha\cos\theta_k$ then yield

$$\begin{aligned} p_1 &= -\sinh(0.476)\sin(\pi/6) + j\cosh(0.476)\cos(\pi/6) = -0.2471 + j0.966 \\ p_2 &= -\sinh(0.476)\sin(\pi/2) + j\cosh(0.476)\cos(\pi/2) = -0.4942 \\ p_3 &= -\sinh(0.476)\sin(5\pi/6) + j\cosh(0.476)\cos(5\pi/6) = -0.2471 - j0.966 \end{aligned}$$

The denominator $Q_P(s)$ of the prototype transfer function $H_P(s) = K/Q_P(s)$ is thus

$$Q_P(s) = (s + 0.4942)(s + 0.2471 - j0.966)(s + 0.2471 + j0.966) = s^3 + 0.9883s^2 + 1.2384s + 0.4913$$

Since n is odd, we choose $K = Q_P(0) = 0.4913$ for unit (dc or peak) gain to get

$$H_P(s) = \frac{0.4913}{s^3 + 0.9883s^2 + 1.2384s + 0.4913}$$

Using the attenuation equation, we compute $A_p = A(\nu_p) = 1$ dB and $A_s = A(\nu_s) = 22.456$ dB. The attenuation specification is thus exactly met at the passband edge ν_p and exceeded at the stopband edge ν_s.

PGR2026-00033
Krisp Technologies EX1024 Page 432

Finally, the LP2LP transformation $s \to s/4$ gives the required lowpass filter $H(s)$ as

$$H(s) = H_P(s/4) = \frac{31.4436}{s^3 + 3.9534s^2 + 19.8145s + 31.4436}$$

Comment: To match A_s to the stopband edge, we first find $\tilde{\nu}_s$ and the stretching factor β:

$$\tilde{\nu}_s = \cosh\left\{\frac{1}{n}\cosh^{-1}\left[\frac{1}{\epsilon}(10^{0.1A_s} - 1)^{1/2}\right]\right\} = 1.8441 \qquad \beta = \frac{\nu_s}{\tilde{\nu}_s} = \frac{2}{1.8841} = 1.0845$$

The filter $H_S(s) = H_P(s/\beta)$ matches the stopband specs and is given by

$$H_S(s) = H_P(s/\beta) = \frac{0.6267}{s^3 + 1.0719s^2 + 1.4566s + 0.6267}$$

The LP2LP transformation $s \to s/4$ on $H_S(s)$ gives the required filter $H_2(s) = H_S(s/4)$. We can also combine the two scaling operations and obtain $H_2(s) = H_P(s/4\beta) = H_P(s/4.3381)$, to get

$$H_2(s) = H_S(s/4) = H_P(s/4.3381) = \frac{40.1096}{s^3 + 4.2875s^2 + 23.3057s + 40.1096}$$

The linear and decibel magnitude of these filters is plotted in Figure E13.5 and shows an exact match at the passband edge for $H(s)$ (shown light) and at the stopband edge for $H_2(s)$ (shown dark).

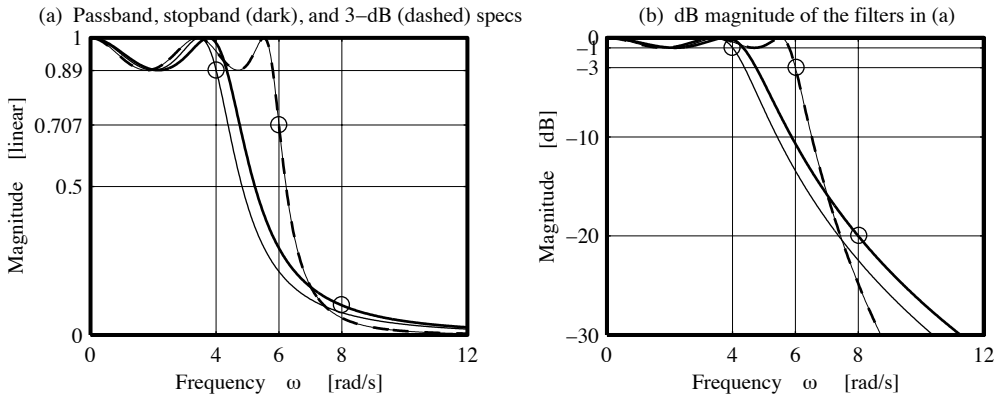


Figure E13.5 Chebyshev lowpass filters for Example 13.5(a and b)

(b) We design a Chebyshev lowpass filter to meet the specifications: $A_p \leq 1$ dB for $\omega \leq 4$ rad/s, $A_s \geq 20$ dB for $\omega \geq 8$ rad/s, and $\omega_3 = 6$ rad/s.

We find an initial estimate of the order (by ignoring the half-power frequency specification) as

$$\epsilon^2 = 10^{0.1A_p} - 1 = 0.2589 \qquad n = 2.783 \Rightarrow n = 3$$

Next, we normalize the stopband edge with respect to the half-power frequency to give $\nu_s = 8/6$ rad/s. From the estimated order ($n = 3$), we compute ν_3 as

$$\nu_3 = \cosh[\tfrac{1}{3}\cosh^{-1}(1/\epsilon)] = 1.0949$$

We use ν_3 to find the true stopband edge $\tilde{\nu}_s = \nu_s/\nu_3$ and check if the attenuation $A(\tilde{\nu}_s)$ exceeds A_s. If not, we increase n by 1, recompute ν_3 for the new order, and repeat the previous step until it does. Starting with $n = 3$, we successively find

PGR2026-00033
Krisp Technologies EX1024 Page 433

n	$\tilde{\nu}_3$	$\tilde{\nu}_s = \omega_s/\omega_N$	$\tilde{A}_s = A(\tilde{\nu}_s)$	$\tilde{A}_s \geq 20$ dB?
3	1.0949	1.4598	12.5139	No
4	1.0530	1.4040	18.4466	No
5	1.0338	1.3784	24.8093	Yes

Thus, $n = 5$. With this value of n, we compute the prototype transfer function $H_P(s)$ as

$$H_P(s) = \frac{0.1228}{s^5 + 0.9368s^4 + 1.6888s^3 + 0.9744s^2 + 0.5805s + 0.1228}$$

We rescale $H_P(s)$ to $H_3(s) = H_P(s\nu_3)$ (for unit half-power frequency) and scale it to $H_A(s) = H_3(s/\omega_3) = H_3(s/6)$ to get the required filter that exactly meets 3-dB specifications. We can implement these two steps together to get $H_A(s) = H_P(1.0338s/6) = H_P(s/5.8037)$ as

$$H_A(s) = \frac{808.7896}{s^5 + 5.4371s^4 + 56.8852s^3 + 190.4852s^2 + 658.6612s + 808.7896}$$

Its linear and decibel magnitude, shown dashed in Figure E13.5, exactly meets the half-power (3-dB) frequency requirement, and exceeds the attenuation specifications at the other band edges.

(c) (A Bandpass Filter) Let us design a Chebyshev bandpass filter for which we are given:

Passband edges: $[\omega_1,\ \omega_2,\ \omega_3,\ \omega_4] = [0.89,\ 1.019,\ 2.221,\ 6.155]$ rad/s

Maximum passband attenuation: $A_p = 2$ dB Minimum stopband attenuation: $A_s = 20$ dB

The frequencies are not geometrically symmetric. So, we assume fixed passband edges and compute $\omega_0^2 = \omega_2\omega_3 = 1.5045$. Since $\omega_1\omega_4 > \omega_0^2$, we decrease ω_4 to $\omega_4 = \omega_0^2/\omega_1 = 2.54$ rad/s.

Then, $B = \omega_3 - \omega_2 = 1.202$ rad/s, $\nu_p = 1$ rad/s, and $\nu_s = \frac{\omega_4 - \omega_1}{B} = \frac{2.54 - 0.89}{1.202} = 1.3738$ rad/s.

The value of ϵ^2 and the order n is given by

$$\epsilon^2 = 10^{0.1A_p} - 1 = 0.5849 \qquad n = \frac{\cosh^{-1}[(10^{0.1A_s} - 1)/\epsilon^2]^{1/2}}{\cosh^{-1}\nu_s} = 3.879 \Rightarrow n = 4$$

The half-power frequency is $\nu_3 = \cosh[\frac{1}{3}\cosh^{-1}(1/\epsilon)] = 1.018$. To find the LHP poles of the prototype filter, we need

$$\alpha = (1/n)\sinh^{-1}(1/\epsilon) = 0.2708 \qquad \theta_k(\text{rad}) = \frac{(2k-1)\pi}{2n} = \frac{(2k-1)\pi}{8}^{\circ}, \quad k = 1, 2, 3, 4$$

From the LHP poles $p_k = -\sinh\alpha\sin\theta_k + j\cosh\alpha\cos\theta_k$, we compute $p_1,\ p_3 = -0.1049 \pm j0.958$ and $p_2,\ p_4 = -0.2532 \pm j0.3968$. The denominator $Q_P(s)$ of the prototype $H_P(s) = K/Q_P(s)$ is thus

$$Q_P(s) = (s - p_1)(s - p_2)(s - p_3)(s - p_4) = s^4 + 0.7162s^3 + 1.2565s^2 + 0.5168s + 0.2058$$

Since n is even, we choose $K = Q_P(0)/\sqrt{1 + \epsilon^2} = 0.1634$ for *peak unit gain*, and thus

$$H_P(s) = \frac{0.1634}{s^4 + 0.7162s^3 + 1.2565s^2 + 0.5168s + 0.2058}$$

PGR2026-00033
Krisp Technologies EX1024 Page 434

We transform this using the LP2BP transformation $s \longrightarrow (s^2+\omega_0^2)/sB$ to give the eighth-order analog bandpass filter $H(s)$ as

$$H(s) = \frac{0.34s^4}{s^8 + 0.86s^7 + 10.87s^6 + 6.75s^5 + 39.39s^4 + 15.27s^3 + 55.69s^2 + 9.99s + 26.25}$$

The linear and decibel magnitude of this filter is shown in Figure E13.5C.

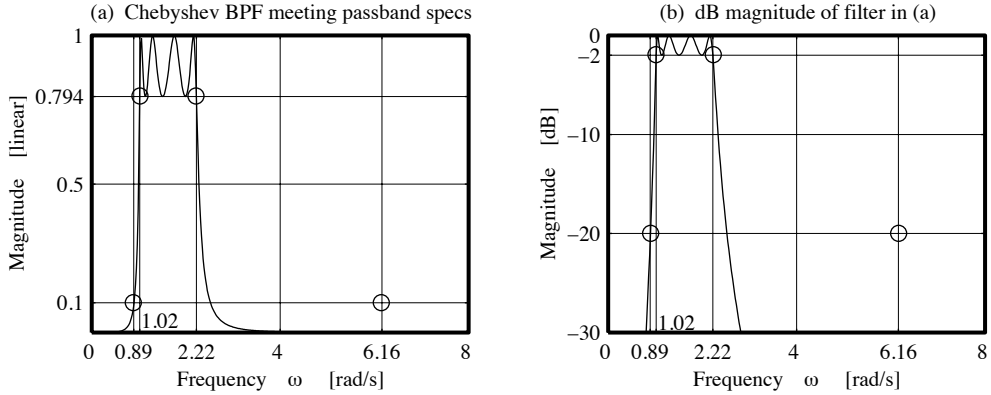


Figure E13.5C Chebyshev I bandpass filter for Example 13.5(c)

13.4.5 Phase and Group Delay of Chebyshev I Filters

A closed-form rational function expression for the group delay of Chebyshev filters involves the so-called *Chebyshev polynomial of the second kind*, $U_n(x)$, defined by

$$U_n(x) = \frac{\sin[(n+1)\cos^{-1}x]}{\sin[\cos^{-1}x]} \tag{13.48}$$

In terms of this function, the group delay $t_g(\nu)$ is given by

$$t_g(\nu) = \frac{1}{1+\epsilon^2 T_n^2(\nu)} \sum_{k=1}^{n} \frac{U_{2k-2}(\nu)\sinh[(2n-2k+1)\nu]}{\epsilon^2 \sin\theta_k} \tag{13.49}$$

Since $T_n(\nu) \longrightarrow 2^{n-1}\nu^n$ for large ν, the phase at high frequencies approaches $-n\pi/2$ radians. The phase response of the Chebyshev filter is not nearly as linear as that of a Butterworth filter of the same order.

REVIEW PANEL 13.13

The Group Delay of a Chebyshev I Lowpass Prototype

$$t_g(\nu) = \frac{1}{1+\epsilon^2 T_n^2(\nu)} \sum_{k=1}^{n} \frac{U_{2k-2}(\nu)\sinh[(2n-2k+1)\nu]}{\epsilon^2 \sin\theta_k} \qquad \text{where} \qquad \theta_k = \frac{(2k-1)\pi}{2n}$$

13.5 The Inverse Chebyshev Approximation

The motivation for the inverse Chebyshev approximation stems from the need to develop a filter magnitude characteristic with a steep transition like the Chebyshev filter, and a phase response with good delay properties like the Butterworth filter. We therefore seek a magnitude response with a maximally flat passband

PGR2026-00033
Krisp Technologies EX1024 Page 435

to improve the delay performance, and somehow retain equiripple behavior to ensure a steep transition. The equiripple property clearly dictates the use of Chebyshev polynomials. To improve the delay characteristics, we need, somehow, to transfer the ripples to a region outside the passband. This is achieved by a frequency transformation that reverses the characteristics of the normal Chebyshev response. This results in the **inverse Chebyshev filter**, or **Chebyshev II filter**, with equiripple behavior in the stopband and a monotonic response (which is also maximally flat) in the stopband. Consider an nth-order Chebyshev lowpass prototype, with ripple μ, described by

$$|H_C(\nu)|^2 = \frac{1}{1 + L(\nu^2)} = \frac{1}{1 + \mu^2 T_n^2(\nu)} \tag{13.50}$$

The highpass transformation $\nu \to 1/\nu$ results in

$$|H_C(1/\nu)|^2 = \frac{1}{1 + \mu^2 T_n^2(1/\nu)} \tag{13.51}$$

It translates the ripples to a band extending from $\nu = 1$ rad/s to $\nu = \infty$ and the monotonic response to the region $\nu \leq 1$ rad/s, as shown in Figure 13.13.

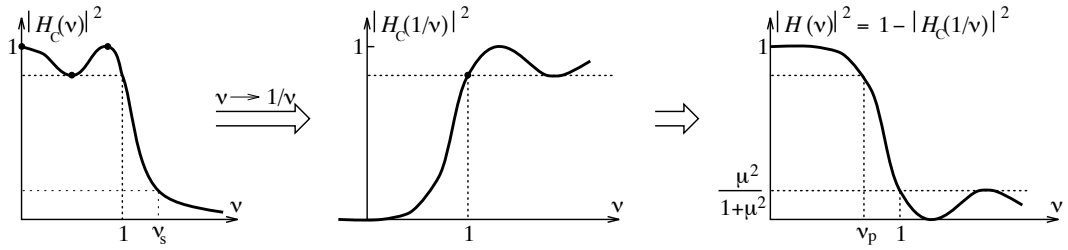


Figure 13.13 The genesis of the inverse Chebyshev lowpass filter

What we need next is a means to convert this to a lowpass form. We cannot use the inverse transformation $\nu \to 1/\nu$ again because it would relocate the ripples in the passband. Subtracting $|H_C(\nu)|^2$ from unity, however, results in

$$|H(\nu)|^2 = 1 - |H_C(1/\nu)|^2 = 1 - \frac{1}{1 + \mu^2 T_n^2(1/\nu)} = \frac{\mu^2 T_n^2(1/\nu)}{1 + \mu^2 T_n^2(1/\nu)} \tag{13.52}$$

The function $|H(\nu)|^2$ (also shown in Figure 13.13) now possesses a lowpass form that is monotonic in the passband and rippled in the stopband, just as required. The ripples in $|H(\nu)|^2$ start at $\nu = 1$ rad/s, which defines the start of the stopband, the frequency where the response *first* reaches the maximum stopband magnitude. This suggests that we normalize the frequency with respect to ω_s, the stopband edge. In keeping with previous forms, we may express the inverse Chebyshev filter characteristic as

$$|H(\nu)|^2 = \frac{\mu^2 T_n^2(1/\nu)}{1 + \mu^2 T_n^2(1/\nu)} = \frac{1}{1 + [1/\mu^2 T_n^2(1/\nu)]} = \frac{1}{1 + L_n^2(\nu)} \tag{13.53}$$

The function $L_n^2(\nu)$ is now a rational function rather than a polynomial. The general form of a maximally flat rational function is $N(\nu)/[N(\nu) + A\nu^{2n}]$, where the degree of $N(\nu)$ is less than $2n$. To show that $|H(\nu)|^2$ yields a maximally flat response at $\nu = 0$, we start with $T_n(1/\nu)$ (using $n = 2$ and $n = 3$ as examples):

$$T_2(1/\nu) = 2\nu^{-2} - 1 = \frac{2 - \nu^2}{\nu^2} \qquad T_3(1/\nu) = 4\nu^{-3} - \nu^{-1} = \frac{4 - 3\nu^2}{\nu^3} \tag{13.54}$$

PGR2026-00033
Krisp Technologies EX1024 Page 436

Substituting into $|H(\nu)|^2$, we obtain

$$|H_2(\nu)|^2 = \frac{\mu^2(4-4\nu^2+\nu^4)}{\mu^2(4-4\nu^2+\nu^4)+\nu^4} \qquad |H_3(\nu)|^2 = \frac{\mu^2(16-24\nu^2+9\nu^4)}{\mu^2(16-24\nu^2+9\nu^4)+\nu^6} \tag{13.55}$$

Both have the required form for maximal flatness. The gain at the stopband edge $\nu_s = 1$ rad/s is given by

$$|H(1)|^2 = \frac{1}{1+(1/\mu^2)} = \frac{\mu^2}{1+\mu^2} \tag{13.56}$$

In contrast to the Chebyshev I filter, the dc gain $|H(0)|$ of the inverse Chebyshev is always 1, regardless of filter order, since $T_n(1/\nu) \to \infty$ for any n.

The inverse Chebyshev filter inherits the improved delay of the maximally flat Butterworth filter and the steeper transition of the Chebyshev I filter by permitting ripples in the stopband. Its rational function form, however, results in a more complex realization (with more elements).

REVIEW PANEL 13.14

The Magnitude Spectrum of a Chebyshev II Filter Has a Rippled Stopband

$$|H(\nu)|^2 = \frac{1}{1+\dfrac{1}{\mu^2 T_n^2(1/\nu)}} \qquad |H(0)| = 1 \qquad |H(1)| = \frac{1}{\sqrt{1+\mu^{-2}}} \qquad H_{\text{dB}} = -10\log\left[1+\frac{1}{\mu^2 T_n^2(1/\nu)}\right]$$

13.5.1 Inverse Chebyshev Filter Design

For the inverse Chebyshev filter, we normalize the *stopband edge* to unity such that $\nu_s = 1$ rad/s and $\nu_p = \omega_p/\omega_s$ rad/s. The attenuation equation is

$$A(\nu) = 10\log|H(\nu)|^{-2} = 10\log\left[1+\frac{1}{\mu^2 T_n^2(1/\nu)}\right] \tag{13.57}$$

We find μ by evaluating the attenuation at the stopband $\nu_s = 1$ rad/s to give

$$A_s = A(\nu_s) = 10\log(1+\mu^{-2}) \quad \text{or} \qquad \mu^{-2} = 10^{0.1A_s} - 1 \tag{13.58}$$

Note that μ is a measure of the stopband ripple. The order n is established by evaluating the attenuation at $\nu = \nu_p$:

$$A_p = A(\nu_p) = 10\log\left[1+\frac{1}{\mu^2 T_n^2(1/\nu_p)}\right] \tag{13.59}$$

Since $\nu_p < 1$ and $1/\nu_p > 1$, evaluation of $T_n(1/\nu_p)$ requires the hyperbolic form and yields

$$A_p = 10\log\left[1+\frac{1}{\{\mu\cosh[n(\cosh^{-1}1/\nu_p)]\}^2}\right] \tag{13.60}$$

The order n is then obtained from this relation as

$$n = \frac{\cosh^{-1}[\mu^{-2}/(10^{0.1A_p}-1)]^{1/2}}{\cosh^{-1}(1/\nu_p)} \tag{13.61}$$

The half-power frequency ν_3, at which $|H(\nu)|^2 = 0.5$, can be expressed in terms of μ and n as

$$\frac{1}{\nu_3} = \cosh\left[\frac{\cosh^{-1}(1/\mu)}{n}\right] \tag{13.62}$$

This form is analogous to the *reciprocal* of the relation for finding ν_3 for the Chebyshev I filter.

PGR2026-00033
Krisp Technologies EX1024 Page 437

13.5.2 The Chebyshev II Prototype Transfer Function

To find the prototype transfer function $H_P(s)$, we start by substituting $\nu = s/j$ into $|H(\nu)|^2$ to obtain

$$H_P(s)H_P(-s) = |H(\nu)|^2_{\nu=s/j} = \frac{\mu^2 T_n^2(1/\nu)}{1+\mu^2 T_n^2(1/\nu)}\Big|_{\nu=s/j} \tag{13.63}$$

The prototype transfer function $H_P(s) = KP(s)/Q(s)$ is a rational function, and its poles (the roots of $Q(s)$) are the left half-plane roots p_k of

$$[1+\mu^2 T_n^2(1/\nu)]|_{\nu=s/j} = 0 \tag{13.64}$$

This is similar to the corresponding relation for the Chebyshev I filter. Since the argument of T_n here is $1/\nu$ (instead of ν), we can apply the Chebyshev I relations for evaluating the inverse Chebyshev roots, provided we replace ϵ by μ and invert the resulting relations. The poles p_k may then be written explicitly as $p_k = 1/(\sigma_k + j\omega_k)$, where

$$\sigma_k = -\sinh\alpha\sin\theta_k, \quad \omega_k = \cosh\alpha\cos\theta_k, \quad \alpha = \frac{\sinh^{-1}(1/\mu)}{n}, \quad \theta_k = \frac{(2k-1)\pi}{2n},\ k = 1,2,\ldots,n \tag{13.65}$$

For odd n, we have a real root $s = -1/\sinh[(1/n)\sinh^{-1}(1/\mu)]$. The poles of the inverse Chebyshev filter are not simply reciprocals of the Chebyshev I poles because μ and ϵ are different. Nor do they lie on an ellipse, even though their locus does resemble an ellipse that is elongated, much like a dumbbell. The zeros of $H_P(s)$ are the roots of

$$T_n^2(1/\nu)|_{\nu=s/j} = 0 \tag{13.66}$$

Now, the zeros ν_z of a Chebyshev polynomial are given by

$$\nu_z = \cos\theta_k, \qquad k = 1,2,\ldots,n \tag{13.67}$$

Replacing ν_z by s/j, ignoring the root at infinity for odd n (for $\theta_k = \pi/2$), and invoking conjugate symmetry, we find

$$z_k = \pm j\sec(\theta_k) = \pm j\omega_k, \qquad k = 1,2,\ldots,\text{int}(n/2) \tag{13.68}$$

These conjugate zeros yield factors of the form $s^2+\omega_k^2$, and we may express $H_P(s)$ as

$$H_P(s) = K\frac{P(s)}{Q(s)} = K\frac{(s^2+\omega_1^2)(s^2+\omega_2^2)\cdots(s^2+\omega_m^2)}{(s-p_1)(s-p_2)\cdots(s-p_n)}, \qquad m = \text{int}(n/2) \tag{13.69}$$

For unit peak (or dc) gain, $K = Q(0)/P(0)$. For a given order n and with $\epsilon = \mu$, the coefficients of $Q(s)$ are identical to those of a Chebyshev I transfer function denominator $Q_C(s)$ in reversed order and related by $Q(s) = s^n Q_C(1/s)$.

Exactly Meeting the Passband Attenuation

The Chebyshev II lowpass prototype $H_P(s)$ shows an attenuation of A_s at the stopband edge $\nu_s = 1$ rad/s. Since the filter exceeds specifications everywhere else, the exact value of A_p occurs at a frequency $\tilde{\nu}_p$ that is smaller than the design frequency ν_p. To *exactly* match A_p to the passband edge ν_p, we must frequency scale (stretch) $H_P(s)$ by the factor $\beta = \nu_p/\tilde{\nu}_p$, using $s \to s/\beta$. The frequency $\tilde{\nu}_p$ is found from the attenuation expression for $A(\nu)$ with $\nu = \tilde{\nu}_p$ as

$$\tilde{\nu}_p = \frac{1}{\cosh\left\{\frac{1}{n}\cosh^{-1}\left[\frac{1}{\mu(10^{0.1A_p}-1)^{1/2}}\right]\right\}} \tag{13.70}$$

The new filter $H(s) = H_P(s/\beta)$ will show an exact match to A_p at ν_p, and exceed the attenuation specifications at the stopband edge (where the attenuation will exceed A_s).

PGR2026-00033
Krisp Technologies EX1024 Page 438

EXAMPLE 13.6 (Chebyshev II Lowpass Filter Design)

(a) We design an inverse Chebyshev lowpass filter to meet the specifications $A_p \leq 1$ dB for $\omega \leq 4$ rad/s and $A_s \geq 20$ dB for $\omega \geq 8$ rad/s.

Using the design equations, we compute:

$$\nu_p = \frac{\omega_p}{\omega_s} = 0.5 \qquad \mu^2 = \frac{1}{10^{0.1A_s} - 1} = 0.0101 \qquad n = \frac{\cosh^{-1}[1/(10^{0.1A_p} - 1)\mu^2]^{1/2}}{\cosh^{-1}(1/\nu_p)} = 2.783 \Rightarrow n = 3$$

The half-power frequency is $\nu_3 = 1/\cosh[(1/n)\cosh^{-1}(1/\mu)] = 0.65$.
To find the prototype transfer function $H_P(s) = KP(s)/Q(s)$, we require

$$\alpha = (1/n)\sinh^{-1}(1/\mu) = 0.9977 \quad \theta_k = (2k-1)\pi/2n = (2k-1)\pi/6^\circ, \quad k = 1, 2, 3$$

Then, $1/p_k = -\sinh\alpha\sin\theta_k + j\cosh\alpha\cos\theta_k$, and we compute

$$\begin{aligned} 1/p_1 &= -\sinh(0.9977)\sin(\pi/6) + j\cosh(0.9977)\cos(\pi/6) = -0.5859 + j1.3341 \\ 1/p_2 &= -\sinh(0.9977)\sin(\pi/2) + j\cosh(0.9977)\cos(\pi/2) = -1.1717 \\ 1/p_3 &= -\sinh(0.9977)\sin(5\pi/6) + j\cosh(0.9977)\cos(5\pi/6) = -0.5859 - j1.3341 \end{aligned}$$

The left half-plane poles are $[p_1, p_2, p_3] = [-0.276 - j0.6284, -0.8534, -0.276 + j0.6284]$.
These give the denominator $Q(s)$ as

$$Q(s) = (s + 0.853)(s^2 + 0.552s + 0.471) = s^3 + 1.4054s^2 + 0.9421s + 0.4020$$

Since $n = 3$, there are only two left half-plane zeros given by $\pm j\sec\theta_1 = \pm j1.1547$.
The numerator is thus $P(s) = s^2 + (1.1547)^2 = s^2 + 1.3333$.
For unit peak (or dc) gain, we choose $K = Q(0)/P(0) = 0.402/1.3333 = 0.3015$. Thus,

$$H_P(s) = \frac{0.3015(s^2 + 0.13333)}{s^3 + 1.4054s^2 + 0.9421s + 0.4020} = \frac{0.3015s^2 + 0.4020}{s^3 + 1.4054s^2 + 0.9421s + 0.4020}$$

The lowpass filter $H(s)$ that matches the attenuation at the stopband edge is obtained by using the LP2LP transformation $H(s) = H_P(s/\omega_s) = H_S(s/8)$ to give

$$H(s) = \frac{2.4121s^2 + 205.8317}{s^3 + 11.2431s^2 + 60.2942s + 205.8317}$$

Comment: To exactly match the attenuation at the passband edge, we compute $\beta = \nu_p/\tilde{\nu}_p$, where

$$\tilde{\nu}_p = \frac{1}{\cosh\left\{\frac{1}{n}\cosh^{-1}\left[\frac{1}{\mu(10^{0.1A_p} - 1)^{1/2}}\right]\right\}} = 0.5423$$

Thus, $\beta = \nu_p/\tilde{\nu}_p = 0.9221$. The prototype $H_S(s)$ that matches the passband specifications is

$$H_S(s) = H_P(s/\beta) = \frac{0.2780s^2 + 0.3152}{s^3 + 1.2959s^2 + 0.8010s + 0.3152}$$

The required lowpass filter $H_2(s)$ that matches the attenuation at the passband edge is obtained by using the LP2LP transformation $H_2(s) = H_S(s/\omega_s)$. We could also use the two transformations together in one step to give $H_2(s) = H_P(s/\omega_s\beta) = H_P(s/8\beta)$. The transfer function $H_2(s)$ is

$$H_2(s) = H_P(s/8\beta) = \frac{2.2241s^2 + 161.3602}{s^3 + 10.3668s^2 + 51.2623s + 161.3602}$$

PGR2026-00033
Krisp Technologies EX1024 Page 439

The linear and decibel magnitude of $H(s)$ (dark) and $H_2(s)$ (light) is shown in Figure E13.6.

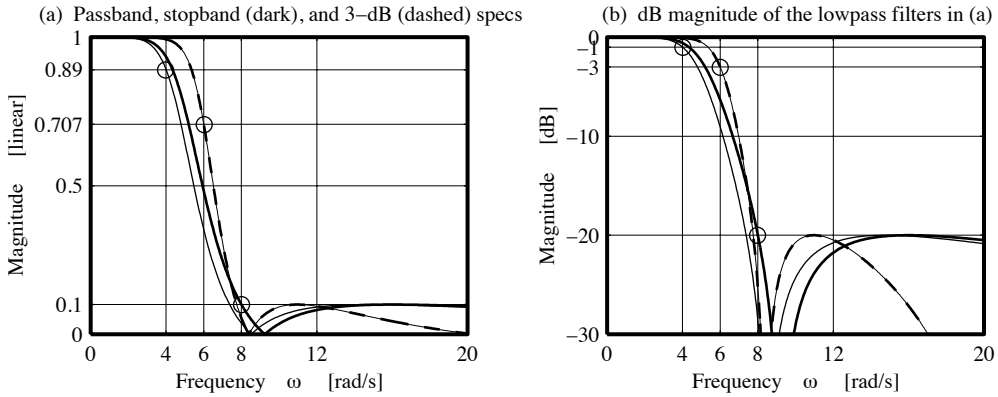


Figure E13.6 Chebyshev II lowpass filters for Example 13.6(a and b)

(b) Consider the design of a Chebyshev II lowpass filter that meets the specifications $A_p \leq 1$ dB for $\omega \leq 4$ rad/s, $A_s \geq 20$ dB for $\omega \geq 8$ rad/s, and $\omega_3 = 6$ rad/s.

We find an initial estimate of the order (by ignoring the half-power frequency specification) as

$$\nu_p = \frac{\omega_p}{\omega_s} = 0.5 \qquad \mu^2 = \frac{1}{10^{0.1A_s} - 1} = 0.0101 \qquad n = \frac{\cosh^{-1}[1/(10^{0.1A_p} - 1)\mu^2]^{1/2}}{\cosh^{-1}(1/\nu_p)} = 2.78 \Rightarrow n = 3$$

We normalize the stopband edge with respect to the half-power frequency to give $\nu_s = 8/6$. The half-power frequency is $\nu_3 = 1/\cosh[(1/n)\cosh^{-1}(1/\mu)] = 0.65$. We use this to find the true stopband edge $\tilde{\nu}_s = \nu_s/\nu_3$ and check that the attenuation $A(\tilde{\nu}_s)$ exceeds A_s. If not, we increase n by 1, recompute ν_3 for the new order, and repeat the previous step until it does. Starting with $n = 3$, we find

n	ν_3	$\tilde{\nu}_s = \omega_s/\omega_N$	$\tilde{A}_s = A(\tilde{\nu}_s)$	$\tilde{A}_s \geq 20$ dB?
3	0.65	0.8667	11.6882	No
4	0.7738	1.0318	25.2554	Yes

Thus, $n = 4$ and $\nu_3 = 0.7738$, and we find the prototype transfer function $H_P(s)$ as

$$H_P(s) = \frac{0.1s^4 + 0.8s^2 + 0.8}{s^4 + 2.2615s^3 + 2.6371s^2 + 1.7145s + 0.8}$$

We first scale $H_P(s)$ to $H_3(s) = H_P(s\nu_3)$ (for unit half-power frequency), and then scale $H_3(s)$ to $H(s) = H_3(s/\omega_3) = H_3(s/6)$ to get the required filter that exactly meets 3-dB specifications. We can implement these two steps together to get $H(s) = H_P(0.7738s/6) = H_P(s/7.7535)$ as

$$H(s) = H_P(s/7.7535) = \frac{0.1s^4 + 48.0940s^2 + 2891.2863}{s^4 + 17.5344s^3 + 158.5376s^2 + 799.1547s + 2891.2863}$$

Its linear and decibel magnitude, shown dashed in Figure E13.6, exactly meets the half-power (3-dB) frequency requirement and exceeds the attenuation specifications at the other band edges.

Comment: We designed a Chebyshev I filter for the same specifications in Example 13.5(a and b). With ω_3 not specified, both types (Chebyshev I and II) yield identical orders ($n = 3$). With ω_3 also specified, the Chebyshev I filter yields $n = 5$, but the Chebyshev II filter yields $n = 4$ (a *lower* order).

PGR2026-00033
Krisp Technologies EX1024 Page 440

13.6 The Elliptic Approximation

The elliptic approximation provides ripples in both bands by using rational functions whose numerator and denominator both display the equiripple property, as sketched in Figure 13.14.

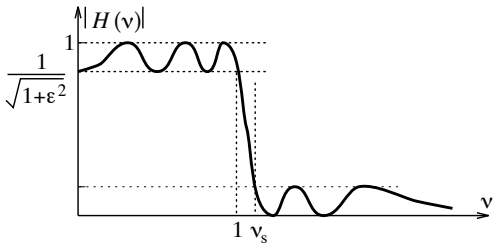


Figure 13.14 Magnitude spectrum of an elliptic lowpass filter

The magnitude squared function of such an approximation can be described by

$$|H(\nu)|^2 = \frac{1}{1 + L_n^2(\nu)} = \frac{1}{1 + \epsilon^2 R_n^2(\nu, \delta)} \tag{13.71}$$

Here, $R_n(\nu, \delta)$ is the so-called **Chebyshev rational function**, ϵ^2 describes the passband ripple, and the additional parameter δ provides a measure of the stopband ripple magnitude. Clearly, $R_n(\nu, \delta)$ must be sought in the rational function form $A(\nu^2)/B(\nu^2)$, with both the numerator and denominator satisfying near optimal constraints and possessing properties analogous to Chebyshev polynomials. This implies that, for a given order n,

1. $R_n(\nu, \delta)$ should exhibit oscillatory behavior with equal extrema in the passband ($|\nu| < 1$), and with all its n zeros within the passband.
2. $R_n(\nu, \delta)$ should exhibit oscillatory behavior, with equal extrema in the stopband ($|\nu| > 1$), and with all its n poles within the stopband.
3. $R_n(\nu, \delta)$ should be even symmetric if n is even and odd symmetric if n is odd.

We also impose the additional simplifying constraint

$$R_n(1/\nu, \delta) \propto \frac{1}{R_n(\nu, \delta)} \tag{13.72}$$

This provides a reciprocal relation between the poles and zeros of R_n and suggests that if we can find a function of this form with equiripple behavior in the passband $0 < \nu < 1$, it will automatically result in equiripple behavior in the reciprocal range $1 < \nu < \infty$ representing the stopband. The functional form for $R_n(\nu, \delta)$ that meets these criteria may be described in terms of its root locations ν_k by

$$R_n(\nu, \delta) = C\nu^N \prod_{k=1}^{\text{int}(n/2)} \frac{\nu^2 - \nu_k^2}{\nu^2 - B/\nu_k^2} \tag{13.73}$$

where $\text{int}(x)$ is the integer part of x, ν_k is a root of the numerator, the constants B and C are chosen to ensure that $R_n(\nu, \delta) = 1$ for even n and $R_n(\nu, \delta) = 0$ for odd n, and $N = 0$ for even n and $N = 1$ for odd n.

Elliptic filters yield the lowest filter order for given specifications by permitting ripples in both the passband and stopband. For a given order, they exhibit the steepest transition region, but the very presence of the ripples also makes for the most nonlinear phase and the worst delay characteristics.

PGR2026-00033
Krisp Technologies EX1024 Page 441

The Elliptic Approximation

An alternative to the trial-and-error method for establishing $R_n(\nu, \delta)$ is to find a transformation $\nu = g(\phi)$ that transforms R_n, a polynomial in ν, to a periodic function in ϕ much as $\nu = \cos\phi$ transformed the Chebyshev polynomial $T_n(\nu)$ into the periodic function $\cos(n\phi)$ with its attendant equiripple character. It turns out that such a transformation involves *elliptic integrals* and *elliptic functions*. Before examining this transformation, we introduce the concept of elliptic functions and integrals.

13.6.1 Elliptic Integrals and Elliptic Functions

The **elliptic integral of the first kind**, u, is actually a function of two arguments, the *amplitude*, ϕ, and the *parameter*, m, and is defined by

$$u(\phi, m) = \int_0^{\phi} (1 - m\sin^2 x)^{-1/2}\, dx \tag{13.74}$$

The dependence of u on ϕ is better visualized by taking the derivative

$$\frac{du}{d\phi} = (1 - m\sin^2\phi)^{-1/2} \tag{13.75}$$

With $m = 0$, we have the trivial result

$$\left.\frac{du}{d\phi}\right|_{m=0} = 1 \qquad u(\phi, 0) = \phi \tag{13.76}$$

If the amplitude ϕ equals $\pi/2$, we get the **complete elliptic integral of the first kind**, $u(\pi/2, m)$, denoted by $K(m)$, which is now a function only of m:

$$K(m) = u(\pi/2, m) = \int_0^{\pi/2} (1 - m\sin^2 x)^{-1/2}\, dx \tag{13.77}$$

The **modulus** k is related to the parameter m by $m = k^2$. The **complementary complete elliptic integral of the first kind** is denoted by $K'(m)$. With $k'^2 = 1 - k^2$ and $m' = 1 - m$, it is defined by

$$K'(m) = K(m') = u(\pi/2, m') = \int_0^{\pi/2} (1 - m'\sin^2 x)^{-1/2}\, dx \tag{13.78}$$

We thus have $K'(m) = K(1 - m)$. The quantity m' is called the **complementary parameter**, and k' is called the **complementary modulus**.

The **Jacobian elliptic functions** also involve two arguments, one being the elliptic integral $u(\phi, m)$, and the other the parameter m. Some of these functions are

Jacobian elliptic sine	$\mathrm{sn}(u, m) = \sin\phi$
Jacobian elliptic cosine	$\mathrm{cn}(u, m) = \cos\phi$
Jacobian elliptic difference	$\mathrm{dn}(u, m) = d\phi/du$

(13.79)

The Jacobian elliptic functions are actually *doubly periodic* for complex arguments. In particular, $\mathrm{sn}(u, m)$, where u is complex, has a real period of $4K$ and an imaginary period of $2K'$. Jacobian elliptic functions resemble trigonometric functions and even satisfy some of the same identities. For example,

$$\mathrm{sn}^2 + \mathrm{cn}^2 = 1 \quad \mathrm{cn}^2 = 1 - \mathrm{sn}^2 \quad \mathrm{dn} = d\phi/du = (1 - m\sin^2\phi)^{1/2} = (1 - m\ \mathrm{sn}^2)^{1/2} \tag{13.80}$$

PGR2026-00033
Krisp Technologies EX1024 Page 442

The Jacobian elliptic sine, $\text{sn}(u, m)$ is illustrated in Figure 13.15. It behaves much like the trigonometric sine, except that it is flattened and elongated. The degree of elongation (called the *period*) and flatness depend on the parameter m. For $m = 0$, $\text{sn}(u, m)$ is identical to $\sin(\phi)$. For small $u, \text{sn}(u, m)$ closely follows the sine and with increasing m, becomes more flattened and elongated, reaching an infinite period when $m = 1$. This ability of $\text{sn}(u, m)$ to change shape with changing m is what provides us with the means to characterize the function R_n.

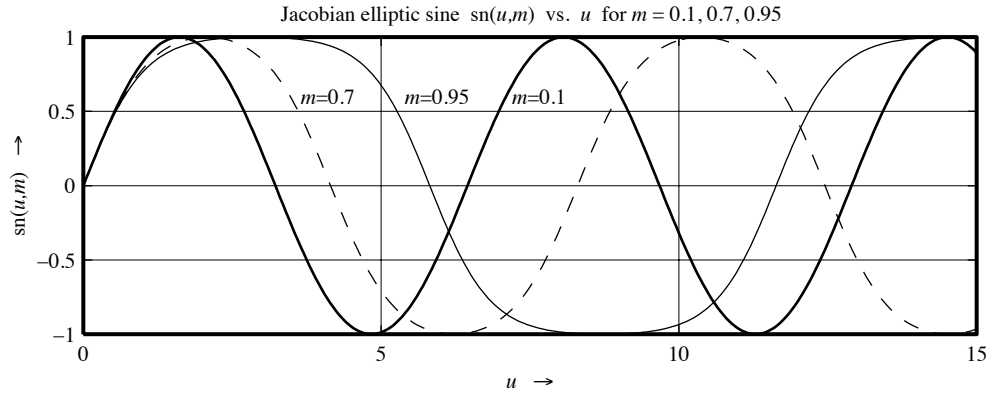


Figure 13.15 The Jacobian elliptic sine function

13.6.2 The Chebyshev Rational Function

The transformation for ν in the rational function $R_n(\nu, \delta)$ that achieves the desired equiripple behavior is of the form $\nu = \text{sn}[\alpha K(m), m]$, where $K(m)$ is itself the complete elliptic integral of the first kind. Without going into the involved details of its development, the transformation $\nu = \text{sn}[\alpha K(m), m]$ yields $R_n(\nu, \delta)$ as a Jacobian elliptic function of the form

$$R_n(\nu, \delta) = C\nu^N \prod_{k=1}^{\text{int}(n/2)} \frac{\nu^2 - z_k^2}{\nu^2 - p_k^2}, \qquad \text{where} \quad N = \begin{cases} 0, & n \text{ even} \\ 1, & n \text{ odd} \end{cases} \tag{13.81}$$

Here, $N = 0$ for even n, and $N = 1$ for odd n. The constant C is found by recognizing that $R_n(\nu, \delta) = 1$ at $\nu = 1$ rad/s, leading to

$$C = \prod_{k=1}^{\text{int}(n/2)} \frac{1 - p_k^2}{1 - z_k^2}, \qquad k = 1, 2, \ldots, n \tag{13.82}$$

The poles p_k and zeros z_k are imaginary, and their locations depend on the order n. They are also related by the reciprocal relation

$$p_k = \frac{\nu_s}{z_k}, \qquad k = 1, 2, \ldots, \text{int}(\tfrac{n}{2}) \tag{13.83}$$

where ν_s is the first value of ν at which $R_n(\nu, \delta) = \delta$ and

$$z_k = \begin{cases} \text{sn}[2kK(m)/n, m], & n \text{ odd} \\ \text{sn}[(2k-1)K(m)/n, m], & n \text{ even} \end{cases} \qquad \text{where} \quad k = 1, 2, \ldots, \text{int}(\tfrac{n}{2}) \tag{13.84}$$

where sn is the Jacobian elliptic sine function, $K(m) = K(1/\nu_s^2)$ is the complete elliptic integral of the first kind, and δ is the maximum magnitude of oscillations for $|\nu| > \nu_s$.

PGR2026-00033
Krisp Technologies EX1024 Page 443

The rational function $R_n(\nu,\delta)$ then assumes the form

$$R_n(\nu,\delta) = C\nu^N \prod_{k=1}^{\text{int}(n/2)} \frac{\nu^2 - [\text{sn}(u_k,m)]^2}{\nu^2 - [\nu_s/\text{sn}(u_k,m)]^2}, \quad \text{where} \quad u_k = \begin{cases} (2k-1)K(m)/n, & n \text{ even} \\ 2kK(m)/n, & n \text{ odd} \end{cases} \tag{13.85}$$

Note that $R_n(\nu,\delta)$ obeys the reciprocal relation $R_n(\nu,\delta) = \delta/R_n(\nu_s/\nu,\delta)$. The frequencies ν_{pass} of the passband maxima (where $R_n(\nu,\delta) = 1$) are given by

$$\nu_{\text{pass}} = \begin{cases} \text{sn}[(2k+1)K(m)/n, m], & n \text{ odd} \\ \text{sn}[2kK(m)/n, m], & n \text{ even} \end{cases} \quad \text{where} \quad k = 1, 2, \ldots, \text{int}(\tfrac{n}{2}) \tag{13.86}$$

The frequencies ν_{stop} of the stopband maxima (where $R_n(\nu,\delta) = \delta$) are related to ν_{pass} by

$$\nu_{\text{stop}} = \frac{\nu_s}{\nu_{\text{pass}}} \tag{13.87}$$

Filter Order

Like the Jacobian elliptic functions, $R_n(\nu,\delta)$ is also doubly periodic. The periods are not independent but are related through the order n, as follows:

$$n = \frac{K(1/\nu_s^2)K'(1/\delta^2)}{K'(1/\nu_s^2)K(1/\delta^2)} = \frac{K(1/\nu_s^2)K(1-1/\delta^2)}{K(1-1/\nu_s^2)K(1/\delta^2)} \tag{13.88}$$

13.6.3 Design Recipe for Elliptic Filters

We start by normalizing the passband edge to unity to obtain $\nu_p = 1$ rad/s and $\nu_s = \omega_s/\omega_p$ rad/s. The parameters ϵ and δ are found from the attenuation relation

$$A(\nu) = 10\log[1 + \epsilon^2 R_n^2(\nu,\delta)] \tag{13.89}$$

At the passband edge $\nu = \nu_p = 1$ rad/s, $R_n(\nu_p,\delta) = 1$, and we obtain

$$A_p = A(\nu_p) = 10\log[1+\epsilon^2] \qquad \text{or} \qquad \epsilon^2 = 10^{0.1A_p} - 1 \tag{13.90}$$

Note that if ω_p corresponds to the half-power frequency ($A_p = 3.01$ dB), then $\epsilon^2 = 1$. At the stopband edge $\nu = \nu_s$ rad/s, $R_n(\nu_s,\delta) = \delta$, and we obtain

$$A_s = A(\nu_s) = 10\log[1+\epsilon^2\delta^2] \qquad \text{or} \qquad \delta^2 = (10^{0.1A_s} - 1)/\epsilon^2 \tag{13.91}$$

The order n is next evaluated from the relation

$$n = \frac{K(1/\nu_s^2)K'(1/\delta^2)}{K'(1/\nu_s^2)K(1/\delta^2)} = \frac{K(1/\nu_s^2)K(1-1/\delta^2)}{K(1-1/\nu_s^2)K(1/\delta^2)} \tag{13.92}$$

If n does not work out to an integer, we choose the next higher integer. To ensure that the design *exactly* meets passband specifications, however, we must (iteratively) find a value of the stopband frequency ν_s and $K(1/\nu_s^2)$ that exactly satisfies the above relation for the chosen integer n.

To find the prototype transfer function $H_P(s) = KP(s)/Q(s)$, we start with $|H(\nu)|^2$:

$$|H(\nu)|^2 = \frac{1}{1+\epsilon^2 R_n^2(\nu,\delta)} = \frac{\prod^{\text{int}(n/2)} (\nu^2 - p_k^2)^2}{\prod^{\text{int}(n/2)} (\nu^2 - p_k^2)^2 + (\epsilon C\nu^N)^2 \prod^{\text{int}(n/2)} (\nu^2 - z_k^2)^2} \tag{13.93}$$

PGR2026-00033
Krisp Technologies EX1024 Page 444

With $\nu^2 \to -s^2$, we obtain $H_P(s)H_P(-s)$ in the form

$$H_P(s)H_P(-s) = \frac{\prod^{\text{int}(n/2)} (s^2+p_k^2)^2}{\prod^{\text{int}(n/2)} (s^2+p_k^2)^2 + (-s^2)^N \epsilon^2 C^2 \prod^{\text{int}(n/2)} (s^2+z_k^2)^2}, \quad \text{where} \quad N = \begin{cases} 0, & n \text{ even} \\ 1, & n \text{ odd} \end{cases} \tag{13.94}$$

The numerator $P(s)$ is found from the conjugate symmetric roots $\pm jp_k$ and equals

$$P(s) = \prod_{k=1}^{\text{int}(n/2)} (s^2+p_k^2) \tag{13.95}$$

The denominator $Q(s)$ is extracted from the left half-plane roots of

$$Q(s)Q(-s) = 0 = \prod_{k=1}^{\text{int}(n/2)} (s^2+p_k^2)^2 + (-s^2)^N \epsilon^2 C^2 \prod_{k=1}^{\text{int}(n/2)} (s^2+z_k^2)^2, \quad \text{where} \quad N = \begin{cases} 0, & n \text{ even} \\ 1, & n \text{ odd} \end{cases} \tag{13.96}$$

For $n > 2$, finding the roots of this equation becomes tedious and often requires numerical methods. For unit dc gain, we choose $K = Q(0)/P(0)$. Since $R_n(0,\delta) = 0$ for odd n and $R_n(0,\delta) = 1$ for even n, $H(0) = 1$ for odd n and $H(0) = 1/\sqrt{1+\epsilon^2}$ for even n. Therefore, for a peak gain of unity, we choose $K = Q(0)/P(0)$ for odd n and $K = Q(0)/[P(0)\sqrt{1+\epsilon^2}]$ for even n.

EXAMPLE 13.7 (Elliptic Filter Design)

Consider the design of an elliptic lowpass filter to meet the following specifications: $A_p = 2$ dB, $A_s = 20$ dB, $\omega_p = 4$ rad/s, and $\omega_s = 8$ rad/s.

Following the design procedure, we normalize by ω_p to get $\nu_p = 1$ rad/s, $\nu_s = \omega_s/\omega_p = 2$ rad/s, and

$$\epsilon^2 = 10^{0.1A_p} - 1 = 0.5849 \qquad \delta^2 = \frac{10^{0.1A_s}-1}{\epsilon^2} = 169.2617$$

We let $m = 1/\nu_s^2 = 0.25$ and $p = 1/\delta^2 = 0.0059$. Then, $K(m) = 1.6858$, $K'(m) = K(1-m) = 2.1565$, $K(p) = 1.5731$, and $K'(p) = K(1-p) = 3.9564$. The filter order is given by

$$n = \frac{K(m)K'(p)}{K'(m)K(p)} = \frac{(1.6858)(3.9564)}{(2.1565)(1.5731)} = 1.97$$

We choose $n = 2$. Note that to exactly meet passband specifications, we (iteratively) find that the actual ν_s that yields $n = 2$ is not $\nu_s = 2$ rad/s but $\nu_s = 1.942$ rad/s. However, we perform the following computations using $\nu_s = 2$ rad/s. With $n = 2$, we compute

$$\text{sn}[(2k-1)K(m)/n, m] = \text{sn}[K(m)/n, m] = \text{sn}(0.8429, 0.25) = 0.7321$$

The zeros and poles of $R_n(\nu,\delta)$ are then $z_k = 0.7321$ and $p_k = \nu_s/z_k = 2.7321$.
With $H_P(s) = KP(s)/Q(s)$, we find $P(s)$ from the poles p_k as

$$P(s) = \prod_{k=1}^{\text{int}(n/2)} (s^2+p_k^2) = s^2 + (2.7321)^2 = s^2 + 7.4641$$

PGR2026-00033
Krisp Technologies EX1024 Page 445

To find $Q(s)$, we first compute the constant C as

$$C = \prod_{k=1}^{\text{int}(n/2)} \frac{1-p_k^2}{1-z_k^2} = \frac{1-7.4641}{1-0.5359} = -13.9282$$

With $N = 0$, we evaluate $Q(s)$ from the left half-plane roots of

$$\prod^{\text{int}(n/2)} (s^2+p_k^2)^2 + \epsilon^2 C^2 \prod^{\text{int}(n/2)} (s^2+z_k^2)^2 = 114.47s^4 + 136.54s^2 + 88.299 = 0$$

The two left half-plane roots are $r_{1,2} = -0.3754 \pm j0.8587$, and we get

$$Q(s) = (s+0.3754+j0.8587)(s+0.3754-j0.8587) = s^2 + 0.7508s + 0.8783$$

The prototype transfer function may now be written as

$$H_P(s) = K\frac{s^2+7.4641}{s^2+0.7508s+0.8783}$$

Since n is even, we choose $K = 0.8783/(7.4641\sqrt{1+\epsilon^2}) = 0.09347$ for a *peak* gain of unity. The required filter $H(s)$ is found using the LP2LP transformation $s \Rightarrow s/\omega_p = s/4$, to give

$$H(s) = H_P(s/\omega_p) = H_P(s/4) = \frac{0.09347s^2 + 11.1624}{s^2 + 3.0033s + 14.0527}$$

The linear and decibel magnitude of $H(s)$ is shown in Figure E13.7.

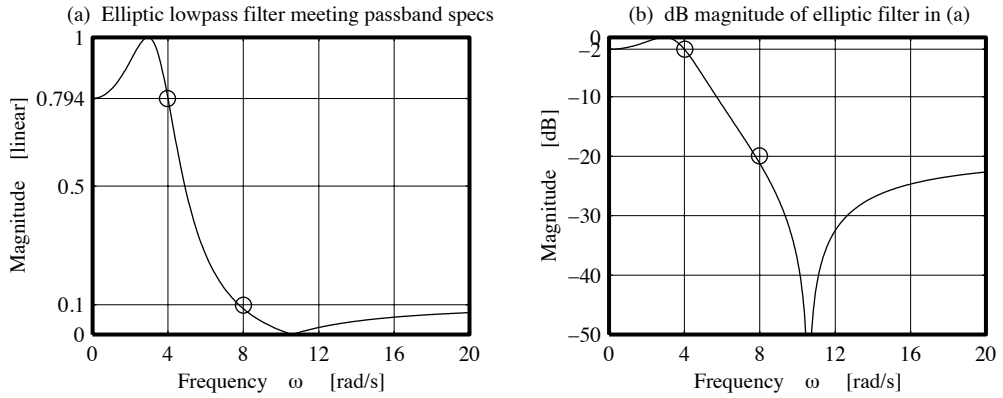


Figure E13.7 Elliptic lowpass filter for Example 13.7

13.7 The Bessel Approximation

The Bessel approximation is used in filter design as a means of obtaining linear phase or flat delay in the passband. Unlike other filter forms, it is much easier to start with a transfer function $H(s)$ directly rather than derive it from a parametric model. Unlike other filters, however, we work with a normalized *unit delay*. Denormalization to a desired delay t_g is then accomplished by the transformation $s \to t_g s$ in the transfer function $H(s)$. We start with the nth-order lowpass transfer function $H(s)$ given by

$$H(s) = \frac{K}{q_0 + q_1 s + q_2 s^2 + \cdots + q_n s^n} \tag{13.97}$$

PGR2026-00033
Krisp Technologies EX1024 Page 446

Its frequency response $H(\omega)$ may then be written as

$$H(\omega) = \frac{K}{(q_0 - q_2\omega^2 + q_4\omega^4 + \cdots) + j(q_1\omega - q_3\omega^3 + q_5\omega^5 + \cdots)} \tag{13.98}$$

And the phase $\phi(\omega)$ readily follows as

$$\phi(\omega) = -\tan^{-1}\left[\frac{q_1\omega - q_3\omega^3 + q_5\omega^5 + \cdots}{q_0 - q_2\omega^2 + q_4\omega^4 + \cdots}\right] \tag{13.99}$$

Recall that even though the phase function is transcendental in nature, its derivative, which represents the time delay, is a rational function. It is also an even function (of ω^2) with no pole at the origin. We can use several approaches to find the coefficients q_k that result in a constant (or unit) delay.

13.7.1 The Maximally Flat Approach

One method is to approximate the delay $t_g(\omega)$ by an nth order maximally flat rational function whose first $n-1$ derivatives are zero at $\omega = 0$, and whose general form is described by

$$t_g(\omega) = \frac{N(\omega)}{N(\omega) + A\omega^n} \tag{13.100}$$

Here, $N(\omega)$ is a polynomial of order less than n. The polynomial coefficients q_k in the denominator of the transfer function $H(s)$ may be obtained by approximating the phase as $\phi(\omega) \approx -\omega$ (for unit delay) and setting all higher-order derivatives of $\phi(\omega)$ (except the nth) to zero (for maximal flatness).

First-Order Case: Consider the first-order transfer function $H(s)$ described by

$$H(s) = \frac{K}{q_0 + s} \qquad H(\omega) = \frac{K}{q_0 + j\omega} \qquad \phi(\omega) = -\tan^{-1}(\omega/q_0) \tag{13.101}$$

To determine q_0, we seek a linear phase function $\tilde{\phi}(\omega) \approx -\omega$ for unit delay.

With $\tan^{-1}(x) \approx x - \frac{1}{3}x^3 + \frac{1}{5}x^5 - \frac{1}{7}x^7 + \cdots$, the phase $\phi(\omega)$ may be approximated by

$$\phi(\omega) = -\tan^{-1}(\omega/q_0) \approx -\left[\frac{\omega}{q_0} - \frac{1}{3}\Big(\frac{\omega}{q_0}\Big)^3 + \cdots\right]$$

Comparison with $\tilde{\phi}(\omega) = -\omega$ gives $q_0 = 1$ and

$$H(s) = \frac{K}{1+s} \qquad H(\omega) = \frac{K}{1+j\omega} \qquad t_g(\omega) = -\frac{d\phi}{d\omega} = \frac{d}{d\omega}[\tan^{-1}\omega] = \frac{1}{1+\omega^2} \tag{13.102}$$

For $K = 1$, the dc gain equals 1. The delay $t_g(\omega)$ clearly represents a maximally flat function.

Second-Order Case: Consider a second-order transfer function described by

$$H(s) = \frac{K}{q_0 + q_1 s + s^2} \qquad H(\omega) = \frac{K}{(q_0 - \omega^2) + j\omega q_1} \tag{13.103}$$

$$\phi(\omega) = -\tan^{-1}\left[\frac{q_1\omega}{q_0 - \omega^2}\right] = -\tan^{-1}\left[\frac{\omega q_1/q_0}{1 - \omega^2/q_0}\right] \tag{13.104}$$

Using $\tan^{-1}(x) \approx x - \frac{1}{3}x^3 + \frac{1}{5}x^5 - \frac{1}{7}x^7 + \cdots$, its phase $\phi(\omega)$ can be approximated by

$$\phi(\omega) \approx -\left[\frac{\omega q_1/q_0}{1 - \omega^2/q_0}\right] + \frac{1}{3}\left[\frac{\omega q_1/q_0}{1 - \omega^2/q_0}\right]^3 - \cdots \tag{13.105}$$

PGR2026-00033
Krisp Technologies EX1024 Page 447

If the denominators of each term are expressed in terms of the first-order binomial approximation given by $(1-x)^{-n} \approx (1+nx)$ for $x \ll 1$, we obtain

$$\phi(\omega) \approx -\frac{\omega q_1}{q_0}\left[1+\frac{\omega^2}{q_0}\right] + \tfrac{1}{3}\left(\frac{\omega q_1}{q_0}\right)^3\left[1+\frac{3\omega^2}{q_0}\right] - \cdots \tag{13.106}$$

Retaining terms to ω^3, and comparing with the required phase $\tilde{\phi}(\omega) \approx -\omega$, we get

$$\phi(\omega) \approx -\omega\frac{q_1}{q_0} - \omega^3\left[\frac{q_1}{q_0^2} - \frac{q_1^3}{3q_0^3}\right] \qquad \frac{q_1}{q_0} = 1 \qquad \frac{q_1}{q_0^2} - \frac{q_1^3}{3q_0^3} = 0 \tag{13.107}$$

Although the two equations in q_0 and q_1 are *nonlinear*, they are easy enough to solve. Substitution of the first equation into the second yields $q_0 = 3$ and $q_1 = 3$, and the transfer function $H(s)$ becomes

$$H(s) = \frac{K}{3+3s+s^2} \tag{13.108}$$

We choose $K = 3$ for dc gain. The delay $t_g(\omega)$ is given by

$$t_g(\omega) = -\frac{d\phi}{d\omega} = \frac{d}{d\omega}\left[\tan^{-1}\left(\frac{3\omega}{3-\omega^2}\right)\right] = \frac{9+3\omega^2}{(9+3\omega^2)+\omega^4} \tag{13.109}$$

Once again, the delay is a maximally flat function.

This approach becomes quite tedious for higher-order filters due to the nonlinear equations involved (even for the second-order case).

13.7.2 The Continued Fraction Approach

This approach is based on the fact that ideally, the filter output $x_0(t)$ should represent a replica of the input $x_i(t)$ but delayed or shifted by a normalized delay of 1 second such that

$$x_0(t) = x_i(t-1) \tag{13.110}$$

The transfer function of the filter that achieves this must be given by

$$H(s) = \frac{X_0(s)}{X_i(s)} = \frac{X_i(s)e^{-s}}{X_i(s)} = \frac{1}{e^s} \tag{13.111}$$

We approximate e^s to obtain $H(s) = P(s)/Q(s)$ in polynomial form. Now, for a stable, physically realizable filter, it turns out that if $Q(s)$ can be written as the sum of an even polynomial $E(s)$ and an odd polynomial $O(s)$, the ratio $E(s)/O(s)$ results in positive coefficients. We thus seek a form for e^s in terms of odd and even functions. Such a form is given by $e^s = \cosh(s) + \sinh(s)$, where $\cosh(s)$ and $\sinh(s)$ may be approximated by the even and odd series

$$\cosh(s) = 1 + \frac{s^2}{2!} + \frac{s^4}{4!} + \cdots \qquad \sinh(s) = s + \frac{s^3}{3!} + \frac{s^5}{5!} + \cdots \tag{13.112}$$

Their ratio, written as a continued fraction, yields

$$\frac{\cosh(s)}{\sinh(s)} = \frac{E(s)}{O(s)} = \frac{1}{s} + \cfrac{1}{\frac{3}{s} + \cfrac{1}{\frac{5}{s} + \cfrac{1}{\frac{7}{s} + \cfrac{1}{\ddots \; + \frac{2n-1}{s} + \cfrac{1}{\ddots}}}}} \tag{13.113}$$

PGR2026-00033
Krisp Technologies EX1024 Page 448

All terms in the continued fraction expansion are positive. We truncate this expansion to the required order, reassemble $E(s)/O(s)$, and establish the polynomial form $Q(s) = E(s) + O(s)$, and $H(s) = K/Q(s)$. Finally, we choose K so as to normalize the dc gain to unity.

EXAMPLE 13.8 (The Continued Fraction Method)

(a) For a second-order polynomial approximation to unit delay, we have

$$\frac{E(s)}{O(s)} = \frac{1}{s} + \frac{1}{\frac{3}{s}} = \frac{1}{s} + \frac{s}{3} = \frac{3+s^2}{3s}$$

This leads to the polynomial representation

$$Q(s) = E(s) + O(s) = (3+s^2) + (3s) = 3 + 3s + s^2$$

and the transfer function (normalized to unit dc gain)

$$H(s) = \frac{3}{3+3s+s^2}$$

This is identical to the $H(s)$ based on maximal flatness.

(b) A third-order polynomial would require the truncation

$$\frac{E(s)}{O(s)} = \frac{1}{s} + \frac{1}{\frac{3}{s} + \frac{1}{\frac{5}{s}}} = \frac{1}{s} + \frac{1}{\frac{3}{s} + \frac{s}{5}} = \frac{1}{s} + \frac{5s}{15+s^2} = \frac{15+6s^2}{15s+s^3}$$

leading to the polynomial approximation

$$Q(s) = E(s) + O(s) = (15+6s^2) + (15s+s^3) = 15 + 15s + 6s^2 + s^3$$

The transfer function $H(s)$, normalized to unit dc gain, is then

$$H(s) = \frac{15}{15+15s+6s^2+s^3}$$

13.7.3 Bessel Polynomials

The continued fraction approach can actually be developed into systematic procedures for evaluating either $Q(s)$ or its coefficients and results in the family of so-called **Bessel polynomials** or **Thomson polynomials**, all of which result in linear phase over some range and in delays that satisfy the constraint of maximal flatness at the origin.

Table 13.2 shows the first few Bessel polynomials $Q(s)$ in both unfactored and factored form. The numerator K of the transfer function $H(s) = K/Q(s)$ simply equals the constant term in $Q(s)$ for unit dc gain. It is interesting to note that the coefficients of the last two terms in the denominator are always equal (as long as the order exceeds unity).

PGR2026-00033
Krisp Technologies EX1024 Page 449

Table 13.2 Bessel Polynomials

Order n	**Denominator Polynomial** $Q_n(s)$
0	1
1	$s+1$
2	s^2+3s+3
3	$s^3+6s^2+15s+15$
4	$s^4+10s^3+45s^2+105s+105$
5	$s^5+15s^4+105s^3+420s^2+945s+945$
6	$s^6+21s^5+210s^4+1260s^3+4725s^2+10,395s+10,395$

Bessel polynomials may be generated in one of two ways.

1. Starting with $Q_0(s)=1$ and $Q_1(s)=s+1$, we use recursion to generate

$$Q_0(s)=1 \quad Q_1(s)=s+1 \qquad Q_n(s)=(2n-1)Q_{n-1}(s)+s^2Q_{n-2}(s), \quad n>1 \tag{13.114}$$

2. Starting with $Q_n(s)=q_0+q_1s+q_2s^2+\cdots+q_ns^n$, we find the q_k from

$$q_n=1 \qquad q_k=\frac{(2n-k)!}{2^{n-k}\,k!(n-k)!}, \qquad k=0,1,2,\ldots,n-1 \tag{13.115}$$

As a check, the value of q_0 equals the product $(1)(3)(5)\cdots(2n-1)$.

13.7.4 Magnitude and Group Delay from the Transfer Function

One approach for computing the magnitude and delay of Bessel filters is to start with the transfer function

$$H(s)=\frac{Q_n(0)}{Q_n(s)}=\frac{Q_n(0)}{q_0+q_1s+q_2s^2+\cdots+q_ns^n} \tag{13.116}$$

The denominator of the steady-state transfer function $H(\omega)$ may then be written as the sum of an even part $E(\omega)$ and an odd part $O(\omega)$ as

$$H_n(\omega)=\frac{Q_n(0)}{(q_0-q_2\omega^2+q_4\omega^4+\cdots)+j(q_1\omega-q_3\omega^3+q_5\omega^5+\cdots)}=\frac{Q_n(0)}{E_n(\omega)+jO_n(\omega)} \tag{13.117}$$

The magnitude squared function $|H_N(\omega)|^2$ may be expressed as

$$|H_N(\omega)|^2=\frac{Q_n^2(0)}{E_n^2(\omega)+O_n^2(\omega)} \tag{13.118}$$

Using this result, the delay for first- and second-order filters can then be described by

$$t_{g1} = \frac{1}{1+\omega^2}=1-\frac{\omega^2}{1+\omega^2}=1-\frac{\omega^2}{E_1^2+O_1^2}=1-\frac{\omega^2|H(\omega)|^2}{Q_1^2(0)} \tag{13.119}$$

$$t_{g2} = \frac{(9+3\omega^2)}{(9+3\omega^2)+\omega^4}=1-\frac{\omega^4}{(9+3\omega^2)+\omega^4}=1-\frac{\omega^4}{E_2^2+O_2^2}=1-\frac{\omega^4|H(\omega)|^2}{Q_2^2(0)} \tag{13.120}$$

PGR2026-00033
Krisp Technologies EX1024 Page 450

By induction, the delay of an nth-order transfer function equals

$$t_g = 1 - \frac{\omega^{2n}}{E_n^2 + O_n^2} = 1 - \frac{\omega^{2n}|H(\omega)|^2}{Q_n^2(0)} \tag{13.121}$$

13.7.5 Magnitude and Group Delay from Bessel Functions

A second approach to finding magnitude and delay is based on the relationship between *Bessel polynomials* and the more familiar *Bessel functions*. In fact, Bessel polynomials are so named because of this relationship. The **spherical Bessel function of the first kind**, $\mathcal{J}_n(x)$, is related to the integer-order Bessel function $J_n(x)$ by

$$\mathcal{J}(n,x) = \mathcal{J}_n(x) = \sqrt{\frac{\pi}{2x}}\, J_{n+1/2}(x) \tag{13.122}$$

This forms the basis for relating Bessel polynomials to Bessel functions. In particular, the Bessel polynomial $Q_n(1/s)$ may be expressed as

$$Q_n(1/j\omega) = j^{-n}\left[\cos(n\pi)\mathcal{J}_{-n-1}(\omega) - j\mathcal{J}_n(\omega)\right]\omega e^{j\omega} \tag{13.123}$$

The expressions for the magnitude, phase, and group delay of the Bessel filter in terms of Bessel functions derive from this relationship and yield the following results:

$$|H(\omega)|^2 = \frac{Q_n^2(0)}{\omega^{2n+2}[\mathcal{J}_n^2(\omega) + \mathcal{J}_{-n-1}^2(\omega)]} \tag{13.124}$$

$$\phi(\omega) = -\omega + \tan^{-1}\left[\cos(n\pi)\frac{\mathcal{J}_n(\omega)}{\mathcal{J}_{-n-1}(\omega)}\right] \tag{13.125}$$

$$t_g = -\frac{d\phi}{d\omega} = 1 - \frac{1}{\omega^2[\mathcal{J}_n^2(\omega) + \mathcal{J}_{-n-1}^2(\omega)]} = 1 - \frac{\omega^{2n}|H(\omega)|^2}{Q_n^2(0)} \tag{13.126}$$

Magnitude Response

As the filter order increases, the magnitude response of Bessel filters tends to a Gaussian shape as shown in Figure 13.16. On this basis, it turns out that the attenuation A_x at any frequency ω_x for reasonably large filter order ($n > 3$, say) can be approximated by

$$A_x \approx \frac{10\omega_x^2}{(2n-1)\ln 10} \tag{13.127}$$

The frequency ω_p corresponding to a specified passband attenuation A_p equals

$$\omega_p \approx \sqrt{0.1A_p(2n-1)\ln 10} \qquad \omega_{3\text{dB}} \approx \sqrt{(2n-1)\ln 2}, \quad n \geq 3 \tag{13.128}$$

13.7.6 Design Recipe for Bessel Filters

The design of Bessel filters involves finding the order n for the specifications, which can include

- A constant delay t_g (with a prescribed tolerance), and a maximum passband attenuation A_p, or
- A constant delay t_g (to within a prescribed tolerance), and a rise time smaller than T_r. Specifying T_r is equivalent to specifying the half-power frequency ω_3, since the time-bandwidth product $T_r\omega_3$ is a constant, with $T_r\omega_3 \approx 2.2$.

PGR2026-00033
Krisp Technologies EX1024 Page 451

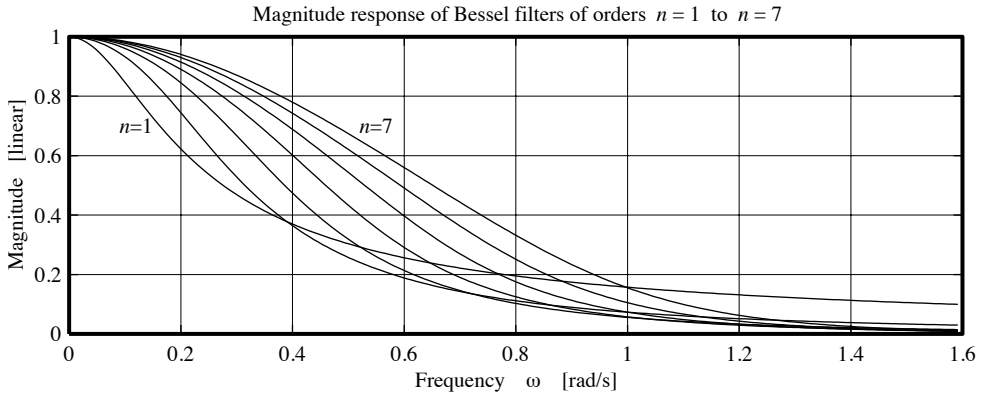


Figure 13.16 The magnitude spectrum of lowpass Bessel filters

A formal design procedure involves the following steps:

1. We normalize the passband frequency ω_p to $\nu = \omega_p t_g$ for unit delay.
2. We either assume a low enough filter order n, or estimate n from

$$A_p \approx \frac{10\nu^2}{(2n-1)\ln 10} \qquad n \approx \frac{5\nu^2}{A_p \ln 10} \tag{13.129}$$

3. We find the normalized transfer function $H_N(s)$ and calculate $A(\nu) = -10\log|H(\nu)|^2$ from either of the following relations

$$|H(\nu)|^2 = \frac{Q_n^2(0)}{\nu^{2n+2}[\mathcal{J}_n^2(\nu) + \mathcal{J}_{-n-1}^2(\nu)]} \qquad \text{or} \qquad |H(\nu)|^2 = \frac{Q_n^2(0)}{E^2(\nu) + O^2(\nu)} \tag{13.130}$$

4. The delay is then computed from

$$t_n = 1 - \frac{\nu^{2n}|H(\nu)|^2}{Q_n^2(0)} \tag{13.131}$$

5. If $1 - t_n$ exceeds the given tolerance, or $A(\nu)$ exceeds A_p, we increase the filter order n by one and repeat steps 3 and 4.
6. Finally, we establish the prototype transfer function $H_P(s)$ and the required filter using $s \to st_g$ to obtain $H(s) = H_P(st_g)$, which provides the required delay t_g.

Another popular but much less flexible approach resorts to nomograms of delay and attenuation for various filter orders to establish the smallest order that best matches specifications.

EXAMPLE 13.9 (Design of a Lowpass Bessel Filter)
We wish to design a Bessel filter with a constant delay of $t_0 = 0.02\,\text{s}$ to within 0.1% up to $\omega = 100$ rad/s and an attenuation of no more than 2.5 dB in the passband.

The normalized frequency for unit delay is $\nu = \omega t_0 = (100)(0.02) = 2$ rad/s.

PGR2026-00033
Krisp Technologies EX1024 Page 452

A first estimate of the filter order is given by

$$n \approx \frac{5\nu^2}{A_p \ln 10} = \frac{5(4)}{2.5 \ln 10} \approx 3.47$$

Thus, we pick $n = 4$.

Next, we compute the actual attenuation and normalized delay. We illustrate by using the Bessel function approach and successively computing:

1. Coefficients of $Q_4(s) = [1,\ 18,\ 45,\ 105,\ 105]$; $\quad Q_4(0) = 105$
2. $\mathcal{J}_n(\omega_n) = \mathcal{J}_4(2) = 0.0141, \quad \mathcal{J}_{-n-1}(\omega_n) = \mathcal{J}_{-5}(2) = 4.4613$
3. $|H(\nu)|^2 = \dfrac{(105)^2}{2^{10}[(0.0141)^2 + (4.4613)^2]} = 0.541 \qquad A_x = -10\log(0.541) = 2.67 \text{ dB}$
4. $t_n = 1 - \dfrac{(0.541)(2^8)}{(105)^2} = 0.987$

Since $1 - t_n > 0.001$ and $A(\nu) > A_p$, neither the attenuation nor the normalized delay meet requirements. We therefore increase n to 5 and recompute:

1. Coefficients of $Q_5(s) = [1,\ 15,\ 105,\ 420,\ 945,\ 945]$; $\qquad Q_5(0) = 945$
2. $\mathcal{J}_n(\omega_n) = \mathcal{J}_5(2) = 0.0026, \qquad \mathcal{J}_{-n-1}(\omega_n) = \mathcal{J}_{-6}(2) = -18.5914$
3. $|H(\nu)|^2 = \dfrac{(945)^2}{2^{14}[(0.0026)^2 + (18.59)^2]} = 0.631 \qquad A_x = -10\log(0.631) = 2.001 \text{ dB}$
4. $t_n = 1 - \dfrac{(0.631)(2^{10})}{(945)^2} = 0.9993$

Since t_n is now within 0.1% of unity and $A_x \approx 2$ dB, we are done and choose $n = 5$.
The prototype transfer function $H_P(s)$ equals

$$H_P(s) = \frac{945}{s^5 + 15s^4 + 105s^3 + 420s^2 + 945s + 945}$$

With $t_0 = 0.02\,\text{s}$, the required transfer function is $H(s) = H_P(st_0) = H_P(s/50)$.
The magnitude and delay of $H(s)$ are shown in Figure E13.9.

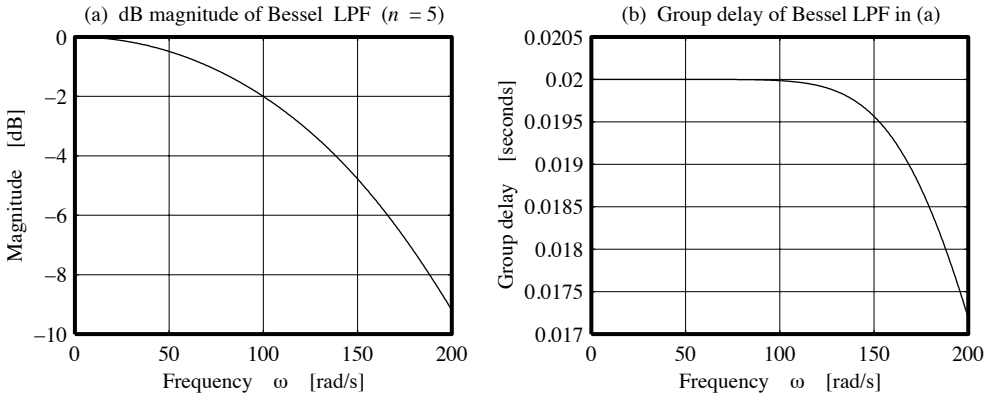


Figure E13.9 Magnitude and delay of Bessel lowpass filter for Example 13.9

PGR2026-00033
Krisp Technologies EX1024 Page 453

CHAPTER 13 PROBLEMS

DRILL AND REINFORCEMENT

13.1 (Lowpass Prototypes) For each set of filter specifications, find the normalized lowpass prototype frequencies. Assume a fixed passband, where necessary.

(a) Lowpass: passband edge at 2 kHz; stopband edge at 3 kHz
(b) Highpass: passband edge at 3 kHz; stopband edge at 2 kHz
(c) Bandpass: passband edges at [10, 15] kHz; stopband edges at [5, 20] kHz
(d) Bandstop: passband edges at [20, 40] kHz; stopband edges at [30, 34] kHz

13.2 (Lowpass Prototypes) For each set of bandpass filter specifications, find the normalized lowpass prototype frequencies. Assume a fixed center frequency f_0 (if given) or a fixed passband (if given) or fixed stopband, in that order.

(a) Passband edges at [10, 14] Hz; stopband edges at [5, 20] Hz; $f_0 = 12$ Hz
(b) Passband edges at [10, 40] Hz; lower stopband edge at 5 Hz
(c) Passband edges at [10, 40] Hz; lower stopband edge at 5 Hz; $f_0 = 25$ Hz
(d) Stopband edges at [5, 50] Hz; lower passband edge at 15 Hz; $f_0 = 25$ Hz

13.3 (Lowpass Prototypes) For each set of bandstop filter specifications, find the normalized lowpass prototype frequencies. Assume a fixed center frequency f_0 (if given) or a fixed passband (if given) or a fixed stopband, in that order.

(a) Passband edges at [20, 40] Hz; stopband edges at [26, 36] Hz; $f_0 = 30$ Hz
(b) Passband edges at [20, 80] Hz; lower stopband edge at 25 Hz
(c) Stopband edges at [40, 80] Hz; lower passband edge at 25 Hz
(d) Passband edges at [20, 60] Hz; lower stopband edge at 30 Hz; $f_0 = 40$ Hz

13.4 (Prototype Transformations) Let $H(s) = \dfrac{1}{s^2 + s + 1}$ describe the transfer function of a lowpass filter with a passband of 1 rad/s. Use frequency transformations to find the transfer function of the following filters.

(a) A lowpass filter with a passband of 10 rad/s
(b) A highpass filter with a cutoff frequency of 1 rad/s
(c) A highpass filter with a cutoff frequency of 10 rad/s
(d) A bandpass filter with a passband of 1 rad/s and a center frequency of 1 rad/s
(e) A bandpass filter with a passband of 10 rad/s and a center frequency of 100 rad/s
(f) A bandstop filter with a stopband of 1 rad/s and a center frequency 1 rad/s
(g) A bandstop filter with a stopband of 2 rad/s and a center frequency of 10 rad/s

13.5 (Butterworth Filter Poles) Sketch the pole locations of a third-order and a fourth-order Butterworth lowpass filter assuming the following information.

(a) A half-power frequency of 1 rad/s
(b) $\epsilon = 0.707$ and a passband edge at 1 rad/s
(c) $\epsilon = 0.707$ and a passband edge at 100 Hz

PGR2026-00033
Krisp Technologies EX1024 Page 454

13.6 (Butterworth Filters) Set up the form of $|H(\nu)|^2$ for a third-order Butterworth lowpass filter with $\epsilon = 0.7$ and find the following:

(a) The attenuation at $\nu = 0.5$ rad/s and $\nu = 2$ rad/s
(b) The half-power frequency ν_3
(c) The pole orientations and magnitudes
(d) The transfer function $H(s)$
(e) The high-frequency decay rate

13.7 (Butterworth Filters) Consider a fifth-order lowpass Butterworth filter with a passband of 1 kHz and a maximum passband attenuation of 1 dB. What is the attenuation, in decibels, of this filter at a frequency of 2 kHz?

13.8 (Butterworth Filters) We wish to design a Butterworth lowpass filter with a 2-dB bandwidth of 1 kHz and an attenuation of at least 25 dB beyond 2 kHz. What is the actual attenuation, in decibels, of the designed filter at $f = 1$ kHz and $f = 2$ kHz?

13.9 (Butterworth Filter Design) Design a Butterworth lowpass filter with a peak gain of 2 that meets the following sets of specifications.

(a) A half-power frequency of 4 rad/s and high-frequency decay rate of 60 dB/dec
(b) A 1-dB passband edge at 10 Hz and filter order $n = 2$
(c) Minimum passband gain of 1.8 at 5 Hz and filter order $n = 2$

13.10 (Butterworth Filter Design) Design a Butterworth filter that meets the following sets of specifications. Assume $A_p = 2$ dB, $A_s = 40$ dB, and a fixed passband where necessary.

(a) Lowpass filter: passband edge at 2 kHz; stopband edge at 6 kHz
(b) Highpass filter: passband edge at 4 kHz; stopband edge at 1 kHz
(c) Lowpass filter: passband edge at 10 Hz; stopband edge at 50 Hz, $f_{3\text{dB}} = 15$ Hz
(d) Bandpass filter: passband edges at [20, 30] Hz; stopband edges at [10, 50] Hz
(e) Bandstop filter: passband edges at [10, 60] Hz; stopband edges at [20, 25] Hz

13.11 (Chebyshev Filter Poles) Analytically evaluate the poles of the following filters.

(a) A third-order Chebyshev lowpass filter with $\epsilon = 0.5$
(b) A fourth-order Chebyshev lowpass filter with a passband ripple of 2 dB

13.12 (Chebyshev Filter Poles) Use a geometric construction to locate the poles of a Chebyshev lowpass filter with the following characteristics.

(a) $\epsilon = 0.4$ and order $n = 3$
(b) A passband ripple of 1 dB and filter order $n = 4$

13.13 (Chebyshev Filters) Set up $|H(\nu)|^2$ for a third-order Chebyshev lowpass prototype with $\epsilon = 0.7$ and find the following:

(a) The attenuation at $\nu = 0.5$ rad/s and $\nu = 2$ rad/s
(b) The half-power frequency ν_3
(c) The pole orientations and magnitudes
(d) The transfer function $H(s)$
(e) The high-frequency decay rate

PGR2026-00033
Krisp Technologies EX1024 Page 455

13.14 (Chebyshev Filters) Consider a fifth-order lowpass Chebyshev filter with a passband of 1 kHz and a passband ripple of 1 dB. What is the attenuation of this filter (in dB) at $f = 1$ kHz and $f = 2$ kHz?

13.15 (Chebyshev Filters) What is the order n, the value of the ripple factor ϵ, and the normalized passband attenuation A_p of the Chebyshev lowpass filter of Figure P13.15?

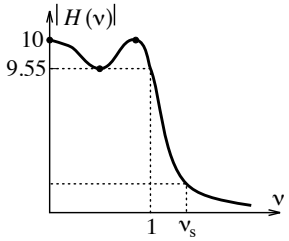


Figure P13.15 Chebyshev lowpass filter for Problem 13.15

13.16 (Chebyshev Filters) A fourth-order Chebyshev filter shows a passband ripple of 1 dB up to 2 kHz. What is the half-power frequency of the lowpass prototype? What is the actual half-power frequency?

13.17 (Chebyshev Filter Design) Design a Chebyshev lowpass filter with a peak gain of unity that meets the following sets of specifications.

(a) A half-power frequency of 4 rad/s and high-frequency decay rate of 60 dB/dec
(b) A 1-dB passband edge at 10 Hz and filter order $n = 2$
(c) A minimum passband gain of 0.9 at the passband edge at 5 Hz and filter order $n = 2$

13.18 (Chebyshev Filter Design) Design a Chebyshev filter to meet the following sets of specifications. Assume $A_p = 2$ dB, $A_s = 40$ dB, and a fixed passband where necessary.

(a) Lowpass filter: passband edge at 2 kHz; stopband edge at 6 kHz
(b) Highpass filter: passband edge at 4 kHz; stopband edge at 1 kHz
(c) Lowpass filter: passband edge at 10 Hz; stopband edge at 50 Hz; $f_{3\text{dB}} = 15$ Hz
(d) Bandpass filter: passband edges at [20, 30] Hz; stopband edges at [10, 50] Hz
(e) Bandstop filter: passband edges at [10, 60] Hz; stopband edges at [20, 25] Hz

13.19 (Inverse Chebyshev Filter Poles) Find the poles and pole locations of a third-order inverse Chebyshev lowpass filter with $\epsilon = 0.1$.

13.20 (Inverse Chebyshev Filters) Set up $|H(\nu)|^2$ for a third-order inverse Chebyshev lowpass filter with $\epsilon = 0.1$ and find the following:

(a) The attenuation at $\nu = 0.5$ rad/s and $\nu = 2$ rad/s
(b) The half-power frequency ν_3
(c) The pole orientations and magnitudes
(d) The transfer function $H(s)$
(e) The high-frequency decay rate

13.21 (Inverse Chebyshev Filters) Consider a fourth-order inverse Chebyshev lowpass filter with a stopband edge of 2 kHz and a minimum stopband ripple of 40 dB. What is the attenuation of this filter, in decibels, at $f = 1$ kHz and $f = 2$ kHz?

13.22 (Inverse Chebyshev Filter Design) Design an inverse Chebyshev filter to meet the following sets of specifications. Assume $A_p = 2$ dB, $A_s = 40$ dB, and a fixed passband where necessary.

PGR2026-00033
Krisp Technologies EX1024 Page 456

(a) Lowpass filter: passband edge at 2 kHz; stopband edge at 6 kHz
(b) Highpass filter: passband edge at 4 kHz; stopband edge at 1 kHz
(c) Lowpass filter: passband edge at 10 Hz; stopband edge at 50 Hz, $f_{3\text{dB}} = 15$ Hz
(d) Bandpass filter: passband edges at [20, 30] Hz; stopband edges at [10, 50] Hz
(e) Bandstop filter: passband edges at [10, 60] Hz; stopband edges at [20, 25] Hz

13.23 (Elliptic Filter Design) Design an elliptic filter that meets the following sets of specifications. Assume $A_p = 2$ dB, $A_s = 40$ dB, and a fixed passband where necessary.

(a) Lowpass filter: passband edge at 2 kHz; stopband edge at 6 kHz
(b) Highpass filter: passband edge at 6 kHz; stopband edge at 2 kHz

13.24 (Bessel Filters) Design a Bessel lowpass filter with a passband edge of 100 Hz and a passband delay of 1 ms with a 1% tolerance for each of the following cases.

(a) The passband edge corresponds to the half-power frequency.
(b) The maximum attenuation in the passband is 0.3 dB.

REVIEW AND EXPLORATION

13.25 (The Maximally Flat Concept) The general form of a maximally flat rational function is

$$R(\nu) = \frac{N(\nu)}{N(\nu) + A\nu^n}$$

where $N(\nu)$ is a polynomial of order less than n.

(a) What is $N(\nu)$ for a Butterworth lowpass filter?
(b) Show that $R(\nu)$ is maximally flat for $n = 2$, using $N(\nu) = b_0 + b_1\nu$.
(c) Show that $R(\nu)$ is maximally flat for $n = 3$, using $N(\nu) = b_1\nu$.

13.26 (Butterworth Filter Design) Design a Butterworth bandpass filter to meet the following sets of specifications. Assume $A_p = 2$ dB and $A_s = 40$ dB. Assume a fixed center frequency (if given) or a fixed passband (if given) or fixed stopband, in that order.

(a) Passband edges at [30, 50] Hz; stopband edges at [5, 400] Hz; $f_0 = 40$ Hz
(b) Passband edges at [20, 40] Hz; lower stopband edge at 5 Hz
(c) Stopband edges at [5, 50] Hz; lower passband edge at 15 Hz; $f_0 = 20$ Hz

13.27 (Butterworth Filter Design) Design a Butterworth bandstop filter to meet the following sets of specifications. Assume $A_p = 2$ dB and $A_s = 40$ dB. Assume a fixed center frequency (if given) or a fixed passband (if given) or fixed stopband, in that order.

(a) Passband edges at [20, 50] Hz; stopband edges at [26, 36] Hz; $f_0 = 30$ Hz
(b) Stopband edges at [30, 100] Hz; lower passband edge at 50 Hz
(c) Stopband edges at [50, 200] Hz; lower passband edge at 80 Hz; $f_0 = 90$ Hz

13.28 (Chebyshev Filter Design) Design a Chebyshev bandpass filter that meets the following sets of specifications. Assume $A_p = 2$ dB and $A_s = 40$ dB. Assume a fixed center frequency (if given) or a fixed passband (if given) or fixed stopband, in that order.

PGR2026-00033
Krisp Technologies EX1024 Page 457

(a) Passband edges at [30, 50] Hz; stopband edges at [5, 400] Hz; $f_0 = 40$ Hz
(b) Passband edges at [20, 40] Hz; lower stopband edge at 5 Hz
(c) Stopband edges at [5, 50] Hz; lower passband edge at 15 Hz; $f_0 = 20$ Hz

13.29 (Chebyshev Filter Design) Design a Chebyshev bandstop filter that meets the following sets of specifications. Assume $A_p = 2$ dB and $A_s = 40$ dB. Assume a fixed center frequency (if given) or a fixed passband (if given) or fixed stopband, in that order.

(a) Passband edges at [20, 50] Hz; stopband edges at [26, 36] Hz; $f_0 = 30$ Hz
(b) Stopband edges at [30, 100] Hz; lower passband edge at 50 Hz
(c) Stopband edges at [50, 200] Hz; lower passband edge at 80 Hz; $f_0 = 90$ Hz

13.30 (Inverse Chebyshev Filter Design) Design an inverse Chebyshev bandpass filter that meets the following sets of specifications. Assume $A_p = 2$ dB and $A_s = 40$ dB. Assume a fixed center frequency (if given) or a fixed passband (if given) or fixed stopband, in that order.

(a) Passband edges at [30, 50] Hz; stopband edges at [5, 400] Hz; $f_0 = 40$ Hz
(b) Passband edges at [20, 40] Hz; lower stopband edge at 5 Hz

13.31 (Inverse Chebyshev Filter Design) Design an inverse Chebyshev bandstop filter that meets the following sets of specifications. Assume $A_p = 2$ dB and $A_s = 40$ dB. Assume a fixed center frequency (if given) or a fixed passband (if given) or fixed stopband, in that order.

(a) Passband edges at [20, 50] Hz; stopband edges at [26, 36] Hz; $f_0 = 30$ Hz
(b) Stopband edges at [30, 100] Hz; lower passband edge at 50 Hz

13.32 (Group Delay) Find a general expression for the group delay of a second-order lowpass filter described by the transfer function $H(s) = \dfrac{K}{s^2 + As + B}$. Use this result to compute the group delay of the following filters. Do any of the group delays show a maximally flat form?

(a) A second-order normalized Butterworth lowpass filter
(b) A second-order normalized Chebyshev lowpass filter with $A_p = 1$ dB
(c) A second-order normalized Bessel lowpass filter

13.33 (Identifying Analog Filters) Describe how you would go about identifying a lowpass transfer function $H(s)$ as a Butterworth, Chebyshev I, Chebyshev II, or elliptic filter. Consider both the pole locations (and orientations) and the zero locations as possible clues.

COMPUTATION AND DESIGN

afdgui A GUI for Analog Filter Design and Visualization

The graphical user interface `afdgui` allows you to design analog filters and visualize their magnitude and phase spectrum.
You can select various forms (highpass, lowpass, etc.) and types (Butterworth, elliptic, etc.).
The plots are displayed along with the filter order and design attenuation at the band edges.
To explore this routine, type `afdgui` at the MATLAB prompt.

PGR2026-00033
Krisp Technologies EX1024 Page 458

13.34 (Subsonic Filters) Some audio amplifiers or equalizers include a subsonic filter to remove or reduce unwanted low-frequency noise (due, for example, to warped records) that might otherwise modulate the audible frequencies causing intermodulation distortion. The audio equalizer of one manufacturer is equipped with a subsonic filter that is listed as providing an 18-dB/octave cut for frequencies below 15 Hz. Sketch the Bode plot of such a filter. Design a Butterworth filter that can realize the design specifications.

13.35 (A Multi-Band Filter) A requirement exists for a multi-band analog filter with the following specifications.

Passband 1: From dc to 20 Hz
Peak gain = 0 dB
Maximum passband attenuation = 1 dB (from peak)
Minimum attenuation at 40 Hz = 45 dB (from peak)
No ripples allowed in the passband or stopband

Passband 2: From 150 Hz to 200 Hz
Peak gain = 12 dB
Maximum passband ripple = 1 dB
Minimum attenuation at 100 Hz and 300 Hz = 50 dB (from peak)
Ripples permitted in the passband but not in the stopband

(a) Design each filter separately using the ADSP routine `afd`. Then adjust the gain(s) and combine the two stages to obtain the filter transfer function $H(s)$. To adjust the gain, you need to scale *only* the numerator array of a transfer function by the *linear* (not dB) gain.

(b) Plot the filter magnitude spectrum of $H(s)$ over $0 \leq f \leq 400$ Hz in increments of 1 Hz. Tabulate the attenuation in dB at $f = 20, 40, 100, 150, 200, 300$ Hz and compare with expected design values. What is the *maximum* attenuation in dB in the range [0, 400] Hz, and at what frequency does it occur?

(c) Find the poles and zeros of $H(s)$. Is the filter stable (with *all* its poles in the left half of the s-plane)? Should it be? Is it minimum phase (with *all* its poles and zeros in the left half of the s-plane)? Should it be? If not, use the ADSP routine `minphase` to generate the minimum-phase transfer function $H_M(s)$. Check the poles and zeros of $H_M(s)$ to verify that it is a stable, minimum-phase filter.

(d) Plot the magnitude of $H(s)$ and $H_M(s)$ in dB against the frequency on a log scale on the same plot. Are the magnitudes identical? Should they be? Explain. Use the ADSP routine `bodelin` to obtain a Bode plot from $H_M(s)$. Can you identify the break frequencies from the plot? How does the asymptotic Bode magnitude plot compare with the actual magnitude in dB?

PGR2026-00033
Krisp Technologies EX1024 Page 459

Chapter 14

SAMPLING AND QUANTIZATION

14.0 Scope and Objectives

Sampling and quantization form the critical link between analog and digital signals. This chapter begins with various sampling operations and develops the key concept that the spectrum of ideally sampled signals is periodic. It examines the sampling theorem, which allows us to sample band-limited signals without loss of information, and describes schemes for the recovery of analog signals from their samples using idealized and practical methods. It then provides an introduction to signal quantization and its effects and concludes with some applications.

14.1 Ideal Sampling

Ideal sampling describes a sampled signal as a weighted sum of impulses, the weights being equal to the values of the analog signal at the impulse locations. An ideally sampled signal $x_I(t)$ may be regarded as the product of an analog signal $x(t)$ and a periodic impulse train $i(t)$, as illustrated in Figure 14.1.

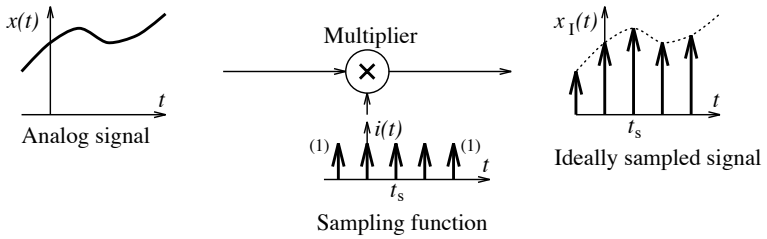


Figure 14.1 The ideal sampling operation

The ideally sampled signal may be mathematically described as

$$x_I(t) = x(t)i(t) = x(t)\sum_{n=-\infty}^{\infty}\delta(t-nt_s) = \sum_{n=-\infty}^{\infty}x(nt_s)\delta(t-nt_s) = \sum_{n=-\infty}^{\infty}x[n]\delta(t-nt_s) \tag{14.1}$$

Here, the discrete signal $x[n]$ simply represents the sequence of sample values $x(nt_s)$. Clearly, the sampling operation leads to a potential loss of information in the ideally sampled signal $x_I(t)$, when compared with its underlying analog counterpart $x(t)$. The smaller the sampling interval t_s, the less is this loss of information.

PGR2026-00033
Krisp Technologies EX1024 Page 460

Intuitively, there must always be some loss of information, no matter how small an interval we use. Fortunately, our intuition notwithstanding, it is indeed possible to sample signals without any loss of information. The catch is that the signal $x(t)$ must be band-limited to some finite frequency B.

REVIEW PANEL 14.1

The Ideally Sampled Signal Is a Train of Impulses

$$x_I(t) \;=\; x(t)\sum_{n=-\infty}^{\infty}\delta(t-nt_s) \;=\; \sum_{n=-\infty}^{\infty}x(nt_s)\delta(t-nt_s) = \sum_{n=-\infty}^{\infty}x[n]\delta(t-nt_s)$$

The discrete signal $x[n]$ corresponds to the sequence of the sample values $x(nt_s)$.

The spectra associated with the various signals in ideal sampling are illustrated in Figure 14.2. The impulse train $i(t)$ is a periodic signal with period $T = t_s = 1/S$ and Fourier series coefficients $I[k] = S$. Its Fourier transform is a train of impulses (at $f = kS$) whose strengths equal $I[k]$.

$$I(f) = \sum_{k=-\infty}^{\infty} I[k]\delta(f-kS) = S\sum_{k=-\infty}^{\infty}\delta(f-kS) \tag{14.2}$$

The ideally sampled signal $x_I(t)$ is the product of $x(t)$ and $i(t)$. Its spectrum $X_I(f)$ is thus described by the convolution

$$X_I(f) = X(f)\star I(f) = X(f)\star S\sum_{k=-\infty}^{\infty}\delta(f-kS) = S\sum_{k=-\infty}^{\infty}X(f-kS) \tag{14.3}$$

The spectrum $X_I(f)$ consists of $X(f)$ and its shifted replicas or **images**. It is periodic in frequency, with a period that equals the sampling rate S.

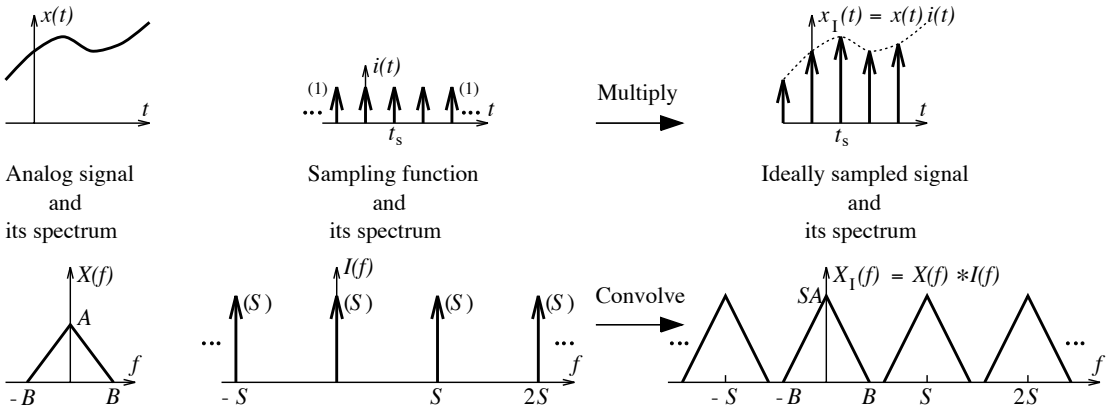


Figure 14.2 Spectra of the signals for ideal sampling

REVIEW PANEL 14.2

The Spectrum of an Ideally Sampled Signal Is Periodic with Period S

$$X_I(f) \;=\; X(f)\star S\sum_{k=-\infty}^{\infty}\delta(f-kS) \;=\; S\sum_{k=-\infty}^{\infty}X(f-kS)$$

The spectrum is the periodic extension of $SX(f)$ with period S.

PGR2026-00033
Krisp Technologies EX1024 Page 461

Since the spectral image at the origin extends over $(-B, B)$, and the next image (centered at S) extends over $(S - B, S + B)$, the images will not overlap if

$$S - B > B \qquad \text{or} \qquad S > 2B \tag{14.4}$$

Figure 14.3 illustrates the spectra of an ideally sampled band-limited signal for three choices of the sampling frequency S. As long as the images do not overlap, each period is a replica of the scaled analog signal spectrum $SX(f)$. We can thus extract $X(f)$ (and hence $x(t)$) as the principal period of $X_I(f)$ (between $-0.5S$ and $0.5S$) by passing the ideally sampled signal through an ideal lowpass filter with a cutoff frequency of $0.5S$ and a gain of $1/S$ over the frequency range $-0.5S \leq f \leq 0.5S$.

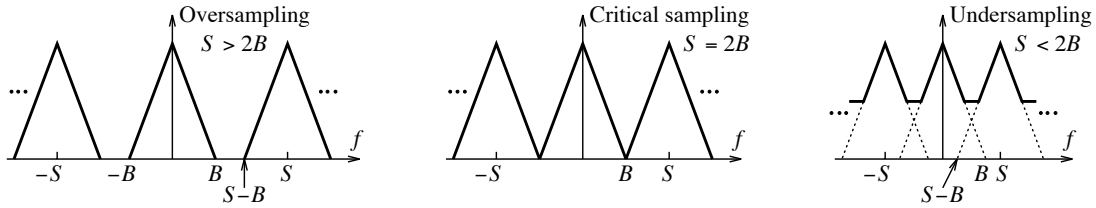


Figure 14.3 Spectrum of an ideally sampled signal for three choices of the sampling frequency

This is the celebrated **sampling theorem**, which tells us that an analog signal band-limited to a frequency B can be sampled without loss of information if the sampling rate S exceeds $2B$ (or the sampling interval t_s is smaller than $\frac{1}{2B}$). The *critical* sampling rate $S_N = 2B$ is often called the **Nyquist rate** or **Nyquist frequency** and the critical sampling interval $t_N = 1/S_N = 1/2B$ is called the **Nyquist interval**.

REVIEW PANEL 14.3

Band-Limited Analog Signals Can Be Sampled Without Loss of Information

If $x(t)$ is band-limited to B, it must be sampled at $S > 2B$ to prevent loss of information.
The images of $X(f)$ do not overlap in the periodic spectrum $X_I(f)$ of the ideally sampled signal.
We can recover $x(t)$ *exactly* from the principal period $(-0.5S,\ 0.5S)$, using an ideal lowpass filter.

If the sampling rate S is less than $2B$, the spectral images overlap and the principal period $(-0.5S,\ 0.5S)$ of $X_I(f)$ is no longer an exact replica of $X(f)$. In this case, we cannot exactly recover $x(t)$, and there is loss of information due to **undersampling**. Undersampling results in spectral overlap. Components of $X(f)$ outside the principal range $(-0.5S,\ 0.5S)$ fold back into this range (due to the spectral overlap from adjacent images). Thus, frequencies higher than $0.5S$ appear as *lower frequencies* in the principal period. This is *aliasing*. The frequency $0.5S$ is also called the **folding frequency**.

REVIEW PANEL 14.4

Undersampling Causes Spectral Overlap, Aliasing, and Irreversible Loss of Information

If $S < 2B$, the images of $X(f)$ overlap in the periodic spectrum and we cannot recover $x(t)$.
Aliasing: A frequency $|f_0| > 0.5S$ gets *aliased* to a lower frequency f_a in the range $(-0.5S,\ 0.5S)$.

Sampling is a band-limiting operation in the sense that in practice we typically extract only the principal period of the spectrum, which is band-limited to the frequency range $(-0.5S,\ 0.5S)$. Thus, the highest frequency we can recover or identify is $0.5S$ and depends only on the sampling rate S.

PGR2026-00033
Krisp Technologies EX1024 Page 462

EXAMPLE 14.1 (The Nyquist Rate)
Let the signal $x_1(t)$ be band-limited to 2 kHz and $x_2(t)$ be band-limited to 3 kHz. Using properties of the Fourier transform, we find the Nyquist rate for the following signals.

The spectrum of $x_1(2t)$ (time compression) stretches to 4 kHz. Thus, $S_N = 8$ kHz.
The spectrum of $x_2(t-3)$ extends to 3 kHz (a time shift changes only the phase). Thus, $S_N = 6$ kHz.
The spectrum of $x_1(t) + x_2(t)$ (sum of the spectra) extends to 3 kHz. Thus, $S_N = 6$ kHz.
The spectrum of $x_1(t)x_2(t)$ (convolution in the frequency domain) extends to 5 kHz. Thus, $S_N = 10$ kHz.
The spectrum of $x_1(t) \star x_2(t)$ (product of the spectra) extends only to 2 kHz. Thus, $S_N = 4$ kHz.
The spectrum of $x_1(t)\cos(1000\pi t)$ (modulation) is stretched by 500 Hz to 2.5 kHz. Thus, $S_N = 5$ kHz.

14.1.1 Sampling of Sinusoids and Periodic Signals

The Nyquist frequency for a sinusoid $x(t) = \cos(2\pi f_0 t + \theta)$ is $S_N = 2f_0$. The Nyquist interval is $t_N = 1/2f_0$, or $t_N = T/2$. This amounts to taking *more than* two samples per period. If, for example, we acquire just two samples per period, starting at a zero crossing, all sample values will be zero, and will yield no information.

If a signal $x(t) = \cos(2\pi f_0 t + \theta)$ is sampled at S, the sampled signal is $x[n] = \cos(2\pi n f_0/S + \theta)$. Its spectrum is periodic, with principal period $(-0.5S,\ 0.5S)$. If $f_0 < 0.5S$, there is no aliasing, and the principal period shows a pair of impulses at $\pm f_0$ (with strength 0.5). If $f_0 > 0.5S$, we have aliasing. The components at $\pm f_0$ are aliased to a lower frequency $\pm f_a$ in the principal range. To find the aliased frequency $|f_a|$, we subtract integer multiples of the sampling frequency from f_0 until the result $f_a = f_0 - NS$ lies in the principal range $(-0.5S,\ 0.5S)$. The spectrum then describes a sampled version of the *lower-frequency* aliased signal $x_a(t) = \cos(2\pi f_a t + \theta)$. The relation between the aliased frequency, original frequency, and sampling frequency is illustrated in Figure 14.4. The aliased frequency always lies in the principal range.

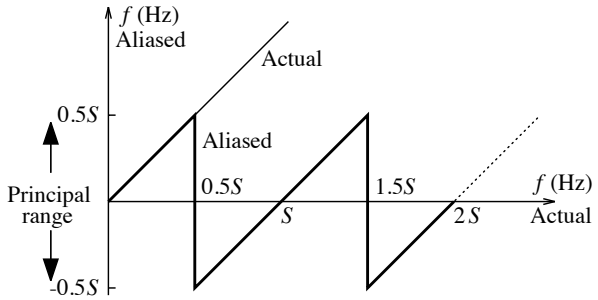


Figure 14.4 Relation between the actual and aliased frequency

REVIEW PANEL 14.5
Finding Aliased Frequencies and Aliased Signals
The signal $x(t) = \cos(2\pi f_0 t + \theta)$ is recovered as $x_a(t) = \cos(2\pi f_a t + \theta)$ if $S < 2f_0$.
$f_a = f_0 - NS$, where N is an integer that places f_a in the principal period $(-0.5S,\ 0.5S)$.

A periodic signal $x_p(t)$ with period T can be described by a sum of sinusoids at the fundamental frequency $f_0 = 1/T$ and its harmonics kf_0. In general, such a signal may not be band-limited and cannot be sampled without aliasing for any choice of sampling rate.

PGR2026-00033
Krisp Technologies EX1024 Page 463

EXAMPLE 14.2 (Aliasing and Sampled Sinusoids)

(a) Consider the sinusoid $x(t) = A\cos(2\pi f_0 t + \theta)$ with $f_0 = 100$ Hz.

1. If $x(t)$ is sampled at $S = 300$ Hz, no aliasing occurs, since $S > 2f_0$.
2. If $x(t)$ is sampled at $S = 80$ Hz, we obtain $f_a = f_0 - S = 20$ Hz, The sampled signal describes a sampled version of the aliased signal $A\cos[2\pi(20)t + \theta]$.
3. If $S = 60$ Hz, we obtain $f_0 - 2S = -20$ Hz. The aliased sinusoid corresponds to $A\cos[2\pi(-20)t + \theta] = A\cos[2\pi(20)t - \theta]$. Note the phase reversal.

(b) Let $x_p(t) = 8\cos(2\pi t) + 6\cos(8\pi t) + 4\cos(22\pi t) + 6\sin(32\pi t) + \cos(58\pi t) + \sin(66\pi t)$.

If it is sampled at $S = 10$ Hz, the last four terms will be aliased. The reconstruction of the sampled signal will describe an analog signal whose first two terms are identical to $x_p(t)$, but whose other components are at the aliased frequencies. The following table shows what to expect.

f_0 (Hz)	**Aliasing?**	**Aliased Frequency** f_a	**Analog Equivalent**
1	No ($f_0 < 0.5S$)	No aliasing	$8\cos(2\pi t)$
4	No ($f_0 < 0.5S$)	No aliasing	$6\cos(8\pi t)$
11	Yes	$11 - S = 1$	$4\cos(2\pi t)$
16	Yes	$16 - 2S = -4$	$6\sin(-8\pi t) = -6\sin(8\pi t)$
29	Yes	$29 - 3S = -1$	$\cos(-2\pi t) = \cos(2\pi t)$
33	Yes	$33 - 3S = 3$	$\sin(6\pi t)$

The reconstructed signal corresponds to $x_S(t) = 13\cos(2\pi t) + \sin(6\pi t) + 6\cos(8\pi t) - 6\sin(8\pi t)$, which cannot be distinguished from $x_p(t)$ at the sampling instants $t = nt_s$, where $t_s = 0.1$ s. To avoid aliasing and recover $x_p(t)$, we must choose $S > 2B = 66$ Hz.

(c) Suppose we sample a sinusoid $x(t)$ at 30 Hz and obtain the periodic spectrum of the sampled signal, as shown in Figure E14.2C. Is it possible to uniquely identify $x(t)$?

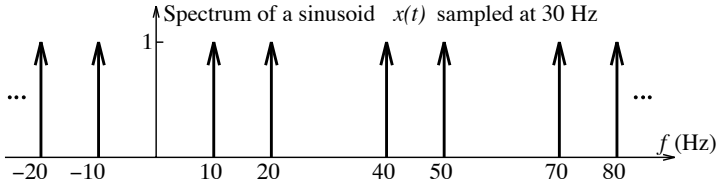


Figure E14.2C Spectrum of sampled sinusoid for Example 14.2(c)

We can certainly identify the period as 30 Hz, and thus $S = 30$ Hz. But we cannot *uniquely* identify $x(t)$ because it could be a sinusoid at 10 Hz (with no aliasing) or a sinusoid at 20 Hz, 50 Hz, 80 Hz, etc. (all aliased to 10 Hz). However, the analog signal $y(t)$ reconstructed from the samples will describe a 10-Hz sinusoid because reconstruction extracts only the principal period, $(-15,\ 15)$ Hz, of the periodic spectrum. In the absence of any a priori information, we almost invariably use the *principal period* as a means to uniquely identify the underlying signal from its periodic spectrum, for better or worse.

PGR2026-00033
Krisp Technologies EX1024 Page 464

14.1.2 Application Example: The Sampling Oscilloscope

The sampling theorem tells us that in order to sample an analog signal without aliasing and loss of information, we must sample above the Nyquist rate. However, some applications depend on the very act of aliasing for their success. One example is the sampling oscilloscope. A conventional oscilloscope cannot directly display a waveform whose bandwidth exceeds the bandwidth of the oscilloscope. However, if the signal is periodic (or can be repeated periodically) and band-limited, a new, time-expanded waveform can be built up and displayed by sampling the periodic signal at successively later instants in successive periods. The idea is illustrated in Figure 14.5 for a periodic signal $x(t) = 1 + \cos(2\pi f_0 t)$ with fundamental frequency f_0.

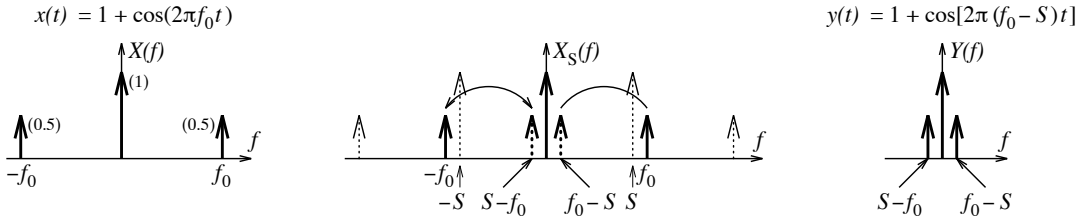


Figure 14.5 The principle of the sampling oscilloscope

If the sampling rate is chosen to be less than f_0 (i.e., $S < f_0$), the spectral component at f_0 will alias to the smaller frequency $f_a = f_0 - S$. To ensure no phase reversal, the aliased frequency must be less than $0.5S$. Thus, $f_0 - S < 0.5S$ or $S > f_0/1.5$. The choice of sampling frequency S is thus bounded by $\frac{f_0}{1.5} < S < f_0$. Subsequent recovery by a lowpass filter with a cutoff frequency of $f_C = 0.5S$ will yield the signal $y(t) = 1 + \cos[2\pi(f_0 - S)t]$, which represents a time-stretched version of $x(t)$. With $y(t) = x(t/\alpha)$, the stretching factor is $\alpha = f_0/(f_0 - S)$. The closer S is to the fundamental frequency f_0, the larger is the stretching factor and the more slowed down the signal $y(t)$. This reconstructed signal $y(t)$ is what is displayed on the oscilloscope (with appropriate scale factors to reflect the parameters of the original signal). As an example, if we wish to see a 100-MHz sinusoid slowed down to 2 MHz, then $f_0 - S = 2\,\text{MHz}$, $S = 98\,\text{MHz}$, $\alpha = 50$, and $f_C = 49\,\text{MHz}$. We may use an even lower sampling rate $S_L = S/L$, where L is an integer, as long as the aliased component appears at 2 MHz and shows no phase reversal. For example, $S_L = 49\,\text{MHz}$ (for $L = 2$), $S_L = 14\,\text{MHz}$ (for $L = 7$) or $S_L = 7\,\text{MHz}$ (for $L = 14$) will all yield a 2-MHz aliased signal with no phase reversal. The only disadvantage is that at these lower sampling rates we acquire less than one sample per period and must wait much longer in order to acquire enough samples to build up a one-period display.

More generally, if a periodic signal $x(t)$ has a fundamental frequency of f_0 and is band-limited to the Nth harmonic frequency Nf_0, the sampling rate must satisfy $\frac{Nf_0}{N+0.5} < S < f_0$, while the stretching factor remains $\alpha = f_0/(f_0 - S)$. For the same stretching factor, the sampling rate may also be reduced to $S_L = S/L$ (where L is an integer) as long as none of the aliased components shows phase reversal.

14.1.3 Sampling of Bandpass Signals

The spectrum of **baseband** signals includes the origin whereas the spectrum of **bandpass signals** occupies a range of frequencies between f_L and f_H, where f_L is greater than zero. The quantity $B = f_H - f_L$ is a measure of the bandwidth of the bandpass signal. Even though a Nyquist rate of $2f_H$ can be used to recover such signals, we can often get by with a lower sampling rate. This is especially useful for narrowband signals (or modulated signals) centered about a very high frequency. To retain all the information in a bandpass signal, we actually need a sampling rate that aliases the entire spectrum to a lower frequency range without overlap. The smallest such frequency is $S = 2f_H/N$, where $N = \text{int}(f_H/B)$ is the integer part of f_H/B.

PGR2026-00033
Krisp Technologies EX1024 Page 465

For recovery, we must sample at exactly S. However, other choices also result in bounds on the sampling frequency S. These bounds are given by

$$\frac{2f_H}{k} \le S \le \frac{2f_L}{k-1}, \qquad k = 1, 2, \ldots, N \tag{14.5}$$

Here, the integer k can range from 1 to N, with $k = N$ yielding the smallest value $S = 2f_H/N$, and $k = 1$ corresponding to the Nyquist rate $S \ge 2f_H$.

EXAMPLE 14.3 (Sampling of Bandpass Signals)
Consider a bandpass signal $x_C(t)$ with $f_L = 4$ kHz and $f_H = 6$ kHz. Then $B = f_H - f_L = 2$ kHz and we compute $N = \text{int}(f_H/B) = 3$. Possible choices for the sampling rate S (in kHz) are given by

$$\frac{12}{k} \le S \le \frac{8}{k-1}, \qquad k = 1,\ 2,\ 3$$

This yields the following bounds on S

k	**Bounds on S**	**Remarks**
3	$S = 4$ kHz	The smallest S; requires a bandpass filter for recovery
2	6 kHz $\le S \le$ 8 kHz	Requires bandpass filter for recovery
1	$S \ge 12$ kHz	No aliasing; we can use a lowpass filter for recovery

The signal and the spectra of the sampled signal are shown in Figure E14.3 for $S = 4,\ 7,\ 14$ kHz.

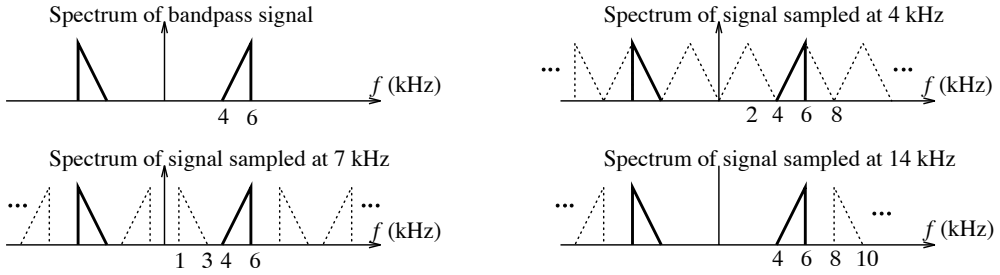


Figure E14.3 Spectra of bandpass signal and its sampled versions for Example 14.3

14.1.4 Natural Sampling

Conceptually, natural sampling, is equivalent to passing the signal $x(t)$ through a switch that opens and closes every t_s seconds. The action of the switch can be modeled as a periodic pulse train $p(t)$ of unit height, with period t_s and pulse width t_d. The sampled signal $x_N(t)$ equals $x(t)$ for the t_d seconds that the switch remains closed and is zero when the switch is open, as illustrated in Figure 14.6.

The naturally sampled signal is described by $x_N(t) = x(t)p(t)$. The Fourier transform $P(f)$ is a train of impulses whose strengths $P[k]$ equal the Fourier series coefficients of $p(t)$:

$$P(f) = \sum_{k=-\infty}^{\infty} P[k]\delta(f - kS) \tag{14.6}$$

The spectrum $X_N(f)$ of the sampled signal $x_S(t) = x(t)p(t)$ is described by the convolution

$$X_N(f) = X(f) \star P(f) = X(f) \star \sum_{k=-\infty}^{\infty} P[k]\delta(f - kS) = \sum_{k=-\infty}^{\infty} P[k]X(f - kS) \tag{14.7}$$

PGR2026-00033
Krisp Technologies EX1024 Page 466

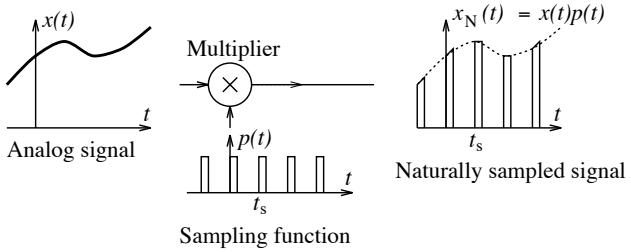


Figure 14.6 Natural sampling

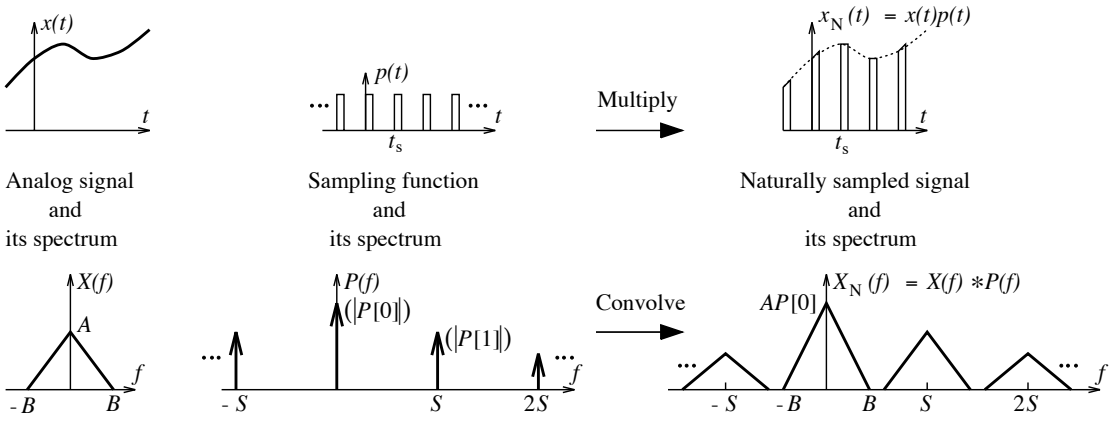


Figure 14.7 Spectra of the signals for natural sampling

The various spectra for natural sampling are illustrated in Figure 14.7. Again, $X_N(f)$ is a superposition of $X(f)$ and its amplitude-scaled (by $P[k]$), shifted replicas. Since the $P[k]$ decay as $1/k$, the spectral images get smaller in height as we move away from the origin. In other words, $X_N(f)$ does not describe a periodic spectrum. However, if there is no spectral overlap, the image centered at the origin equals $P[0]X(f)$, and $X(f)$ can still be recovered by passing the sampled signal through an ideal lowpass filter with a cutoff frequency of $0.5S$ and a gain of $1/P[0]$, $-0.5S \leq f \leq 0.5S$, where $P[0]$ is the dc offset in $p(t)$. In theory, natural sampling can be performed by any periodic signal with a nonzero dc offset.

REVIEW PANEL 14.6

Naturally Sampling Uses a Periodic Pulse Train $p(t)$

$$x_N(t) = x(t)p(t) \qquad X_N(f) = \sum_{k=-\infty}^{\infty} P[k]X(f - kS) \qquad P[k] = \frac{1}{t_s}\int_{t_s} p(t)e^{-j2\pi kSt}\,dt$$

14.1.5 Zero-Order-Hold Sampling

In practice, analog signals are sampled using **zero-order-hold** (ZOH) devices that hold a sample value constant until the next sample is acquired. This is also called **flat-top sampling**. This operation is

PGR2026-00033
Krisp Technologies EX1024 Page 467

equivalent to ideal sampling followed by a system whose impulse response $h(t) = \text{rect}[(t - 0.5t_s)/t_s]$ is a pulse of unit height and duration t_s (to stretch the incoming impulses). This is illustrated in Figure 14.8.

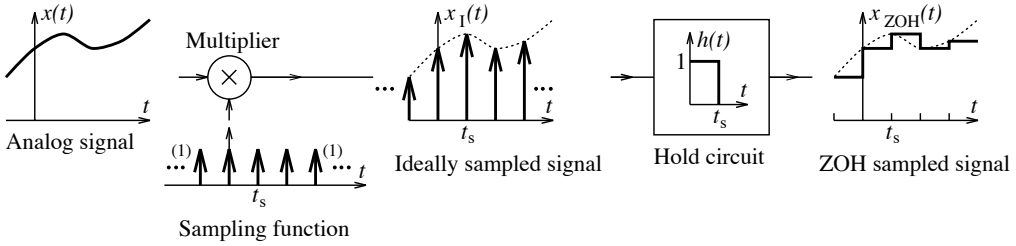


Figure 14.8 Zero-order-hold sampling

The sampled signal $x_{\text{ZOH}}(t)$ can be regarded as the convolution of $h(t)$ and an ideally sampled signal:

$$x_{\text{ZOH}}(t) = h(t) \star x_I(t) = h(t) \star \left[\sum_{n=-\infty}^{\infty} x(nt_s)\delta(t - nt_s) \right] \tag{14.8}$$

The transfer function $H(f)$ of the zero-order-hold circuit is the sinc function

$$H(f) = t_s \,\text{sinc}(ft_s)e^{-j\pi f t_s} = \frac{1}{S}\,\text{sinc}\Big(\frac{f}{S}\Big)e^{-j\pi f/S} \tag{14.9}$$

Since the spectrum of the ideally sampled signal is $S\sum X(f - kS)$, the spectrum of the zero-order-hold sampled signal $x_{\text{ZOH}}(t)$ is given by the product

$$X_{\text{ZOH}}(f) = \text{sinc}\Big(\frac{f}{S}\Big)e^{-j\pi f/S} \sum_{k=-\infty}^{\infty} X(f - kS) \tag{14.10}$$

This spectrum is illustrated in Figure 14.9. The term $\text{sinc}(f/S)$ attenuates the spectral images $X(f - kS)$ and causes *sinc distortion*. The higher the sampling rate S, the less is the distortion in the spectral image $X(f)$ centered at the origin.

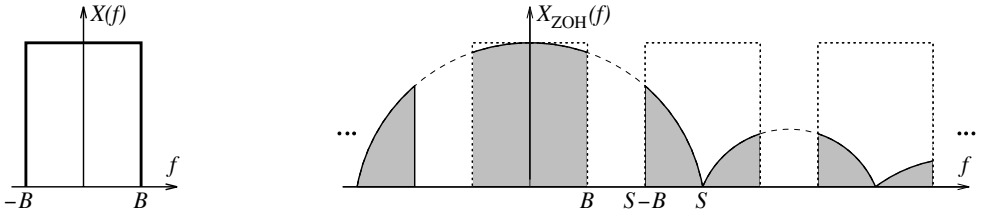


Figure 14.9 Spectrum of a zero-order-hold sampled signal

REVIEW PANEL 14.7

The Zero-Order-Hold (ZOH) Sampled Signal Is a Staircase Approximation of $x(t)$

ZOH sampling: Ideal sampling followed by a *hold* system with $h(t) = \text{rect}[(t - 0.5t_s)/t_s]$

$$x_{\text{ZOH}}(t) \;=\; h(t) \star \left[\sum_{n=-\infty}^{\infty} x(nt_s)\delta(t - nt_s) \right] \qquad X_{\text{ZOH}}(f) = \left[\text{sinc}\Big(\frac{f}{S}\Big) \sum_{k=-\infty}^{\infty} X(f - kS) \right] e^{-j\pi f/S}$$

PGR2026-00033
Krisp Technologies EX1024 Page 468

An ideal lowpass filter with unity gain over $-0.5S \le f \le 0.5S$ recovers the *distorted* signal

$$\tilde{X}(f) = X(f)\text{sinc}\Big(\frac{f}{S}\Big)e^{-j\pi f/S}, \qquad -0.5S \le f \le 0.5S \tag{14.11}$$

To recover $X(f)$ with no amplitude distortion, we must use a *compensating* filter that negates the effects of the sinc distortion by providing a concave-shaped magnitude spectrum corresponding to the reciprocal of the sinc function over the principal period $|f| \le 0.5S$, as shown in Figure 14.10.



Figure 14.10 Spectrum of a filter that compensates for sinc distortion

The magnitude spectrum of the compensating filter is given by

$$|H_r(f)| = \frac{1}{\text{sinc}(f/S)}, \qquad |f| \le 0.5S \tag{14.12}$$

EXAMPLE 14.4 (Sampling Operations)

(a) The signal $x(t) = \text{sinc}(4000t)$ is sampled at a sampling frequency $S = 5$ kHz. The Fourier transform of $x(t) = \text{sinc}(4000t)$ is $X(f) = (1/4000)\text{rect}(f/4000)$ (a rectangular pulse over ± 2 kHz). The signal $x(t)$ is band-limited to $f_B = 2$ kHz. The spectrum of the ideally sampled signal is periodic with replication every 5 kHz, and may be written as

$$X_I(f) = \sum_{k=-\infty}^{\infty} SX(f - kS) = 1.25 \sum_{k=-\infty}^{\infty} \text{rect}\Big(\frac{f - 5000k}{4000}\Big)$$

(b) The magnitude spectrum of the zero-order-hold sampled signal is a version of the ideally sampled signal distorted (multiplied) by $\text{sinc}(f/S)$, and described by

$$X_{\text{ZOH}}(f) = \text{sinc}\Big(\frac{f}{S}\Big) \sum_{k=-\infty}^{\infty} X(f - kS) = 1.25\,\text{sinc}(0.0002f) \sum_{k=-\infty}^{\infty} \text{rect}\Big(\frac{f - 5000k}{4000}\Big)$$

(c) The spectrum of the naturally sampled signal, assuming a rectangular sampling pulse train $p(t)$ with unit height, and a duty ratio of 0.5, is given by

$$X_N(f) = \sum_{k=-\infty}^{\infty} P[k]X(f - kS) = \frac{1}{4000} \sum_{k=-\infty}^{\infty} P[k]\text{rect}\Big(\frac{f - 5000k}{4000}\Big)$$

Here, $P[k]$ are the Fourier series coefficients of $p(t)$, with $P[k] = 0.5\,\text{sinc}(0.5k)$.

PGR2026-00033
Krisp Technologies EX1024 Page 469

14.2 Sampling, Interpolation, and Signal Recovery

For a sampled sequence obtained from an analog signal $x(t)$, an important aspect is the recovery of the original signal from its samples. This requires "filling in" the missing details, or interpolating between the sampled values. The nature of the interpolation that recovers $x(t)$ may be discerned by considering the sampling operation in the time domain. Of the three sampling operations considered earlier, only ideal (or impulse) sampling leads to a truly discrete signal $x[n]$ whose samples equal the strengths of the impulses $x(nt_s)\delta(t - nt_s)$ at the sampling instants nt_s. This is the only case we pursue.

14.2.1 Ideal Recovery and the Sinc Interpolating Function

The ideally sampled signal $x_I(t)$ is the product of the impulse train $i(t) = \sum \delta(t - nt_s)$ and the analog signal $x(t)$ and may be written as

$$x_I(t) = x(t) \sum_{n=-\infty}^{\infty} \delta(t - nt_s) = \sum_{n=-\infty}^{\infty} x(nt_s)\delta(t - nt_s) = \sum_{n=-\infty}^{\infty} x[n]\delta(t - nt_s) \tag{14.13}$$

The discrete signal $x[n]$ is just the sequence of samples $x(nt_s)$. We can recover $x(t)$ by passing $x_I(t)$ through an ideal lowpass filter with a gain of t_s and a cutoff frequency of $0.5S$. The frequency-domain and time-domain equivalents are illustrated in Figure 14.11.

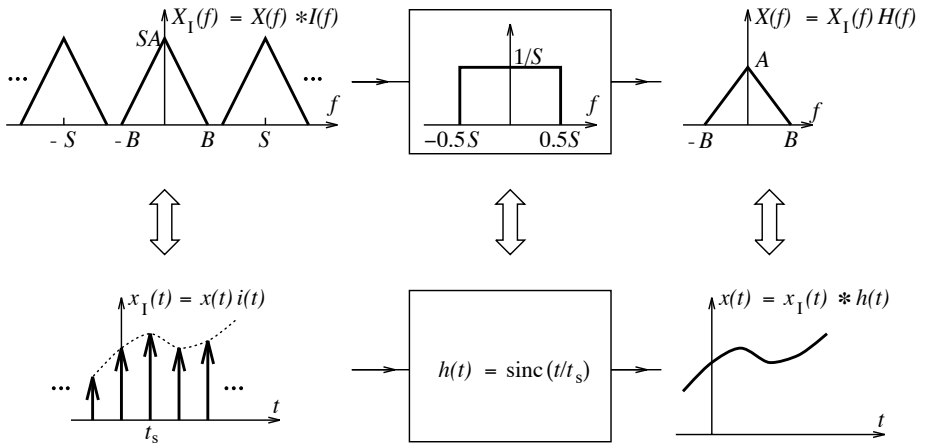


Figure 14.11 Recovery of an analog signal from its sampled version

The impulse response of the ideal lowpass filter is a sinc function given by $h(t) = \text{sinc}(t/t_s)$. The recovered signal $x(t)$ may therefore be described as the convolution

$$x(t) = x_I(t) \star h(t) = \left[\sum_{n=-\infty}^{\infty} x(nt_s)\delta(t - nt_s) \right] \star h(t) = \sum_{n=-\infty}^{\infty} x[n]h(t - nt_s) \tag{14.14}$$

This describes the superposition of shifted versions of $h(t)$ weighted by the sample values $x[n]$. Substituting for $h(t)$, we obtain the following result that allows us to recover $x(t)$ exactly from its samples $x[n]$ as a sum of scaled shifted versions of sinc functions:

$$x(t) = \sum_{n=-\infty}^{\infty} x[n]\text{sinc}\Big(\frac{t - nt_s}{t_s}\Big) \tag{14.15}$$

PGR2026-00033
Krisp Technologies EX1024 Page 470

The signal $x(t)$ equals the superposition of shifted versions of $h(t)$ weighted by the sample values $x[n]$. At each sampling instant, we replace the sample value $x[n]$ by a sinc function whose peak value equals $x[n]$ and whose zero crossings occur at all the other sampling instants. The sum of these sinc functions yields the analog signal $x(t)$, as illustrated in Figure 14.12.

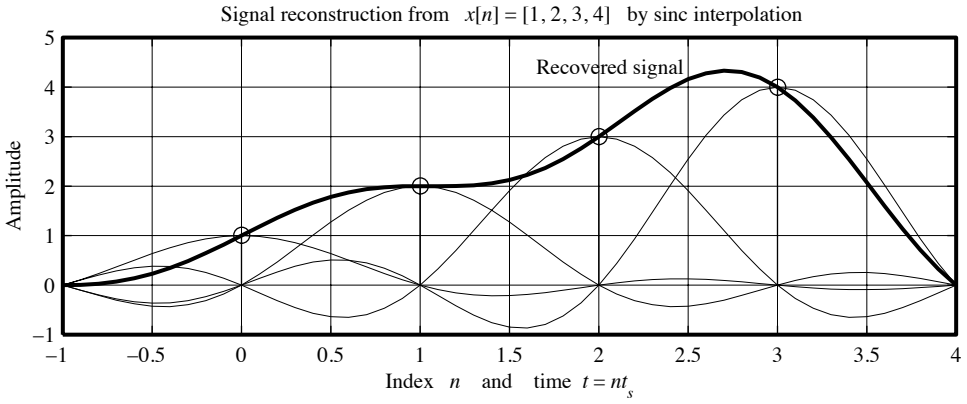


Figure 14.12 Ideal recovery of an analog signal by sinc interpolation

If we use a lowpass filter whose impulse response is $h_f(t) = 2t_s B\,\text{sinc}(2Bt)$ (i.e., whose cutoff frequency is B instead of $0.5S$), the recovered signal $x(t)$ may be described by the convolution

$$x(t) = x_I(t) \star h_f(t) = \sum_{n=-\infty}^{\infty} 2t_s B\, x[n]\text{sinc}[2B(t - nt_s)] \tag{14.16}$$

This general result is valid for *any oversampled signal* with $t_s \leq 0.5/B$ and reduces to the previously obtained result if the sampling rate S equals the Nyquist rate (i.e., $S = 2B$).

Sinc interpolation is unrealistic from a practical viewpoint. The infinite extent of the sinc means that it cannot be implemented on-line, and perfect reconstruction requires all its past and future values. We could truncate it on either side after its magnitude becomes small enough. But, unfortunately, it decays very slowly and must be preserved for a fairly large duration (covering many past and future sampling instants) in order to provide a reasonably accurate reconstruction. Since the sinc function is also smoothly varying, it cannot properly reconstruct a discontinuous signal at the discontinuities even with a large number of values. Sinc interpolation is also referred to as **band-limited interpolation** and forms the yardstick by which all other schemes are measured in their ability to reconstruct band-limited signals.

14.2.2 Interpolating Functions

Since the sinc interpolating function is a poor choice in practice, we must look to other interpolating signals. If an analog signal $x(t)$ is to be recovered from its sampled version $x[n]$ using an interpolating function $h_i(t)$, what must we require of $h_i(t)$ to obtain a good approximation to $x(t)$? At the very least, the interpolated approximation $\tilde{x}(t)$ should match $x(t)$ at the sampling instants nt_s. This suggests that $h_i(t)$ should equal zero at all sampling instants, except the origin where it must equal unity, such that

$$h_i(t) = \begin{cases} 1, & t = 0 \\ 0, & t = nt_s,\ (n = \pm 1, \pm 2, \pm 3, \ldots) \end{cases} \tag{14.17}$$

PGR2026-00033
Krisp Technologies EX1024 Page 471

In addition, we also require $h_i(t)$ to be absolutely integrable to ensure that it stays finite between the sampling instants. The interpolated signal $\tilde{x}(t)$ is simply the convolution of $h_i(t)$ with the ideally sampled signal $x_I(t)$, or a summation of shifted versions of the interpolating function

$$\tilde{x}(t) = h_i(t) \star x_I(t) = \sum_{n=-\infty}^{\infty} x[n] h_i(t - nt_s) \tag{14.18}$$

At each instant nt_s, we erect the interpolating function $h_i(t - nt_s)$, scale it by $x[n]$, and sum to obtain $\tilde{x}(t)$. At a sampling instant $t = kt_s$, the interpolating function $h_i(kt_s - nt_s)$ equals zero, unless $n = k$ when it equals unity. As a result, $\tilde{x}(t)$ *exactly* equals $x(t)$ at each sampling instant. At all other times, the interpolated signal $\tilde{x}(t)$ is only an approximation to the actual signal $x(t)$.

14.2.3 Interpolation in Practice

The nature of $h_i(t)$ dictates the nature of the interpolating system both in terms of its causality, stability, and physical realizability. It also determines how good the reconstructed approximation is. There is no "best" interpolating signal. Some are better in terms of their accuracy, others are better in terms of their cost effectiveness, still others are better in terms of their numerical implementation.

Step Interpolation

Step interpolation is illustrated in Figure 14.13 and uses a rectangular interpolating function of width t_s, given by $h(t) = \text{rect}[(t - 0.5t_s)/t_s]$, to produce a stepwise or staircase approximation to $x(t)$. Even though it appears crude, it is quite useful in practice.

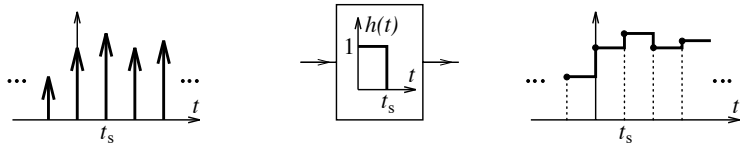


Figure 14.13 Step interpolation

At any time between two sampling instants, the reconstructed signal equals the previously sampled value and does not depend on any future values. This is useful for *on-line* or *real-time* processing where the output is produced at the same rate as the incoming data. Step interpolation results in exact reconstruction of signals that are piecewise constant.

A system that performs step interpolation is just a *zero-order-hold*. A practical digital-to-analog converter (DAC) for sampled signals uses a zero-order-hold for a staircase approximation (step interpolation) followed by a lowpass (anti-imaging) filter (for smoothing the steps).

Linear Interpolation

Linear interpolation is illustrated in Figure 14.14 and uses the interpolating function $h(t) = \text{tri}(t/t_s)$ to produce a linear approximation to $x(t)$ between the sample values.

At any instant t between adjacent sampling instants nt_s and $(n+1)t_s$, the reconstructed signal equals $x[n]$ plus an increment that depends on the slope of the line joining $x[n]$ and $x[n+1]$. We have

$$\tilde{x}(t) = x[n] + (t - nt_s)\frac{x[n+1] - x[n]}{t_s}, \qquad nt_s \le t < (n+1)t_s \tag{14.19}$$

PGR2026-00033
Krisp Technologies EX1024 Page 472

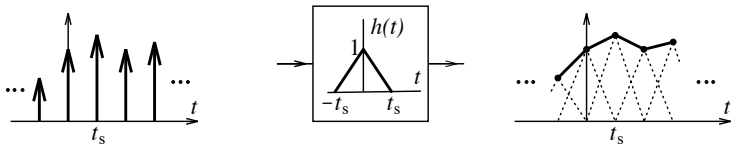


Figure 14.14 Linear interpolation

This operation requires one future value of the input and cannot actually be implemented on-line. It can, however, be realized with a delay of one sampling interval t_s, which is tolerable in many situations. Systems performing linear interpolation are also called **first-order-hold systems**. They yield exact reconstructions of piecewise linear signals.

Raised Cosine Interpolating Function

The sinc interpolating function forms the basis for several others described by the generic relation

$$h(t) = g(t)\text{sinc}(t/t_s), \qquad \text{where} \qquad g(0) = 1 \tag{14.20}$$

One of the more commonly used of these is the **raised cosine interpolating function** described by

$$h_{\text{rc}}(t) = \frac{\cos(\pi Rt/t_s)}{1-(2Rt/t_s)^2}\,\text{sinc}(t/t_s), \qquad 0 \le R \le 1 \tag{14.21}$$

Here, R is called the **roll-off factor**. Like the sinc interpolating function, $h_{\text{rc}}(t)$ equals 1 at $t = 0$ and 0 at the other sampling instants. It exhibits faster decaying oscillations on either side of the origin for $R > 0$ as compared to the sinc function. This faster decay results in improved reconstruction if the samples are not acquired at exactly the sampling instants (in the presence of *jitter*, that is). It also allows fewer past and future values to be used in the reconstruction as compared with the sinc interpolating function. The terminology *raised cosine* is actually based on the shape of its spectrum. For $R = 0$, the raised cosine interpolating function reduces to the sinc interpolating function.

EXAMPLE 14.5 (Signal Reconstruction from Samples)
Let $x[n] = \{-1, 2, 3, 2\}$, $t_s = 1$. What is the value of the reconstructed signal $x(t)$ at 2.5 s that results from step, linear, sinc, and raised cosine (with $R = 0.5$) interpolation?

(a) For step interpolation, the signal value at $t = 2.5$ s is simply the value at $t = 2$. Thus, $\tilde{x}(2.5) = 3$.

(b) If we use linear interpolation, the signal value at $t = 2.5$ s is simply the average of the values at $t = 2$ and $t = 3$. Thus, $\tilde{x}(2.5) = 0.5(3 + 2) = 2.5$.

(c) If we use sinc interpolation, we obtain

$$\tilde{x}(t) = \sum_{k=0}^{3} x[k]\text{sinc}(t - kt_s) = -\text{sinc}(t) + 2\ \text{sinc}(t-1) + 3\ \text{sinc}(t-2) + 2\ \text{sinc}(t-3)$$

So, $\tilde{x}(2.5) = 0.1273 - 0.4244 + 1.9099 + 2.5465 = 4.1592$.

PGR2026-00033
Krisp Technologies EX1024 Page 473

(d) If we use raised cosine interpolation (with $R = 0.5$), we obtain

$$\tilde{x}(t) = \sum_{k=0}^{3} x[k]\text{sinc}(t-k)\frac{\cos[0.5\pi(t-k)]}{1-(t-k)^2}$$

We use this equation to compute

$$\tilde{x}(t) = \frac{\text{sinc}(t)\cos(0.5\pi t)}{1-t^2} + 2\frac{\text{sinc}(t-1)\cos[0.5\pi(t-1)]}{1-(t-1)^2} + 3\frac{\text{sinc}(t-2)\cos[0.5\pi(t-2)]}{1-(t-2)^2} + 4\frac{\text{sinc}(t-3)\cos[0.5\pi(t-3)]}{1-(t-3)^2}$$

Thus, $\tilde{x}(2.5) = 0.0171 - 0.2401 + 1.8006 + 2.4008 = 3.9785$.

14.3 Quantization

The importance of digital signals stems from the proliferation of high-speed digital computers for signal processing. Due to the finite memory limitations of such machines, we can process only finite data sequences. We must not only sample an analog signal in time but also *quantize* (round or truncate) the signal amplitudes to a finite set of values. Since quantization affects only the signal amplitude, both analog and discrete-time signals can be quantized. Quantized discrete-time signals are called *digital* signals.

Each quantized sample is represented as a group (*word*) of zeros and ones (*bits*) that can be processed digitally. The finer the quantization, the longer the word. Like sampling, improper quantization leads to loss of information. But unlike sampling, no matter how fine the quantization, its effects are irreversible, since word lengths must necessarily be finite. The systematic treatment of quantization theory is very difficult because finite word lengths appear as nonlinear effects. Quantization always introduces some noise, whose effects can be described only in statistical terms, and is usually considered only in the final stages of any design, and many of its effects (such as overflow and limit cycles) are beyond the realm of this text.

14.3.1 Uniform Quantizers

Quantizers are devices that operate on a signal to produce a finite number of amplitude levels or quantization levels. It is common practice to use **uniform quantizers** with equal quantization levels.

The number of levels L in most quantizers used in an analog-to-digital converter (ADC) is invariably a power of 2. If $L = 2^B$, each of the L levels is coded to a binary number, and each signal value is represented in binary form as a B-bit word corresponding to its quantized value. A 4-bit quantizer is thus capable of 2^4 (or 16) levels and a 12-bit quantizer yields 2^{12} (or 4096) levels.

A signal may be quantized by **rounding** to the nearest quantization level, by **truncation** to a level smaller than the next higher one, or by **sign-magnitude truncation**, which is rather like truncating absolute values and then using the appropriate sign. These operations are illustrated in Figure 14.15.

For a quantizer with a full-scale amplitude of $\pm X$, input values outside the range $|X|$ will get clipped and result in overflow. The observed value may be set to the full-scale value (*saturation*) or zero (*zeroing*), leading to the overflow characteristics shown in Figure 14.16.

The quantized signal value is usually represented as a group (word) with a specified number of bits called the word length. Several number representations are in use and illustrated in Table 14.1 for $B = 3$. Some forms of number representation are better suited to some combinations of quantization and overflow characteristics and, in practice, certain combinations are preferred. In any case, the finite length of binary words leads to undesirable, and often irreversible, effects collectively known as finite-word-length effects (some of which are beyond our scope).

PGR2026-00033
Krisp Technologies EX1024 Page 474

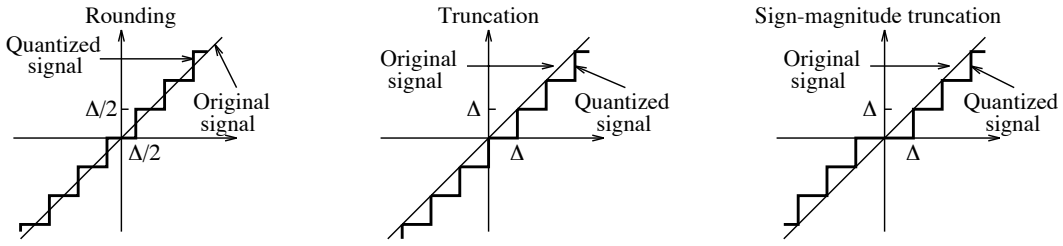


Figure 14.15 Various ways of quantizing a signal

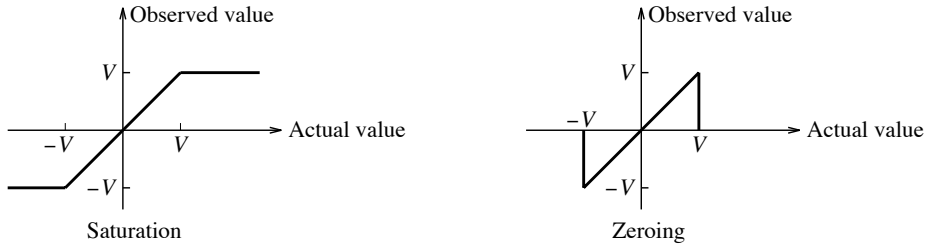


Figure 14.16 Two overflow characteristics

Table 14.1 Various Number Representations for $B = 3$ Bits

Decimal Value	Sign and Magnitude	One's Complement	Two's Complement	Offset Binary
+4				111
+3	011	011	011	110
+2	010	010	010	101
+1	001	001	001	000
+0	000	000	000	
−0	100	111		011
−1	101	110	111	010
−2	110	101	110	001
−3	111	100	101	000
−4			100	

14.3.2 Quantization Error and Quantization Noise

The **quantization error**, naturally enough, depends on the number of levels. If the quantized signal corresponding to a discrete signal $x[n]$ is denoted by $x_Q[n]$, the quantization error $\epsilon[n]$ equals

$$\epsilon[n] = x[n] - x_Q[n] \tag{14.22}$$

It is customary to define the **quantization signal-to-noise ratio** (SNR_Q) as the ratio of the power P_S in the signal and the power P_N in the error $\epsilon[n]$ (or noise). This is usually measured in decibels, and we get

$$P_S = \frac{1}{N}\sum_{n=0}^{N-1} x^2[n] \qquad P_N = \frac{1}{N}\sum_{n=0}^{N-1} \epsilon^2[n] \qquad \text{SNR}_Q(\text{dB}) = 10\log\frac{P_S}{P_N} = 10\log\frac{\sum x^2[n]}{\sum \epsilon^2[n]} \tag{14.23}$$

PGR2026-00033
Krisp Technologies EX1024 Page 475

The effect of quantization errors due to rounding or truncation is quite difficult to quantify analytically unless statistical estimates are used. The **dynamic range** or full-scale range of a signal $x(t)$ is defined as its maximum variation $D = x_{\max} - x_{\min}$. If $x(t)$ is sampled and quantized to L levels using a quantizer with a full-scale range of D, the **quantization step size**, or *resolution*, Δ, is defined as

$$\Delta = D/L \tag{14.24}$$

This step size also corresponds to the least significant bit (LSB). The dynamic range of a quantizer is often expressed in decibels. For a 16-bit quantizer, the dynamic range is $20 \log 2^{16} \approx 96$ dB.

For quantization by rounding, the maximum value of the quantization error must lie between $-\Delta/2$ and $\Delta/2$. If L is large, the error is equally likely to take on any value between $-\Delta/2$ and $\Delta/2$ and is thus *uniformly distributed*. Its probability density function $f(\epsilon)$ has the form shown in Figure 14.17.

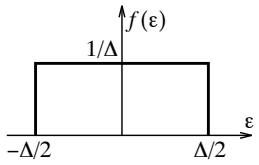


Figure 14.17 Probability density function of a signal quantized by rounding

The noise power P_N equals its variance σ^2 (the second central moment), and is given by

$$P_N = \sigma^2 = \int_{-\Delta/2}^{\Delta/2} \epsilon^2 f(\epsilon)\, d\epsilon = \frac{1}{\Delta} \int_{-\Delta/2}^{\Delta/2} \epsilon^2\, d\epsilon = \frac{\Delta^2}{12} \tag{14.25}$$

The quantity $\sigma = \Delta/\sqrt{12}$ defines the **rms quantization error**. With $\Delta = D/L$, we compute

$$10 \log P_N = 10 \log \frac{D^2}{12L^2} = 20 \log D - 20 \log L - 10.8 \tag{14.26}$$

A *statistical estimate* of the SNR in decibels, denoted by SNR_S, is thus provided by

$$\text{SNR}_S(\text{dB}) = 10 \log P_S - 10 \log P_N = 10 \log P_S + 10.8 + 20 \log L - 20 \log D \tag{14.27}$$

For a B-bit quantizer with $L = 2^B$ levels (and $\Delta = D/2^B$), we obtain

$$\text{SNR}_S(\text{dB}) = 10 \log P_S + 10.8 + 6B - 20 \log D \tag{14.28}$$

This result suggests a 6-dB improvement in the SNR for each additional bit. It also suggests a reduction in the SNR if the dynamic range D is chosen to exceed the signal limits. In practice, signal levels do not often reach the extreme limits of their dynamic range. This allows us to increase the SNR by choosing a smaller dynamic range D for the quantizer but at the expense of some distortion (at very high signal levels).

EXAMPLE 14.6 (Quantization Effects)

(a) A sampled signal that varies between -2 V and 2 V is quantized using B bits. What value of B will ensure an rms quantization error of less than 5 mV?

The full-scale range is $D = 4$ V. The rms error is given by $\sigma = \Delta/\sqrt{12}$. With $\Delta = D/2^B$, we obtain

$$2^B = \frac{D}{\sigma\sqrt{12}} = \frac{4}{0.005\sqrt{12}} = 230.94 \qquad B = \log_2 230.94 = 7.85$$

Rounding up this result, we get $B = 8$ bits.

PGR2026-00033
Krisp Technologies EX1024 Page 476

(b) Consider the ramp $x(t) = 2t$ over $(0, 1)$. For a sampling interval of 0.1 s, and $L = 4$, we obtain the sampled signal, quantized (by rounding) signal, and error signal as

$$\begin{aligned} x[n] &= \{0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0\} \\ x_Q[n] &= \{0, 0.0, 0.5, 0.5, 1.0, 1.0, 1.0, 1.5, 1.5, 2.0, 2.0\} \\ e[n] &= \{0, 0.2, -0.1, 0.1, -0.2, 0.0, 0.2, -0.1, 0.1, -0.2, 0.0\} \end{aligned}$$

We can now compute the SNR in several ways.

1. $\text{SNR}_Q = 10 \log \dfrac{P_S}{P_N} = 10 \log \dfrac{\sum x^2[n]}{\sum e^2[n]} = 10 \log \dfrac{15.4}{0.2} = 18.9$ dB
2. We could also use $\text{SNR}_Q = 10 \log P_S + 10.8 + 20 \log L - 20 \log D$. With $D = 2$ and $N = 11$,

$$\text{SNR}_S = 10 \log \left[\frac{1}{N} \sum_{n=0}^{N-1} x^2[n] \right] + 10.8 + 20 \log 4 - 20 \log 2 = 18.7 \text{ dB}$$

3. If $x(t)$ forms one period of a periodic signal with $T = 1$, we can also find P_S and SNR_S as

$$P_S = \frac{1}{T} \int x^2(t)\, dt = \frac{4}{3} \qquad \text{SNR}_S = 10 \log(\tfrac{4}{3}) + 10 \log 12 + 20 \log 4 - 20 \log 2 = 18.062 \text{ dB}$$

Why the differences between the various results? Because SNR_S is a *statistical estimate*. The larger the number of samples N, the less SNR_Q and SNR_S differ. For $N = 500$, for example, we find that $\text{SNR}_Q = 18.0751$ dB and $\text{SNR}_S = 18.0748$ dB are very close indeed.

(c) Consider the sinusoid $x(t) = A \cos(2\pi f t)$. The power in $x(t)$ is $P_S = 0.5A^2$. The dynamic range of $x(t)$ is $D = 2A$ (the peak-to-peak value). For a B-bit quantizer, we obtain the widely used result

$$\text{SNR}_S = 10 \log P_S + 10.8 + 6B - 20 \log D = 6B + 1.76 \text{ dB}$$

14.3.3 Quantization and Oversampling

Even though sampling and quantization are independent operations, it turns out that oversampling allows us to use quantizers with fewer bits per sample. The idea is that the loss of accuracy in the sample values (by using fewer bits) is offset by the larger number of samples (by using a higher sampling rate). This is a consequence of how the quantization noise power is distributed.

The quantized signal is $x_Q[n] = x[n] + \epsilon[n]$. The quantization error $\epsilon[n]$ may be assumed to be a white-noise sequence that is uncorrelated with the signal $x[n]$, with a uniform probability density over the range $(-0.5\Delta,\ 0.5\Delta)$. The variance (average power) of $\epsilon[n]$ is $\sigma^2 = \Delta^2/12$. The spectrum of $\epsilon[n]$ is flat over the principal range $(0.5S,\ 0.5S)$, and the average power σ^2 is equally distributed over this range. The power spectral density of $\epsilon[n]$ is thus $P_{ee}(f) = \sigma^2/S$. If a signal is oversampled at the rate NS, the noise spectrum is spread over a wider frequency range $(-0.5NS,\ 0.5NS)$. The power spectral density is thus N times less for the same total quantization noise power. In fact, the in-band quantization noise in the region of interest, $(-0.5S,\ 0.5S)$, is also decreased by a factor of N. This is illustrated in Figure 14.18.

If we were to keep the in-band noise power the same as before, we can recover the signal using fewer bits. For such a quantizer with $B_2 = (B - \Delta B)$ bits per sample, an average power of σ_2^2, and a sampling rate of

PGR2026-00033
Krisp Technologies EX1024 Page 477

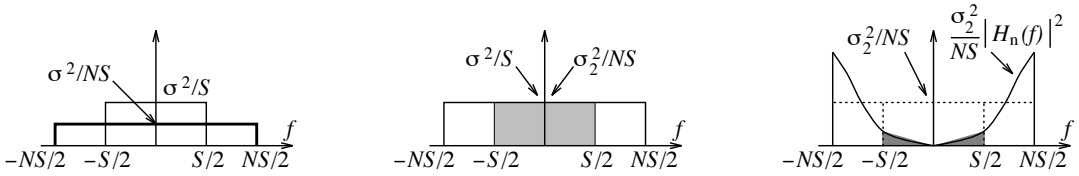


Figure 14.18 Quantization noise spectrum of an oversampled signal

$S_2 = NS$ Hz, we obtain the same in-band quantization noise power when its power spectral density is also σ^2/S, as illustrated in Figure 14.18. Thus,

$$\frac{\sigma^2}{S} = \frac{\sigma_2^2}{NS} \qquad \sigma^2 = \frac{\sigma_2^2}{N} \tag{14.29}$$

Assuming the same full-scale range D for each quantizer, we obtain

$$\frac{D^2}{(12)2^{2B}} = \frac{D^2}{(12N)2^{2(B-\Delta B)}} \qquad N = 2^{2\Delta B} \qquad \Delta B = 0.5 \log_2 N \tag{14.30}$$

This result suggests that we gain 0.5 bits for every doubling of the sampling rate. For example, $N = 4$ (four-times oversampling) leads to a gain of 1 bit. In practice, a better trade-off between bits and samples is provided by quantizers that not only use oversampling but also *shape* the noise spectrum (using filters) to further reduce the in-band noise, as shown in Figure 14.18. A typical noise shape is the sine function, and a pth-order noise-shaping filter has the form $H_{\text{NS}}(f) = |2\sin(\pi f/NS)|^p,\ -0.5NS \le f \le 0.5NS$, where N is the oversampling factor. Equating the filtered in-band noise power to σ^2, we obtain

$$\sigma^2 = \left(\frac{\sigma_2^2}{N}\right)\frac{1}{S}\int_{-S/2}^{S/2} |H_{\text{NS}}(f)|^2 df \tag{14.31}$$

If N is large, $H_{\text{NS}}(f) \approx |(2\pi f/NS)|^p$ over the much smaller in-band range $(-0.5S,\ 0.5S)$, and we get

$$\sigma^2 = \left(\frac{\sigma_2^2}{N}\right)\frac{1}{S}\int_{-S/2}^{S/2} \left(\frac{2\pi f}{NS}\right)^{2p} df = \frac{\sigma_2^2 \pi^{2p}}{(2p+1)N^{2p+1}} \tag{14.32}$$

Simplification, with $\sigma_2^2/\sigma^2 = 2^{2\Delta B}$, gives

$$\Delta B = (p + 0.5)\log_2 N - 0.5 \log_2\left(\frac{\pi^{2p}}{2p+1}\right) \tag{14.33}$$

This shows that noise shaping (or *error-spectrum shaping*) results in a savings of $p \log_2 N$ additional bits. With $p = 1$ and $N = 4$, for example, $\Delta B \approx 2$ bits. This means that we can make do with a $(B-2)$-bit DAC during reconstruction if we use noise shaping with four-times oversampling. In practical implementation, noise shaping is achieved by using an oversampling *sigma-delta* ADC. State-of-the-art CD players make use of this technology. What does it take to achieve 16-bit quality using a 1-bit quantizer (which is just a sign detector)? Since $\Delta B = 15$, we could, for example, use oversampling by $N = 64$ (to 2.8 MHz for audio signals sampled at 44.1 kHz) and $p = 3$ (third-order noise shaping).

PGR2026-00033
Krisp Technologies EX1024 Page 478

14.4 Digital Processing of Analog Signals

The crux of the sampling theorem is not just the choice of an appropriate sampling rate. More important, the processing of an analog signal is equivalent to the processing of its Nyquist sampled version, because it retains the same information content as the original. This is how the sampling theorem is used in practice. It forms the link between analog and digital signal processing, and allows us to use digital techniques to manipulate analog signals. When we sample a signal $x(t)$ at the instants nt_s, we imply that the spectrum of the sampled signal is periodic with period $S = 1/t_s$ and band-limited to a highest frequency $B = 0.5S$. Figure 14.19 illustrates a typical system for analog-to-digital conversion.

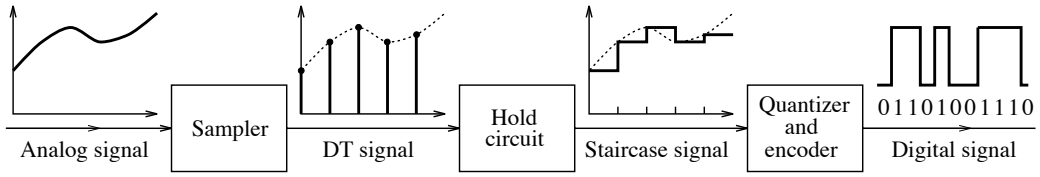


Figure 14.19 Block diagram of a system for analog-to-digital conversion

An analog lowpass **pre-filter** or **anti-aliasing filter** (not shown) limits the highest analog signal frequency to allow a suitable choice of the sampling rate and ensure freedom from aliasing. The **sampler** operates above the Nyquist sampling rate and is usually a zero-order-hold device. The **quantizer** limits the sampled signal values to a finite number of levels (16-bit quantizers allow a signal-to-noise ratio close to 100 dB). The **encoder** converts the quantized signal values to a string of *binary* **bits** or zeros and ones (**words**) whose length is determined by the number of quantization levels of the quantizer.

A digital signal processing system, in hardware or software (consisting of digital filters), processes the encoded digital signal (or **bit stream**) in a desired fashion. Digital-to-analog conversion essentially reverses the process and is accomplished by the system shown in Figure 14.20.

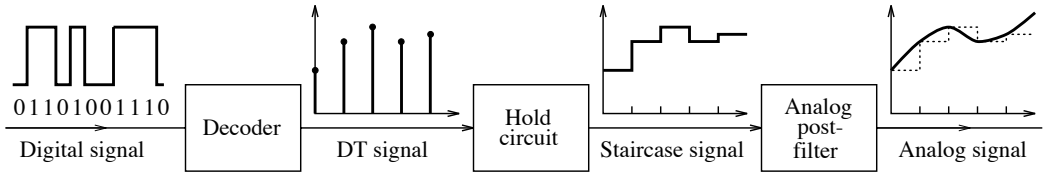


Figure 14.20 Block diagram of a system for analog-to-digital conversion

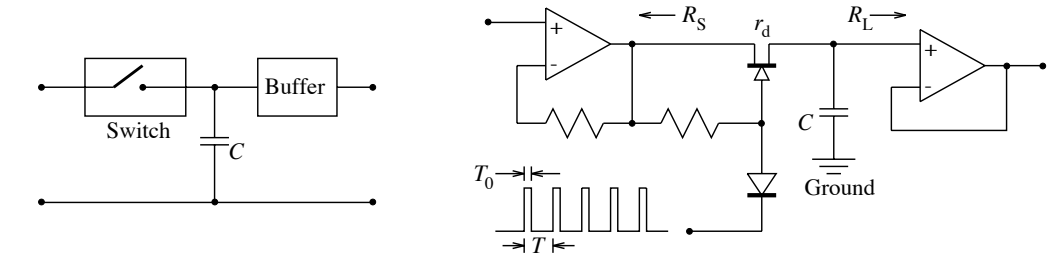
A **decoder** converts the processed bit stream to a discrete signal with quantized signal values. The zero-order-hold device reconstructs a staircase approximation of the discrete signal. The lowpass analog **post-filter**, or **anti-imaging filter**, extracts the central period from the periodic spectrum, removes the unwanted replicas (images), and results in a smoothed reconstructed signal.

14.4.1 Practical ADC Considerations

An analog-to-digital converter (ADC) for converting analog to digital signals consists of a sample-and-hold circuit followed by a quantizer and encoder. A block diagram of a typical sample-and-hold circuit and its practical realization is shown in Figure 14.21.

A clock signal controls a switch (an FET, for example) that allows the hold capacitor C to charge very rapidly to the sampled value (when the switch is closed) and to discharge very slowly through the

PGR2026-00033
Krisp Technologies EX1024 Page 479



Block diagram of sample-and-hold system Implementation using an FET as a switch

Figure 14.21 Block diagram of a sample-and-hold system and a practical realization

high input resistance of the buffer amplifier (when the switch is open). Ideally, the sampling should be as instantaneous as possible and, once acquired, the level should be held constant until it can be quantized and encoded. Practical circuits also include an operational amplifier at the input to isolate the source from the hold capacitor and provide better tracking of the input. In practice, the finite **aperture time** T_A (during which the signal is being measured), the finite **acquisition time** T_H (to switch from the hold mode to the sampling mode), the **droop** in the capacitor voltage (due to the leakage of the hold capacitor), and the finite **conversion time** T_C (of the quantizer) are all responsible for less than perfect performance.

A finite aperture time limits both the accuracy with which a signal can be measured and the highest frequency that can be handled by the ADC. Consider the sinusoid $x(t) = A\sin(2\pi f_0 t)$. Its derivative $x'(t) = 2\pi A f_0 \cos(2\pi f_0 t)$ describes the rate at which the signal changes. The fastest rate of change equals $2\pi A f_0$ at the zero crossings of $x(t)$. If we assume that the signal level can change by no more than ΔX during the aperture time T_A, we must satisfy

$$2\pi A f_0 \le \frac{\Delta X}{T_A} \qquad \text{or} \qquad f_0 \le \frac{\Delta X}{\pi D T_A} \tag{14.34}$$

where $D = 2A$ corresponds to the full-scale range of the quantizer. Typically, ΔX may be chosen to equal the rms quantization error (which equals $\Delta/\sqrt{12}$) or 0.5 LSB (which equals 0.5Δ). In the absence of a sample-and-hold circuit (using only a quantizer), the aperture time must correspond to the conversion time of the quantizer. This time may be much too large to permit conversion of frequencies encountered in practice. An important reason for using a sample-and-hold circuit is that it holds the sampled value constant and allows much higher frequencies to be handled by the ADC. The maximum sampling rate S that can be used for an ADC (with a sample-and-hold circuit) is governed by the aperture time T_A, and hold time T_H, of the sample-and-hold circuit, as well as the conversion time T_C of the quantizer, and corresponds to

$$S \le \frac{1}{T_A + T_H + T_C} \tag{14.35}$$

Naturally, this sampling rate must also exceed the Nyquist rate. During the hold phase, the capacitor discharges through the high input resistance R of the buffer amplifier, and the capacitor voltage is given by $v_C(t) = V_0 e^{-t/\tau}$, where V_0 is the acquired value and $\tau = RC$. The voltage during the hold phase is not strictly constant but shows a droop. The maximum rate of change occurs at $t = 0$ and is given by $v'(t)|_{t=0} = -V_0/\tau$. A proper choice of the holding capacitor C can minimize the droop. If the maximum droop is restricted to ΔV during the hold time T_H, we must satisfy

$$\frac{V_0}{RC} \le \frac{\Delta V}{T_H} \qquad \text{or} \qquad C \ge \frac{V_0 T_H}{R\,\Delta V} \tag{14.36}$$

PGR2026-00033
Krisp Technologies EX1024 Page 480

To impose a lower bound, V_0 is typically chosen to equal the full-scale range, and ΔV may be chosen to equal the rms quantization error (which equals $\Delta/\sqrt{12}$) or 0.5 LSB (which corresponds to 0.5Δ). Naturally, an upper bound on C is also imposed by the fact that during the capture phase the capacitor must be able to charge very rapidly to the input level.

A digital-to-analog converter (DAC) allows the coded and quantized signal to be converted to an analog signal. In its most basic form, it consists of a summing operational amplifier, as shown in Figure 14.22.

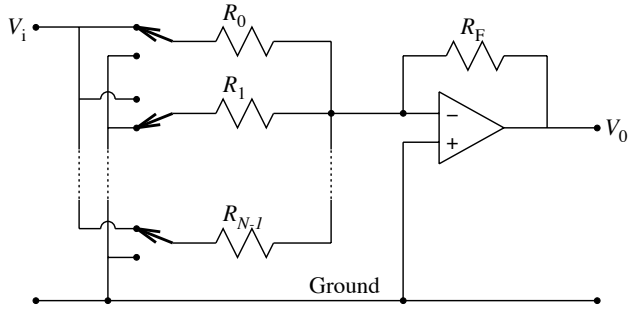


Figure 14.22 A system for digital-to-analog conversion

The switches connect either to ground or to the input. The output voltage V_0 is a weighted sum of only those inputs that are switched in. If the resistor values are selected as

$$R_0 = R \quad R_1 = \frac{R}{2} \quad R_2 = \frac{R}{2^2} \quad \cdots \; R_{N-1} = \frac{R}{2^{N-1}} \tag{14.37}$$

the output is given by

$$V_0 = -V_i \frac{R_F}{R} \left(b_{N-1} 2^{N-1} + b_{N-2} 2^{N-2} + \cdots + b_1 2^1 + b_0 2^0 \right) \tag{14.38}$$

where the coefficients b_k correspond to the position of the switches and equal 1 (if connected to the input) or 0 (if connected to the ground). Practical circuits are based on modifications that use only a few different resistor values (such as R and $2R$) to overcome the problem of choosing a wide range of resistor values (especially for high-bit converters). Of course, a 1-bit DAC simply corresponds to a constant gain.

EXAMPLE 14.7 (ADC Considerations)

(a) Suppose we wish to digitize a signal band-limited to 15 kHz using only a 12-bit quantizer. If we assume that the signal level can change by no more than the rms quantization error during capture, the aperture time T_A must satisfy

$$T_A \leq \frac{\Delta X}{\pi D f_0} = \frac{D/2^B\sqrt{12}}{\pi D f_0} = \frac{1}{2^{12}\pi(15)(10^3)} \approx 5.18 \text{ ns}$$

The conversion time of most practical quantizers is much larger (in the microsecond range), and as a result, we can use such a quantizer only if it is preceded by a sample-and-hold circuit whose aperture time is less than 5 ns.

PGR2026-00033
Krisp Technologies EX1024 Page 481

(b) If we digitize a signal band-limited to 15 kHz, using a sample-and-hold circuit (with a capture time of 4 ns and a hold time of 10 μs) followed by a 12-bit quantizer, the conversion time of the quantizer can be computed from

$$S \leq \frac{1}{T_A + T_H + T_C} \qquad T_A + T_H + T_C \leq \frac{1}{S} = \frac{1}{(30)10^3} \qquad T_C = 23.3\ \mu\text{s}$$

The value of T_C is well within the capability of practical quantizers.

(c) Suppose the sample-and-hold circuit is buffered by an amplifier with an input resistance of 1 MΩ. To ensure a droop of no more that 0.5 LSB during the hold phase, we require

$$C \geq \frac{V_0 T_H}{R\Delta V}$$

Now, if V_0 corresponds to the full-scale value, $\Delta V = 0.5\,\text{LSB} = V_0/2^{B+1}$, and

$$C \geq \frac{V_0 T_H}{R(V_0/2^{B+1})} = \frac{(10)10^{-6}2^{13}}{10^6} \approx 81.9\ \text{nF}$$

14.4.2 Anti-Aliasing Filter Considerations

The purpose of the anti-aliasing filter is to band-limit the input signal. In practice, however, we cannot design brick-wall filters, and some degree of aliasing is inevitable. The design of anti-aliasing filters must ensure that the effects of aliasing are kept small. One way to do this is to attenuate components above the Nyquist frequency to a level that cannot be detected by the ADC. The choice of sampling frequency S is thus dictated not only by the highest frequency of interest but also the resolution of the ADC. If the aliasing level is ΔV and the maximum passband level is V, we require a filter with a stopband attenuation of $A_s > 20\log\frac{V}{\Delta V}$ dB. If the peak signal level equals A, and the passband edge is defined as the half-power (or 3-dB) frequency, and ΔV is chosen as the rms quantization error for a B-bit quantizer, we have

$$A_s > 20\log\frac{\text{maximum rms passband level}}{\text{minimum rms stopband level}} = \frac{A/\sqrt{2}}{\Delta/\sqrt{12}} = \frac{A/\sqrt{2}}{A/(2^B\sqrt{12})} = 20\log(2^B\sqrt{6}) \quad \text{dB} \tag{14.39}$$

EXAMPLE 14.8 (Anti-Aliasing Filter Considerations)
Suppose we wish to process a noisy speech signal with a bandwidth of 4 kHz using an 8-bit ADC.
We require a minimum stopband attenuation of $A_s = 20\log(2^B\sqrt{6}) \approx 56$ dB.
If we use a fourth-order Butterworth anti-aliasing filter with a 3-dB passband edge of f_p, the normalized frequency $\nu_s = f_s/f_p$ at which the stopband attenuation of 56 dB occurs is computed from

$$A_s = 10\log(1+\nu_s^{2n}) \qquad \nu_s = \left(10^{0.1A_s} - 1\right)^{1/2n} \approx 5$$

If the passband edge f_p is chosen as 4 kHz, the actual stopband frequency is $f_s = \nu_s f_p = 20$ kHz.
If the stopband edge also corresponds to highest frequency of interest, then $S = 40$ kHz.
The frequency f_a that gets aliased to the passband edge f_p corresponds to $f_a = S - f_p = 36$ kHz.
The attenuation A_a (in dB) at this frequency f_a is

$$A_a = 10\log(1+\nu_a^{2n}) = 10\log(1+9^8) \approx 76.3\ \text{dB}$$

This corresponds to a signal level (relative to unity) of $(1.5)10^{-4}$, well below the rms quantization error (which equals $1/(2^B\sqrt{12}) = 0.0011$).

PGR2026-00033
Krisp Technologies EX1024 Page 482

14.4.3 Anti-Imaging Filter Considerations

The design of reconstruction filters requires that we extract only the central image and minimize the sinc distortion caused by practical zero-order-hold sampling. The design of such filters must meet stringent specifications unless oversampling is employed during the digital-to-analog conversion. In theory, **oversampling** (at a rate much higher than the Nyquist rate) may appear wasteful, because it can lead to the manipulation of a large number of samples. In practice, however, it offers several advantages. First, it provides adequate separation between spectral images and thus allows filters with less stringent cutoff requirements to be used for signal reconstruction. Second, it minimizes the effects of the sinc distortion caused by practical zero-order-hold sampling. Figure 14.23 shows staircase reconstructions at two sampling rates.

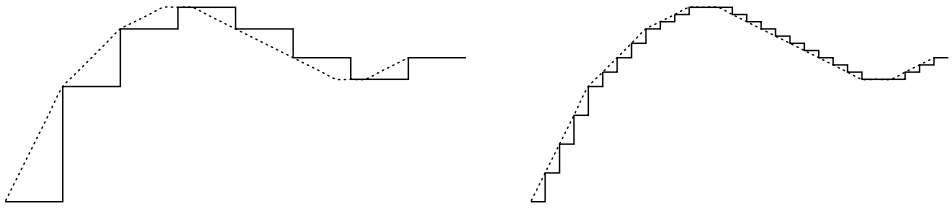


Figure 14.23 Staircase reconstruction of a signal at low (left) and high (right) sampling rates

The smaller steps in the staircase reconstruction for higher sampling rates lead to a better approximation of the analog signal, and these smaller steps are much easier to smooth out using lower-order filters with less stringent specifications.

EXAMPLE 14.9 (Anti-Imaging Filter Considerations)

(a) Consider the reconstruction of an audio signal band-limited to 20 kHz, assuming ideal sampling. The reconstruction filter is required to have a maximum signal attenuation of 0.5 dB and to attenuate all images by at least 60 dB.

The passband and stopband attenuations are $A_p = 0.5$ dB at the passband edge $f_p = 20$ kHz and $A_s = 60$ dB at the stopband edge $f_s = S - f_p = S - 20$ kHz, where S is the sampling rate in kilohertz. This is illustrated in Figure E14.9A.

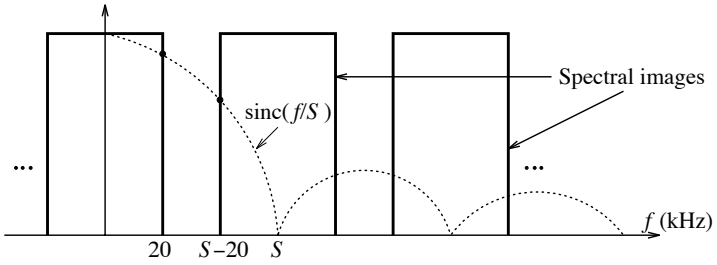


Figure E14.9A Spectra for Example 14.9(a)

If we design a Butterworth filter, its order n is given by

$$n = \frac{\log[(10^{0.1A_s} - 1)/\epsilon^2]^{1/2}}{\log(f_s/f_p)} \qquad \epsilon^2 = 10^{0.1A_p} - 1$$

1. If we choose $S = 44.1$ kHz, we compute $f_s = 44.1 - 20 = 22.1$ kHz and $n = 80$.

PGR2026-00033
Krisp Technologies EX1024 Page 483

2. If we choose $S = 176.4$ kHz (four-times oversampling), we compute $f_s = 176.4 - 20 = 156.4$ kHz and $n = 4$. This is an astounding reduction in the filter order

---

(b) If we assume zero-order-hold sampling and $S = 176.4$ kHz, the signal spectrum is multiplied by $\text{sinc}(f/S)$. This already provides a signal attenuation of $-20\log[\text{sinc}(20/176.4)] = 0.184$ dB at the passband edge of 20 kHz and $-20\log[\text{sinc}(156.4/176.4)] = 18.05$ dB at the stopband edge of 156.4 kHz. The new filter attenuation specifications are thus $A_p = 0.5 - 0.184 = 0.316$ dB and $A_s = 60 - 18.05 = 41.95$ dB, and we require a Butterworth filter whose order is given by

$$n = \frac{\log[(10^{0.1A_s} - 1)/\epsilon^2]^{1/2}}{\log(f_s/f_p)} = 2.98 \quad \Rightarrow \quad n = 3$$

We see that oversampling allows us to use reconstruction filters of much lower order. In fact, the earliest commercial CD players used four-times oversampling (during the DSP stage) and second-order or third-order Bessel reconstruction filters (for linear phase).

The effects of sinc distortion may also be minimized by using digital filters with a 1/sinc response during the DSP phase itself (prior to reconstruction).

14.4.4 Finite-Word-Length Effects

Signal quantization uses a finite number of bits and this less than ideal representation leads to problems that are collectively referred to as *finite-word-length effects* and are outlined below. Such problems afflict both IIR and FIR filters. A complete discussion is beyond the scope of this text.

Quantization noise: Quantization noise limits the signal-to-noise ratio. One way to improve the SNR is to increase the number of bits. Another is to use oversampling (as discussed earlier).
Coefficient quantization: This refers to representation of the filter coefficients by a limited number of bits. Its effects can produce benign effects such as a slight change in the frequency response of the resulting filter (typically, a larger passband ripple and/or a smaller stopband attenuation) or disastrous effects (for IIR filters) that can lead to instability.
Roundoff errors: When lower-order bits are discarded before storing results (of a multiplication, say), there is roundoff error. The amount of error depends on the type of arithmetic used and the filter structure. The effects of roundoff errors are similar to the effects of quantization noise and lead to a reduction in the signal-to-noise ratio.
Overflow errors: Overflow errors occur when the filter output or the result of arithmetic operations (such as the sum of two large numbers with the same sign) exceeds the permissible wordlength. Such errors are avoided in practice by scaling the filter coefficients and/or the input in a manner that the output remains within the permissible word length.

14.5 Compact Disc Digital Audio

One application of digital signal processing that has had a profound effect on the consumer market is in compact disc (CD) digital audio systems. The human ear is an incredibly sensitive listening device that depends on the pressure of the air molecules on the eardrum for the perception of sound. It covers a whopping dynamic range of 120 dB, from the Brownian motion of air molecules (the threshold of hearing) all the way up to the threshold of pain. The technology of recorded sound has come a long way since

PGR2026-00033
Krisp Technologies EX1024 Page 484

Edison's invention of the phonograph in 1877. The analog recording of audio signals on long-playing (LP) records suffers from poor signal-to-noise ratio (about 60 dB), inadequate separation between stereo channels (about 30 dB), wow and flutter, and wear due to mechanical tracking of the grooves. The CD overcomes the inherent limitations of LP records and cassette tapes and yields a signal-to-noise ratio, dynamic range, and stereo separation, all in excess of 90 dB. It makes full use of digital signal processing techniques during both recording and playback.

14.5.1 Recording

A typical CD recording system is illustrated in Figure 14.24. The analog signal recorded from each microphone is passed through an anti-aliasing filter, sampled at the industry standard of 44.1 kHz and quantized to 16 bits in each channel. The two signal channels are then multiplexed, the multiplexed signal is encoded, and *parity bits* are added for later error correction and detection. Additional bits are also added to provide information for the listener (such as playing time and track number). The encoded data is then modulated for efficient storage, and more bits (synchronization bits) are added for subsequent recovery of the sampling frequency. The modulated signal is used to control a laser beam that illuminates the photosensitive layer of a rotating glass disc. As the laser turns on or off, the digital information is etched on the photosensitive layer as a pattern of *pits* and *lands* in a spiral track. This master disk forms the basis for mass production of the commercial CD from thermoplastic material.

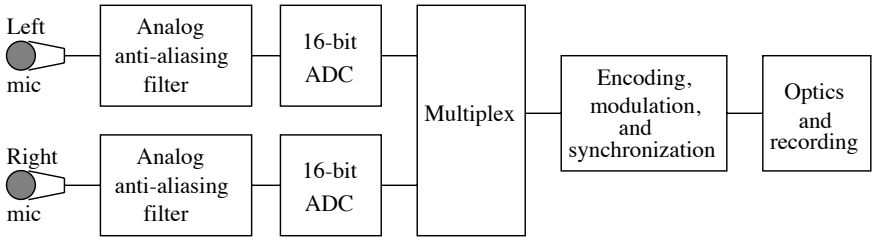


Figure 14.24 Components of a compact disc recording system

How much information can a CD store? With a sampling rate of 44.1 kHz and 32 bits per sample (in stereo), the bit rate is $(44.1)(32)10^3 = (1.41)10^6$ audio bits per second. After encoding, modulation, and synchronization, the number of bits roughly triples to give a bit rate of $(4.32)10^6$ channel bits per second. For a recording time of an hour, this translates to about $(15)10^9$ bits, or 2 gigabytes (with 8 bits corresponding to a byte).

14.5.2 Playback

The CD player reconstructs the audio signal from the information stored on the compact disc in a series of steps that essentially reverses the process at the recording end. A typical CD player is illustrated in Figure 14.25. During playback, the tracks on a CD are optically scanned by a laser to produce a digital signal. This digital signal is demodulated, and the parity bits are used for the detection of any errors (due to manufacturing defects or dust, for example) and to correct the errors by interpolation between samples (if possible) or to mute the signal (if correction is not possible). The demodulated signal is now ready for reconstruction using a DAC. However, the analog reconstruction filter following the DAC must meet tight specifications in order to remove the images that occur at multiples of 44.1 kHz. Even though the images are well above the audible range, they must be filtered out to prevent overloading of the amplifier and speakers. What is done in practice is to digitally oversample the signal (by a factor of 4) to a rate of 176.4 kHz and

PGR2026-00033
Krisp Technologies EX1024 Page 485

pass it through the DAC. A digital filter that compensates for the sinc distortion of the hold operation is also used prior to digital-to-analog conversion. Oversampling relaxes the requirements of the analog filter, which must now smooth out much smaller steps. The sinc compensating filter also provides an additional attenuation of 18 dB for the spectral images and further relaxes the stopband specifications of the analog reconstruction filter. The earliest systems used a third-order Bessel filter with a 3-dB passband of 30 kHz. Another advantage of oversampling is that it reduces the noise floor and spreads the quantization noise over a wider bandwidth. This allows us to round the oversampled signal to 14 bits and use a 14-bit DAC to provide the same level of performance as a 16-bit DAC.

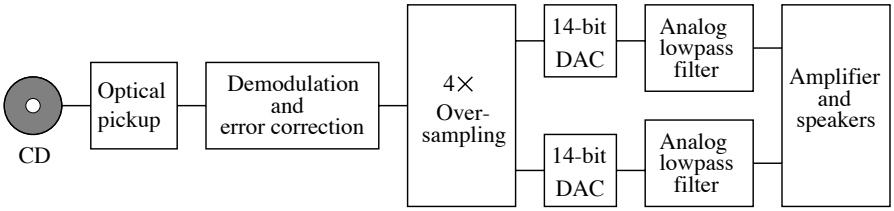


Figure 14.25 Components of a compact disc playback system

14.6 Dynamic Range Processors

There are many instances in which an adjustment of the dynamic range of a signal is desirable. Sound levels above 100 dB appear uncomfortably loud, and the dynamic range should be matched to the listening environment for a pleasant listening experience. Typically, compression is desirable when listening to music with a large dynamic range in small enclosed spaces (unlike a concert hall) such as a living room or an automobile where the ambient (background) noise is not very low. For example, if the music has a dynamic range of 80 dB and the background noise is 40 dB above the threshold of hearing, it can lead to sound levels as high as 120 dB (close to the threshold of pain). Dynamic range compression is also desirable for background music (in stores or elevators). It is also used to prevent distortion when recording on magnetic tape, and in studios to adjust the dynamic range of the individual tracks in a piece of recorded music.

There are also situations where we may like to expand the dynamic range of a signal. For example, the dynamic range of LP records and cassette tapes is not very high (typically, between 50 dB and 70 dB) and can benefit from dynamic range expansion, which in effect makes loud passages louder and soft passages softer and in so doing also reduces the record or tape hiss.

A compressor is a variable gain device whose gain is unity for low signal levels and decreases at higher signal levels. An expander is an amplifier with variable gain (which never exceeds unity). For high signal levels it provides unity gain, whereas for low signal levels it decreases the gain and makes the signal level even lower. Typical compression and expansion characteristics are shown in Figure 14.26 and are usually expressed as a ratio, such as a 2:1 compression or a 1:4 expansion. A 10:1 compression ratio (or higher) describes a **limiter** and represents an extreme form of compression. A 1:10 expansion ratio describes a **noise gate** whose output for low signal levels may be almost zero. Noise gating is one way to eliminate noise or hiss during moments of silence in a recording.

Control of the dynamic range requires variable gain devices whose gain is controlled by the signal level, as shown in Figure 14.27. Dynamic range processing uses a control signal $c(t)$ to adjust the gain. If the gain decreases with increasing $c(t)$, we obtain compression; if the gain increases with increasing $c(t)$, we obtain expansion. Following the signal too closely is undesirable because it would eliminate the dynamic variation altogether. Once the signal level exceeds the threshold, a typical compressor takes up to 0.1 s (called the *attack time*) to respond, and once the level drops below threshold, it takes another second or

PGR2026-00033
Krisp Technologies EX1024 Page 486

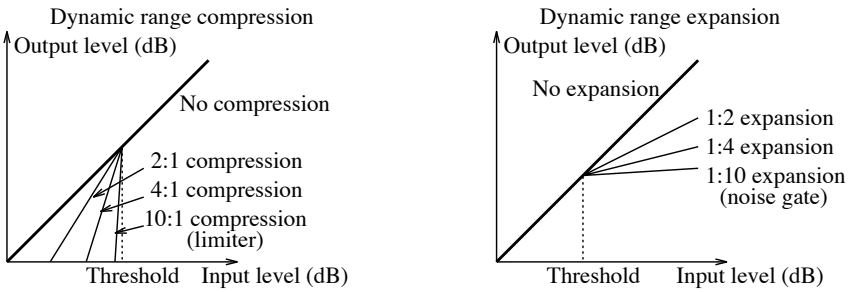


Figure 14.26 Input-output characteristics of dynamic range processors

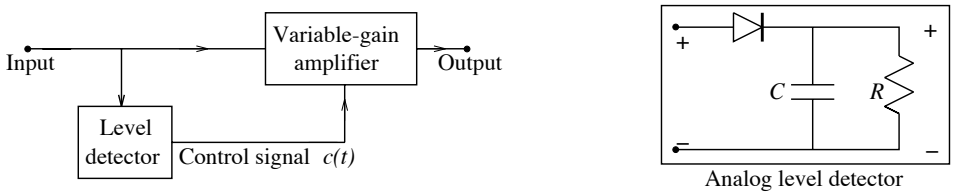


Figure 14.27 Block digram of a dynamic range processor

two (called the *release time*) to restore the gain to unity. Analog circuits for dynamic range processing may use a peak detector (much like the one used for the detection of AM signals) that provides a control signal. Digital circuits replace the rectifier by simple binary operations and simulate the control signal and the attack characteristics and release time by using digital filters. In concept, the compression ratio (or gain), the delay, attack, and release characteristics may be adjusted independently.

14.6.1 Companders

Dynamic range expanders and compressors are often used to combat the effects of noise during transmission of signals, especially if the dynamic range of the channel is limited. A compander is a combination of a compressor and expander. Compression allows us to increase the signal level relative to the noise level. An expander at the receiving end returns the signal to its original dynamic range. This is the principle behind noise reduction systems for both professional and consumer use. An example is the Dolby noise reduction system.

In the professional Dolby A system, the input signal is split into four bands by a lowpass filter with a cutoff frequency of 80 Hz, a bandpass filter with band edges at [80, 3000] Hz, and two highpass filters with cutoff frequencies of 3 kHz and 8 kHz. Each band is compressed separately before being mixed and recorded. During playback, the process is reversed. The characteristics of the compression and expansion are shown in Figure 14.28.

During compression, signal levels below −40 dB are boosted by a constant 10 dB, signal levels between −40 dB and −20 dB are compressed by 2:1, and signal levels above −20 dB are not affected. During playback (expansion), signal levels above −20 dB are cut by 10 dB, signal levels between −30 dB and −20 dB face a 1:2 expansion, and signal levels below −30 dB are not affected. In the immensely popular Dolby B system found in consumer products (and also used by some FM stations), the input signal is not split but a pre-emphasis circuit is used to provide a high-frequency boost above 600 Hz. Another popular system is dbx, which uses pre-emphasis above 1.6 kHz with a maximum high-frequency boost of 20 dB.

PGR2026-00033
Krisp Technologies EX1024 Page 487

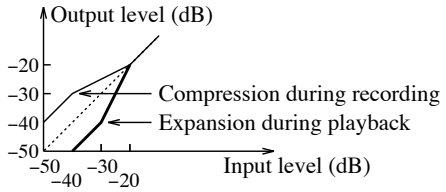


Figure 14.28 Compression and expansion characteristics for Dolby A noise reduction

Voice-grade telephone communication systems also make use of dynamic range compression because the distortion caused is not significant enough to affect speech intelligibility. Two commonly used compressors are the *μ-law compander* (used in North America and Japan) and the *A-law compander* (used in Europe). For a signal $x(t)$ whose peak level is normalized to unity, the two compression schemes are defined by

$$y_\mu(x) = \frac{\ln(1+\mu|x|)}{\ln(1+\mu)}\,\mathrm{sgn}(x) \qquad y_A(x) = \begin{cases} \dfrac{A|x|}{1+\ln A}\,\mathrm{sgn}(x), & 0 \le |x| \le \frac{1}{A} \\ \dfrac{1+\ln(A|x|)}{1+\ln A}\,\mathrm{sgn}(x), & \frac{1}{A} \le |x| \le 1 \end{cases} \tag{14.40}$$

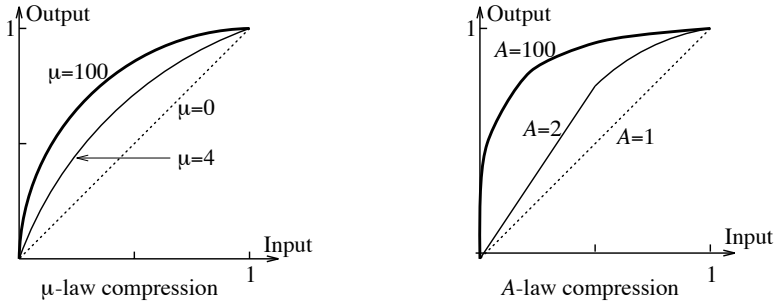


Figure 14.29 Characteristics of μ-law and A-law compressors

The characteristics of these compressors are illustrated in Figure 14.29. The value $\mu = 255$ has become the standard in North America, and $A = 100$ is typically used in Europe. For $\mu = 0$ (and $A = 1$), there is no compression or expansion. The μ-law compander is nearly linear for $\mu|x| \ll 1$. In practice, compression is based on a piecewise linear approximation of the theoretical μ-law characteristic and allows us to use fewer bits to digitize the signal. At the receiving end, an expander (ideally, a true inverse of the compression law) restores the dynamic range of the original signal (except for the effects of quantization). The inverse for μ-law compression is

$$|x| = \frac{(1+\mu)^{|y|} - 1}{\mu} \tag{14.41}$$

The quantization of the compressed signal using the same number of bits as the uncompressed signal results in a higher quantization SNR. For example, the value of $\mu = 255$ increases the SNR by about 24 dB. Since the SNR improves by 6 dB per bit, we can use a quantizer with fewer (only $B - 4$) bits to achieve the same performance as a B-bit quantizer with no compression.

PGR2026-00033
Krisp Technologies EX1024 Page 488

CHAPTER 14 PROBLEMS

DRILL AND REINFORCEMENT

14.1 (Sampling Operations) The signal $x(t) = \cos(4000\pi t)$ is sampled at intervals of t_s. Sketch the sampled signal over $0 \le t \le 2$ ms and also sketch the spectrum of the sampled signal over a frequency range of at least -5 kHz $\le f \le 5$ kHz for the following choices of the sampling function and t_s.

(a) Impulse train (ideal sampling function) with $t_s = 0.2$ ms
(b) Impulse train (ideal sampling function) with $t_s = 0.25$ ms
(c) Impulse train (ideal sampling function) with $t_s = 0.8$ ms
(d) Rectangular pulse train with pulse width $t_d = 0.1$ ms and $t_s = 0.2$ ms
(e) Flat-top (zero-order-hold) sampling with $t_s = 0.2$ ms

14.2 (Sampling Operations) The signal $x(t) = \text{sinc}(4000t)$ is sampled at intervals of t_s. Sketch the sampled signal over $0 \le t \le 2$ ms and also sketch the spectrum of the sampled signal for the following choices of the sampling function and t_s

(a) Impulse train (ideal sampling function) with $t_s = 0.2$ ms
(b) Impulse train (ideal sampling function) with $t_s = 0.25$ ms
(c) Impulse train (ideal sampling function) with $t_s = 0.4$ ms
(d) Rectangular pulse train with pulse width $t_d = 0.1$ ms and $t_s = 0.2$ ms
(e) Flat-top (zero-order-hold) sampling with $t_s = 0.2$ ms

14.3 (Digital Frequency) Express the following signals using a digital frequency $|F| < 0.5$.

(a) $x[n] = \cos(\frac{4n\pi}{3})$
(b) $x[n] = \cos(\frac{4n\pi}{7}) + \sin(\frac{8n\pi}{7})$

14.4 (Sampling Theorem) Establish the Nyquist sampling rate for the following signals.

(a) $x(t) = 5\sin(300\pi t + \pi/3)$
(b) $x(t) = \cos(300\pi t) - \sin(300\pi t + 51^\circ)$
(c) $x(t) = 3\cos(300\pi t) + 5\sin(500\pi t)$
(d) $x(t) = 3\cos(300\pi t)\sin(500\pi t)$
(e) $x(t) = 4\cos^2(100\pi t)$
(f) $x(t) = 6\,\text{sinc}(100t)$
(g) $x(t) = 10\,\text{sinc}^2(100t)$
(h) $x(t) = 6\,\text{sinc}(100t)\cos(200\pi t)$

14.5 (Sampling Theorem) A sinusoid $x(t) = A\cos(2\pi f_0 t)$ is sampled at three times the Nyquist rate for six periods. How many samples are acquired?

14.6 (Sampling Theorem) The sinusoid $x(t) = A\cos(2\pi f_0 t)$ is sampled at twice the Nyquist rate for 1 s. A total of 100 samples is acquired. What is f_0 and the digital frequency of the sampled signal?

14.7 (Sampling Theorem) A sinusoid $x(t) = \sin(150\pi t)$ is sampled at a rate of five samples per three periods. What fraction of the Nyquist sampling rate does this correspond to? What is the digital frequency of the sampled signal?

14.8 (Spectrum of Sampled Signals) Given the spectrum $X(f)$ of an analog signal $x(t)$, sketch the spectrum of its sampled version $x[n]$, assuming a sampling rate of 50, 40, and 30 Hz.

(a) $X(f) = \text{rect}(f/40)$ **(b)** $X(f) = \text{tri}(f/20)$

PGR2026-00033
Krisp Technologies EX1024 Page 489

14.9 (Spectrum of Sampled Signals) Sketch the spectrum of the following sampled signals against the digital frequency F.

(a) $x(t) = \cos(200\pi t)$, ideally sampled at 450 Hz
(b) $x(t) = \sin(400\pi t - \frac{\pi}{4})$, ideally sampled at 300 Hz
(c) $x(t) = \cos(200\pi t) + \sin(350\pi t)$, ideally sampled at 300 Hz
(d) $x(t) = \cos(200\pi t + \frac{\pi}{4}) + \sin(250\pi t - \frac{\pi}{4})$, ideally sampled at 120 Hz

14.10 (Sampling and Aliasing) The signal $x(t) = e^{-t}u(t)$ is sampled at a rate S such that the maximum aliased magnitude (at $f = 0.5S$) is less than 5% of the peak magnitude of the un-aliased image. Estimate the sampling rate S.

14.11 (Signal Recovery) A sinusoid $x(t) = \sin(150\pi t)$ is ideally sampled at 80 Hz. Describe the signal $y(t)$ that is recovered if the sampled signal is passed through the following filters.

(a) An ideal lowpass filter with cutoff frequency $f_C = 10$ Hz
(b) An ideal lowpass filter with cutoff frequency $f_C = 100$ Hz
(c) An ideal bandpass filter with a passband between 60 Hz and 80 Hz
(d) An ideal bandpass filter with a passband between 60 Hz and 100 Hz

14.12 (Interpolation) The signal $x[n] = \{\overset{\Downarrow}{-1},\ 2,\ 3,\ 2\}$ (with $t_s = 1\,\text{s}$) is passed through an interpolating filter.

(a) Sketch the output if the filter performs step interpolation.
(b) Sketch the output if the filter performs linear interpolation.
(c) What is the interpolated value at $t = 2.5$ s if the filter performs sinc interpolation?
(d) What is the interpolated value at $t = 2.5$ s if the filter performs raised cosine interpolation (assume that $R = 0.5$)?

14.13 (Quantization SNR) Consider the signal $x(t) = t^2,\ 0 \le t \le 2$. Choose $t_s = 0.1$ s, four quantization levels, and rounding to find the following:

(a) The sampled signal $x[n]$
(b) The quantized signal $x_Q[n]$
(c) The actual quantization signal to noise ratio SNR_Q
(d) The statistical estimate of the quantization SNR_S
(e) An estimate of the SNR, assuming $x(t)$ to be periodic, with period $T = 2$ s

REVIEW AND EXPLORATION

14.14 (Sampling Theorem) A periodic square wave with period $T = 1$ ms whose value alternates between $+1$ and -1 for each half-period is passed through an ideal lowpass filter with a cutoff frequency of 4 kHz. The filter output is to be sampled. What is the smallest sampling rate we can choose? Consider both the symmetry of the signal as well as the Nyquist criterion.

14.15 (Sampling and Reconstruction) A periodic signal whose one full period is $x(t) = \text{tri}(20t)$ is band-limited by an ideal analog lowpass filter whose cutoff frequency is f_C. It is then ideally sampled at 80 Hz. The sampled signal is reconstructed using an ideal lowpass filter whose cutoff frequency is 40 Hz to obtain the signal $y(t)$. Find $y(t)$ if $f_C = 20,\ 40$, and 60 Hz.

PGR2026-00033
Krisp Technologies EX1024 Page 490

14.16 (Sampling and Reconstruction) Sketch the spectra at the intermediate points and at the output of the following cascaded systems. Assume that the input is $x(t) = 5\text{sinc}(5t)$, the sampler operates at $S = 10$ Hz and performs ideal sampling, and the cutoff frequency of the ideal lowpass filter is 5 Hz.

(a) $x(t) \longrightarrow \boxed{\text{sampler}} \longrightarrow \boxed{\text{ideal LPF}} \longrightarrow y(t)$

(b) $x(t) \longrightarrow \boxed{\text{sampler}} \longrightarrow \boxed{h(t)=u(t)-u(t-0.1)} \longrightarrow \boxed{\text{ideal LPF}} \longrightarrow y(t)$

(c) $x(t) \longrightarrow \boxed{\text{sampler}} \longrightarrow \boxed{h(t)=u(t)-u(t-0.1)} \longrightarrow \boxed{\text{ideal LPF}} \longrightarrow \boxed{|H(f)|=\frac{1}{|\text{sinc}(0.1f)|}} \longrightarrow y(t)$

14.17 (Sampling and Aliasing) A signal $x(t)$ is made up of the sum of pure sines with unit peak value at the frequencies 10, 40, 200, 220, 240, 260, 300, 320, 340, 360, 380, and 400 Hz.

(a) Sketch the magnitude and phase spectra of $x(t)$.

(b) If $x(t)$ is sampled at $S = 140$ Hz, which components, if any, will show aliasing?

(c) The sampled signal is passed through an ideal reconstruction filter whose cutoff frequency is $f_C = 0.5S$. Write an expression for the reconstructed signal $y(t)$ and sketch its magnitude spectrum and phase spectrum. Is $y(t)$ identical to $x(t)$? Should it be? Explain.

(d) What is the minimum sampling rate S that will allow ideal reconstruction of $x(t)$ from its samples?

14.18 (Sampling and Aliasing) The signal $x(t) = \cos(100\pi t)$ is applied to the following systems. Is it possible to find a minimum sampling rate required to sample the system output $y(t)$? If so, find the Nyquist sampling rate.

(a) $y(t) = x^2(t)$ **(b)** $y(t) = x^3(t)$
(c) $y(t) = |x(t)|$ **(d)** $h(t) = \text{sinc}(200t)$
(e) $h(t) = \text{sinc}(500t)$ **(f)** $h(t) = \delta(t-1)$
(g) $y(t) = x(t)\cos(400\pi t)$ **(h)** $y(t) = u[x(t)]$

14.19 (Sampling and Aliasing) The sinusoid $x(t) = \cos(2\pi f_0 t + \theta)$ is sampled at 400 Hz and shows up as a 150-Hz sinusoid upon reconstruction. When the signal $x(t)$ is sampled at 500 Hz, it again shows up as a 150-Hz sinusoid upon reconstruction. It is known that $f_0 < 2.5$ kHz.

(a) If each sampling rate exceeds the Nyquist rate, what is f_0?

(b) By sampling $x(t)$ again at a different sampling rate, explain how you might establish whether aliasing has occurred?

(c) If aliasing occurs and the reconstructed phase is θ, find all possible values of f_0.

(d) If aliasing occurs but the reconstructed phase is $-\theta$, find all possible values of f_0.

14.20 (Sampling and Aliasing) An even symmetric periodic triangular waveform $x(t)$ with period $T = 4$ is ideally sampled by the impulse train $i(t) = \sum_{k=-\infty}^{\infty} \delta(t-k)$ to obtain the impulse sampled signal $x_s(t)$ described over one period by $x_s(t) = \delta(t) - \delta(t-2)$.

(a) Sketch the signals $x(t)$ and $x_s(t)$.

(b) If $x_s(t)$ is passed through an ideal lowpass filter with a cutoff frequency of 0.6 Hz, sketch the filter output $y(t)$.

(c) How do $x_s(t)$ and $y(t)$ change if the sampling function is $i(t) = \sum_{k=-\infty}^{\infty} (-1)^k \delta(t-k)$?

PGR2026-00033
Krisp Technologies EX1024 Page 491

14.21 (Sampling Oscilloscopes) It is required to ideally sample a signal $x(t)$ at S Hz and pass the sampled signal $s(t)$ through an ideal lowpass filter with a cutoff frequency of $0.5S$ Hz such that its output $y(t) = x(t/\alpha)$ is a stretched-by-α version of $x(t)$.

(a) Suppose $x(t) = 1 + \cos(20\pi t)$. What value of S will ensure that the output of the lowpass filter is $y(t) = x(0.1t)$. Sketch the spectra of $x(t)$, $s(t)$, and $y(t)$ for the chosen value of S.

(b) Suppose $x(t) = 2\cos(80\pi t) + \cos(160\pi t)$ and the sampling rate is chosen as $S = 48$ Hz. Sketch the spectra of $x(t)$, $s(t)$, and $y(t)$. Will the output $y(t)$ be a stretched version of $x(t)$ with $y(t) = x(t/\alpha)$? If so, what will be the value of α?

(c) Suppose $x(t) = 2\cos(80\pi t) + \cos(100\pi t)$. What value of S will ensure that the output of the lowpass filter is $y(t) = x(t/20)$? Sketch the spectra of $x(t)$, $s(t)$, and $y(t)$ for the chosen value of S to confirm your results.

14.22 (Bandpass Sampling) The signal $x(t)$ is band-limited to 500 Hz. The smallest frequency present in $x(t)$ is f_0. Find the minimum rate S at which we can sample $x(t)$ without aliasing and explain how we can recover the signal if

(a) $f_0 = 0$.

(b) $f_0 = 300$ Hz.

(c) $f_0 = 400$ Hz.

14.23 (Bandpass Sampling) A signal $x(t)$ is band-limited to 40 Hz and modulated by a 320-Hz carrier to generate the modulated signal $y(t)$. The modulated signal is processed by a square law device that produces $g(t) = y^2(t)$.

(a) What is the minimum sampling rate for $x(t)$ to prevent aliasing?

(b) What is the minimum sampling rate for $y(t)$ to prevent aliasing?

(c) What is the minimum sampling rate for $g(t)$ to prevent aliasing?

14.24 (Interpolation) We wish to sample a speech signal band-limited to 4 kHz using zero-order-hold sampling.

(a) Select the sampling frequency if the spectral magnitude of the sampled signal at 4 kHz is to be within 90% of its peak magnitude.

(b) On recovery, the signal is filtered using a Butterworth filter with an attenuation of less than 1 dB in the passband and more than 30 dB for all image frequencies. Compute the total attenuation in decibels due to both the sampling and filtering operations at 4 kHz and 12 kHz.

(c) What is the order of the Butterworth filter?

14.25 (Quantization SNR) A sinusoid with a peak value of 4 V is sampled and then quantized by a 12-bit quantizer whose full-scale range is ± 5 V. What is the quantization SNR of the quantized signal?

14.26 (Quantization Noise Power) The quantization noise power based on quantization by rounding is $\sigma^2 = \Delta^2/12$, where Δ is the quantization step size. Find similar expressions for the quantization noise power based on quantization by truncation and quantization by sign-magnitude truncation.

14.27 (Anti-Aliasing Filters) A speech signal is to be band-limited using an anti-aliasing third-order Butterworth filter with a half-power frequency of 4 kHz, and then sampled and quantized by an 8-bit quantizer. Determine the minimum stopband attenuation and the corresponding stopband frequency to ensure that the maximum stopband aliasing level (relative to the passband edge) is less than the rms quantization error.

PGR2026-00033
Krisp Technologies EX1024 Page 492

14.28 (Anti-Aliasing Filters) A speech signal with amplitude levels of ± 1 V is to be band-limited using an anti-aliasing second-order Butterworth filter with a half-power frequency of 4 kHz, and then sampled and quantized by an 8-bit quantizer. What minimum sampling rate S will ensure that the maximum aliasing error at the passband edge is less than the rms quantization error?

14.29 (Anti-Aliasing Filters) A noise-like signal with a flat magnitude spectrum is filtered using a third-order Butterworth filter with a half-power frequency of 3 kHz, and the filtered signal is sampled at 10 kHz. What is the aliasing level (relative to the signal level) at the half-power frequency in the sampled signal?

14.30 (Sampling and Quantization) A sinusoid with amplitude levels of ± 1 V is quantized by rounding, using a 12-bit quantizer. What is the rms quantization error and the quantization SNR?

14.31 (Sampling and Quantization) The signal $x(t) = 2\cos(2000\pi t) - 4\sin(4000\pi t)$ is quantized by rounding, using a 12-bit quantizer. What is the rms quantization error and the quantization SNR?

14.32 (Sampling and Quantization) A speech signal, band-limited to 4 kHz, is to be sampled and quantized by rounding, using an 8-bit quantizer. What is the conversion time of the quantizer if it is preceded by a sample-and-hold circuit with an aperture time of 20 ns and an acquisition time of $2\,\mu$s? Assume sampling at the Nyquist rate.

14.33 (Sampling and Quantization) A 10-kHz sinusoid with amplitude levels of ± 1 V is to be sampled and quantized by rounding. How many bits are required to ensure a quantization SNR of 45 dB? What is the bit rate (number of bits per second) of the digitized signal if the sampling rate is chosen as twice the Nyquist rate?

14.34 (Anti-Imaging Filters) A digitized speech signal, band-limited to 4 kHz, is to be reconstructed using a zero-order-hold. What minimum reconstruction sampling rate will ensure that the signal level in the passband is attenuated by less than 1.2 dB due to the sinc distortion of the zero-order-hold? What will be the image rejection at the stopband edge?

14.35 (Anti-Imaging Filters) A digitized speech signal, band-limited to 4 kHz, is to be reconstructed using a zero-order-hold followed by an analog Butterworth lowpass filter. The signal level in the passband should be attenuated less than 1.5 dB, and an image rejection of better than 45 dB in the stopband is required. What are the specifications for the Butterworth filter if the reconstruction sampling rate is 16 kHz? What is the order of the Butterworth filter?

14.36 (Anti-Aliasing and Anti-Imaging Filters) A speech signal is to be band-limited using an anti-aliasing Butterworth filter with a half-power frequency of 4 kHz. The sampling frequency is 20 kHz. What filter order will ensure that the in-band aliasing level is be less 1% of the signal level? The processed signal is to be reconstructed using a zero-order-hold. What is the stopband attenuation required of an anti-imaging filter to ensure image rejection of better than 50 dB?

COMPUTATION AND DESIGN

14.37 (Interpolating Functions) Consider the signal $x(t) = \cos(0.5\pi t)$ sampled at $S = 1$ Hz to generate the sampled signal $x[n]$. We wish to compute the value of $x(t)$ at $t = 0.5$ by interpolating between its samples.

(a) Superimpose a plot of $x(t)$ and its samples $x[n]$ over one period. What is the value of $x(t)$ predicted by step interpolation and linear interpolation of $x[n]$? How good are these estimates?

PGR2026-00033
Krisp Technologies EX1024 Page 493

Can these estimates be improved by taking more signal samples (using the same interpolation schemes)?

(b) Use the sinc interpolation formula

$$x(t) = \sum_{n=-\infty}^{\infty} x[n]\text{sinc}[(t - nt_s)/t_s]$$

to obtain an estimate of $x(0.5)$. With $t = 0.5$ and $t_s = 1/S = 1$, compute the summation for $|n| \leq 10,\ 20,\ 50$ to generate three estimates of $x(0.5)$. How good are these estimates? Would you expect the estimate to converge to the actual value as more terms are included in the summation (i.e., as more signal samples are included)? Compare the advantages and disadvantages of sinc interpolation with the schemes in part (a).

14.38 (Interpolating Functions) To interpolate a signal $x[n]$ by N, we use an up-sampler (that places $N - 1$ zeros after each sample) followed by a filter that performs the appropriate interpolation as shown:

$$x[n] \longrightarrow \boxed{\text{up-sample} \uparrow N} \longrightarrow \boxed{\text{interpolating filter}} \longrightarrow y[n]$$

The filter impulse response for step interpolation, linear interpolation, and ideal (sinc) interpolation is

$$h_S[n] = u[n] - u[n - (N - 1)] \qquad h_L[n] = \text{tri}(n/N) \qquad h_I[n] = \text{sinc}(n/N),\ |n| \leq M$$

Note that the ideal interpolating function is actually of infinite length but must be truncated in practice. Generate the test signal $x[n] = \cos(0.5n\pi),\ 0 \leq n \leq 3$. Up-sample this by $N = 8$ (seven zeros after each sample) to obtain the signal $x_U[n]$. Use the MATLAB routine `filter` to filter $x_U[n]$, using

(a) The step interpolation filter to obtain the filtered signal $x_S[n]$. Plot $x_U[n]$ and $x_S[n]$ on the same plot. Does the system perform the required interpolation? Does the result look like a sine wave?

(b) The linear interpolation filter to obtain the filtered signal $x_L[n]$. Plot $x_U[n]$ and a delayed (by 8) version of $x_L[n]$ (to account for the noncausal nature of $h_L[n]$) on the same plot. Does the system perform the required interpolation? Does the result look like a sine wave?

(c) The ideal interpolation filter (with $M = 4,\ 8,\ 16$) to obtain the filtered signal $x_I[n]$. Plot $x_U[n]$ and a delayed (by M) version of $x_I[n]$ (to account for the noncausal nature of $h_I[n]$) on the same plot. Does the system perform the required interpolation? Does the result look like a sine wave? What is the effect of increasing M on the interpolated signal? What is the effect of increasing both M and the signal length? Explain.

14.39 (Sampling and Quantization) The signal $x(t) = \cos(2\pi t) + \cos(6\pi t)$ is sampled at $S = 20$ Hz.

(a) Generate 200 samples of $x[n]$ and superimpose plots of $x(t)$ vs. t and $x[n]$ vs. n/S. Use a 4-bit quantizer to quantize $x[n]$ by rounding and obtain the signal $x_R[n]$. Superimpose plots of $x(t)$ vs. t and $x_R[n]$ vs. n/S. Obtain the quantization error signal $e[n] = x[n] - x_R[n]$ and plot $e[n]$ and its 10-bin histogram. Does it show a uniform distribution? Compute the quantization SNR in decibels. Repeat for 800 samples. Does an increase in the signal length improve the SNR?

(b) Generate 200 samples of $x[n]$ and superimpose plots of $x(t)$ vs. t and $x[n]$ vs. n/S. Use an 8-bit quantizer to quantize $x[n]$ by rounding and obtain the signal $x_R[n]$. Superimpose plots of $x(t)$ vs. t and $x_R[n]$ vs. n/S. Obtain the quantization error signal $e[n] = x[n] - x_R[n]$ and

PGR2026-00033
Krisp Technologies EX1024 Page 494

plot $e[n]$ and its 10-bin histogram. Compute the quantization SNR in decibels. Compare the quantization error, the histogram, and the quantization SNR for the 8-bit and 4-bit quantizer and comment on any differences. Repeat for 800 samples. Does an increase in the signal length improve the SNR?

(c) Based on your knowledge of the signal $x(t)$, compute the theoretical SNR for a 4-bit and 8-bit quantizer. How does the theoretical value compare with the quantization SNR obtained in parts (a) and (b)?

(d) Repeat parts (a) and (b), using quantization by truncation of $x[n]$. How does the quantization SNR for truncation compare with the the quantization SNR for rounding?

PGR2026-00033
Krisp Technologies EX1024 Page 495

Chapter 15

THE DISCRETE-TIME FOURIER TRANSFORM

15.0 Scope and Objectives

The spectrum of a sampled signal is not only continuous but also periodic. This periodicity is a consequence of the duality and reciprocity between time and frequency and leads to the formulation of the *discrete-time Fourier transform* (DTFT). This chapter develops the DTFT as a tool for the frequency domain description of discrete-time signals and systems. It describes the properties of the DTFT and concludes with applications of the DTFT to system analysis and signal processing.

15.1 The Discrete-Time Fourier Transform

The **discrete-time Fourier transform** describes the spectrum of discrete-time signals and formalizes the concept that discrete-time signals have periodic spectra. Ideal sampling of an analog signal $x(t)$ leads to the ideally sampled signal $x_I(t)$ whose spectrum $X_p(f)$ is periodic. We have

$$x_I(t) = \sum_{k=-\infty}^{\infty} x(kt_s)\delta(t - kt_s) \qquad X_p(f) = S\sum_{k=-\infty}^{\infty} X(f - kS) \tag{15.1}$$

Using the Fourier transform pair $\delta(t-\alpha) \Leftrightarrow \exp(-j2\pi\alpha f)$, the spectrum $X_p(f)$ may also be described by

$$x_I(t) = \sum_{k=-\infty}^{\infty} x(kt_s)\delta(t - kt_s) \qquad X_p(f) = \sum_{k=\infty}^{\infty} x(kt_s)e^{-j2\pi kt_s f} \tag{15.2}$$

Note that $X_p(f)$ is periodic with period S and its central period equals $SX(f)$. To recover the analog signal $x(t)$, we pass the sampled signal through an ideal lowpass filter whose gain equals $1/S$ over $-0.5S \le f \le 0.5S$. Formally, we obtain $x(t)$ (or its samples $x(nt_s)$) from the inverse Fourier transform result

$$x(t) = \frac{1}{S}\int_{-S/2}^{S/2} X_p(f)e^{j2\pi ft}\,df \qquad x(nt_s) = \frac{1}{S}\int_{-S/2}^{S/2} X_p(f)e^{j2\pi fnt_s}\,df \tag{15.3}$$

Equations (15.2) and (15.3) define a transform pair. They allow us to obtain the periodic spectrum $X_p(f)$ of an ideally sampled signal from its samples $x(nt_s)$, and to recover the samples $x(nt_s)$ from the spectrum. We point out that these relations are the exact duals of the Fourier series relations for a periodic signal $x_p(t)$ and its discrete spectrum $X[k]$ (the Fourier series coefficients). We can revise these relations for discrete-time signals if we use the *digital* frequency $F = f/S$ and replace $x(nt_s)$ by the discrete sequence $x[n]$ to obtain

$$X_p(F) = \sum_{k=-\infty}^{\infty} x[k]e^{-j2\pi kF} \qquad x[n] = \int_{-1/2}^{1/2} X_p(F)e^{j2\pi nF}\,dF \qquad \text{(the } F\text{-form)} \tag{15.4}$$

PGR2026-00033
Krisp Technologies EX1024 Page 496

The first result defines $X_p(F)$ as the discrete-time Fourier transform (DTFT) of $x[n]$. The second result is the inverse DTFT (IDTFT), which allows us to recover $x[n]$ from its spectrum. The DTFT $X_p(F)$ is periodic with *unit period* because it assumes unit spacing between samples of $x[n]$. The interval $-0.5 \le F \le 0.5$ (or $0 \le F \le 1$) defines the **principal period**.

The DTFT relations may also be written in terms of the radian frequency Ω as

$$X_p(\Omega) = \sum_{k=-\infty}^{\infty} x[k]e^{-jk\Omega} \qquad x[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} X_p(\Omega)e^{jn\Omega}\,d\Omega \qquad \text{(the } \Omega\text{-form)} \tag{15.5}$$

The quantity $X_p(\Omega)$ is now periodic with period $\Omega = 2\pi$ and represents a scaled (stretched by 2π) version of $X_p(F)$. The principal period of $X_p(\Omega)$ corresponds to the interval $-\pi \le \Omega \le \pi$ or $0 \le \Omega \le 2\pi$. We will find it convenient to work with the F-form because, as in the case of Fourier transforms, it rids us of factors of 2π in many situations.

REVIEW PANEL 15.1

The DTFT Is a Frequency-Domain Representation of Discrete-Time Signals

Form	**DTFT**	**Inverse DTFT**
F-form	$X_p(F) = \sum_{k=-\infty}^{\infty} x[k]e^{-j2\pi kF}$	$x[n] = \int_{-1/2}^{1/2} X_p(F)e^{j2\pi nF}\,dF$
Ω-form	$X_p(\Omega) = \sum_{k=-\infty}^{\infty} x[k]e^{-jk\Omega}$	$x[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} X_p(\Omega)e^{jn\Omega}\,d\Omega$

15.2 Connections: The DTFT and the Fourier Transform

If the signal $x[n]$ is obtained by ideal sampling of an analog signal $x(t)$ at a sampling rate S, the Fourier transform $X_I(f)$ of the *analog impulse train* $x_I(t) = \sum x(t)\delta(t - k/S)$ equals $X_p(F)|_{F\to f/S}$ and represents a frequency-scaled version of $X_p(F)$ with principal period $(-0.5S,\ 0.5S)$. If S exceeds the Nyquist sampling rate, the Fourier transform $X(f)$ of $x(t)$ equals the *principal period* of $SX_p(F)|_{F\to f/S}$. If S is below the Nyquist rate, $SX_p(F)$ matches the *periodic extension* of $X(f)$. In other words, the DTFT of a discrete-time signal $x[n]$ is related to both the Fourier transform $X_I(f)$ of the underlying impulse-sampled analog signal $x_I(t)$ and to the Fourier transform $X(f)$ of $x(t)$.

REVIEW PANEL 15.2

The DTFT of a Sampled Analog Signal $x(t)$ Is Related to Its Fourier Transform $X(f)$

The Fourier transform of the analog impulse train $x_I(t) = \sum x(t)\delta(t - k/S)$ equals $X_I(f) = X_p(F)|_{F\to f/S}$.
If $x(t)$ is sampled above the Nyquist rate, then $X(f) = SX_p(F)|_{F\to f/S},\ -0.5S \le f \le 0.5S$.
If $x(t)$ is sampled below the Nyquist rate, $SX_p(F)$ matches the *periodic extension* of $X(f)$.

15.2.1 Symmetry of the DTFT

The DTFT is, in general, a complex quantity, which may be expressed as

$$X_p(F) = R(F) + jI(F) = |X_p(F)|e^{j\phi(F)} = |X_p(F)|\angle\phi(F) \qquad \text{(the } F\text{-form)} \tag{15.6}$$

$$X_p(\Omega) = R(\Omega) + jI(\Omega) = |X_p(\Omega)|e^{j\phi(\Omega)} = |X_p(\Omega)|\angle\phi(\Omega) \qquad \text{(the } \Omega\text{-form)} \tag{15.7}$$

PGR2026-00033
Krisp Technologies EX1024 Page 497

For real signals, the DTFT shows *conjugate symmetry* about $F = 0$ (or $\Omega = 0$) with

$$X_p(F) = X_p^*(-F) \qquad |X_p(F)| = |X_p(-F)| \qquad \phi(F) = -\phi(-F) \qquad \text{(the } F\text{-form)} \tag{15.8}$$

$$X_p(\Omega) = X_p^*(-\Omega) \qquad |X_p(\Omega)| = |X_p(-\Omega)| \qquad \phi(\Omega) = -\phi(-\Omega) \qquad \text{(the } \Omega\text{-form)} \tag{15.9}$$

This means that its magnitude spectrum has *even symmetry* and its phase spectrum has *odd symmetry* about the origin. The DTFT $X_p(F)$ also shows conjugate symmetry about $F = 0.5$. Since it is periodic, we find it convenient to plot just one period of $X_p(F)$ over the principal period $(-0.5 \le F \le 0.5)$ with conjugate symmetry about $F = 0$. We may even plot $X_p(F)$ over $(0 \le F \le 1)$ with conjugate symmetry about $F = 0.5$. These ideas are illustrated in Figure 15.1.

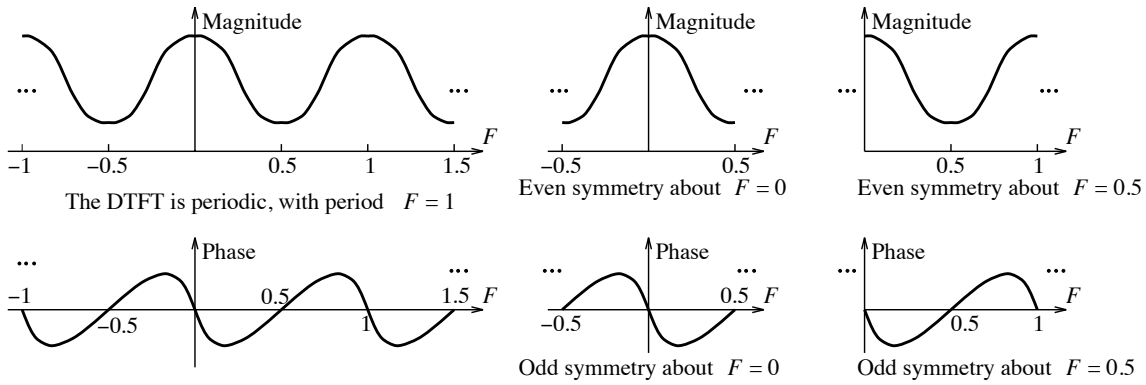


Figure 15.1 Illustrating the symmetry in the DTFT spectrum of real signals

Naturally, if we work with $X_p(\Omega)$, it shows conjugate symmetry about the origin $\Omega = 0$, and about $\Omega = \pi$, and may be plotted only over its principal period $(-\pi \le \Omega \le \pi)$ (with conjugate symmetry about $\Omega = 0$) or over $(0 \le \Omega \le 2\pi)$ (with conjugate symmetry about $\Omega = \pi$). The principal range for each form is illustrated in Figure 15.2.



Figure 15.2 Various ways of plotting the DTFT spectrum

REVIEW PANEL 15.3

The DTFT Is Always Periodic and Shows Conjugate Symmetry for Real Signals

F-form: $X(F)$ is periodic with unit period and conjugate symmetric about $F = 0$ and $F = 0.5$.
Ω-form: $X(\Omega)$ is periodic with period 2π and conjugate symmetric about $\Omega = 0$ and $\Omega = \pi$.
Plotting: It is sufficient to plot the DTFT over one period ($-0.5 \le F \le 0.5$ or $-\pi \le \Omega \le \pi$).

PGR2026-00033
Krisp Technologies EX1024 Page 498

If a real signal $x[n]$ is even symmetric, its DTFT is always real and even symmetric in F, and has the form $X_p(F) = A(F)$. If a real signal $x[n]$ is odd symmetric, its DTFT is always imaginary and odd symmetric in F, and has the form $X_p(F) = \pm jA(F)$. A real symmetric signal is called a **linear-phase signal**. The quantity $A(F)$ is called the **amplitude spectrum**. For linear-phase signals, it is much more convenient to plot just the amplitude spectrum (rather than the magnitude spectrum and phase spectrum separately).

REVIEW PANEL 15.4

The DTFT of Real Symmetric Signals Is Purely Real or Purely Imaginary

Even symmetric $x[n] \Leftrightarrow A(F)$ or $X(\Omega)$ (real) Odd symmetric $x[n] \Leftrightarrow jA(F)$ or $jA(\Omega)$ (imaginary)
Plotting: It is convenient to plot just the amplitude spectrum $A(F)$ or $A(\Omega)$.

15.2.2 Some DTFT Pairs

The DTFT is a summation. It always exists if the summand $x[n]e^{j2\pi nF} = x[n]e^{jn\Omega}$ is absolutely integrable. Since $|e^{j2\pi nF}| = |e^{jn\Omega}| = 1$, the DTFT of absolutely summable signals always exists. In analogy with the Fourier transform, the DTFT of sinusoids, steps, or constants, which are not absolutely summable, includes impulses (of the continuous kind). The DTFT of many energy (square summable) signals is a rational function of (ratio of polynomials in) $e^{j2\pi F}$ (or $e^{j\Omega}$). The DTFT of signals that grow exponentially or faster does not exist. A list of DTFT pairs appears in Table 15.1. Both the F-form and Ω-form are shown. The transforms are identical (with $\Omega = 2\pi F$), except for the extra factors of 2π in the impulsive terms of the Ω-form (in the transform of the constant, the step function, and the sinusoid). The reason for these extra factors is the scaling property of impulses that says $\delta(F) = \delta(\Omega/2\pi) = 2\pi\delta(\Omega)$.

REVIEW PANEL 15.5

Differences Between the F-Form and Ω-Form of the DTFT

If the DTFT contains no impulses: $H(F)$ and $H(\Omega)$ are related by $\Omega = 2\pi F$.
If the DTFT contains impulses: Replace $\delta(F)$ by $2\pi\delta(\Omega)$ (and $2\pi F$ by Ω elsewhere) to get $H(\Omega)$.

EXAMPLE 15.1 (DTFT from the Defining Relation)

(a) The DTFT of $x[n] = \delta[n]$ follows immediately from the definition as

$$X_p(F) = \sum_{k=-\infty}^{\infty} x[k]e^{-j2\pi kF} = \sum_{k=-\infty}^{\infty} \delta[k]e^{-j2\pi kF} = 1$$

(b) The DTFT of the sequence $x[n] = \{1, 0, 3, -2\}$ also follows from the definition as

$$X_p(F) = \sum_{k=-\infty}^{\infty} x[k]e^{-j2\pi kF} = 1 + 3e^{-j4\pi F} - 2e^{-j6\pi F}$$

In the Ω-form, we have

$$X_p(\Omega) = \sum_{k=-\infty}^{\infty} x[k]e^{-jk\Omega} = 1 + 3e^{-j2\Omega} - 2e^{-j3\Omega}$$

For finite sequences, the DTFT can be written just by inspection. Each term is the product of a sample value at index n and the exponential $e^{-j2\pi nF}$ (or $e^{-jn\Omega}$).

PGR2026-00033
Krisp Technologies EX1024 Page 499

Table 15.1 Some Useful DTFT Pairs

Note: In all cases, we assume $\|\alpha\| < 1$.			
Entry	**Signal** $x[n]$	**The** F**-Form:** $X_p(F)$	**The** Ω**-Form:** $X_p(\Omega)$
1	$\delta[n]$	1	1
2	$\alpha^n u[n],\ \|\alpha < 1\|$	$\dfrac{1}{1-\alpha e^{-j2\pi F}}$	$\dfrac{1}{1-\alpha e^{-j\Omega}}$
3	$n\alpha^n u[n],\ \|\alpha < 1\|$	$\dfrac{\alpha e^{-j2\pi F}}{(1-\alpha e^{-j2\pi F})^2}$	$\dfrac{\alpha e^{-j\Omega}}{(1-\alpha e^{-j\Omega})^2}$
4	$(n+1)\alpha^n u[n],\ \|\alpha < 1\|$	$\dfrac{1}{(1-\alpha e^{-j2\pi F})^2}$	$\dfrac{1}{(1-\alpha e^{-j\Omega})^2}$
5	$\alpha^{\|n\|},\ \|\alpha < 1$	$\dfrac{1-\alpha^2}{1-2\alpha\cos(2\pi F)+\alpha^2}$	$\dfrac{1-\alpha^2}{1-2\alpha\cos\Omega+\alpha^2}$
6	1	$\delta(F)$	$2\pi\delta(\Omega)$
7	$\cos(2n\pi F_0) = \cos(n\Omega_0)$	$0.5[\delta(F+F_0)+\delta(F-F_0)]$	$\pi[\delta(\Omega+\Omega_0)+\delta(\Omega-\Omega_0)]$
8	$\sin(2n\pi F_0) = \sin(n\Omega_0)$	$j0.5[\delta(F+F_0)-\delta(F-F_0)]$	$j\pi[\delta(\Omega+\Omega_0)-\delta(\Omega-\Omega_0)]$
9	$2F_C\,\mathrm{sinc}(2nF_C) = \dfrac{\sin(n\Omega_C)}{n\pi}$	$\mathrm{rect}\Big(\dfrac{F}{2F_C}\Big)$	$\mathrm{rect}\Big(\dfrac{\Omega}{2\Omega_C}\Big)$
10	$u[n]$	$0.5\delta(F) + \dfrac{1}{1-e^{-j2\pi F}}$	$\pi\delta(\Omega) + \dfrac{1}{1-e^{-j\Omega}}$

(c) The DTFT of the exponential signal $x[n] = \alpha^n u[n]$ follows from the definition and the closed form for the resulting geometric series:

$$X_p(F) = \sum_{k=0}^{\infty} \alpha^k e^{-j2\pi kF} = \sum_{k=0}^{\infty} \left(\alpha e^{-j2\pi F}\right)^k = \frac{1}{1-\alpha e^{-j2\pi F}}, \qquad |\alpha| < 1$$

The sum converges only if $|\alpha e^{-j2\pi F}| < 1$, or $|\alpha| < 1$ (since $|e^{-j2\pi F}| = 1$). In the Ω-form,

$$X_p(\Omega) = \sum_{k=0}^{\infty} \alpha^k e^{-jk\Omega} = \sum_{k=0}^{\infty} \left(\alpha e^{-j\Omega}\right)^k = \frac{1}{1-\alpha e^{-j\Omega}}, \qquad |\alpha| < 1$$

(d) The signal $x[n] = u[n]$ is a limiting form of $\alpha^n u[n]$ as $\alpha \to 1$ but must be handled with care, since $u[n]$ is not absolutely summable. In analogy with the analog pair $u(t) \Leftrightarrow 0.5\delta(f) + \frac{1}{j2\pi f}$, $X_p(F)$ also includes an impulse (now an impulse train due to the periodic spectrum). Over the principal period,

$$X_p(F) = \frac{1}{1-e^{-j2\pi F}} + 0.5\delta(F) \quad (F\text{-form}) \qquad\qquad X_p(\Omega) = \frac{1}{1-e^{-j\Omega}} + \pi\delta(\Omega) \quad (\Omega\text{-form})$$

PGR2026-00033
Krisp Technologies EX1024 Page 500

15.3 Properties of the DTFT

The properties of the DTFT are summarized in Table 15.2 for both the F-form and the Ω-form. Many of these are duals of the corresponding Fourier series properties. The differences in how we apply the properties are based on the fact that for Fourier series we typically deal with real periodic signals and discrete spectra with conjugate symmetry while for the DTFT, the discrete signal is typically real, while the periodic spectrum shows conjugate symmetry. As with the Fourier series (and transform) relations, the proofs of most of the properties follow from the defining relations if we start with the basic transform pair $x[n] \Leftrightarrow X_p(F)$.

Table 15.2 Properties of the DTFT

Property	**DT Signal**	**Result (F-Form)**	**Result (Ω-Form)**
Folding	$x[-n]$	$X_p(-F) = X_p^*(F)$	$X_p(-\Omega) = X_p^*(\Omega)$
Time shift	$x[n-m]$	$e^{-j2\pi mF}X_p(F)$	$e^{-j\Omega m}X_p(\Omega)$
Frequency shift	$e^{j2\pi nF_0}x[n]$	$X_p(F-F_0)$	$X_p(\Omega-\Omega_0)$
Half-period shift	$(-1)^n x[n]$	$X_p(F-0.5)$	$X_p(\Omega-\pi)$
Modulation	$\cos(2\pi nF_0)x[n]$	$0.5[X_p(F+F_0)+X_p(F-F_0)]$	$0.5[(\Omega+\Omega_0)+X_p(\Omega-\Omega_0)]$
Convolution	$x[n] \star y[n]$	$X_p(F)Y_p(F)$	$X_p(\Omega)Y_p(\Omega)$
Product	$x[n]y[n]$	$X_p(F) \circledast Y_p(F)$	$\frac{1}{2\pi}[X_p(\Omega) \circledast Y_p(\Omega)]$
Times-n	$nx[n]$	$\frac{j}{2\pi}\frac{dX_p(F)}{dF}$	$j\frac{dX_p(\Omega)}{d\Omega}$
Parseval's relation	$\sum_{k=-\infty}^{\infty} x^2[k] = \int_1 \lvert X_p(F)\rvert^2\, dF = \frac{1}{2\pi}\int_{2\pi} \lvert X_p(\Omega)\rvert^2\, d\Omega$		
Central ordinates	$x[0] = \int_1 X_p(F)\, dF = \frac{1}{2\pi}\int_{2\pi} X_p(\Omega)\, d\Omega \qquad X_p(0) = \sum_{n=-\infty}^{\infty} x[n]$		
	$X_p(F)\Big\rvert_{F=0.5} = X_p(\Omega)\Big\rvert_{\Omega=\pi} = \sum_{n=-\infty}^{\infty} (-1)^n x[n]$		

Folding: With $x[n] \Leftrightarrow X_p(F)$, the DTFT of the signal $y[n] = x[-n]$ may be written (using a change of variable $m = -n$) as

$$Y_p(F) = \sum_{n=-\infty}^{\infty} x[-n]e^{-j2n\pi F} = \sum_{m=-\infty}^{\infty} x[l]e^{j2m\pi F} = X_p(-F) \tag{15.10}$$

A folding of $x[n]$ to $x[-n]$ results in a folding of $X_p(F)$ to $X_p(-F)$. The magnitude spectrum of the folded signal stays the same, but the phase is reversed.

REVIEW PANEL 15.6

Folding $x[n]$ to $x[-n]$ Results in Folding $X_p(F)$ to $X_p(-F)$

The magnitude spectrum stays the same, and only the phase is reversed (changes sign).

PGR2026-00033
Krisp Technologies EX1024 Page 501

Time shift: With $x[n] \Leftrightarrow X_p(F)$, the DTFT of the signal $y[n] = x[n-m]$ may be written (using a change of variable $l = n - m$) as

$$Y_p(F) = \sum_{n=-\infty}^{\infty} x[n-m]e^{-j2n\pi F} = \sum_{l=-\infty}^{\infty} x[l]e^{-j2(m+l)\pi F} = X_p(F)e^{-j2\pi mF} \tag{15.11}$$

A time shift of $x[n]$ to $x[n-m]$ does not affect the magnitude spectrum. It augments the phase spectrum by $\theta(F) = -2\pi mF$ (or $\theta(\Omega) = -m\Omega$), which varies linearly with frequency.

Frequency shift: By duality, a frequency shift of $X_p(F)$ to $X_p(F - F_0)$ yields the signal $x[n]e^{j2\pi nF_0}$.

Half-period shift: If $X_p(F)$ is shifted by 0.5 to $X_p(F \pm 0.5)$, then $x[n]$ changes to $e^{\mp jn\pi} = (-1)^n x[n]$. Thus, samples of $x[n]$ at odd index values ($n = \pm 1, \pm 3, \pm 5, \ldots$) change sign.

REVIEW PANEL 15.7

The Shift Properties of the DTFT

$x[n-m] \Longleftrightarrow X_p(F)e^{-j2\pi mF}$ or $X_p(\Omega)e^{-jm\Omega}$ $\qquad$ $x[n]e^{j2\pi nF_0} \Longleftrightarrow X_p(F-F_0)$ or $X_p(\Omega-\Omega_0)$ $(\Omega_0 = 2\pi F_0)$

$(-1)^n x[n] \Longleftrightarrow X_p(F \pm 0.5)$ or $X_p(\Omega \pm \pi)$

A time delay by m adds a linear-phase component ($-2\pi mF$ or $-m\Omega$) to the phase.

Modulation: Using the frequency-shift property and superposition gives the modulation property

$$\left(\frac{e^{j2\pi nF_0} + e^{-j2\pi nF_0}}{2}\right)x[n] = \cos(2\pi nF_0)x[n] \Longleftrightarrow \frac{X_p(F+F_0) + X_p(F-F_0)}{2} \tag{15.12}$$

REVIEW PANEL 15.8

Modulation by $\cos(2\pi nF_0)$: The DTFT Gets Halved, Centered at $\pm F_0$, and Added

$$\cos(2\pi nF_0)x[n] \Longleftrightarrow \frac{X_p(F+F_0) + X_p(F-F_0)}{2} \quad \text{or} \quad \frac{X_p(\Omega+\Omega_0) + X_p(\Omega-\Omega_0)}{2} \quad (\Omega_0 = 2\pi F_0)$$

Convolution: The *regular* convolution of discrete-time signals results in the product of their DTFTs. The product of discrete-time signals results in the *periodic* convolution of their DTFTs.

REVIEW PANEL 15.9

Multiplication in One Domain Corresponds to Convolution in the Other

$x[n] \star h[n] \Longleftrightarrow X_p(F)H_p(F)$ or $X_p(\Omega)H_p(\Omega)$ $\qquad$ $x[n]h[n] \Longleftrightarrow X_p(F) \circledast H_p(F)$ or $\frac{1}{2\pi}X_p(\Omega) \circledast H_p(\Omega)$

The times-n property: With $x[n] \Leftrightarrow X_p(F)$, differentiation of the defining DTFT relation gives

$$\frac{dX_p(F)}{dF} = \sum_{n=-\infty}^{\infty} (-j2n\pi)x[n]e^{-j2n\pi F} \tag{15.13}$$

The corresponding signal is $(-j2n\pi)x[n]$, and thus the DTFT of $y[n] = nx[n]$ is $Y_p(F) = \frac{j}{2\pi}\frac{dX_p(F)}{dF}$.

PGR2026-00033
Krisp Technologies EX1024 Page 502

REVIEW PANEL 15.10

The Times-n Property: Multiply $x[n]$ by n $\Leftrightarrow$ Differentiate the DTFT

$$nx[n] \Longleftrightarrow \frac{j}{2\pi}\frac{dX_p(F)}{dF} \qquad \text{or} \qquad j\frac{dX_p(\Omega)}{d\Omega}$$

Parseval's relation: The DTFT is an energy-conserving transform, and the signal energy may be found in either domain by Parseval's theorem:

$$\sum_{k=-\infty}^{\infty} x^2[k] = \int_1 |X_p(F)|^2\,dF = \frac{1}{2\pi}\int_{2\pi} |X_p(\Omega)|^2\,d\Omega \quad \text{(Parseval's relation)} \tag{15.14}$$

The notation $\int_1$ (or $\int_{2\pi}$) means integration over a one-period duration (typically, the principal period).

REVIEW PANEL 15.11

Parseval's Theorem: We can Find the Signal Energy from $x[n]$ or Its Magnitude Spectrum

$$\sum_{k=-\infty}^{\infty} x^2[k] = \int_1 |X_p(F)|^2\,dF = \frac{1}{2\pi}\int_{2\pi} |X_p(\Omega)|^2\,d\Omega$$

Central ordinate theorems: The DTFT obeys the *central ordinate* relations (found by substituting $F = 0$ (or $\Omega = 0$) in the DTFT or $n = 0$ in the IDTFT.

$$x[0] = \int_1 X_p(F)\,dF = \frac{1}{2\pi}\int_{2\pi} X_p(\Omega)\,d\Omega \qquad X_p(0) = \sum_{n=-\infty}^{\infty} x[n] \qquad \text{(central ordinates)} \tag{15.15}$$

With $F = 0.5$ (or $\Omega = \pi$), we also have the useful result

$$X_p(F)\Big|_{F=0.5} = X_p(\Omega)\Big|_{\Omega=\pi} = \sum_{n=-\infty}^{\infty} (-1)^n x[n] \tag{15.16}$$

REVIEW PANEL 15.12

Central Ordinate Theorems

$$x[0] = \int_1 X_p(F)dF = \frac{1}{2\pi}\int_{2\pi} X_p(\Omega)d\Omega \qquad X_p(0) = \sum_{n=-\infty}^{\infty} x[n] \qquad X_p(F)\Big|_{F=0.5} = X_p(\Omega)\Big|_{\Omega=\pi} = \sum_{n=-\infty}^{\infty} (-1)^n x[n]$$

EXAMPLE 15.2 (Some DTFT Pairs Using the Properties)

(a) The DTFT of $x[n] = n\alpha^n u[n],\ |\alpha < 1|$ may be found using the times-n property as

$$X_p(F) = \frac{j}{2\pi}\frac{d}{dF}\left[\frac{1}{1-\alpha e^{-j2\pi F}}\right] = \frac{\alpha e^{-j2\pi F}}{(1-\alpha e^{-j2\pi F})^2}$$

In the Ω-form,

$$X_p(\Omega) = j\frac{d}{d\Omega}\left[\frac{1}{1-\alpha e^{-j\Omega}}\right] = \frac{\alpha e^{-j\Omega}}{(1-\alpha e^{-j\Omega})^2}$$

PGR2026-00033
Krisp Technologies EX1024 Page 503

(b) The DTFT of the signal $x[n] = (n+1)\alpha^n u[n]$ may be found if we write $x[n] = n\alpha^n u[n] + \alpha^n u[n]$, and use superposition, to give

$$X_p(F) = \frac{\alpha e^{-j2\pi F}}{(1-\alpha e^{-j2\pi F})^2} + \frac{1}{1-\alpha e^{-j2\pi F}} = \frac{1}{(1-\alpha e^{-j2\pi F})^2}$$

In the Ω-form,

$$X_p(\Omega) = \frac{\alpha e^{-j\Omega}}{(1-\alpha e^{-j\Omega})^2} + \frac{1}{1-\alpha e^{-j\Omega}} = \frac{1}{(1-\alpha e^{-j\Omega})^2}$$

By the way, if we recognize that $x[n] = \alpha^n u[n] \star \alpha^n u[n]$, we can also use the convolution property to obtain the same result.

(c) To find DTFT of the N-sample exponential pulse $x[n] = \alpha^n$, $0 \le n < N$, express it as $x[n] = \alpha^n(u[n] - u[n-N]) = \alpha^n u[n] - \alpha^N \alpha^{n-N} u[n-N]$ and use the shifting property to get

$$X_p(F) = \frac{1}{1-\alpha e^{-j2\pi F}} - \alpha^N \frac{e^{-j2\pi FN}}{1-\alpha e^{-j2\pi F}} = \frac{1-(\alpha e^{-j2\pi F})^N}{1-\alpha e^{-j2\pi F}}$$

In the Ω-form,

$$X_p(\Omega) = \frac{1}{1-\alpha e^{-j\Omega}} - \alpha^N \frac{e^{-j\Omega N}}{1-\alpha e^{-j\Omega}} = \frac{1-(\alpha e^{-j\Omega})^N}{1-\alpha e^{-j\Omega}}$$

(d) The DTFT of the two-sided decaying exponential $x[n] = \alpha^{|n|}$, $|\alpha| < 1$, may be found by rewriting this signal as $x[n] = \alpha^n u[n] + \alpha^{-n} u[-n] - \delta[n]$ and using the folding property to give

$$X_p(F) = \frac{1}{1-\alpha e^{-j2\pi F}} + \frac{1}{1-\alpha e^{j2\pi F}} - 1$$

Simplification leads to the result

$$X_p(F) = \frac{1-\alpha^2}{1-2\alpha\cos(2\pi F)+\alpha^2} \qquad \text{or} \qquad X_p(\Omega) = \frac{1-\alpha^2}{1-2\alpha\cos\Omega+\alpha^2}$$

(e) (Properties of the DTFT)

Find the DTFT of $x[n] = 4(0.5)^{n+3}u[n]$ and $y[n] = n(0.4)^{2n}u[n]$.

1. For $x[n]$, we rewrite it as $x[n] = 4(0.5)^3(0.5)^n u[n]$ to get

$$X_p(F) = \frac{0.5}{1-0.5e^{-j2\pi F}} \qquad \text{or} \qquad X_p(\Omega) = \frac{0.5}{1-0.5e^{-j\Omega}}$$

2. For $y[n]$, we rewrite it as $y[n] = n(0.16)^n u[n]$ to get

$$Y_p(F) = \frac{0.16e^{-j2\pi F}}{(1-0.16e^{-j2\pi F})^2} \qquad \text{or} \qquad Y_p(\Omega) = \frac{0.16e^{-j\Omega}}{(1-0.16e^{-j\Omega})^2}$$

(f) (Properties of the DTFT)

Let $x[n] \Longleftrightarrow \dfrac{4}{2-e^{-j2\pi F}} = X_p(F)$.

Find the DTFT of $y[n] = nx[n]$, $c[n] = x[-n]$, $g[n] = x[n] \star x[n]$, and $h[n] = (-1)^n x[n]$.

PGR2026-00033
Krisp Technologies EX1024 Page 504

1. By the times-n property,

$$Y_p(F) = \frac{j}{2\pi}\frac{d}{dF}X_p(F) = \frac{4(-j/2\pi)(-j2\pi e^{-j2\pi F})}{(2-e^{-j2\pi F})^2} = \frac{-4e^{-j2\pi F}}{(2-e^{-j2\pi F})^2}$$

In the Ω-form,

$$Y_p(\Omega) = j\frac{d}{d\Omega}X_p(\Omega) = \frac{4(-j/2\pi)(-j2\pi e^{-j\Omega})}{(2-e^{-j\Omega})^2} = \frac{-4e^{-j\Omega}}{(2-e^{-j\Omega})^2}$$

2. By the folding property,

$$C_p(F) = X_p(-F) = \frac{4}{2-e^{-j2\pi F}} \qquad \text{or} \qquad C_p(\Omega) = X_p(-\Omega) = \frac{4}{2-e^{-j\Omega}}$$

3. By the convolution property,

$$G_p(F) = X_p^2(F) = \frac{16}{(2-e^{-j2\pi F})^2} \qquad \text{or} \qquad G_p(\Omega) = X_p^2(\Omega) = \frac{16}{(2-e^{-j\Omega})^2}$$

4. By the modulation property,

$$H_p(F) = X_p(F-0.5) = \frac{4}{2-e^{-j2\pi(F-0.5)}} = \frac{4}{2+e^{-j2\pi F}}$$

In the Ω-form,

$$H_p(\Omega) = X_p(\Omega-\pi) = \frac{4}{2-e^{-j(\Omega-\pi)}} = \frac{4}{2+e^{-j\Omega}}$$

(g) (Properties of the DTFT)

Let $X_p(F) \Longleftrightarrow (0.5)^n u[n] = x[n]$. Find the time signals corresponding to

$$Y_p(F) = X_p(F) \circledast X_p(F) \qquad H_p(F) = X_p(F+0.4) + X_p(F-0.4) \qquad G_p(F) = X_p^2(F)$$

1. By the convolution property, $y[n] = x^2[n] = (0.25)^n u[n]$.
2. By the modulation property, $h[n] = 2\cos(2n\pi F_0)x[n] = 2(0.5)^n \cos(0.8n\pi)u[n]$ (where $F_0 = 0.4$).
3. By the convolution property, $g[n] = x[n] \star x[n] = (0.5)^n u[n] \star (0.5)u[n] = (n+1)(0.5)^n u[n]$.

15.3.1 The DTFT of Discrete-Time Periodic Signals

Recall (from Chapter 9) that if $X_1(f)$ describes the Fourier transform of one period $x_1(t)$ of a periodic signal $x_p(t)$, the Fourier transform $X(f)$ of $x_p(t)$ may be found from the sampled version of $X_1(f)$ as

$$X(f) = \frac{1}{T}\sum_{k=-\infty}^{\infty} X_1(kf_0)\delta(f-kf_0) \tag{15.17}$$

where T is the period and $f_0 = 1/T$ is the fundamental frequency. This result is a direct consequence of the fact that periodic extension in one domain leads to sampling in the other. By analogy, if $X_1(F)$ describes

PGR2026-00033
Krisp Technologies EX1024 Page 505

the DTFT of one period $x_1[n]$ (over $0 \le n \le N-1$) of a periodic signal $x_p[n]$, the DTFT $X_p(F)$ of $x_p[n]$ may be found from the sampled version of $X_1(F)$ as

$$X_p(F) = \frac{1}{N}\sum_{k=0}^{N-1} X_1(kF_0)\delta(f - kF_0) \qquad \text{(over one period } 0 \le F < 1) \tag{15.18}$$

where N is the period and $F_0 = 1/N$ is the fundamental frequency. The summation index is shown from $n = 0$ to $n = N-1$ because the DTFT is periodic and because the result is usually expressed for one period over $0 \le F < 1$. The DTFT of a periodic signal with period N is a periodic impulse train with N impulses per period. The impulse strengths correspond to N samples of $X_1(F)$, the DTFT of one period of $x_p[n]$. Note that $X_p(F)$ exhibits conjugate symmetry about $k = 0$ (corresponding to $F = 0$ or $\Omega = 0$) and $k = \frac{N}{2}$ (corresponding to $F = 0.5$ or $\Omega = \pi$). By analogy with analog signals, the quantity

$$X_{\text{DFS}}[k] = \frac{1}{N}X_1(kF_0) = \frac{1}{N}\sum_{k=0}^{N-1} x_1[n]e^{-j2\pi nk/N} \tag{15.19}$$

may be regarded as the Fourier series coefficients of the periodic signal $x_p[n]$. The sequence $X_{\text{DFS}}[k]$ describes the **discrete Fourier series** (DFS) coefficients (which we study in the next chapter). In fact, if the samples correspond to a *band-limited* periodic signal $x_p(t)$ sampled above the Nyquist rate for an integer number of periods, the coefficients $X_{\text{DFS}}[k]$ are an exact match to the Fourier series coefficients $X[k]$ of $x_p(t)$.

REVIEW PANEL 15.13

The DTFT of $x_p[n]$ (Period N) Is a Periodic Impulse Train (N Impulses per Period)

If $x_p[n]$ is periodic with period N and its one-period DTFT is $x_1[n] \Leftrightarrow X_1(F)$, then

$$x_p[n] \iff X_p(F) = \frac{1}{N}\sum_{k=0}^{N-1} X_1(kF_0)\delta(f - kF_0) \qquad (N \text{ impulses per period } 0 \le F < 1)$$

EXAMPLE 15.3 (DTFT of Periodic Signals)

(a) Let one period of $x_p[n]$ be given by $x_1[n] = \{3,\ 2,\ 1,\ 2\}$, with $N = 4$.

Its DTFT is $X_1(F) = 3 + 2e^{-j2\pi F} + e^{-j4\pi F} + 2e^{-j6\pi F}$.

The four samples of $X_1(kF_0)$ over $0 \le k \le 3$ are

$$X_1(kF_0) = 3 + 2e^{-j2\pi k/4} + e^{-j4\pi k/4} + 2e^{-j6\pi k/4} = \{8,\ 2,\ 0,\ 2\}$$

The DTFT of the periodic signal $x_p[n]$ for one period $0 \le F < 1$ is thus

$$X_p(F) = \frac{1}{4}\sum_{k=0}^{3} X_1(kF_0)\delta(f - \tfrac{k}{4}) = 2\delta(F) + 0.5\delta(F - \tfrac{1}{4}) + 0.5\delta(F - \tfrac{3}{4}) \quad \text{(over one period } 0 \le F < 1)$$

The signal $x_p[n]$ and its DTFT $X_p(F)$ are shown in Figure E15.3A. Note that the DTFT is conjugate symmetric about $F = 0$ (or $k = 0$) and $F = 0.5$ (or $k = 0.5N = 2$).

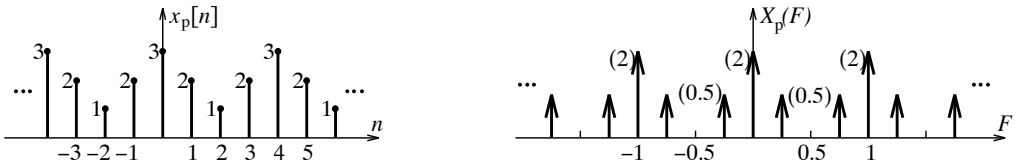


Figure E15.3A Periodic signal for Example 15.3(a) and its DTFT

PGR2026-00033
Krisp Technologies EX1024 Page 506

(b) Let $x_1[n] = \{1,\ 0,\ 2,\ 0,\ 3\}$ describe one period of a periodic signal $x_p[n]$.
Its period is $N = 5$. The discrete Fourier series (DFS) coefficients of $x_p[n]$ are

$$X_{\text{DFS}}[k] = \tfrac{1}{N} X_1(kF_0) = 0.2X_1(kF_0)$$

where

$$X_1(kF_0) = 1 + 2e^{-j4\pi F} + 3e^{-j8\pi F}\Big|_{F=k/5}, \qquad k = 0, 1, \ldots, 5$$

We find that

$$X_{\text{DFS}}[k] = \{1.2,\ 0.0618 + j0.3355,\ -0.1618 + j0.7331,\ -0.1618 - j0.7331,\ 0.0618 - j0.3355\}$$

The DTFT $X_p(F)$ of the periodic signal $x_p[n]$, for one period $0 \le F < 1$, is then

$$X_p(F) = \sum_{k=0}^{4} X_p[k]\delta(f - \tfrac{k}{5}) \qquad \text{(over one period } 0 \le F < 1)$$

Both $X_{\text{DFS}}[k]$ and $X_p(F)$ are conjugate symmetric about $k = 0$ and about $k = 0.5N = 2.5$.

15.3.2 The Inverse DTFT

We can find the inverse DTFT in one of several ways. For finite sequences whose DTFT is a polynomial in $e^{j2\pi F}$ (or $e^{j\Omega}$), the signal $x[n]$ is just the sequence of the polynomial coefficients. For other signals, we use the defining relation whose form is similar to the expression for finding the Fourier series of a periodic signal. If the DTFT is a ratio of polynomials in $e^{j2\pi F}$ (or $e^{j\Omega}$), we can use partial fraction expansion and table look-up, much as we do with Laplace transforms.

EXAMPLE 15.4 (The Inverse DTFT)

(a) Let $X(F) = 1 + 3e^{-j4\pi F} - 2e^{-j6\pi F}$.

Its IDFT is simply $x[n] = \delta[n] + 3\delta[n-2] - 2\delta[n-3]$ or $x[n] = \{\overset{\Downarrow}{1}, 0, 3, -2\}$.

(b) Let $X_p(\Omega) = \dfrac{2e^{-j\Omega}}{1 - 0.25e^{-j2\Omega}}$. We factor the denominator and use partial fractions to get

$$X_p(\Omega) = \frac{2e^{-j\Omega}}{(1 - 0.5e^{-j\Omega})(1 + 0.5e^{-j\Omega})} = \frac{2}{1 - 0.5e^{-j\Omega}} - \frac{2}{1 + 0.5e^{-j\Omega}}$$

We then find $x[n] = 2(0.5)^n u[n] - 2(-0.5)^n u[n]$.

(c) An **ideal differentiator** is described by $H_p(F) = j2\pi F,\ |F| < 0.5$. Its magnitude and phase spectrum are shown in Figure E15.4C.

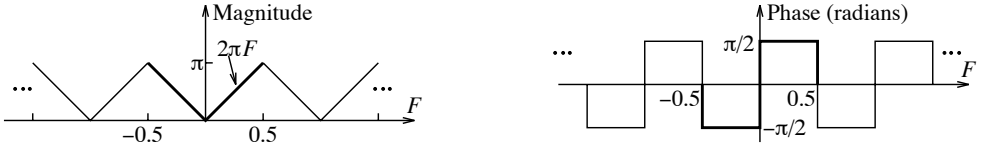


Figure E15.4C DTFT of the ideal differentiator for Example 15.4(c)

PGR2026-00033
Krisp Technologies EX1024 Page 507

To find its inverse $h[n]$, we note that $h[0] = 0$ since $H_p(F)$ is odd. For $n \neq 0$, we also use the odd symmetry of $H_p(F)$ in the IDTFT to obtain

$$h[n] = \int_{-1/2}^{1/2} j2\pi F[\cos(2\pi nF) + j\sin(2\pi nF)]\,dF = -4\pi \int_0^{1/2} F\sin(2\pi nF)\,dF$$

Using tables and simplifying the result, we get

$$h[n] = \frac{-4\pi[\sin(2\pi nF) - 2\pi nF\cos(2\pi nF)]}{(2\pi n)^2}\bigg|_0^{1/2} = \frac{\cos(n\pi)}{n}$$

Since $H_p(F)$ is odd and imaginary, $h[n]$ is odd symmetric, as expected.

(d) A **Hilbert transformer** shifts the phase of a signal by $-90°$. Its magnitude and phase spectrum are shown in Figure E15.4D.

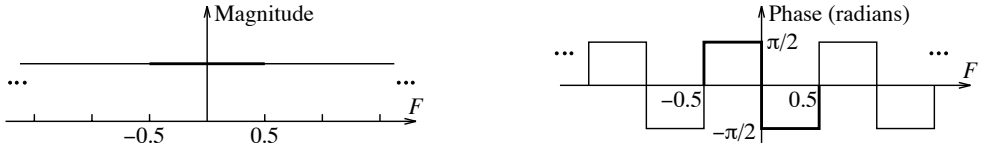


Figure E15.4D DTFT of the Hilbert transformer for Example 15.4(d)

Its DTFT given by $H_p(F) = -j\,\mathrm{sgn}(F)$, $|F| < 0.5$. This is imaginary and odd. To find its inverse $h[n]$, we note that $h[0] = 0$ and

$$h[n] = \int_{-1/2}^{1/2} -j\,\mathrm{sgn}(F)[\cos(2\pi nF) + j\sin(2\pi nF)]\,dF = 2\int_0^{1/2} \sin(2\pi nF)\,dF = \frac{1-\cos(n\pi)}{n\pi}$$

15.4 The Transfer Function

The response $y[n]$ of a relaxed discrete-time LTI system with impulse response $h[n]$ to the input $x[n]$ is given by the convolution

$$y[n] = x[n] \star h[n] \tag{15.20}$$

Since convolution transforms to multiplication, transformation results in

$$Y_p(F) = X_p(F)H_p(F) \qquad \text{or} \qquad Y_p(\Omega) = X_p(\Omega)H_p(\Omega) \tag{15.21}$$

The **transfer function** then equals

$$H_p(F) = \frac{Y_p(F)}{X_p(F)} \qquad \text{or} \qquad H_p(\Omega) = \frac{Y_p(\Omega)}{X_p(\Omega)} \tag{15.22}$$

We emphasize that the transfer function is defined only for a *relaxed* LTI system—either as the ratio $Y_p(F)/X_p(F)$ (or $Y_p(\Omega)/X_p(\Omega)$) of the DTFT of the output $y[n]$ and input $x[n]$, or as the DTFT of the impulse response $h[n]$. The equivalence between the time-domain and frequency-domain operations is illustrated in Figure 15.3.

PGR2026-00033
Krisp Technologies EX1024 Page 508

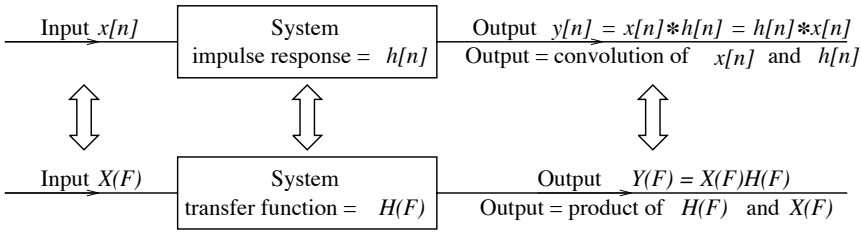


Figure 15.3 The equivalence between the time domain and frequency domain

A relaxed LTI system may also be described by the difference equation

$$y[n] + A_1 y[n-1] + \cdots + A_N y[n-N] = B_0 x[n] + B_1 x[n-1] + \cdots + B_M x[n-M] \tag{15.23}$$

The DTFT results in the transfer function, or **frequency response**, given by

$$H_p(F) = \frac{Y_p(F)}{X_p(F)} = \frac{B_0 + B_1 e^{-j2\pi F} + \cdots + B_M e^{-j2\pi MF}}{1 + A_1 e^{-j2\pi F} + \cdots + A_N e^{-j2\pi NF}} \tag{15.24}$$

In the Ω-form,

$$H_p(\Omega) = \frac{Y_p(\Omega)}{X_p(\Omega)} = \frac{B_0 + B_1 e^{-j\Omega} + \cdots + B_M e^{-jM\Omega}}{1 + A_1 e^{-j\Omega} + \cdots + A_N e^{-jN\Omega}} \tag{15.25}$$

The transfer function is thus a ratio of polynomials in $e^{j2\pi F}$ (or $e^{j\Omega}$).

REVIEW PANEL 15.14

The Transfer Function Is a Frequency-Domain Description of a Relaxed LTI System

The transfer function $H_p(F)$ equals the transform of the system impulse response $h[n]$.
It is also defined as $H_p(F) = Y_p(F)/X_p(F)$, the ratio of the *transformed output* and *transformed input*.

15.4.1 System Representation in Various Forms

A relaxed LTI system may be described in various forms: by its difference equation, by its impulse response, or by its transfer function. Given one form, it is fairly easy to obtain the others using time-domain methods and/or frequency-domain transformations. For example, given the impulse response $h[n]$, we can find the transfer function from its DTFT $H_p(F)$. If we express the transfer function as $H_p(F) = Y_p(F)/X_p(F)$, cross-multiplication and inverse transformation leads to the system difference equation.

REVIEW PANEL 15.15

Three Representations of a Relaxed LTI System

Transfer function: $H_p(F) = \frac{Y_p(F)}{X_p(F)}$ **Impulse response:** $h[n] = \text{IDTFT}\{H_p(F)\}$

Difference equation: from $H_p(F) = \frac{Y_p(F)}{X_p(F)}$ by cross-multiplication and IDTFT

PGR2026-00033
Krisp Technologies EX1024 Page 509

EXAMPLE 15.5 (System Representation in Various Forms)

(a) Let $y[n] - 0.6y[n-1] = x[n]$. Its DTFT gives

$$Y(F) - 0.6e^{-j2\pi F}Y(F) = X(F) \qquad \text{or} \qquad Y(\Omega) - 0.6e^{-j\Omega}Y(\Omega) = X(\Omega)$$

We thus get the transfer function as

$$H(F) = \frac{Y(F)}{X(F)} = \frac{1}{1 - 0.6e^{-j2\pi F}} \qquad \text{or} \qquad H(\Omega) = \frac{Y(\Omega)}{X(\Omega)} = \frac{1}{1 - 0.6e^{-j\Omega}}$$

The system impulse response is thus $h[n] = (0.6)^n u[n]$.

(b) Let $h[n] = \delta[n] - (0.5)^n u[n]$. Its DTFT gives

$$H(F) = 1 - \frac{1}{1 - 0.5e^{-j2\pi F}} = \frac{-0.5e^{-j2\pi F}}{1 - 0.5e^{-j2\pi F}}$$

In the Ω-form,

$$H(\Omega) = 1 - \frac{1}{1 - 0.5e^{-j\Omega}} = \frac{-0.5e^{-j\Omega}}{1 - 0.5e^{-j\Omega}}$$

From this we find

$$H(F) = \frac{Y(F)}{X(F)} \quad \text{or} \quad Y(F)(1 - 0.5e^{-j2\pi F}) = X(F)(-0.5e^{-j2\pi F})$$

In the Ω-form,

$$H(\Omega) = \frac{Y(\Omega)}{X(\Omega)} \quad \text{or} \quad Y(\Omega)(1 - 0.5e^{-j\Omega}) = X(\Omega)(-0.5e^{-j\Omega})$$

Inverse transformation gives $y[n] - 0.5y[n-1] = -0.5x[n-1]$.

(c) Let $H(F) = 1 + 2e^{-j2\pi F} + 3e^{-j4\pi F}$.

Its impulse response is $h[n] = \delta[n] + 2\delta[n-1] + 3\delta[n-2] = \{1,\ 2,\ 3\}$.

Since $H(F) = \frac{Y(F)}{X(F)} = 1 + 2e^{-j2\pi F} + 3e^{-j4\pi F}$, we find

$$Y(F) = (1 + 2e^{-j2\pi F} + 3e^{-j4\pi F})X(F) \qquad y[n] = x[n] + 2x[n-1] + 3x[n-2]$$

15.4.2 Frequency Response

Plots of the magnitude and phase of the transfer function against frequency are referred to collectively as the **frequency response**. The frequency response is a very useful way of describing (and identifying) digital filters. For example, the frequency response of a lowpass filter displays a magnitude that decreases with frequency. The DTFT evaluated at $F = 0$ (or $\Omega = 0$), the smallest frequency, and at $F = 0.5$ (or $\Omega = \pi$), the highest useful frequency of the principal period, provides useful clues for identifying traditional filters (such as lowpass and highpass). For example, for a lowpass filter $H(F)$, the dc gain (at $F = 0$) should exceed the gain $|H(F)|$ at $F = 0.5$.

PGR2026-00033
Krisp Technologies EX1024 Page 510

REVIEW PANEL 15.16

Identify Traditional Filters by Finding the Gain at DC and $F = 0$ (or $\Omega = \pi$)

From transfer function: Evaluate $H(F)$ at $F = 0$ and $F = 0.5$, or evaluate $H(\Omega)$ at $\Omega = 0$ and $\Omega = \pi$.
From impulse response: Evaluate $\sum h[n] = H(0)$ and $\sum(-1)^n h[n] = H(F)|_{F=0.5} = H(\Omega)|_{\Omega=\pi}$.

15.4.3 Linear-Phase Filters

The DTFT of a filter whose impulse response is symmetric about the origin is purely real or purely imaginary. If the symmetric sequence is shifted (to make it causal, for example), its phase is augmented by a linear-phase term. A filter whose impulse response is symmetric about its midpoint is termed a (generalized) *linear-phase filter*. Linear phase is important in filter design because it results in a pure time delay with no amplitude distortion. The transfer function of a linear-phase filter may be written as $H(F) = A(F)e^{j2\pi\alpha F}$ (for even symmetry) or $H(F) = jA(F)e^{j2\pi\alpha F} = A(F)e^{j(2\pi\alpha F + \frac{\pi}{2})}$ (for odd symmetry), where the real quantity $A(F)$ is the amplitude spectrum and α is the (integer or half-integer) index corresponding to the midpoint of its impulse response $h[n]$. The easiest way to obtain this form is to first set up an expression for $H(F)$, then extract the factor $e^{j2\pi\alpha F}$ from $H(F)$, and finally simplify using Euler's relation.

REVIEW PANEL 15.17

A Linear-Phase Sequence Is Symmetric About Its Midpoint

If $h[n]$ is a linear-phase sequence with its midpoint at the (integer or half-integer) index α, then $H(F) = A(F)e^{j2\pi\alpha F}$ (for even symmetric $h[n]$) and $H(F) = jA(F)e^{j2\pi\alpha F}$ (for odd symmetric $h[n]$).
To find $A(F)$: Set up $H(F)$, extract $e^{j2\pi\alpha F}$ from $H(F)$, and simplify using Euler's relation.

EXAMPLE 15.6 (Frequency Response and Filter Characteristics)

(a) Let $h[n] = \{\overset{\Downarrow}{-1},\ -2,\ 2,\ 1\}$. Is this a linear-phase sequence? What type of filter is this?

The sequence $h[n]$ is linear-phase. It shows odd symmetry about the half-integer index $n = 1.5$.

We have $H_p(F) = -1 - 2e^{-j2\pi F} + 2e^{-j4\pi F} + e^{-j6\pi F}$.

Extract the factor $(-e^{-j3\pi F})$ to get $H_p(F) = -e^{-j3\pi F}[e^{j3\pi F} + 2e^{j\pi F} - 2e^{-j\pi F} - e^{-j3\pi F}]$.

Using Euler's relation: $H_p(F) = -je^{-j3\pi F}[2\sin(3\pi F) + 4\sin(\pi F)]$ $\quad H_p(0) = 0$ $\quad H_p(0.5) = 2$

Amplitude $A(F) = 2\sin(3\pi F) + 4\sin(\pi F)$ $\quad$ Phase $= -je^{-j3\pi F} = -3\pi F - \frac{\pi}{2}$.

Since the magnitude at high frequencies increases, this appears to be a highpass filter.

Comment: We can also use the central ordinate relations to compute $H_p(0)$ and $H_p(0.5)$ directly from $h[n]$. Thus, $H_p(0) = \sum_{n=0}^{3} h[n] = 0$ and $H_p(0.5) = \sum_{n=0}^{3} (-1)^n h[n] = -1 + 2 + 2 - 1 = 2$.

(b) Let $h[n] = (-0.8)^n u[n]$. Identify the filter type and establish whether the impulse response is a linear-phase sequence.

The sequence $h[n]$ is not linear-phase because it shows no symmetry.

We have $H_p(F) = \dfrac{1}{1 + 0.8e^{-j2\pi F}}$.

PGR2026-00033
Krisp Technologies EX1024 Page 511

We find that $H_p(0) = \dfrac{1}{1+0.8} = 0.556 and H_p(0.5) = \dfrac{1}{1+0.8e^{-j\pi}} = \dfrac{1}{1-0.8} = 5.$

Since the magnitude at high frequencies increases, this appears to be a highpass filter.

(c) Consider a system described by $y[n] = \alpha y[n-1] + x[n],\ 0 < \alpha < 1$. This is an example of a *reverb filter* whose response equals the input plus a delayed version of the output. Its frequency response may be found by taking the DTFT of both sides to give $Y_p(F) = \alpha Y_p(F)e^{-j2\pi F} + X_p(F)$. Rearranging this equation, we obtain

$$H_p(F) = \frac{Y_p(F)}{X_p(F)} = \frac{1}{1-\alpha e^{-j2\pi F}}$$

Using Euler's relation, we rewrite this as

$$H_p(F) = \frac{1}{1-\alpha e^{-j2\pi F}} = \frac{1}{1-\alpha\cos(2\pi F) - j\alpha\sin(2\pi F)}$$

Its magnitude and phase are given by

$$|H_p(F)| = \frac{1}{[1-2\alpha\cos(2\pi F)+\alpha^2]^{1/2}} \qquad \phi(F) = -\tan^{-1}\left[\frac{\alpha\sin(2\pi F)}{1-\alpha\cos(2\pi F)}\right]$$

A typical magnitude and phase plot for this system (for $0 < \alpha < 1$) is shown in Figure E15.6C. The impulse response of this system equals $h[n] = \alpha^n u[n]$.

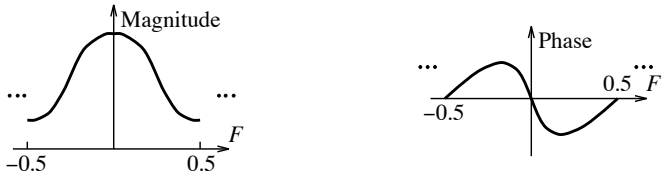


Figure E15.6C Frequency response of the system for Example 15.6(c)

(d) Consider the 3-point moving average filter, $y[n] = \frac{1}{3}\{x[n-1] + x[n] + x[n+1]\}$. The filter replaces each input value $x[n]$ by an average of itself and its two neighbors. Its impulse response $h_1[n]$ is simply $h_1[n] = \{\frac{1}{3}, \underset{\Downarrow}{\frac{1}{3}}, \frac{1}{3}\}$. The frequency response is given by

$$H_1(F) = \sum_{n=-1}^{1} h[n]e^{-j2\pi Fn} = \tfrac{1}{3}\left[e^{j2\pi F} + 1 + e^{-j2\pi F}\right] = \tfrac{1}{3}[1 + 2\cos(2\pi F)]$$

The magnitude $|H_1(F)|$ decreases until $F = \frac{2}{3}$ (when $H_1(F) = 0$) and then increases to $\frac{1}{3}$ at $F = \frac{1}{2}$, as shown in Figure E15.6D. This filter thus does a poor job of smoothing past $F = \frac{2}{3}$.

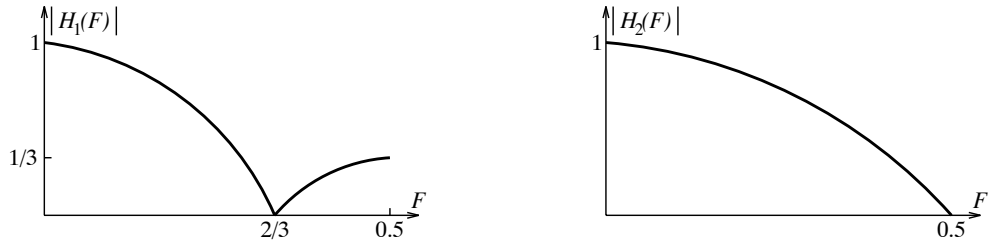


Figure E15.6D Frequency response of the filters for Example 15.6(d and e)

PGR2026-00033
Krisp Technologies EX1024 Page 512

(e) Consider the tapered 3-point averager $h_2[n] = \{\frac{1}{4}, \underset{\Downarrow}{\frac{1}{2}}, \frac{1}{4}\}$. Its frequency response is given by

$$H_2(F) = \sum_{n=-1}^{1} h[n]e^{-j2\pi Fn} = \tfrac{1}{4}e^{j2\pi F} + \tfrac{1}{2} + \tfrac{1}{4}e^{-j2\pi F} = \tfrac{1}{2} + \tfrac{1}{2}\cos(2\pi F)$$

Figure E15.6D shows that $|H_2(F)|$ decreases monotonically to zero at $F = 0.5$ and shows a much better smoothing performance. This is a lowpass filter and actually describes the von Hann (or Hanning) smoothing window. Other tapering schemes lead to other window types.

(f) Consider the difference operator described by $y[n] = x[n] - x[n-1]$. Its impulse response may be written as $h[n] = \delta[n] - \delta[n-1] = \{1, -1\}$. This is actually odd symmetric about its midpoint ($n = 0.5$). Its DTFT (in the Ω-form) is given by

$$H(\Omega) = 1 - e^{j\Omega} = e^{-j\Omega/2}(e^{j\Omega/2} - e^{-j\Omega/2}) = 2j\sin(0.5\Omega)e^{-j\Omega/2}$$

Its phase is $\phi(\Omega) = \frac{\pi}{2} - \frac{\Omega}{2}$ and shows a linear variation with Ω. Its amplitude $A(\Omega) = 2\sin(\Omega/2)$ increases from zero at $\Omega = 0$ to two at $\Omega = \pi$. In other words, the difference operator enhances high frequencies and acts as a highpass filter.

15.5 System Analysis Using the DTFT

In concept, the DTFT may be used to find the zero-state response (ZSR) of relaxed LTI systems to arbitrary inputs. All it requires is the system transfer function $H_p(F)$ and the DTFT $X_p(F)$ of the input $x[n]$. We first find the response as $Y_p(F) = H_p(F)X_p(F)$ in the frequency domain, and obtain the time-domain response $y[n]$ by using the inverse DTFT. We emphasize, once again, that the DTFT cannot handle the effect of initial conditions.

EXAMPLE 15.7 (The DTFT in System Analysis)

(a) Consider a system described by $y[n] = \alpha y[n-1] + x[n]$. To find the response of this system to the input $\alpha^n u[n]$, we first set up the transfer function as $H_p(\Omega) = 1/(1 - e^{-j\Omega})$. Next, we find the DTFT of $x[n]$ as $X_p(\Omega)$ and multiply the two to obtain

$$Y_p(\Omega) = H_p(\Omega)X_p(\Omega) = \frac{1}{(1 - \alpha e^{-j\Omega})^2}$$

Its inverse transform gives the response as $y[n] = (n+1)\alpha^n u[n]$. We could, of course, also use convolution to obtain $y[n] = h[n] \star x[n]$ directly in the time domain.

(b) Consider the system described by $y[n] = 0.5y[n-1] + x[n]$. Its response to the step $x[n] = 4u[n]$ is found using $Y_p(F) = H_p(F)X_p(F)$:

$$Y_p(F) = H_p(F)X_p(F) = \frac{1}{1 - 0.5e^{-j2\pi F}}\left[\frac{4}{1 - e^{-j2\pi F}} + 2\delta(F)\right]$$

PGR2026-00033
Krisp Technologies EX1024 Page 513

We separate terms and use the product property of impulses to get

$$Y_p(F) = \frac{4}{(1-0.5e^{-j2\pi F})(1-e^{-j2\pi F})} + 4\delta(F)$$

Splitting the first term into partial fractions, we obtain

$$Y_p(F) = \frac{-4}{1-0.5e^{-j2\pi F}} + \left[\frac{8}{1-e^{-j2\pi F}} + 4\delta(F)\right]$$

The response $y[n]$ then equals $y[n] = -4(0.5)^n u[n] + 8u[n]$. The first term represents the *transient* response, and the second term describes the *steady-state* response, which can be found much more easily, as we now show.

15.5.1 The Steady-State Response to Discrete-Time Harmonics

The DTFT is much better suited to finding the *steady-state* response to discrete-time harmonics. Since everlasting harmonics are eigensignals of discrete-time linear systems, the response is simply a harmonic at the input frequency whose magnitude and phase is changed by the system function $H_p(F)$. We evaluate $H_p(F)$ at the input frequency, multiply its magnitude by the input magnitude to obtain the output magnitude, and add its phase to the input phase to obtain the output phase. The steady-state response is useful primarily for stable systems for which the natural response does indeed decay with time.

The response of an LTI system to a sinusoid (or harmonic) is called the **steady-state response** and is a sinusoid (or harmonic) at the input frequency. If the input is $x[n] = A\cos(2\pi F_0 n + \theta)$, the steady-state response is $y_{ss}[n] = AH_0\cos[2\pi Fn + \theta + \phi_0)]$, where H_0 and ϕ_0 are the magnitude and phase of the transfer function $H_p(F)$ evaluated at the input frequency $F = F_0$.

REVIEW PANEL 15.18

Finding the Steady-State Response of an LTI System to a Sinusoidal Input

Input: $x[n] = A\cos(2\pi nF_0 + \theta)$ **Transfer function:** Evaluate $H_p(F)$ at $F = F_0$ as $H_0\angle\phi_0$.
Steady-state output: $y_{\text{ss}}[n] = AH_0\cos(2\pi nF_0 + \theta + \phi_0)$

EXAMPLE 15.8 (The DTFT and Steady-State Response)

(a) Consider a system described by $y[n] = 0.5y[n-1] + x[n]$. We find its steady-state response to the sinusoidal input $x[n] = 10\cos(0.5n\pi + 60^\circ)$. The transfer function $H_p(F)$ is given by

$$H_p(F) = \frac{1}{1-0.5e^{-j2\pi F}}$$

We evaluate $H_p(F)$ at the input frequency, $F = 0.25$:

$$H_p(F)\Big|_{F=0.25} = \frac{1}{1-0.5e^{-j\pi/2}} = \frac{1}{1+0.5j} = 0.4\sqrt{5}\angle -26.6^\circ = 0.8944\angle -26.6^\circ$$

The steady-state response then equals

$$y_{\text{ss}}[n] = 10(0.4\sqrt{5})\cos(0.5n\pi + 60^\circ - 26.6^\circ) = 8.9443\cos(0.5n\pi + 33.4^\circ)$$

Note that if the input was $x[n] = 10\cos(0.5n\pi + 60^\circ)u[n]$ (switched on at $n = 0$), the steady-state component would still be identical to what we calculated but the *total* response would differ.

PGR2026-00033
Krisp Technologies EX1024 Page 514

(b) Consider a system described by $h[n] = (0.8)^n u[n]$. We find its steady-state response to the step input $x[n] = 4u[n]$. The transfer function $H_p(F)$ is given by

$$H_p(F) = \frac{1}{1 - 0.8e^{-j2\pi F}}$$

We evaluate $H_p(F)$ at the input frequency $F = 0$ (corresponding to dc):

$$H_p(F)\Big|_{F=0} = \frac{1}{1 - 0.8} = 5$$

The steady-state response is then $y_{ss}[n] = (5)(4) = 20$.

(c) Design a 3-point FIR filter with impulse response $h[n] = \{\alpha, \underset{\Downarrow}{\beta}, \alpha\}$ that completely blocks the frequency $F = \frac{1}{3}$ and passes the frequency $F = 0.125$ with unit gain. What is the dc gain of this filter?

The filter transfer function is $H_p(F) = \alpha e^{j2\pi F} + \beta + \alpha e^{-j2\pi F} = \beta + 2\alpha\cos(2\pi F)$.

From the information given, we have

$$H(\tfrac{1}{3}) = 0 = \beta + 2\alpha\cos(\tfrac{2\pi}{3}) = \beta - \alpha \qquad H(0.125) = 1 = \beta + 2\alpha\cos(\tfrac{2\pi}{8}) = \beta - \alpha\sqrt{2}$$

This gives $\alpha = \beta = 0.4142$ and $h[n] = \{0.4142,\ \underset{\Downarrow}{0.4142},\ 0.4142\}$.

The dc gain of this filter is $H(0) = \sum h[n] = 3(0.4142) = 1.2426$.

15.6 Connections

The DTFT and Fourier series are duals. This allows us to carry over the operational properties of the Fourier series virtually unchanged to the DTFT. A key difference is that while the discrete sequence $x[n]$ is usually real, the discrete spectrum $X[k]$ is usually complex. Table 15.3 lists some of the analogies between the properties of the DTFT and Fourier series.

Table 15.3 Relating the DTFT and Fourier Series

Properties Arising from Duality with the Fourier Series		
Property	**DTFT Result**	**FS Result with $T = 1$**
Time delay	$x[n-\alpha] \Leftrightarrow X_p(F)e^{-j2\pi F\alpha}$	$x_p(t-\alpha) \Leftrightarrow X[k]e^{-jk\omega_0\alpha}$
Reversal	$x[-n] \Leftrightarrow X_p(-F) = X_p^*(F)$	$x_p(-t) \Leftrightarrow X[-k] = X^*[k]$
Modulation	$x[n]e^{j2\pi n\alpha} \Leftrightarrow X_p(F-\alpha)$	$x_p(t)e^{jm\omega_0 t} \Leftrightarrow X[k-m]$
Product	$x[n]y[n] \Leftrightarrow X_p(F) \circledast Y_p(F)$	$x_p(t)y_p(t) \Leftrightarrow X[k] \star Y[k]$
Convolution	$x[n] \star y[n] \Leftrightarrow X_p(F)Y_p(F)$	$x_p(t) \circledast y_p(t) \Leftrightarrow X[k]Y[k]$

The DTFT can also be related to the Fourier transform by recognizing that sampling in one domain results in a periodic extension in the other. If we sample an analog signal $x(t)$ to get $x[n]$, we obtain $X_p(F)$ as a *periodic extension* of $X(f)$ with unit period. For signals whose Fourier transforms $X(f)$ are not band-limited, the results are often unwieldy.

PGR2026-00033
Krisp Technologies EX1024 Page 515

EXAMPLE 15.9 (Connecting the DTFT, Fourier Series, and Fourier Transform)

(a) To find the DTFT of $x[n] = \text{rect}(n/2N)$, with $M = 2N + 1$, we start with the Fourier series result

$$x(t) = \frac{M\,\text{sinc}(Mt)}{\text{sinc}(t)} \Longleftrightarrow X[k] = \text{rect}\Big(\frac{k}{2N}\Big)$$

Using duality, we obtain $\text{rect}\Big(\frac{n}{2N}\Big) \Longleftrightarrow \frac{M\,\text{sinc}(MF)}{\text{sinc}\,F}$.

(b) To find the DTFT of $x[n] = \cos(2\pi n\alpha)$, we can start with the Fourier transform pair

$$\cos(2\pi\alpha t) \Longleftrightarrow 0.5[\delta(f+\alpha) + \delta(f-\alpha)]$$

Sampling $x(t),\ (t \to n)$ results in the periodic extension of $X(f)$ with period 1 $(f \to F)$, and we get

$$\cos(2\pi n\alpha) \Longleftrightarrow 0.5[\delta(F+\alpha) + \delta(F-\alpha)]$$

(c) To find the DTFT of $x[n] = \text{sinc}(2n\alpha)$, we can start with the Fourier transform pair

$$x(t) = 2\alpha\,\text{sinc}(2\alpha t) \Longleftrightarrow X(f) = \text{rect}(f/2\alpha)$$

Sampling $x(t),\ (t \to n)$ results in the periodic extension of $X(f)$ with period 1 $(f \to F)$, and we get

$$2\alpha\,\text{sinc}(2n\alpha) \Longleftrightarrow \text{rect}(F/2\alpha)$$

(d) The DTFT of $x[n] = e^{-n}u[n] = (e^{-1})^n u[n]$ may be found readily from the defining DTFT relation as $X_p(F) = \dfrac{1}{1 - e^{-1}e^{j2\pi F}}$. Suppose we start with the analog Fourier transform pair $e^{-t}u(t) \Leftrightarrow \dfrac{1}{1 + j2\pi f}$. Sampling the time signal to $x[n] = e^{-n}u[n]$ leads to periodic extension of $X(f)$ and gives the DTFT as $X_p(F) = \displaystyle\sum_k \frac{1}{1 + j2\pi(F+k)}$. But does this equal $\dfrac{1}{1 - e^{-1}e^{j2\pi F}}$? The closed-form result is pretty hard to find in a handbook of mathematical tables! However, if we start with the ideally sampled signal $x(t) = \sum e^{-n}\delta(t-n)$, its Fourier transform (using the shifting property) is

$$X(f) = 1 + e^{-1}e^{-j2\pi f} + e^{-2}e^{-j4\pi f} + e^{-3}e^{-j6\pi f} + \cdots = \frac{1}{1 - e^{-1}e^{j2\pi f}}$$

Since the DTFT corresponds to $X(f)$ with $f \to F$, we have the required equivalence!

15.6.1 The DFS and the DFT

Sampling and duality provide the basis for the connection between all of the frequency-domain transforms described in this book, and the concepts are worth repeating. Sampling in one domain induces a periodic extension in the other. The sample spacing in one domain is the reciprocal of the period in the other. Periodic analog signals have discrete spectra, and discrete-time signals have continuous periodic spectra. A consequence of these concepts is that a sequence that is both discrete and periodic in one domain is also discrete and periodic in the other. This leads to the development of the **discrete Fourier transform** (DFT) and **discrete Fourier series** (DFS), allowing us a practical means of arriving at the sampled spectrum of sampled signals using digital computers. The connections between the various transforms are illustrated in Figure 15.4 and summarized in Table 15.4.

PGR2026-00033
Krisp Technologies EX1024 Page 516

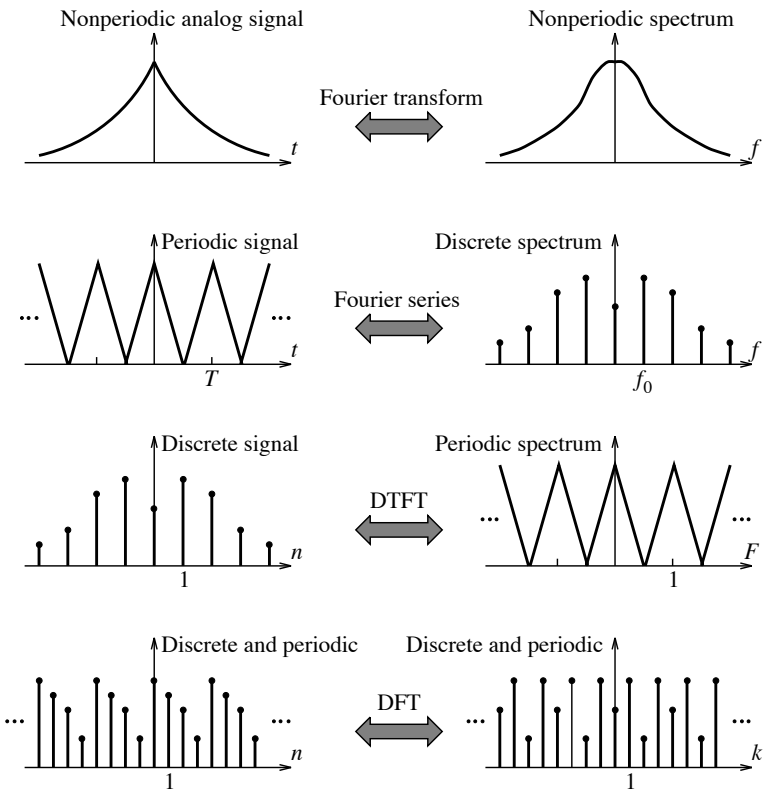


Figure 15.4 Features of the various transforms

Table 15.4 Connections Between Various Transforms

Operation in the Time Domain	**Result in the Frequency Domain**	**Transform**
Aperiodic, continuous $x(t)$	Aperiodic, continuous $X(f)$	FT
Periodic extension of $x(t) \Rightarrow x_p(t)$ Period $= T$	Sampling of $X(f) \Rightarrow X[k]$ Sampling interval $= 1/T = f_0$	FS
Sampling of $x_p(t) \Rightarrow x_p[n]$ Sampling interval $= t_s$	Periodic extension of $X[k] \Rightarrow X_{\text{DFS}}[k]$ Period $= S = 1/t_s$	DFS
Sampling of $x(t) \Rightarrow x[n]$ Sampling interval $= 1$	Periodic extension of $X(f) \Rightarrow X_p(F)$ Period $= 1$	DTFT
Periodic extension of $x[n] \Rightarrow x_p[n]$ Period $= N$	Sampling of $X_p(F) \Rightarrow X_{\text{DFT}}[k]$ Sampling interval $= 1/N$	DFT

15.6.2 The DTFT and the z-Transform

The DTFT is quite useful for signal analysis, but its use in system analysis is restricted to *relaxed* systems (with zero initial conditions) and often becomes unwieldy because the DTFT of even simple signals has

PGR2026-00033
Krisp Technologies EX1024 Page 517

a complicated form. Of course, the DTFT of signals that are not absolutely summable usually includes impulses ($u[n]$, for example) or does not exist (growing exponentials, for example). These shortcomings of the DTFT can be removed as follows.

Since the DTFT describes a discrete-time signal as a sum of weighted harmonics, we include a real weighting factor r^{-k} and redefine the transform as the sum of exponentially weighted harmonics of the form $z = r\exp(-j2\pi F)$ to obtain

$$X(z) = \sum_{k=-\infty}^{\infty} x[k]e^{-j2\pi kF}r^{-k} = \sum_{k=-\infty}^{\infty} x[k]\left(re^{j2\pi F}\right)^{-k} = \sum_{k=-\infty}^{\infty} x[k]z^{-k} \tag{15.26}$$

This defines the **two-sided z-transform** of $x[k]$. The weighting r^{-k} acts as a *convergence factor* to allow transformation of exponentially growing signals. The DTFT of $x[n]$ may thus be viewed as its z-transform $X(z)$ evaluated for $r = 1$ (along the unit circle in the z-plane).

For causal signals, we change the lower limit in the sum to zero to obtain the **one-sided z-transform**. This allows us to analyze discrete LTI systems with arbitrary initial conditions in much the same way that the Laplace transform does for analog LTI systems. The z-transform is discussed in later chapters.

15.7 Ideal Filters

An *ideal* lowpass filter is described by $H_{\rm LP}(F) = 1,\ |F| < F_C$ over its principal period $|F| \le 0.5$, as shown in Figure 15.5.

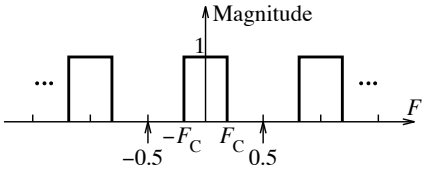


Figure 15.5 Spectrum of an ideal lowpass filter

Its impulse response $h_{\rm LP}[n]$ is found using the IDTFT to give

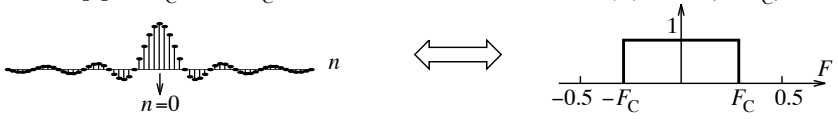
$$h_{\rm LP}[n] = \int_{0.5}^{-0.5} H_{\rm LP}(F)e^{j2\pi nF}dF = \int_{-F_C}^{F_C} e^{j2\pi nF}dF = \int_{-F_C}^{F_C} \cos(2\pi nF)\,dF = \left.\frac{\sin(2\pi nF)}{2\pi n}\right|_{-F_C}^{F_C} \tag{15.27}$$

Simplifying this result, we obtain

$$h_{\rm LP}[n] = \frac{\sin(2\pi nF_C) - \sin(-2\pi nF_C)}{2\pi n} = \frac{2\sin(2\pi nF_C)}{2\pi n} = \frac{2F_C\sin(2\pi nF_C)}{2\pi nF_C} = 2F_C\,\mathrm{sinc}(2nF_C) \tag{15.28}$$

REVIEW PANEL 15.19

The Impulse Response of an Ideal Lowpass Filter Is a Sinc Function



PGR2026-00033
Krisp Technologies EX1024 Page 518

15.7.1 Frequency Transformations

The impulse response and transfer function of highpass, bandpass, and bandstop filters may be related to those of a lowpass filter using frequency transformations based on the properties of the DTFT. The impulse response of an ideal lowpass filter with a cutoff frequency of F_C and unit passband gain is given by

$$h[n] = 2F_C \,\mathrm{sinc}(2nF_C) \tag{15.29}$$

Figure 15.6 shows two ways to obtain a highpass transfer function from this lowpass filter.

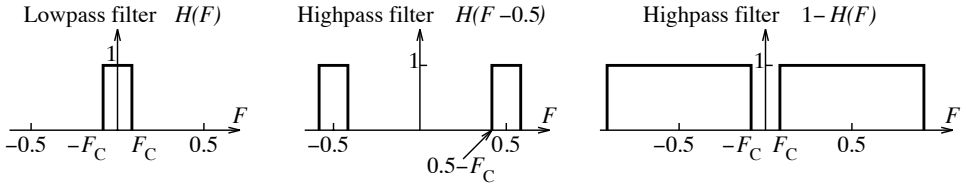


Figure 15.6 Two lowpass-to-highpass transformations

The first way is to shift $H(F)$ by 0.5 to obtain $H_{\mathrm{H1}}(F) = H(F-0.5)$, a highpass filter whose cutoff frequency is given by $F_{\mathrm{H1}} = 0.5 - F_C$. This leads to

$$h_{\mathrm{H1}}[n] = (-1)^n h[n] = 2(-1)^n F_C \,\mathrm{sinc}(2nF_C) \quad \text{(with cutoff frequency } 0.5 - F_C) \tag{15.30}$$

Alternatively, we see that $H_{\mathrm{H2}}(F) = 1 - H(F)$ also describes a highpass filter, but with a cutoff frequency given by $F_{\mathrm{H2}} = F_C$, and this leads to

$$h_{\mathrm{H2}}[n] = \delta[n] - 2F_C \,\mathrm{sinc}(2nF_C) \quad \text{(with cutoff frequency } F_C) \tag{15.31}$$

Figure 15.7 shows how to transform a lowpass filter to a bandpass filter or to a bandstop filter.

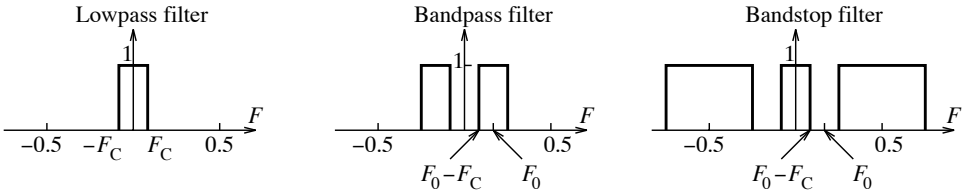


Figure 15.7 Transforming a lowpass filter to a bandpass or bandstop filter

To obtain a bandpass filter with a center frequency of F_0 and band edges $[F_0 - F_C, F_0 + F_C]$, we simply shift $H(F)$ by F_0 to get $H_{\mathrm{BP}}(F) = H(F + F_0) + H(F - F_0)$, and obtain

$$h_{\mathrm{BP}}[n] = 2h[n]\cos(2\pi nF_0) = 4F_C \cos(2\pi nF_0)\mathrm{sinc}(2nF_C) \tag{15.32}$$

A bandstop filter with center frequency F_0 and band edges $[F_0 - F_C, F_0 + F_C]$ can be obtained from the bandpass filter using $H_{\mathrm{BS}}(F) = 1 - H_{\mathrm{BP}}(F)$, to give

$$h_{\mathrm{BS}}[n] = \delta[n] - h_{\mathrm{BP}}[n] = \delta[n] - 4F_C \cos(2\pi nF_0)\mathrm{sinc}(2nF_C) \tag{15.33}$$

PGR2026-00033
Krisp Technologies EX1024 Page 519

REVIEW PANEL 15.20

Transformation of a Lowpass Filter $h[n]$ and $H(F)$ to Other Forms

$h[n] = 2F_C\,\text{sinc}(2nF_C);\ H(F) = \text{rect}(F/2F_C)$ $\quad h_{\text{HP}}[n] = \delta[n] - h[n];\ H_{\text{HP}}(F) = 1 - H(F)$

$h_{\text{BP}}[n] = 2h[n]\cos(2\pi nF_0);\ H_{\text{BP}}(F)$ by modulation $\quad h_{\text{BS}}[n] = \delta[n] - h_{\text{BP}}[n];\ H_{\text{BS}}(F) = 1 - H_{\text{BP}}(F)$

Note: F_C = cutoff frequency of lowpass prototype; F_0 = center frequency (for BPF and BSF)

EXAMPLE 15.10 (Frequency Transformations)

Use frequency transformations to find the transfer function and impulse response of

1. An ideal *lowpass* filter with a digital cutoff frequency of 0.25.
2. An ideal *highpass* filter with a digital cutoff frequency of 0.3.
3. An ideal *bandpass* filter with a passband between $F = 0.1$ and $F = 0.3$.
4. An ideal *bandstop* filter with a stopband between $F = 0.2$ and $F = 0.4$.

(a) For the lowpass filter, pick the LPF cutoff $F_C = 0.25$. Then, $h[n] = 0.5\ \text{sinc}(0.5n)$.

(b) For a highpass filter whose cutoff is $F_{\text{HP}} = 0.3$:

Pick an LPF with $F_C = 0.5 - 0.3 = 0.2$. Then, $h_{\text{HP}}[n] = (-1)^n(0.4)\text{sinc}(0.4n)$.

Alternatively, pick $F_C = 0.3$. Then, $h_{\text{HP}}[n] = \delta[n] - (0.6)\text{sinc}(0.6n)$.

(c) For a bandpass filter with band edges $[F_1, F_2] = [0.1,\ 0.3]$:

Pick the LPF cutoff as $F_C = F_2 - F_0 = 0.1,\ F_0 = 0.2$. Then, $h_{\text{BP}}[n] = 0.4\cos(0.4n\pi)\text{sinc}(0.2n)$.

(d) For a bandstop filter with band edges $[F_1, F_2] = [0.2,\ 0.4]$:

Pick the LPF cutoff as $F_C = F_2 - F_0 = 0.1,\ F_0 = 0.3$. Then, $h_{\text{BS}}[n] = \delta[n] - 0.4\cos(0.6n\pi)\text{sinc}(0.2n)$.

15.7.2 Truncation and Windowing

An ideal filter possesses linear phase because its impulse response $h[n]$ is a symmetric sequence. It can never be realized in practice because $h[n]$ is a sinc function that goes on forever, making the filter *noncausal*. Moreover, the ideal filter is *unstable* because $h[n]$ (a sinc function) is *not absolutely summable*. One way to approximate an ideal lowpass filter is by symmetric truncation (about $n = 0$) of its impulse response $h[n]$ (which ensures linear phase). Truncation of $h[n]$ to $|n| \le N$ is equivalent to multiplying $h[n]$ by a rectangular **window** $w_D[n] = 1,\ |n| \le N$. To obtain a causal filter, we must also delay the impulse response by N samples so that its first sample appears at the origin.

15.7.3 The DTFT of Two Window Functions

The DTFT of the rectangular window $x[n] = \text{rect}(n/2N)$ may be computed directly from the defining relation to give

$$X_p(F) = \sum_{k=-N}^{N} x[k]e^{-j2\pi kF} = M\frac{\text{sinc}(MF)}{\text{sinc}(F)} = W_D(F), \qquad M = 2N + 1 \tag{15.34}$$

The closed form for the summation is readily available. The quantity $W_D(F)$ describes the **Dirichlet kernel** or the **aliased sinc** function and also equals the *periodic extension* of $\text{sinc}(MF)$ with period $F = 1$. Figure 15.8 shows the Dirichlet kernel for $N = 3,\ 5$, and 10.

PGR2026-00033
Krisp Technologies EX1024 Page 520

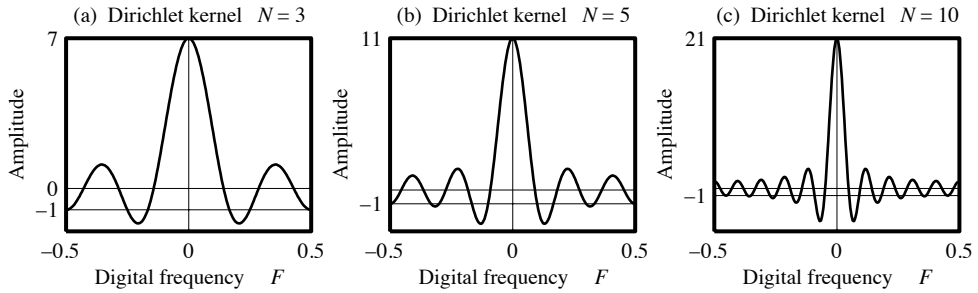


Figure 15.8 The Dirichlet kernel is the spectrum of a rectangular window

The Dirichlet kernel has some very interesting properties. Over one period, we observe the following:

1. It shows N maxima, a positive main lobe of width $2/M$, decaying positive and negative sidelobes of width $1/M$, with $2N$ zeros at $k/M, k = 1, 2, \ldots, 2N$.

2. Its area equals unity, and it attains a maximum peak value of M (at the origin) and a minimum peak value of unity (at $F = 0.5$ for odd N). Increasing M increases the mainlobe height and compresses the sidelobes. The ratio R of the main lobe and peak sidelobe magnitudes stays nearly constant (between 4 and 4.7) for finite M, and $R \to 1.5\pi \approx 4.71$ (or 13.5 dB) for very large M. As $M \to \infty, X_p(F)$ approaches a unit impulse.

To find the DTFT of the triangular window $x[n] = \text{tri}(n/M)$, we start with the DTFT pair $x[n] = \text{rect}(n/2N) \Leftrightarrow M\frac{\text{sinc}(MF)}{\text{sinc}F}$, $M = 2N+1$. Since $\text{rect}(n/2N) \star \text{rect}(n/2N) = (2N+1)\text{tri}(\frac{n}{2N+1}) = M\text{tri}(n/M)$, the convolution property gives

$$M\text{tri}(n/M) \Longleftrightarrow M^2 \frac{\text{sinc}^2(MF)}{\text{sinc}^2 F} \tag{15.35}$$

This leads to the transform pair

$$\text{tri}(n/M) \Leftrightarrow M\frac{\text{sinc}^2(MF)}{\text{sinc}^2 F} = W_F(F) \tag{15.36}$$

The quantity $W_F(F)$ is called the **Fejer kernel**. Figure 15.9 shows the Fejer kernel for $N = 3,\ 5$, and 10. It is always positive and shows M maxima over one period with a peak value of M at the origin.

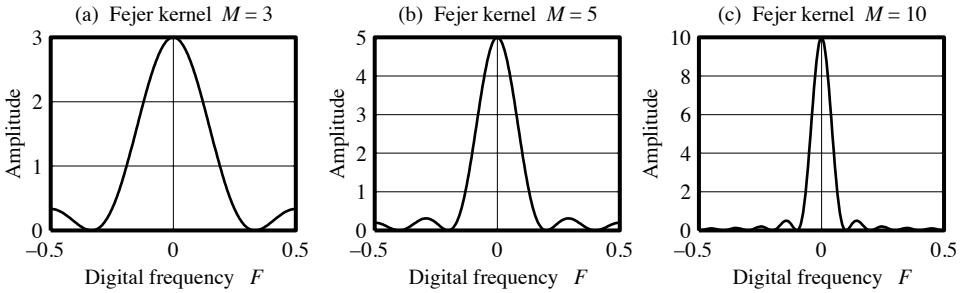


Figure 15.9 The Fejer kernel is the spectrum of a triangular window

PGR2026-00033
Krisp Technologies EX1024 Page 521

15.7.4 The Consequences of Windowing

Windowing of $h[n]$ by a time-domain window $w[n]$ is equivalent to the *periodic convolution* of $H(F)$ and $W(F)$ in the frequency domain. The spectrum of an ideal filter using a rectangular window is shown in Figure 15.10 and exhibits overshoot and oscillations reminiscent of the **Gibbs effect**.

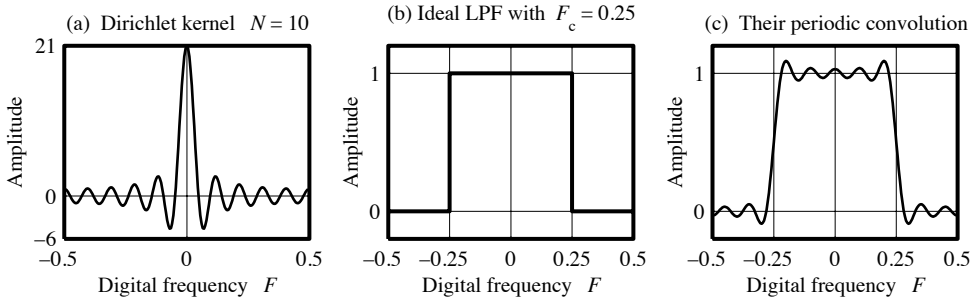


Figure 15.10 The spectrum of an abruptly truncated impulse response

What causes the overshoot and ringing is the *slowly* decaying sidelobes of $W_D(F)$ (the Dirichlet kernel corresponding to the rectangular window). The overshoot and oscillations persist no matter how large the truncation index N. To eliminate overshoot and reduce the oscillations, we multiply $h_{\text{LP}}[n]$ by a *tapered* window whose DTFT sidelobes decay much faster (as we also did for Fourier series smoothing). The triangular window (whose DTFT describes the Fejer kernel) for example, is one familiar example whose periodic convolution with the ideal filter spectrum leads to a monotonic response and the complete absence of overshoot and oscillations, as shown in Figure 15.11. All windows are designed to minimize (or eliminate) the overshoot in the spectrum while maintaining as abrupt a transition as possible.

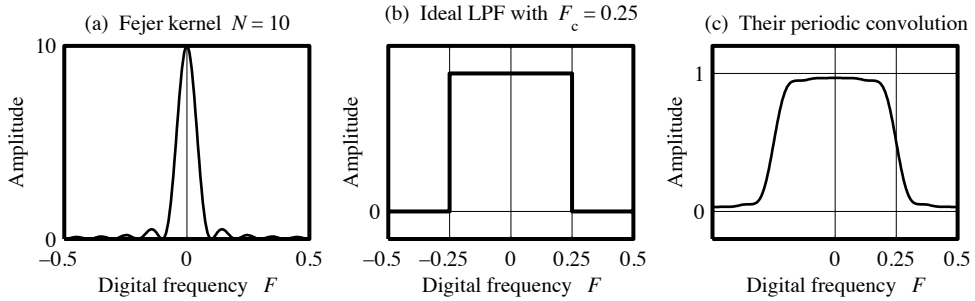


Figure 15.11 The spectrum of an ideal filter windowed by a triangular window

REVIEW PANEL 15.21

Windowing of $h[n]$ in Time Domain Means Periodic Convolution in Frequency Domain

Windowing makes the transition in the spectrum more gradual (less abrupt).
A rectangular window (truncation) leads to overshoot and oscillations in spectrum (Gibbs effect).
Any other window leads to reduced (or no) overshoot in spectrum.

PGR2026-00033
Krisp Technologies EX1024 Page 522

15.8 Some Traditional and Non-Traditional Filters

In the frequency domain, performance measures of digital filters are based on the magnitude, phase (in radians or degrees), phase delay, and group delay. The **phase delay** and **group delay** for a filter whose transfer function is $H_p(F)$ are defined by

$$t_p(F) = -\frac{1}{2\pi}\frac{H_p(F_0)}{F_0} \quad \text{(phase delay)} \qquad t_g(F) = -\frac{1}{2\pi}\frac{dH_p(F)}{dF} \quad \text{(group delay)} \tag{15.37}$$

Performance in the time domain is based on the characteristics of the impulse response and step response.

15.8.1 Lowpass and Highpass Filters

Consider the first-order filter described by

$$H_{\text{LP}}(F) = \frac{1}{1-\alpha e^{-j2\pi F}} \qquad H_{\text{LP}}(0) = \frac{1}{1-\alpha} \quad H_{\text{LP}}(0.5) = \frac{1}{1+\alpha}, \quad 0 < \alpha < 1 \tag{15.38}$$

The filter $H_{\text{LP}}(F)$ describes a lowpass filter because its magnitude at high frequencies (i.e., frequencies closer to $F = 0.5$) decreases and $|H_{\text{LP}}(0)| > |H_{\text{LP}}(0.5)|$. The spectrum and impulse response of this filter are shown in Figure 15.12.

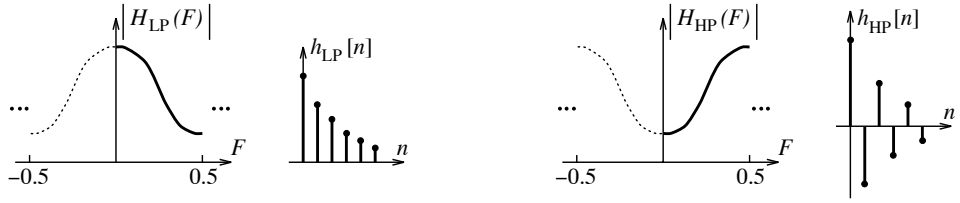


Figure 15.12 Spectrum and impulse response of first-order lowpass and highpass filters

We can find the half-power frequency of the lowpass filter from its magnitude squared function:

$$|H_{\text{LP}}(F)|^2 = \left|\frac{1}{1-\alpha\cos(2\pi F)+j\alpha\sin(2\pi F)}\right|^2 = \frac{1}{1-2\alpha\cos(2\pi F)+\alpha^2} \tag{15.39}$$

At the half-power frequency, we have $|H_{\text{LP}}(F)|^2 = 0.5$, and this gives

$$\frac{1}{1-2\alpha\cos(2\pi F)+\alpha^2} = 0.5 \qquad \text{or} \qquad F = \frac{1}{2\pi}\cos^{-1}\left(\frac{\alpha^2-1}{2\alpha}\right), \quad |F| < 0.5 \tag{15.40}$$

The phase and group delay of the lowpass filter are given by

$$\phi(F) = -\tan^{-1}\left[\frac{\alpha\sin(2\pi F)}{1-\alpha\cos(2\pi F)}\right] \qquad t_g(F) = -\frac{1}{2\pi}\frac{d\phi(F)}{dF} = \frac{\alpha\cos(2\pi F)-\alpha^2}{1-2\alpha\cos(2\pi F)+\alpha^2} \tag{15.41}$$

For low frequencies, the phase is nearly linear and the group delay is nearly constant, and they can be approximated by

$$\phi(F) \approx \frac{-2\pi\alpha F}{1-\alpha}, \quad |F| \ll 1 \qquad t_g(F) = -\frac{1}{2\pi}\frac{d\phi(F)}{dF} \approx \frac{\alpha}{1-\alpha}, \quad |F| \ll 1 \tag{15.42}$$

PGR2026-00033
Krisp Technologies EX1024 Page 523

The impulse response of this lowpass filter is given by

$$h_{\text{LP}}[n] = \alpha^n u[n] \tag{15.43}$$

The **time constant** of the lowpass filter describes how fast the impulse response decays to a specified fraction of its initial value. For a specified fraction of $\epsilon\%$, we obtain the effective time constant η in samples as

$$\alpha^\eta = \epsilon \qquad \eta = \frac{\ln \epsilon}{\ln \alpha} \tag{15.44}$$

This value is rounded up to an integer, if necessary. We obtain the commonly measured 1% or 40-dB time constant if $\epsilon = 0.01$ (corresponding to an attenuation of 40 dB) and the 0.1% or 60-dB time constant if $\epsilon = 0.001$. If the sampling interval t_s is known, the time constant τ in seconds can be computed from $\tau = \eta t_s$. The 60-dB time constant is also called the **reverberation time**. For higher-order filters whose impulse response contains several exponential terms, the effective time constant is dictated by the term with the slowest decay (that dominates the response).

Now, consider the first-order filter where we change the sign of α to give

$$H_{\text{HP}}(F) = \frac{1}{1 + \alpha e^{-j2\pi F}} \qquad H_{\text{HP}}(0) = \frac{1}{1+\alpha} \quad H_{\text{HP}}(0.5) = \frac{1}{1-\alpha}, \quad 0 < \alpha < 1 \tag{15.45}$$

This describes a highpass filter because its magnitude at low frequencies (closer to $F = 0$) is small with $|H_{\text{HP}}(0)| < |H_{\text{HP}}(0.5)|$, as shown in Figure 15.12. The impulse response of this filter is given by

$$h_{\text{HP}}[n] = (-\alpha)^n u[n] = (-1)^n \alpha^n u[n] \tag{15.46}$$

It is the alternating sign changes in the samples of $h_{\text{HP}}[n]$ that cause rapid time variations and lead to its highpass behavior. Its spectrum $H_{\text{HP}}(F)$ is related to $H_{\text{LP}}(F)$ by $H_{\text{HP}}(F) = H_{\text{LP}}(F - 0.5)$. This means that a lowpass cutoff frequency F_0 corresponds to the frequency $0.5 - F_0$ of the highpass filter.

15.8.2 Allpass Filters

Consider the first-order filter described by

$$H_A(F) = \frac{A + Be^{j2\pi F}}{B + Ae^{j2\pi F}} \qquad |H_A(0)| = \left|\frac{A+B}{B+A}\right| = 1 \qquad |H_A(0.5)| = \left|\frac{A-B}{B-A}\right| = 1 \tag{15.47}$$

Its magnitude at $F = 0$ and $F = 0.5$ is equal to unity. In fact, its magnitude is unity for all frequencies, and this describes an **allpass filter**. The fact that its magnitude is constant for all frequencies can be explained by a geometric argument, as illustrated in Figure 15.13.

The numerator and denominator may be regarded as the sum of two vectors of length A and B that subtend the same angle $\Omega = 2\pi F$. The length of their vector sum is thus equal for any Ω. An interesting characteristic that allows us to readily identify an allpass filter is that its numerator and denominator coefficients (associated with powers of $e^{\pm j2\pi F}$ or $e^{j\Omega}$) appear in reversed order. Allpass filters are often used to shape the delay characteristics of digital filters, and its applications are discussed in later chapters.

REVIEW PANEL 15.22

Identifying an Allpass Filter by Checking Coefficients

From transfer function: The coefficients of numerator and denominator appear in reversed order.
From difference equation: The coefficients of RHS and LHS appear in reversed order.

PGR2026-00033
Krisp Technologies EX1024 Page 524

$$H(F) = \frac{A + Be^{j2\pi F}}{B + Ae^{j2\pi F}} = \frac{A\angle 0 + B\angle \Omega}{B\angle 0 + A\angle \Omega} = \frac{\mathbf{X}}{\mathbf{Y}}$$

The vectors **X** and **Y** are of equal length for any avlue of F (or Ω).

The magnitude of $H(F)$ is thus unity.

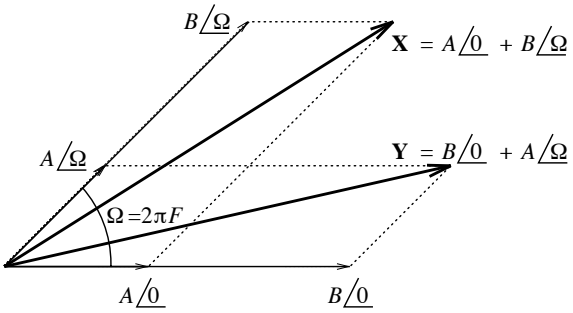


Figure 15.13 The magnitude spectrum of an allpass filter is constant

EXAMPLE 15.11 (Traditional and Non-Traditional Filters)

(a) Let $h[n] = \{\overset{\Downarrow}{5},\ -4,\ 3,\ -2\}$. Identify the filter type and establish whether the impulse response is a linear-phase sequence.

The sequence $h[n]$ is not linear-phase because it shows no symmetry.

We have $H_p(F) = 5 - 4e^{-j2\pi F} + 3e^{-j4\pi F} - 2e^{-j6\pi F}$.

We find $H_p(0) = \sum h[n] = 2$ and $H_p(0.5) = \sum (-1)^n h[n] = 5 + 4 + 3 + 2 = 14$.

Since the magnitude at high frequencies increases, this appears to be a highpass filter.

(b) Let $h[n] = (0.8)^n u[n]$. Identify the filter type, establish whether the impulse response is a linear-phase sequence, and find its 60-dB time constant.

We have $H_p(F) = \dfrac{1}{1 - 0.8e^{-j2\pi F}}$, $H_p(0) = \dfrac{1}{1 - 0.8} = 5$, and $H_p(0.5) = \dfrac{1}{1 - 0.8e^{-j\pi}} = \dfrac{1}{1 + 0.8} = 0.556$

The sequence $h[n]$ is not a linear-phase phase because it shows no symmetry.

Since the magnitude at high frequencies decreases, this appears to be a lowpass filter. The 60-dB time constant of this filter is found as

$$\eta = \frac{\ln \epsilon}{\ln \alpha} = \frac{\ln 0.001}{\ln 0.8} = 30.96 \quad \Rightarrow \quad \eta = 31 \text{ samples}$$

For a sampling frequency of $S = 100$ Hz, this corresponds to $\tau = \eta t_s = \eta / S = 0.31$ s.

(c) Let $h[n] = 0.8\delta[n] + 0.36(-0.8)^{n-1}u[n-1]$. Identify the filter type and establish whether the impulse response is a linear-phase sequence.

We find $H_p(F) = 0.8 + \dfrac{0.36e^{-j2\pi F}}{1 + 0.8e^{-j2\pi F}} = \dfrac{0.8 + e^{-j2\pi F}}{1 + 0.8e^{-j2\pi F}}$. So, $H_p(0) = 1$ and $H_p(0.5) = \dfrac{0.8 - 1}{1 - 0.8} = -1$.

The sequence $h[n]$ is not a linear-phase sequence because it shows no symmetry.

Since the magnitude is identical at low and high frequencies, this could be a bandstop or an allpass filter. Since the numerator and denominator coefficients in $H_p(F)$ appear in reversed order, it is an *allpass* filter.

PGR2026-00033
Krisp Technologies EX1024 Page 525

(d) Let $h[n] = \delta[n] - \delta[n-8]$. Identify the filter type and establish whether the impulse response is a linear-phase sequence.

We find $H_p(F) = 1 - e^{-j16\pi F}$. So, $H_p(0) = 0$ and $H_p(0.5) = 0$.

This suggests a bandpass filter with a maximum between $F = 0$ and $F = 0.5$. On closer examination, we find that $H_p(F) = 0$ at $F = 0,\ 0.125,\ 0.25,\ 0.375,\ 0.5$ when $e^{-j16\pi F} = 1$ and $|H_p(F)| = 2$ at four frequencies halfway between these. This multi-humped response is sketched in Figure E15.11D and describes a **comb filter**.

We may write $h[n] = \{\overset{\Downarrow}{1}, 0, 0, 0, 0, 0, 0, 0, -1\}$. This is a linear-phase sequence because it shows odd symmetry about the index $n = 4$.

Comment: Note that the difference equation $y[n] = x[n] - x[n-8]$ is reminiscent of an echo filter.

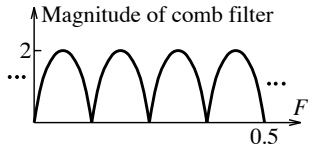


Figure E15.11D Spectrum of the comb filter for Example 15.11(d)

15.9 Frequency Response of Discrete Algorithms

Discrete-time algorithms are used to convert analog systems to discrete systems. The frequency response of an ideal integrator is $H_I(F) = 1/j2\pi F$. For an ideal differentiator, $H_D(F) = j2\pi F$. In the discrete domain, the operations of integration and differentiation are replaced by numerical integration and numerical differences. Table 15.5 lists some of the algorithms and their frequency response $H(F)$.

Numerical differences approximate the slope (derivative) at a point, as illustrated in Figure 15.14.

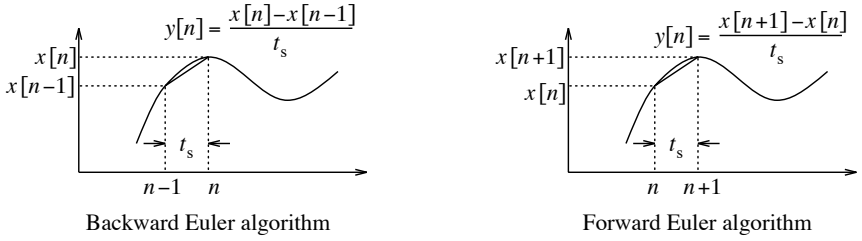


Figure 15.14 Numerical difference algorithms

Most numerical integration algorithms estimate the area $y[n]$ from $y[n-1]$ by using step, linear, or quadratic interpolation between the samples of $x[n]$, as illustrated in Figure 15.15.

Only Simpson's rule finds $y[n]$ over two time steps from $y[n-2]$. Discrete algorithms are good approximations only at low digital frequencies ($F < 0.1$, say). Even when we satisfy the sampling theorem, low digital frequencies mean high sampling rates, well in excess of the Nyquist rate. This is why the sampling rate is a critical factor in the frequency-domain performance of these operators. Another factor is stability. For example, if Simpson's rule is used to convert analog systems to discrete systems, it results in an unstable system. The trapezoidal integration algorithm and the backward difference are two popular choices.

PGR2026-00033
Krisp Technologies EX1024 Page 526

Table 15.5 Discrete Algorithms and their Frequency Response

Algorithm	**Difference Equation**	**Frequency Response**
Numerical Differences		
Backward	$y[n] = x[n] - x[n-1]$	$H_B(F) = 1 - e^{-j2\pi F}$
Central	$y[n] = \frac{1}{2}(x[n+1] - x[n-1])$	$H_C(F) = j\sin(2\pi F)$
Forward	$y[n] = x[n+1] - x[n]$	$H_F(F) = e^{j2\pi F} - 1$
Numerical Integration		
Rectangular	$y[n-1] + x[n]$	$H_R(F) = \dfrac{1}{1 - e^{-j2\pi F}}$
Trapezoidal	$y[n-1] + \frac{1}{2}(x[n] + x[n-1])$	$H_T(F) = \dfrac{1 + e^{-j2\pi F}}{2(1 - e^{-j2\pi F})}$
Simpson's	$y[n-2] + \frac{1}{3}(x[n] + 4x[n-1] + x[n-2])$	$H_S(F) = \dfrac{1 + 4e^{-j2\pi F} + e^{-j4\pi F}}{3(1 - e^{-j4\pi F})}$

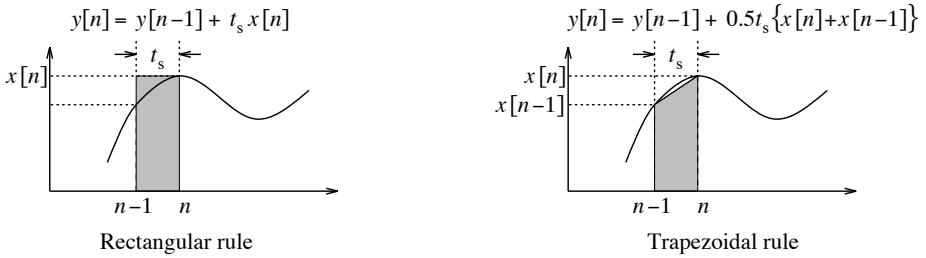


Figure 15.15 Numerical integration algorithms

EXAMPLE 15.12 (DTFT of Numerical Algorithms)

(a) For the trapezoidal numerical integration operator, the DTFT yields

$$Y_p(F) = Y_p(F)e^{-j2\pi F} + 0.5[X_p(F) + e^{-j2\pi F}X_p(F)]$$

A simple rearrangement leads to

$$H_T(F) = \frac{Y_p(F)}{X_p(F)} = 0.5\left[\frac{1 + e^{-j2\pi F}}{1 - e^{-j2\pi F}}\right] = \frac{1}{j2\tan(\pi F)}$$

Normalizing this by $H_I(F) = 1/j2\pi F$ and expanding the result, we obtain

$$\frac{H_T(F)}{H_I(F)} = \frac{j2\pi F}{j2\tan(\pi F)} = \frac{\pi F}{\tan(\pi F)} \approx 1 - \frac{(\pi F)^2}{3} - \frac{(\pi F)^4}{45} - \cdots$$

PGR2026-00033
Krisp Technologies EX1024 Page 527

Figure E15.12 shows the magnitude and phase *error* by plotting the ratio $H_T(F)/H_I(F)$. For an ideal algorithm, this ratio should equal unity at all frequencies. At low frequencies ($F \ll 1$), we have $\tan(\pi F) \approx \pi F$, and $H_T(F)/H_I(F) \approx 1$, and the trapezoidal rule is a valid approximation to integration. The phase response matches the ideal phase at all frequencies.

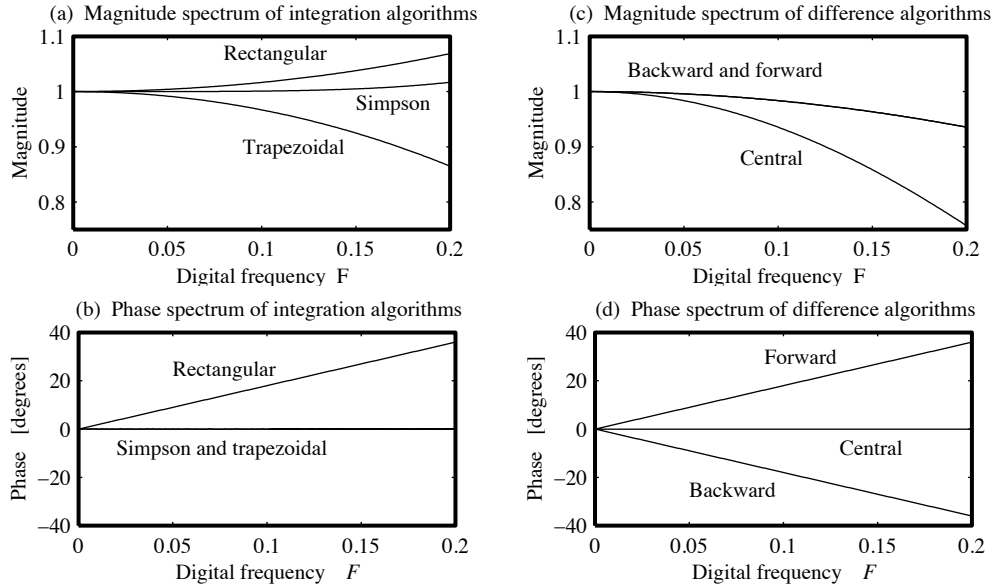


Figure E15.12 Frequency response of the numerical algorithms for Example 15.12

(b) Simpson's numerical integration algorithm yields the following normalized result:

$$\frac{H_S(F)}{H_I(F)} = 2\pi F\left(\frac{2+\cos 2\pi F}{3\sin 2\pi F}\right)$$

It displays a perfect phase match for all frequencies, but has an overshoot in its magnitude response past $F = 0.25$ and thus amplifies high frequencies.

(c) For the forward difference operator, the DTFT yields

$$Y_p(F) = X_p(F)e^{j2\pi F} - X_p(F) = X_p(F)[e^{j2\pi F} - 1] \qquad H_F(F) = \frac{Y_p(F)}{X_p(F)} = e^{j2\pi F} - 1$$

The ratio $H_F(F)/H_D(F)$ may be expanded as

$$\frac{H_F(F)}{H_D(F)} = 1 + j\frac{1}{2!}(2\pi F) - \frac{1}{3!}(2\pi F)^2 + \cdots$$

Again, we observe correspondence only at low digital frequencies (or high sampling rates). The high frequencies are amplified, making the algorithm susceptible to high-frequency noise. The phase response also deviates from the true phase, especially at high frequencies.

PGR2026-00033
Krisp Technologies EX1024 Page 528

(d) For the central difference algorithm, we find $H_C(F)/H_D(F)$ as

$$\frac{H_C(F)}{H_D(F)} = \frac{\sin(2\pi F)}{2\pi F} = 1 - \frac{1}{3!}(2\pi F)^2 + \frac{1}{5!}(2\pi F)^4 + \cdots$$

We see a perfect match only for the phase at all frequencies.

15.10 Oversampling and Sampling Rate Conversion

In practice, different parts of a DSP system are often designed to operate at different sampling rates because of the advantages it offers. For example, oversampling the analog signal prior to digital processing can reduce the errors (sinc distortion) caused by zero-order-hold sampling devices. Since real-time digital filters must complete all algorithmic operations in one sampling interval, oversampling can impose an added computational burden. It is for this reason that the sampling rate is often reduced (by decimation) before performing DSP operations and increased (by interpolation) before reconstruction. Oversampling prior to signal reconstruction allows us to relax the stringent requirements for the design of reconstruction filters.

Figure 15.16 shows the spectrum of sampled signals obtained from a band-limited analog signal whose spectrum is $X(f)$ at a sampling rate S, a higher rate NS, and a lower rate S/M.

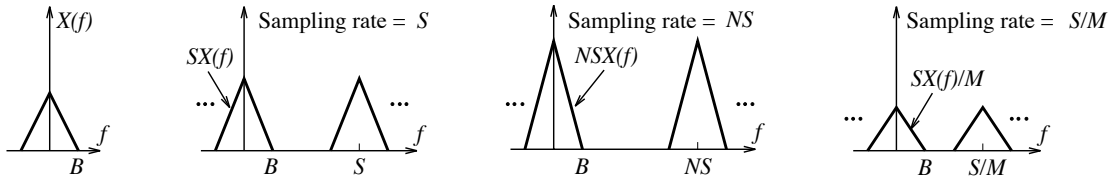


Figure 15.16 Spectra of a signal sampled at three sampling rates

The spectrum of the oversampled signal shows a gain of N but covers a smaller fraction of the principal period. The spectrum of a signal sampled at the lower rate S/M is a stretched version with a gain of $1/M$. In terms of the digital frequency F, the period of all three sampled versions is unity. One period of the spectrum of the signal sampled at S Hz extends to B/S, whereas the spectrum of the same signal sampled at NS Hz extends only to B/NS Hz, and the spectrum of the same signal sampled at S/M Hz extends farther out to BM/S Hz. After an analog signal is first sampled, all subsequent sampling rate changes are typically made by manipulating the signal samples (and not resampling the analog signal). The key to the process lies in interpolation and decimation.

15.10.1 Zero Interpolation and Spectrum Compression

A property of the DTFT that forms the basis for signal interpolation and sampling rate conversion is that M-fold zero interpolation of a discrete-time signal $x[n]$ leads to an M-fold spectrum compression and replication, as illustrated in Figure 15.17.

The zero-interpolated signal $y[n] = x[n/N]$ is nonzero only if $n = kN,\ k = 0, \pm 1, \pm 2, \ldots$ (i.e., if n is an integer multiple of N). The DTFT $Y_p(F)$ of $y[n] = y[kN]$ may be expressed as

$$Y_p(F) = \sum_{n=-\infty}^{\infty} y[n]e^{-j2\pi nF} = \sum_{k=-\infty}^{\infty} y[kN]e^{-j2\pi kNF} = \sum_{k=-\infty}^{\infty} x[k]e^{-j2\pi kNF} = X_p(NF) \tag{15.48}$$

PGR2026-00033
Krisp Technologies EX1024 Page 529

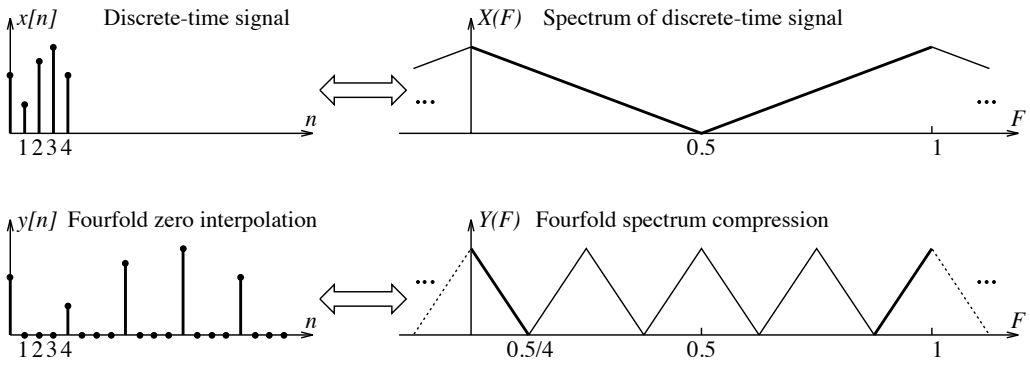


Figure 15.17 Zero interpolation of a signal leads to spectrum compression

This describes $Y_p(F)$ as a scaled (compressed) version of the periodic spectrum $X_p(F)$ and leads to N-fold spectrum replication. The spectrum of the interpolated signal shows N compressed images in the principal period $|F| \leq 0.5$, with the central image occupying the frequency range $|F| \leq 0.5/N$. This is exactly analogous to the Fourier series result where spectrum zero interpolation produced replication (compression) of the periodic signal.

Similarly, the spectrum of a decimated signal $y[n] = x[nM]$ is a stretched version of the original spectrum and described by

$$Y_p(F) = \frac{1}{M} X_p\Big(\frac{F}{M}\Big) \tag{15.49}$$

The factor $1/M$ ensures that we satisfy Parseval's relation. If the spectrum of the original signal $x[n]$ extends to $|F| \leq 0.5/M$ in the central period, the spectrum of the decimated signal extends over $|F| \leq 0.5$, and there is no aliasing or overlap between the spectral images.

REVIEW PANEL 15.23

The Spectra of Zero-Interpolated and Decimated Signals

Zero interpolation by N causes N-fold spectrum replication (with no gain change).
Decimation by M causes M-fold stretching and gain *reduction* by M.

EXAMPLE 15.13 (Zero Interpolation and Spectrum Replication)

The spectrum of a signal $x[n]$ is $X(F) = 2\,\mathrm{tri}(5F)$. Sketch $X(F)$ and the spectra of the following signals and explain how they are related to $X(F)$.

1. The zero interpolated signal $y[n] = x[n/2]$
2. The decimated signal $d[n] = x[2n]$
3. The signal $g[n]$ that equals $x[n]$ for even n, and 0 for odd n

Refer to Figure E15.13 for the spectra.

(a) The spectrum $Y(F)$ is a compressed version of $X(F)$ with $Y(F) = X(2F)$.

(b) The spectrum $D(F) = 0.5X(0.5F)$ is a stretched version of $X(F)$ with a gain factor of 0.5.

(c) The signal $g[n]$ may be expressed as $g[n] = 0.5(x[n] + (-1)^n x[n])$. Its spectrum (or DTFT) is described by $G(F) = 0.5[X(F) + X(F - 0.5)]$. We may also obtain $g[n]$ by first decimating $x[n]$ by 2 (to get $d[n]$), and then zero-interpolating $d[n]$ by 2.

PGR2026-00033
Krisp Technologies EX1024 Page 530

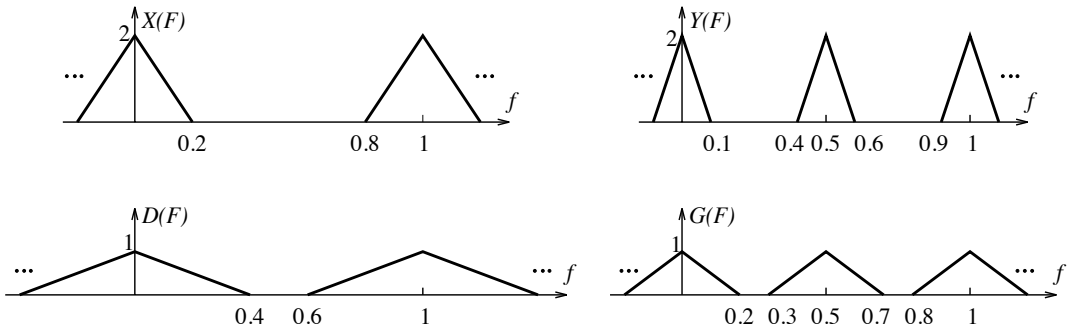


Figure E15.13 The spectra of the signals for Example 15.13

15.10.2 Sampling Rate Increase

A sampling rate increase by an integer factor N involves zero interpolation and digital lowpass filtering, as shown in Figure 15.18. The signals and their spectra at various points in the system are illustrated in Figure 15.19 for $N = 2$.

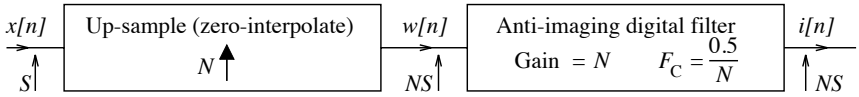


Figure 15.18 Sampling rate increase by an integer factor

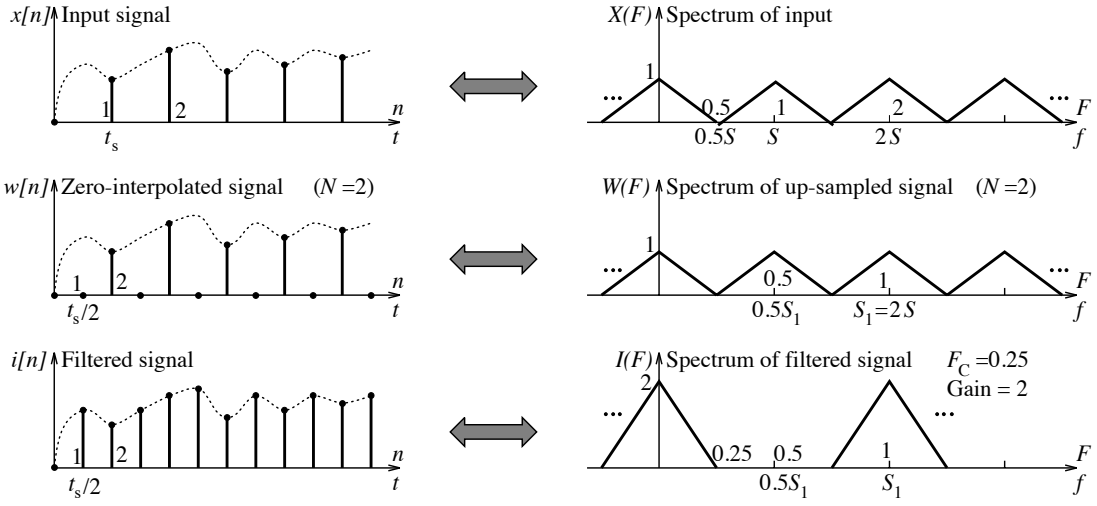


Figure 15.19 Spectra of signals when increasing the sampling rate

An up-sampler inserts $N - 1$ zeros between signal samples and results in an N-fold zero-interpolated signal corresponding to the higher sampling rate NS. Zero interpolation by N results in N-fold replication of the spectrum whose central image over $|F| \leq 0.5/N$ corresponds to the spectrum of the oversampled signal

PGR2026-00033
Krisp Technologies EX1024 Page 531

(except for a gain of N). The spurious images are removed by passing the zero-interpolated signal through a lowpass filter with a gain of N and a cutoff frequency of $F_C = 0.5/N$ to obtain the required oversampled signal. If the original samples were acquired at a rate that *exceeds* the Nyquist rate, the cutoff frequency of the digital lowpass filter can be made smaller than $0.5/N$. This process yields exact results as long as the underlying analog signal has been sampled above the Nyquist rate and the filtering operation is assumed ideal.

REVIEW PANEL 15.24

To Increase the Sampling Rate by N, Up-Sample by N and Lowpass Filter

Up-sampling compresses (and replicates) spectrum. The LPF ($F_C = \frac{0.5}{N}$, gain $= N$) removes the images.

EXAMPLE 15.14 (Up-Sampling and Filtering)

In the system of Figure E15.14, the spectrum of the input signal is given by $X(F) = \text{tri}(2F)$, the up-sampling is by a factor of 2, and the impulse response of the digital filter is given by $h[n] = 0.25\,\text{sinc}(0.25n)$. Sketch $X(F)$, $W(F)$, $H(F)$, and $Y(F)$ over $-0.5 \le F \le 0.5$.

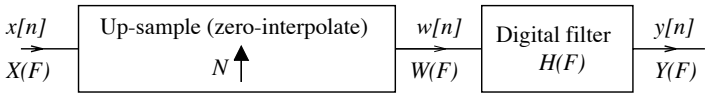


Figure E15.14 The system for Example 15.14

The various spectra are shown in Figure E15.14A.

The spectrum $X(F)$ is a triangular pulse of unit width. Zero interpolation by two results in the compressed spectrum $W(F)$. Now, $h[n] = 2F_C\,\text{sinc}(2nF_C)$ corresponds to an ideal filter with a cutoff frequency of F_C. Thus, $F_C = 0.125$, and the filter passes only the frequency range $|F| \le 0.125$.

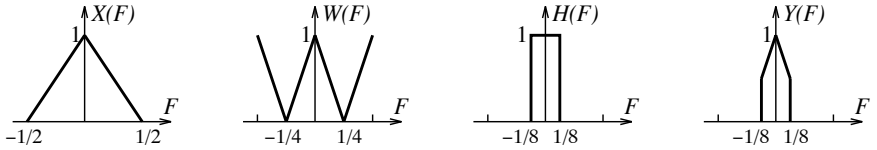


Figure E15.14A The spectra at various points for the system of Example 15.14

15.10.3 Sampling Rate Reduction

A reduction in the sampling rate by an integer factor M involves digital lowpass filtering and decimation, as shown in Figure 15.20.

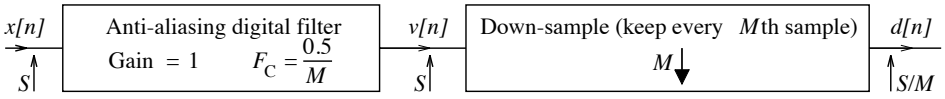


Figure 15.20 Sampling rate reduction by an integer factor

First, the sampled signal is passed through a lowpass filter (with unit gain) to ensure that it is band-limited to $|F| \le 0.5/M$ in the principal period and to prevent aliasing during the decimation stage. It is then decimated by a down-sampler that retains every Mth sample. The spectrum of the decimated signal is a

PGR2026-00033
Krisp Technologies EX1024 Page 532

stretched version that extends to $|F| \leq 0.5$ in the principal period. Note that the spectrum of the decimated signal has a gain of $1/M$ as required, and it is for this reason that we use a lowpass filter with *unit* gain.

Figure 15.21 shows the signals and their spectra at various points in the system for $M = 2$. This process yields exact results as long as the underlying analog signal has been sampled at a rate that is M times the Nyquist rate (or higher) and the filtering operation is assumed ideal.

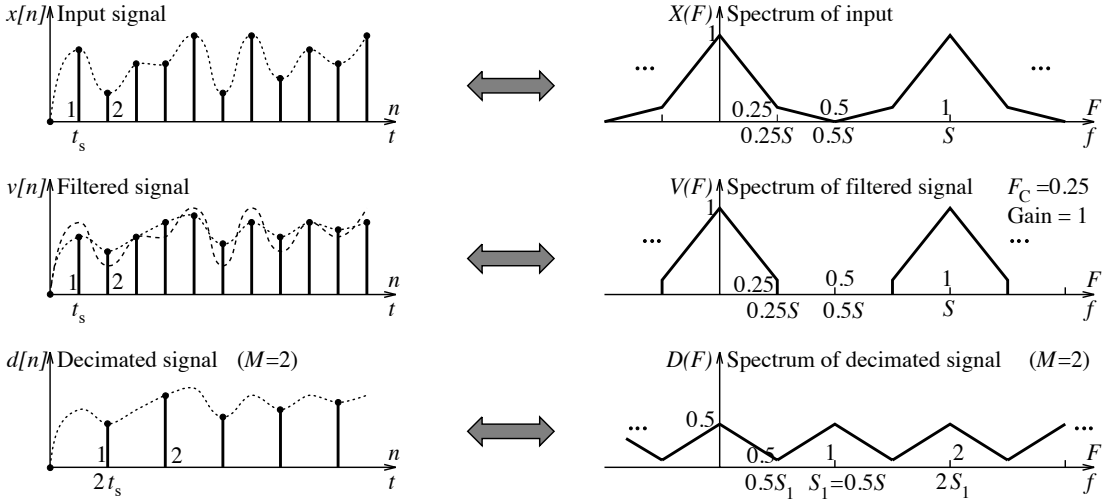


Figure 15.21 The spectra of various signals during sampling rate reduction

REVIEW PANEL 15.25

To Decrease Sampling Rate by M: Unity-Gain LPF ($F_C = \frac{0.5}{M}$) and Down-Sampling by M

The LPF ($F_C = \frac{0.5}{M}$) band-limits the input. Down-sampling stretches the spectrum.

Fractional sampling-rate changes by a factor M/N can be implemented by cascading a system that increases the sampling rate by N (interpolation) and a system that reduces sampling rate by M (decimation). In fact, we can replace the two lowpass filters that are required in the cascade (both of which operate at the sampling rate NS) by a single lowpass filter whose gain is $1/N$ and whose cutoff frequency is the smaller of $0.5/M$ and $0.5/N$, as shown in Figure 15.22.

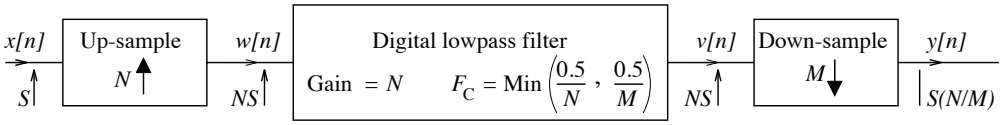


Figure 15.22 Illustrating fractional sampling-rate change

REVIEW PANEL 15.26

How to Implement a Fractional Sampling-Rate Change by M/N

Interpolate by $N \Longrightarrow$ lowpass filter (using $F_C = \frac{0.5}{\max(M,N)}$ and gain $= \frac{1}{N}$) $\Longrightarrow$ decimate by M.

PGR2026-00033
Krisp Technologies EX1024 Page 533

CHAPTER 15 PROBLEMS

DRILL AND REINFORCEMENT

15.1 (DTFT of Sequences) Find the DTFT $X_p(F)$ of the following signals and evaluate $X_p(F)$ at $F = 0$, $F = 0.5$, and $F = 1$.

(a) $x[n] = \{1, 2, \overset{\Downarrow}{3}, 2, 1\}$ **(b)** $x[n] = \{-1, 2, \overset{\Downarrow}{0}, -2, 1\}$
(c) $x[n] = \{\overset{\Downarrow}{1}, 2, 2, 1\}$ **(d)** $x[n] = \{\overset{\Downarrow}{-1}, -2, 2, 1\}$

15.2 (DTFT from Definition) Find the DTFT $X_p(F)$ of the following signals.

(a) $x[n] = (0.5)^{n+2}u[n]$ **(b)** $x[n] = n(0.5)^{2n}u[n]$ **(c)** $x[n] = (0.5)^{n+2}u[n-1]$
(d) $x[n] = n(0.5)^{n+2}u[n-1]$ **(e)** $x[n] = (n+1)(0.5)^n u[n]$ **(f)** $x[n] = (0.5)^{-n}u[-n]$

15.3 (Properties) The DTFT of $x[n]$ is $X(F) = \dfrac{4}{2 - e^{-j2\pi F}}$. Find the DTFT of the following signals without first computing $x[n]$.

(a) $y[n] = x[n-2]$ **(b)** $d[n] = nx[n]$
(c) $p[n] = x[-n]$ **(d)** $g[n] = x[n] - x[n-1]$
(e) $h[n] = x[n] \star x[n]$ **(f)** $r[n] = x[n]e^{jn\pi}$
(g) $s[n] = x[n]\cos(n\pi)$ **(h)** $v[n] = x[n-1] + x[n+1]$

15.4 (Properties) The DTFT of the signal $x[n] = (0.5)^n u[n]$ is $X(F)$. Find the time signal corresponding to the following transforms without first computing $X_p(F)$.

(a) $Y(F) = X(-F)$ **(b)** $G(F) = X(F - 0.25)$
(c) $H(F) = X(F + 0.5) + X(F - 0.5)$ **(d)** $P(F) = X'(F)$
(e) $R(F) = X^2(F)$ **(f)** $S(F) = X(F) \circledast X(F)$
(g) $D(F) = X(F)\cos(4\pi F)$ **(h)** $T(F) = X(F + 0.25) - X(F - 0.25)$

15.5 (Spectrum of Discrete Periodic Signals) Sketch the DTFT magnitude spectrum and phase spectrum of the following signals over $|F| \le 0.5$.

(a) $x[n] = \cos(0.5n\pi)$
(b) $x[n] = \cos(0.5n\pi) + \sin(0.25n\pi)$
(c) $x[n] = \cos(0.5n\pi)\cos(0.25n\pi)$

15.6 (System Representation) Find the transfer function $H(F)$ and the system difference equation for the following systems described by their impulse response $h[n]$.

(a) $h[n] = (\frac{1}{3})^n u[n]$ **(b)** $h[n] = [1 - (\frac{1}{3})^n]u[n]$
(c) $h[n] = n(\frac{1}{3})^n u[n]$ **(d)** $h[n] = 0.5\delta[n]$
(e) $h[n] = \delta[n] - (\frac{1}{3})^n u[n]$ **(f)** $h[n] = [(\frac{1}{3})^n + (\frac{1}{2})^n]u[n]$

15.7 (System Representation) Find the transfer function and impulse response of the following systems described by their difference equation.

(a) $y[n] + 0.4y[n-1] = 3x[n]$ **(b)** $y[n] - \frac{1}{6}y[n-1] - \frac{1}{6}y[n-2] = 2x[n] + x[n-1]$
(c) $y[n] = 0.2x[n]$ **(d)** $y[n] = x[n] + x[n-1] + x[n-2]$

PGR2026-00033
Krisp Technologies EX1024 Page 534

15.8 (System Representation) Set up the system difference equation for the following systems described by their transfer function.

(a) $H_p(F) = \dfrac{6e^{j2\pi F}}{3e^{j2\pi F}+1}$ **(b)** $H_p(F) = \dfrac{3}{e^{-j2\pi F}+2} - \dfrac{e^{-j2\pi F}}{e^{-j2\pi F}+3}$

(c) $H_p(F) = \dfrac{6}{1-0.3e^{-j2\pi F}}$ **(d)** $H_p(F) = \dfrac{6e^{j2\pi F}+4e^{j4\pi F}}{(1+2e^{j2\pi F})(1+4e^{j2\pi F})}$

15.9 (Steady-State Response) Consider the filter $y[n]+0.25y[n-2] = 2x[n]+2x[n-1]$. Find the filter transfer function $H(F)$ of this filter and use this to compute the steady-state response to the following inputs.

(a) $x[n] = 5u[n]$ **(b)** $x[n] = 3\cos(0.5n\pi + \frac{\pi}{4}) - 6\sin(0.5n\pi - \frac{\pi}{4})$

(c) $x[n] = 3\cos(0.5n\pi)u[n]$ **(d)** $x[n] = 2\cos(0.25n\pi) + 3\sin(0.5n\pi)$

15.10 (Response of Digital Filters) Consider the 3-point averaging filter described by the difference equation $y[n] = \frac{1}{3}(x[n]+x[n-1]+x[n-2])$.

(a) Find its impulse response $h[n]$.

(b) Find and sketch its frequency response $H(F)$.

(c) Find its response to $x[n] = \cos(\frac{n\pi}{3} + \frac{\pi}{4})$.

(d) Find its response to $x[n] = \cos(\frac{n\pi}{3} + \frac{\pi}{4}) + \sin(\frac{n\pi}{3} + \frac{\pi}{4})$.

(e) Find its response to $x[n] = \cos(\frac{n\pi}{3} + \frac{\pi}{4}) + \sin(\frac{2n\pi}{3} + \frac{\pi}{4})$.

15.11 (Frequency Response) Sketch the frequency response of the following digital filters and describe the function of each filter.

(a) $y[n]+0.9y[n-1] = x[n]$ **(b)** $y[n]-0.9y[n-1] = x[n]$

(c) $y[n]+0.9y[n-1] = x[n-1]$ **(d)** $y[n] = x[n]-x[n-4]$

15.12 (DTFT of Periodic Signals) Find the DTFT of the following periodic signals described over one period, with N samples per period.

(a) $x[n] = \{\overset{\Downarrow}{1},\ 0,\ 0,\ 0,\ 0\},\ N = 5$

(b) $x[n] = \{\overset{\Downarrow}{1},\ 0,\ -1,\ 0\},\ N = 4$

(c) $x[n] = \{\overset{\Downarrow}{3},\ 2,\ 1,\ 2\},\ N = 4$

(d) $x[n] = \{\overset{\Downarrow}{1},\ 2,\ 3\},\ N = 3$

15.13 (System Response to Periodic Signals) Consider the 3-point moving average filter described by $y[n] = x[n]+x[n-1]+x[n-2]$. Find its response to the following periodic inputs.

(a) $x[n] = \{\overset{\Downarrow}{1},\ 0,\ 0,\ 0,\ 0\},\ N = 5$

(b) $x[n] = \{\overset{\Downarrow}{1},\ 0,\ -1,\ 0\},\ N = 4$

(c) $x[n] = \{\overset{\Downarrow}{3},\ 2,\ 1\},\ N = 3$

(d) $x[n] = \{\overset{\Downarrow}{1},\ 2\},\ N = 2$

15.14 (Sampling, Filtering, and Aliasing) The sinusoid $x(t) = \cos(2\pi f_0 t)$ is sampled at 1 kHz to yield the sampled signal $x[n]$. The signal $x[n]$ is passed through a 2-point averaging filter whose difference equation is $y[n] = 0.5(x[n]+x[n-1])$. The filtered output $y[n]$ is reconstructed using an ideal lowpass

PGR2026-00033
Krisp Technologies EX1024 Page 535

filter with a cutoff frequency of 0.5 kHz to generate the analog signal $y(t)$. Find an expression for $y(t)$ if $f_0 = 0.2,\ 0.5,\ 0.75$ kHz

REVIEW AND EXPLORATION

15.15 (DTFT) Compute the DTFT of the following signals.

(a) $x[n] = \text{sinc}(0.2n)$ **(b)** $h[n] = \sin(0.2n\pi)$ **(c)** $g[n] = \text{sinc}^2(0.2n)$

15.16 (DTFT) Compute the DTFT of the following signals and plot their amplitude and phase spectra.

(a) $x[n] = \delta[n+1] + \delta[n-1]$ **(b)** $x[n] = \delta[n+1] - \delta[n-1]$
(c) $x[n] = \delta[n+1] + \delta[n] + \delta[n-1]$ **(d)** $x[n] = u[n+1] - u[n-1]$
(e) $x[n] = u[n+1] - u[n-2]$ **(f)** $x[n] = u[n] - u[n-3]$

15.17 (DTFT) Compute the DTFT of the following signals and sketch the magnitude and phase spectrum over $-0.5 \le F \le 0.5$.

(a) $x[n] = \cos(0.4n\pi)$ **(b)** $x[n] = \cos(0.2n\pi + \frac{\pi}{4})$
(c) $x[n] = \cos(n)$ **(d)** $x[n] = \cos(1.2n\pi + \frac{\pi}{4})$
(e) $x[n] = \cos(2.4n\pi)$ **(f)** $x[n] = \cos^2(2.4n\pi)$

15.18 (DTFT) Compute the DTFT of the following signals and sketch the magnitude and phase spectrum over $-0.5 \le F \le 0.5$.

(a) $x[n] = \text{sinc}(0.2n)$ **(b)** $x[n] = \text{sinc}(0.2n)\cos(0.4n\pi)$
(c) $x[n] = \text{sinc}^2(0.2n)$ **(d)** $x[n] = \text{sinc}(0.2n)\cos(0.1n\pi)$
(e) $x[n] = \text{sinc}^2(0.2n)\cos(0.4n\pi)$ **(f)** $x[n] = \text{sinc}^2(0.2n)\cos(0.2n\pi)$

15.19 (DTFT) Compute the DTFT of the following signals.

(a) $x[n] = 2^{-n}u[n]$ **(b)** $x[n] = 2^n u[-n]$
(c) $x[n] = 0.5^{|n|}$ **(d)** $x[n] = 0.5^{|n|}(u[n+1] - u[n-2])$
(e) $x[n] = (0.5)^n \cos(0.5n\pi)u[n]$ **(f)** $x[n] = \cos(0.5n\pi)(u[n+5] - u[n-6])$

15.20 (Properties) The DTFT of a real signal $x[n]$ is $X(F)$. How is the DTFT of the following signals related to $X(F)$?

(a) $y[n] = x[-n]$ **(b)** $g[n] = x[n] \star x[-n]$
(c) $r[n] = x[n/4]$ (zero interpolation) **(d)** $s[n] = (-1)^n x[n]$
(e) $h[n] = (j)^n x[n]$ **(f)** $v[n] = \cos(2n\pi F_0)x[n]$
(g) $w[n] = \cos(n\pi)x[n]$ **(h)** $z[n] = [1 + \cos(n\pi)]x[n]$
(i) $b[n] = (-1)^{n/2}x[n]$ **(j)** $p[n] = e^{jn\pi}x[n-1]$

15.21 (Properties) Let $x[n] = \text{tri}(0.2n)$ and let $X(F)$ be its DTFT. Compute the following without evaluating $X(F)$.

(a) The DTFT of the odd part of $x[n]$
(b) The value of $X(F)$ at $F = 0$ and $F = 0.5$
(c) The phase of $X(F)$
(d) The phase of the DTFT of $x[-n]$
(e) The integrals $\displaystyle\int_{-0.5}^{0.5} X(F)\,dF$ and $\displaystyle\int_{-0.5}^{0.5} X(F-0.5)\,dF$

PGR2026-00033
Krisp Technologies EX1024 Page 536

(f) The integrals $\int_{-0.5}^{0.5} |X(F)|^2\,dF$ and $\int_{-0.5}^{0.5} \left|\frac{dX(F)}{dF}\right|^2 dF$

(g) The derivative $\frac{dX(F)}{dF}$

15.22 (DTFT of Periodic Signals) Find the DTFT of the following periodic signals with N samples per period.

(a) $x[n] = \{\overset{\Downarrow}{1},\ 1,\ 1,\ 1,\ 1\},\ N = 5$ **(b)** $x[n] = \{\overset{\Downarrow}{1},\ 1,\ 1,\ 1\},\ N = 4$
(c) $x[n] = (-1)^n$ **(d)** $x[n] = 1$ (n even) and $x[n] = 0$ (n odd)

15.23 (IDTFT) Compute the IDTFT $x[n]$ of the following $X(F)$ described over $|F| \le 0.5$.

(a) $X(F) = \text{rect}(2F)$ **(b)** $X(F) = \cos(\pi F)$ **(c)** $X(F) = \text{tri}(2F)$

15.24 (Properties) Confirm that the DTFT of $x[n] = n\alpha^n u[n]$, using the following methods, gives identical results.

(a) From the defining relation for the DTFT
(b) From the DTFT of $y[n] = \alpha^n u[n]$ and the times-n property
(c) From the convolution result $\alpha^n u[n] \star \alpha^n u[n] = (n+1)\alpha^n u[n+1]$ and the shifting property
(d) From the convolution result $\alpha^n u[n] \star \alpha^n u[n] = (n+1)\alpha^n u[n+1]$ and superposition

15.25 (Properties) Find the DTFT of the following signals using the approach suggested.

(a) Starting with the DTFT of $u[n]$, show that the DTFT of $x[n] = \text{sgn}[n]$ is $X(F) = -j\cot(\pi F)$.
(b) Starting with $\text{rect}(n/2N) \Leftrightarrow (2N+1)\frac{\text{sinc}[(2N+1)F]}{\text{sinc}F}$, use the convolution property to find the DTFT of $x[n] = \text{tri}(n/N)$.
(c) Starting with $\text{rect}(n/2N) \Leftrightarrow (2N+1)\frac{\text{sinc}[(2N+1)F]}{\text{sinc}F}$, use the modulation property to find the DTFT of $x[n] = \cos(n\pi/2N)\text{rect}(n/2N)$.

15.26 (DTFT of Periodic Signals) One period of a periodic signal is given by $x[n] = \{\overset{\Downarrow}{1},\ -2,\ 0,\ 1\}$.

(a) Find the DTFT of this periodic signal.
(b) The signal is passed through a filter whose impulse response is $h[n] = \text{sinc}(0.8n)$. What is the filter output?

15.27 (Frequency Response) Consider a system whose frequency response $H(F)$ in magnitude/phase form is $H(F) = A(F)e^{j\phi(F)}$. Find the response $y[n]$ of this system for the following inputs.

(a) $x[n] = \delta[n]$ **(b)** $x[n] = 1$ **(c)** $x[n] = \cos(2n\pi F_0)$ **(d)** $x[n] = (-1)^n$

15.28 (Frequency Response) Consider a system whose frequency response is $H(F) = 2\cos(\pi F)e^{-j\pi F}$. Find the response $y[n]$ of this system for the following inputs.

(a) $x[n] = \delta[n]$ **(b)** $x[n] = \cos(0.5n\pi)$
(c) $x[n] = \cos(n\pi)$ **(d)** $x[n] = 1$
(e) $x[n] = e^{j0.4n\pi}$ **(f)** $x[n] = (j)^n$

15.29 (Frequency Response) The signal $x[n] = \{\overset{\Downarrow}{1},\ 0.5\}$ is applied to a system with frequency response $H(F)$, and the resulting output is $y[n] = \delta[n] - 2\delta[n-1] - \delta[n-2]$. What is $H(F)$?

PGR2026-00033
Krisp Technologies EX1024 Page 537

15.30 (Frequency Response) Consider the 2-point averager $y[n] = 0.5x[n] + 0.5x[n-1]$.

(a) Sketch its magnitude (or amplitude) and phase spectrum.
(b) Find its response to the input $x[n] = \cos(n\pi/2)$.
(c) Find its response to the input $x[n] = \delta[n]$.
(d) Find its response to the input $x[n] = 1$ and $x[n] = 3 + 2\delta[n] - 4\cos(n\pi/2)$.
(e) Show that its half-power frequency is given by $F_C = \dfrac{\cos^{-1}(\sqrt{0.5})}{\pi}$.
(f) What is the half-power frequency of a cascade of N such 2-point averagers?

15.31 (Frequency Response) Consider the 3-point averaging filter $h[n] = \frac{1}{3}\{1, \overset{\Downarrow}{1}, 1\}$.

(a) Sketch its magnitude (or amplitude) and phase spectrum. What type of filter is this?
(b) Find its phase delay and group delay. Is this a linear-phase filter?
(c) Find its response to the input $x[n] = \cos(n\pi/3)$.
(d) Find its response to the input $x[n] = \delta[n]$.
(e) Find its response to the input $x[n] = (-1)^n$.
(f) Find its response to the input $x[n] = 3 + 3\delta[n] - 6\cos(n\pi/3)$.

15.32 (Frequency Response) Consider a filter described by $h[n] = \frac{1}{3}\{1, \overset{\Downarrow}{-1}, 1\}$.

(a) Sketch its magnitude (or amplitude) and phase spectrum. What type of filter is this?
(b) Find its phase delay and group delay. Is this a linear-phase filter?
(c) Find its response to the input $x[n] = \cos(n\pi/3)$.
(d) Find its response to the input $x[n] = \delta[n]$.
(e) Find its response to the input $x[n] = (-1)^n$.
(f) Find its response to the input $x[n] = 3 + 3\delta[n] - 3\cos(2n\pi/3)$.

15.33 (Frequency Response) Consider the tapered 3-point averaging filter $h[n] = \{\overset{\Downarrow}{0.5}, 1, 0.5\}$.

(a) Sketch its magnitude (or amplitude) and phase spectrum. What type of filter is this?
(b) Find its phase delay and group delay. Is this a linear-phase filter?
(c) Find its response to the input $x[n] = \cos(n\pi/2)$.
(d) Find its response to the input $x[n] = \delta[n-1]$.
(e) Find its response to the input $x[n] = 1 + (-1)^n$.
(f) Find its response to the input $x[n] = 3 + 2\delta[n] - 4\cos(n\pi/2)$.

15.34 (Frequency Response) Consider the 2-point differencing filter $h[n] = \delta[n] - \delta[n-1]$.

(a) Sketch its magnitude (or amplitude) and phase spectrum. What type of filter is this?
(b) Find its phase and group delay. Is this a linear-phase filter?
(c) Find its response to the input $x[n] = \cos(n\pi/2)$.
(d) Find its response to the input $x[n] = u[n]$.
(e) Find its response to the input $x[n] = (-1)^n$.
(f) Find its response to the input $x[n] = 3 + 2u[n] - 4\cos(n\pi/2)$.

15.35 (Frequency Response) Consider the 5-point averaging filter $h[n] = \frac{1}{9}\{\overset{\Downarrow}{1}, 2, 3, 2, 1\}$.

(a) Sketch its magnitude (or amplitude) and phase spectrum. What type of filter is this?

PGR2026-00033
Krisp Technologies EX1024 Page 538

(b) Find its phase delay and group delay. Is this a linear-phase filter?
(c) Find its response to the input $x[n] = \cos(n\pi/4)$.
(d) Find its response to the input $x[n] = \delta[n]$.
(e) Find its response to the input $x[n] = (-1)^n$.
(f) Find its response to the input $x[n] = 9 + 9\delta[n] - 9\cos(n\pi/4)$.

15.36 (Frequency Response) Consider the recursive filter $y[n] + 0.5y[n-1] = 0.5x[n] + x[n-1]$.
(a) Sketch its magnitude (or amplitude) and phase spectrum. What type of filter is this?
(b) Find its phase delay and group delay. Is this a linear-phase filter?
(c) Find its response to the input $x[n] = \cos(n\pi/2)$.
(d) Find its response to the input $x[n] = \delta[n]$.
(e) Find its response to the input $x[n] = (-1)^n$.
(f) Find its response to the input $x[n] = 4 + 2\delta[n] - 4\cos(n\pi/2)$.

15.37 (System Response to Periodic Signals) Consider the filter $y[n] - 0.5y[n-1] = 3x[n-1]$. Find its response to the following periodic inputs.
(a) $x[n] = \{\overset{\Downarrow}{4},\ 0,\ 2,\ 1\},\ N = 4$
(b) $x[n] = (-1)^n$
(c) $x[n] = 1$ (n even) and $x[n] = 0$ (n odd)
(d) $x[n] = 14(0.5)^n(u[n] - u[n-N]),\ N = 3$

15.38 (System Response to Periodic Signals) Find the response if a periodic signal whose one period is $x[n] = \{\overset{\Downarrow}{4},\ 3,\ 2,\ 3\}$ (with period $N = 4$) is applied to the following filters.
(a) $h[n] = \text{sinc}(0.4n)$
(b) $H(F) = \text{tri}(2F)$
(c) $y[n] = x[n] + x[n-1] + x[n-2] + x[n-3]$

15.39 (Frequency Response) Consider the filter realization of Figure P15.39. Find the frequency response $H(F)$ of the overall system if the impulse response $h_1[n]$ of the filter in the forward path is given by

(a) $h_1[n] = \delta[n] - \delta[n-1]$. **(b)** $h_1[n] = 0.5\delta[n] + 0.5\delta[n-1]$.

Find the difference equation relating $y[n]$ and $x[n]$ and check the system stability.

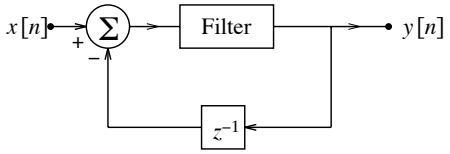


Figure P15.39 Filter realization for Problem 15.39

15.40 (Interconnected Systems) Consider two systems with impulse response $h_1[n] = \delta[n] + \alpha\delta[n-1]$ and $h_2[n] = (0.5)^n u[n]$. Find the frequency response and impulse response of the combination if
(a) The two systems are connected in parallel with $\alpha = 0.5$.
(b) The two systems are connected in parallel with $\alpha = -0.5$.
(c) The two systems are connected in cascade with $\alpha = 0.5$.
(d) The two systems are connected in cascade with $\alpha = -0.5$.

PGR2026-00033
Krisp Technologies EX1024 Page 539

15.41 (Interconnected Systems) Consider two systems with impulse response $h_1[n] = \delta[n] + \alpha\delta[n-1]$ and $h_2[n] = (0.5)^n u[n]$. Find the response $y[n]$ of the overall system to the input $x[n] = (0.5)^n u[n]$ if

(a) The two systems are connected in parallel with $\alpha = 0.5$.
(b) The two systems are connected in parallel with $\alpha = -0.5$.
(c) The two systems are connected in cascade with $\alpha = 0.5$.
(d) The two systems are connected in cascade with $\alpha = -0.5$.

15.42 (Interconnected Systems) Consider two systems with impulse response $h_1[n] = \delta[n] + \alpha\delta[n-1]$ and $h_2[n] = (0.5)^n u[n]$. Find the response of the overall system to the input $x[n] = \cos(0.5n\pi)$ if

(a) The two systems are connected in parallel with $\alpha = 0.5$.
(b) The two systems are connected in parallel with $\alpha = -0.5$.
(c) The two systems are connected in cascade with $\alpha = 0.5$.
(d) The two systems are connected in cascade with $\alpha = -0.5$.

15.43 (The Fourier Transform and DTFT) The analog signal $x(t) = e^{-\alpha t}u(t)$ is ideally sampled at the sampling rate S to yield the analog impulse train $x_i(t) = \sum_{k=-\infty}^{\infty} x(t)\delta(t - kt_s)$ where $t_s = 1/S$. The sampled version of $x(t)$ corresponds to $x[n] = e^{-\alpha n t_s}u[n]$.

(a) Find the Fourier transforms $X(f)$ and $X_i(f)$, and the DTFT $X_p(F)$.
(b) Show that $X_p(F)$ and $X_i(f)$ are identical.
(c) How is $X_p(F)$ related to $X(f)$?

15.44 (Filter Concepts) Let $H_1(F)$ be an ideal lowpass filter with a cutoff frequency of $F_1 = 0.2$ and let $H_2(F)$ be an ideal lowpass filter with a cutoff frequency of $F_2 = 0.4$. Make a block diagram of how you would connect these filters (in cascade and/or parallel) to implement the following filters and express their transfer function in terms of $H_1(F)$ and/or $H_2(F)$.

(a) An ideal highpass filter with a cutoff frequency of $F_C = 0.2$
(b) An ideal highpass filter with a cutoff frequency of $F_C = 0.4$
(c) An ideal bandpass filter with a passband covering $0.2 \le F \le 0.4$
(d) An ideal bandstop filter with a stopband covering $0.2 \le F \le 0.4$

15.45 (Linear-Phase Filters) Consider a filter whose impulse response is $h[n] = \{\overset{\Downarrow}{\alpha},\ \beta,\ \alpha\}$.

(a) Find its frequency response in amplitude/phase form $H(F) = A(F)e^{j\phi(F)}$ and confirm that the phase $\phi(F)$ varies linearly with F.
(b) Determine the values of α and β if this filter is to completely block a signal at the frequency $F = \frac{1}{3}$ while passing a signal at the frequency $F = \frac{1}{8}$ with unit gain.
(c) What will be the output of the filter in part(b) if the input is $x[n] = \cos(0.5n\pi)$?

15.46 (Echo Cancellation) A microphone, whose frequency response is limited to 5 kHz, picks up not only a desired signal $x(t)$ but also its echoes. However, only the first echo, arriving after 1.25 ms, has a significant amplitude (of 0.5) relative to the desired signal $x(t)$.

(a) Set up an equation relating the analog output and analog input of the microphone.
(b) The microphone signal is to be processed digitally in an effort to remove the echo. Set up a difference equation relating the sampled output and sampled input of the microphone using an arbitrary sampling rate S.

PGR2026-00033
Krisp Technologies EX1024 Page 540

(c) Argue that if S equals the Nyquist rate, the difference equation for the microphone will contain fractional delays and be difficult to implement.

(d) If the sampling rate S can be varied only in steps of 1 kHz, choose the smallest S that will ensure integer delays and find the difference equation of the microphone. Use this to set up the difference equation of an echo-canceling system that can recover the original signal.

(e) Sketch the frequency response of each system of the previous part. What is the difference equation, transfer function, and frequency response of the overall system?

15.47 (Sampling and Aliasing) A speech signal $x(t)$ band-limited to 4 kHz is sampled at 10 kHz to obtain $x[n]$. The sampled signal $x[n]$ is filtered by an ideal bandpass filter whose passband extends over $0.03 \leq F \leq 0.3$ to obtain $y[n]$. Will the sampled output $y[n]$ be contaminated if $x(t)$ also includes an undesired signal at the following frequencies?

(a) 60 Hz **(b)** 360 Hz **(c)** 8.8 kHz **(d)** 9.8 kHz

15.48 (DTFT and Sampling) The transfer function of a digital filter is $H(F) = \text{rect}(2F)e^{-j0.5\pi F}$.

(a) What is the impulse response $h[n]$ of this filter?

(b) The analog signal $x(t) = \cos(0.25\pi t)$ applied to the system shown.

$$x(t) \longrightarrow \boxed{\text{sampler}} \longrightarrow \boxed{H(F)} \longrightarrow \boxed{\text{ideal LPF}} \longrightarrow y(t)$$

The sampler is ideal and operates at $S = 1$ Hz. The cutoff frequency of the ideal lowpass filter is $f_C = 0.5$ Hz. How are $X(f)$ and $Y(f)$ related? What is $y(t)$?

15.49 (Sampling and Filtering) The analog signal $x(t) = \cos(2\pi f_0 t)$ is applied to the following system:

$$x(t) \longrightarrow \boxed{\text{sampler}} \longrightarrow \boxed{H(F)} \longrightarrow \boxed{\text{ideal LPF}} \longrightarrow y(t)$$

The sampler is ideal and operates at $S = 60$ Hz. The filter $H(F)$ describes the transfer function of a 3-point averager. The cutoff frequency of the ideal lowpass filter is $f_C = 40$ Hz. Find $y(t)$ if

(a) $f_0 = 20$ Hz. **(b)** $f_0 = 50$ Hz. **(c)** $f_0 = 75$ Hz.

15.50 (Sampling and Filtering) A periodic signal whose one full period is $x(t) = 5\text{tri}(20t)$ is applied to the following system:

$$x(t) \longrightarrow \boxed{\text{pre-filter}} \longrightarrow \boxed{\text{sampler}} \longrightarrow \boxed{H(F)} \longrightarrow \boxed{\text{ideal LPF}} \longrightarrow y(t)$$

The pre-filter band-limits the signal $x(t)$ to B Hz. The sampler is ideal and operates at $S = 80$ Hz. The transfer function of the digital filter is $H(F) = \text{tri}(2F)$. The cutoff frequency of the ideal lowpass filter is $f_C = 40$ Hz. Find $y(t)$ if

(a) $B = 20$ Hz. **(b)** $B = 40$ Hz. **(c)** $B = 80$ Hz.

15.51 (Sampling and Filtering) A speech signal $x(t)$ whose spectrum extends to 5 kHz is filtered by an ideal analog lowpass filter with a cutoff frequency of 4 kHz to obtain the filtered signal $x_f(t)$.

(a) If $x_f(t)$ is to be sampled to generate $y[n]$, what is the minimum sampling rate that will permit recovery of $x_f(t)$ from $y[n]$?

(b) If $x(t)$ is first sampled to obtain $x[n]$ and then digitally filtered to obtain $z[n]$, what must be the minimum sampling rate and the impulse response $h[n]$ of the digital filter such that $z[n]$ is identical to $y[n]$?

PGR2026-00033
Krisp Technologies EX1024 Page 541

15.52 (DTFT and Sampling) A signal is reconstructed using a filter that performs step interpolation between samples. The reconstruction sampling interval is t_s.

(a) What is the impulse response $h(t)$ and transfer function $H(f)$ of the interpolating filter?

(b) What is the transfer function $H_C(f)$ of a filter that can compensate for the non-ideal reconstruction?

15.53 (DTFT and Sampling) A signal is reconstructed using a filter that performs linear interpolation between samples. The reconstruction sampling interval is t_s.

(a) What is the impulse response $h(t)$ and transfer function $H(f)$ of the interpolating filter?

(b) What is the transfer function $H_C(f)$ of a filter that can compensate for the non-ideal reconstruction?

15.54 (Response of Numerical Algorithms) Simpson's and Tick's rules for numerical integration find $y[k]$ (the approximation to the area) over two time steps from $y[k-2]$ and are described by

$$\textbf{Simpson's rule:} \quad y[n] = y[n-2] + \frac{x[n] + 4x[n-1] + x[n-2]}{3}$$

$$\textbf{Tick's rule:} \quad y[n] = y[n-2] + 0.3584x[n] + 1.2832x[n-1] + 0.3584x[n-2]$$

(a) Find the transfer function $H(F)$ corresponding to each rule.

(b) For each rule, sketch $|H(F)|$ over $0 \leq F \leq 0.5$ and compare with the spectrum of an ideal integrator.

(c) It is claimed that the coefficients in Tick's rule optimize $H(F)$ in the range $0 < F < 0.25$. Does your comparison support this claim?

15.55 (Sampling and Filtering) The analog signal $x(t) = \text{sinc}^2(10t)$ is applied to the following system:

$$x(t) \longrightarrow \boxed{\text{sampler}} \longrightarrow \boxed{H(F)} \longrightarrow \boxed{\text{reconstruction}} \longrightarrow y(t)$$

The sampler is ideal and operates at a sampling rate S that corresponds to the Nyquist rate. The transfer function of the digital filter is $H(F) = \text{rect}(2F)$. Sketch the spectra at the output of each subsystem if the reconstruction is based on

(a) An ideal analog filter whose cutoff frequency is $f_C = 0.5S$.

(b) A zero-order-hold followed by an ideal analog filter with $f_C = 0.5S$.

15.56 (Modulation) A signal $x[n]$ is modulated by $\cos(0.5n\pi)$ to obtain the signal $x_1[n]$. The modulated signal $x_1[n]$ is filtered by a filter whose transfer function is $H(F)$ to obtain the signal $y_1[n]$.

(a) Sketch the spectra $X(F)$, $X_1(F)$, and $Y_1(F)$ if $X(F) = \text{tri}(4F)$ and $H(F) = \text{rect}(2F)$.

(b) The signal $y_1[n]$ is modulated again by $\cos(0.5n\pi)$ to obtain the signal $y_2[n]$, and filtered by $H(F)$ to obtain $y[n]$. Sketch $Y_2(F)$ and $Y(F)$.

(c) Are the signals $x[n]$ and $y[n]$ related in any way?

15.57 (Sampling) A speech signal band-limited to 4 kHz is ideally sampled at a sampling rate S, and the sampled signal $x[n]$ is processed by the squaring filter $y[n] = x^2[n]$ whose output is ideally reconstructed to obtain $y(t)$ as follows:

$$x(t) \longrightarrow \boxed{\text{sampler}} \longrightarrow \boxed{y[n] = x^2[n]} \longrightarrow \boxed{\text{ideal reconstruction}} \longrightarrow y(t)$$

What minimum sampling rate S will ensure that $y(t) = x^2(t)$?

PGR2026-00033
Krisp Technologies EX1024 Page 542

15.58 (Sampling) Consider the signal $x(t) = \sin(200\pi t)\cos(120\pi t)$. This signal is sampled at a sampling rate S_1, and the sampled signal $x[n]$ is ideally reconstructed at a sampling rate S_2 to obtain $y(t)$. What is the reconstructed signal if

(a) $S_1 = 400$ Hz and $S_2 = 200$ Hz.
(b) $S_1 = 200$ Hz and $S_2 = 100$ Hz.
(c) $S_1 = 120$ Hz and $S_2 = 120$ Hz.

15.59 (Up-Sampling) The analog signal $x(t) = 4000\text{sinc}^2(4000t)$ is sampled at 12 kHz to obtain the signal $x[n]$. The sampled signal is up-sampled (zero interpolated) by N to obtain the signal $y[n]$ as follows:

$$x(t) \longrightarrow \boxed{\text{sampler}} \rightarrow x[n] \rightarrow \boxed{\text{up-sample} \uparrow N} \longrightarrow y[n]$$

Sketch the spectra of $X(F)$ and $Y(F)$ over $-1 \leq F \leq 1$ for $N = 2$ and $N = 3$.

15.60 (Up-Sampling) The signal $x[n]$ is up-sampled (zero interpolated) by N to obtain the signal $y[n]$. Sketch $X(F)$ and $Y(F)$ over $-1 \leq F \leq 1$ for the following cases.

(a) $x[n] = \text{sinc}(0.4n),\ N = 2$
(b) $X(F) = \text{tri}(4F),\ N = 2$
(c) $X(F) = \text{tri}(6F),\ N = 3$

15.61 (Linear Interpolation) Consider a system that performs linear interpolation by a factor of N. One way to construct such a system is to perform up-sampling by N (zero interpolation between signal samples) and pass the up-sampled signal through an interpolating filter with impulse response $h[n]$ whose output is the linearly interpolated signal $y[n]$ as shown.

$$x[n] \longrightarrow \boxed{\text{up-sample} \uparrow N} \longrightarrow \boxed{\text{filter } h[n]} \longrightarrow y[n]$$

(a) What impulse response $h[n]$ will result in linear interpolation by a factor of $N = 2$?
(b) Sketch the frequency response $H(F)$ of the interpolating filter for $N = 2$.
(c) Let the input to the system be $X(F) = \text{rect}(2F)$. Sketch the spectrum at the output of the up-sampler and the interpolating filter.

15.62 (Interpolation) The input $x[n]$ is applied to a system that up-samples by N followed by an ideal lowpass filter with a cutoff frequency of F_C to generate the output $y[n]$. Draw a block diagram of the system. Sketch the spectra at various points of this system and find $y[n]$ and $Y(F)$ for the following cases.

(a) $x[n] = \text{sinc}(0.4n),\ N = 2,\ F_C = 0.4$
(b) $X(F) = \text{tri}(4F),\ N = 2,\ F_C = 0.375$

15.63 (Decimation) The signal $x(t) = 2\cos(100\pi t)$ is ideally sampled at 400 Hz to obtain the signal $x[n]$, and the sampled signal is decimated by N to obtain the signal $y[n]$. Sketch the spectra $X(F)$ and $Y(F)$ over $-1 \leq F \leq 1$ for the cases $N = 2$ and $N = 3$.

15.64 (Decimation) The signal $x[n]$ is decimated by N to obtain the decimated signal $y[n]$. Sketch $X(F)$ and $Y(F)$ over $-1 \leq F \leq 1$ for the following cases.

(a) $x[n] = \text{sinc}(0.4n),\ N = 2$
(b) $X(F) = \text{tri}(4F),\ N = 2$
(c) $X(F) = \text{tri}(3F),\ N = 2$

PGR2026-00033
Krisp Technologies EX1024 Page 543

15.65 (Interpolation and Decimation) Consider the following system:

$$x[n] \longrightarrow \boxed{\text{up-sample} \uparrow N} \longrightarrow \boxed{\text{digital LPF}} \longrightarrow \boxed{\text{down-sample} \downarrow M} \longrightarrow y[n]$$

The signal $x[n]$ is up-sampled (zero-interpolated) by $N = 2$, the digital lowpass filter is ideal and has a cutoff frequency of F_C, and down-sampling is by $M = 3$. Sketch $X(F)$ and $Y(F)$ and explain how they are related for the following cases.

(a) $X(F) = \text{tri}(4F)$ and $F_C = 0.125$ **(b)** $X(F) = \text{tri}(2F)$ and $F_C = 0.25$

15.66 (Interpolation and Decimation) For each of the following systems, $X(F) = \text{tri}(4F)$. The digital lowpass filter is ideal and has a cutoff frequency of $F_C = 0.25$ and a gain of 2. Sketch the spectra at the various points over $-1 \le F \le 1$ and determine whether any systems produce identical outputs.

(a) $x[n] \longrightarrow \boxed{\text{up-sample} \uparrow N = 2} \longrightarrow \boxed{\text{digital LPF}} \longrightarrow \boxed{\text{down-sample} \downarrow M = 2} \longrightarrow y[n]$

(b) $x[n] \longrightarrow \boxed{\text{down-sample} \downarrow M = 2} \longrightarrow \boxed{\text{digital LPF}} \longrightarrow \boxed{\text{up-sample} \uparrow N = 2} \longrightarrow y[n]$

(c) $x[n] \longrightarrow \boxed{\text{down-sample} \downarrow M = 2} \longrightarrow \boxed{\text{up-sample} \uparrow N = 2} \longrightarrow \boxed{\text{digital LPF}} \longrightarrow y[n]$

15.67 (Interpolation and Decimation) For each of the following systems, $X(F) = \text{tri}(3F)$. The digital lowpass filter is ideal and has a cutoff frequency of $F_C = \frac{1}{3}$ and a gain of 2. Sketch the spectra at the various points over $-1 \le F \le 1$ and determine whether any systems produce identical outputs.

(a) $x[n] \longrightarrow \boxed{\text{up-sample} \uparrow N = 2} \longrightarrow \boxed{\text{digital LPF}} \longrightarrow \boxed{\text{down-sample} \downarrow M = 2} \longrightarrow y[n]$

(b) $x[n] \longrightarrow \boxed{\text{down-sample} \downarrow M = 2} \longrightarrow \boxed{\text{digital LPF}} \longrightarrow \boxed{\text{up-sample} \uparrow N = 2} \longrightarrow y[n]$

(c) $x[n] \longrightarrow \boxed{\text{down-sample} \downarrow M = 2} \longrightarrow \boxed{\text{up-sample} \uparrow N = 2} \longrightarrow \boxed{\text{digital LPF}} \longrightarrow y[n]$

15.68 (Interpolation and Decimation) You are asked to investigate the claim that interpolation by N and decimation by N performed in any order, as shown, will recover the original signal.

Method 1: $x[n] \longrightarrow \boxed{\text{up-sample} \uparrow N} \longrightarrow \boxed{\text{digital LPF}} \longrightarrow \boxed{\text{down-sample} \downarrow N} \longrightarrow y[n]$

Method 2: $x[n] \longrightarrow \boxed{\text{down-sample} \downarrow N} \longrightarrow \boxed{\text{up-sample} \uparrow N} \longrightarrow \boxed{\text{digital LPF}} \longrightarrow y[n]$

(a) Let $X(F) = \text{tri}(4F)$ and $N = 2$. Let the lowpass filter have a cutoff frequency of $F_C = 0.25$ and a gain of 2. Sketch the spectra over $-1 \le F \le 1$ at the various points. For which method does $y[n]$ equal $x[n]$? Do the results justify the claim?

(b) Let $X(F) = \text{tri}(3F)$ and $N = 2$. Let the lowpass filter have a cutoff frequency of $F_C = \frac{1}{3}$ and a gain of 2. Sketch the spectra over $-1 \le F \le 1$ at the various points. For which method does $y[n]$ equal $x[n]$? Do the results justify the claim?

(c) Are any restrictions necessary on the input for $x[n]$ to equal $y[n]$ in each method? Explain.

15.69 (Fractional Delay) The following system is claimed to implement a half-sample delay:

$$x(t) \longrightarrow \boxed{\text{sampler}} \longrightarrow \boxed{H(F)} \longrightarrow \boxed{\text{ideal LPF}} \longrightarrow y(t)$$

The signal $x(t)$ is band-limited to f_C, and the sampler is ideal and operates at the Nyquist rate. The digital filter is described by $H_1(F) = e^{-j\pi F}$, $|F| \le F_C$, and the cutoff frequency of the ideal lowpass filter is f_C.

(a) Sketch the magnitude and phase spectra at the various points in this system.

(b) Show that $y(t) = x(t - 0.5t_s)$ (corresponding to a half-sample delay).

PGR2026-00033
Krisp Technologies EX1024 Page 544

15.70 (Fractional Delay) In practice, the signal $y[n] = x[n - 0.5]$ may be generated from $x[n]$ using interpolation by 2 (to give $x[0.5n]$) followed by a one-sample delay (to give $x[0.5(n-1)]$) and decimation by 2 (to give $x[n - 0.5]$) as follows:

$$x[n] \longrightarrow \boxed{\text{up-sample} \uparrow 2} \longrightarrow \boxed{\text{digital LPF}} \longrightarrow \boxed{\text{1-sample delay}} \longrightarrow \boxed{\text{down-sample} \downarrow 2} \longrightarrow y[n]$$

Let $X(F) = \text{tri}(4F)$. Sketch the magnitude and phase spectra at the various points and show that $Y(F) = X(F)e^{-j\pi F}$ (implying a half-sample delay).

15.71 (Group Delay) Show that the group delay t_g of a filter described by its transfer function $H(F)$ may be expressed as

$$t_g = \frac{H_R'(F)H_I(F) - H_I'(F)H_R(F)}{2\pi|H(F)|^2}$$

Here, the primed quantities describe derivatives with respect to F. For an FIR filter with impulse response $h[n]$, these primed quantities are easily found by recognizing that $H_R'(F) = \text{Re}\{H'(F)\}$ and $H_I'(F) = \text{Im}\{H'(F)\}$ where,

$$H(F) = \sum_{k=0}^{N} h[k]e^{-j2\pi kF} \qquad H'(F) = -j2\pi \sum_{k=0}^{N} kh[k]e^{-j2\pi kF}$$

For an IIR filter described by $H(F) = N(F)/D(F)$, the overall group delay may be found as the difference of the group delays of $N(F)$ and $D(F)$. Use these results to find the group delay of

(a) $h[n] = \{\overset{\Downarrow}{\alpha},\ 1\}$. **(b)** $H(F) = 1 + \alpha e^{-j2\pi F}$. **(c)** $H(F) = \dfrac{\alpha + e^{-j2\pi F}}{1 + \alpha e^{-j2\pi F}}$.

COMPUTATION AND DESIGN

dfreqgui A GUI for Visualization of Frequency Response

The graphical user interface `dfreqgui` allows you to visualize the frequency response.
You can display results against F or Ω, using several options.
To explore this routine, type `dfreqgui` at the MATLAB prompt.

15.72 (Interconnected Systems) The signal $x[n] = \cos(2\pi F_0 n)$ forms the input to a cascade of two systems whose impulse response is described by $h_1[n] = \{\overset{\Downarrow}{0.25},\ 0.5,\ 0.25\}$ and $h_2[n] = \delta[n] - \delta[n-1]$.

(a) Generate and plot $x[n]$ over $0 \le n \le 100$ if $F_0 = 0.1$. Can you identify its period from the plot?

(b) Let the output of the first system be fed to the second system. Plot the output of the each system. Are the two outputs periodic? If so, do they have the same period? Explain.

(c) Reverse the order of cascading and plot the output of the each system. Are the two outputs periodic? If so, do they have the same period? Does the order of cascading alter the overall response? Should it?

(d) Let the first system be described by $y[n] = x^2[n]$. Find and plot the output of each system in the cascade. Repeat after reversing the order of cascading. Does the order of cascading alter the intermediate response? Does the order of cascading alter the overall response? Should it?

PGR2026-00033
Krisp Technologies EX1024 Page 545

15.73 (Frequency Response) This problem deals with the cascade and parallel connection of two FIR filters whose impulse response is given by

$$h_1[n] = \{\overset{\Downarrow}{1},\ 2,\ 1\} \qquad h_2[n] = \{\overset{\Downarrow}{2},\ 0,\ -2\}$$

(a) Plot the frequency response of each filter and identify the filter type.

(b) The frequency response of the parallel connection of $h_1[n]$ and $h_2[n]$ is $H_{P1}(F)$. If the second filter is delayed by one sample and then connected in parallel with the first, the frequency response changes to $H_{P2}(F)$. It is claimed that $H_{P1}(F)$ and $H_{P2}(F)$ have the same magnitude and differ only in phase. Use MATLAB to argue for or against this claim.

(c) Obtain the impulse response $h_{P1}[n]$ and $h_{P2}[n]$ and plot their frequency response. Use MATLAB to compare their magnitude and phase. Do the results justify your argument? What type of filters do $h_{P1}[n]$ and $h_{P2}[n]$ describe?

(d) The frequency response of the cascade of $h_1[n]$ and $h_2[n]$ is $H_{C1}(F)$. If the second filter is delayed by one sample and then cascaded with the first, the frequency response changes to $H_{C2}(F)$. It is claimed that $H_{C1}(F)$ and $H_{C2}(F)$ have the same magnitude and differ only in phase. Use MATLAB to argue for or against this claim.

(e) Obtain the impulse response $h_{C1}[n]$ and $h_{C2}[n]$ and plot their frequency response. Use MATLAB to compare their magnitude and phase. Do the results justify your argument? What type of filters do $h_{C1}[n]$ and $h_{C2}[n]$ represent?

15.74 (Nonrecursive Forms of IIR Filters) We can only approximately represent an IIR filter by an FIR filter by truncating its impulse response to N terms. The larger the truncation index N, the better is the approximation. Consider the IIR filter described by $y[n] - 0.8y[n-1] = x[n]$.

(a) Find its impulse response $h[n]$.

(b) Truncate $h[n]$ to three terms to obtain $h_N[n]$. Plot the frequency response $H(F)$ and $H_N(F)$. What differences do you observe?

(c) Truncate $h[n]$ to ten terms to obtain $h_N[n]$. Plot the frequency response $H(F)$ and $H_N(F)$. What differences do you observe?

(d) If the input to the original filter and truncated filter is $x[n]$, will the greatest mismatch in the response $y[n]$ of the two filters occur at earlier or later time instants n?

15.75 (Compensating Filters) Digital filters are often used to compensate for the sinc distortion of a zero-order-hold DAC by providing a $1/\text{sinc}(F)$ boost. Two such filters are described by

Compensating Filter 1: $y[n] = \frac{1}{16}(x[n] - 18x[n-1] + x[n-2])$

Compensating Filter 2: $y[n] + \frac{1}{8}y[n-1] = \frac{9}{8}x[n]$

(a) For each filter, state whether it is FIR (and if so, linear phase) or IIR.

(b) Plot the frequency response of each filter and compare with $|1/\text{sinc}(F)|$.

(c) Over what digital frequency range does each filter provide the required sinc boost? Which of these filters provides better compensation?

15.76 (Up-Sampling and Decimation) Let $x[n] = \cos(0.2n\pi) + 0.5\cos(0.4n\pi),\ 0 \le n \le 100$.

(a) Plot the spectrum of this signal.

(b) Generate the zero-interpolated signal $y[n] = x[n/2]$ and plot its spectrum. Can you observe the spectrum replication? Is there a correspondence between the frequencies in $y[n]$ and $x[n]$? Should there be? Explain.

PGR2026-00033
Krisp Technologies EX1024 Page 546

(c) Generate the decimated signal $d[n] = x[2n]$ and plot its spectrum. Can you observe the stretching effect in the spectrum? Is there a correspondence between the frequencies in $d[n]$ and $x[n]$? Should there be? Explain.

(d) Generate the decimated signal $g[n] = x[3n]$ and plot its spectrum. Can you observe the stretching effect in the spectrum? Is there a correspondence between the frequencies in $g[n]$ and $x[n]$? Should there be? Explain.

15.77 (Frequency Response of Interpolating Functions) The impulse response of filters for step interpolation, linear interpolation, and ideal (sinc) interpolation by N are given by

$$h_S[n] = u[n] - u[n-(N-1)] \qquad h_L[n] = \text{tri}(n/N) \qquad h_I[n] = \text{sinc}(n/N)$$

Note that the ideal interpolating function is of infinite length.

(a) Plot the frequency response of each interpolating function for $N = 4$ and $N = 8$.

(b) How does the response of the step-interpolation and linear-interpolation schemes compare with ideal interpolation?

15.78 (Interpolating Functions) To interpolate a signal $x[n]$ by N, we use an up-sampler (that places $N-1$ zeros after each sample) followed by a filter that performs the appropriate interpolation. The filter impulse response for step interpolation, linear interpolation, and ideal (sinc) interpolation is chosen as

$$h_S[n] = u[n] - u[n-(N-1)] \qquad h_L[n] = \text{tri}(n/N) \qquad h_I[n] = \text{sinc}(n/N),\ |n| \le M$$

Note that the ideal interpolating function is actually of infinite length and must be truncated in practice. Generate the test signal $x[n] = \cos(0.5n\pi),\ 0 \le n \le 3$. Up-sample this by $N = 8$ (seven zeros after each sample) to obtain the signal $x_U[n]$. Use the MATLAB routine `filter` to filter $x_U[n]$ as follows:

(a) Use the step-interpolation filter to obtain the filtered signal $x_S[n]$. Plot $x_U[n]$ and $x_S[n]$ on the same plot. Does the system perform the required interpolation? Does the result look like a sine wave?

(b) Use the step-interpolation filter followed by the compensating filter $y[n] = \{x[n] - 18x[n-1] + x[n-2]\}/16$ to obtain the filtered signal $x_C[n]$. Plot $x_U[n]$ and $x_C[n]$ on the same plot. Does the system perform the required interpolation? Does the result look like a sine wave? Is there an improvement compared to part (a)?

(c) Use the linear-interpolation filter to obtain the filtered signal $x_L[n]$. Plot $x_U[n]$ and a delayed (by 8) version of $x_L[n]$ (to account for the noncausal nature of $h_L[n]$) on the same plot. Does the system perform the required interpolation? Does the result look like a sine wave?

(d) Use the ideal interpolation filter (with $M = 4,\ 8,\ 16$) to obtain the filtered signal $x_I[n]$. Plot $x_U[n]$ and a delayed (by M) version of $x_I[n]$ (to account for the noncausal nature of $h_I[n]$) on the same plot. Does the system perform the required interpolation? Does the result look like a sine wave? What is the effect of increasing M on the interpolated signal? Explain.

15.79 (FIR Filter Design) A 22.5-Hz sinusoid is contaminated by 60-Hz interference. We wish to sample this signal and design a causal 3-point linear-phase FIR digital filter operating at a sampling frequency of $S = 180$ Hz to eliminate the interference and pass the desired signal with unit gain.

(a) Argue that an impulse response of the form $h[n] = \{\overset{\Downarrow}{\alpha},\ \beta,\ \alpha\}$ can be used. Choose α and β to satisfy the design requirements.

PGR2026-00033
Krisp Technologies EX1024 Page 547

(b) To test your filter, generate two signals $x[n]$ and $s[n]$, $0 \le n \le 50$, by sampling $x(t) = \cos(45\pi t)$ and $s(t) = 3\cos(120\pi t)$ at 180 Hz. Generate the noisy signal $g[n] = x[n] + s[n]$ and pass it through your filter to obtain the filtered signal $y[n]$. Compare $y[n]$ with the noisy signal $g[n]$ and the desired signal $x[n]$ to confirm that the filter meets design requirements. What is the phase of $y[n]$ at the desired frequency? Can you find an exact expression for $y[n]$?

15.80 (Allpass Filters) Consider a lowpass filter with impulse response $h[n] = (0.5)^n u[n]$. The input to this filter is $x[n] = \cos(0.2n\pi)$. We expect the output to be of the form $y[n] = A\cos(0.2n\pi + \theta)$.

(a) Find the values of A and θ.

(b) What should be the transfer function $H_1(F)$ of a first-order allpass filter that can be cascaded with the lowpass filter to correct for the phase distortion and produce the signal $z[n] = B\cos(0.2n\pi)$ at its output?

(c) What should be the gain of the allpass filter in order that $z[n] = x[n]$?

15.81 (Frequency Response of Averaging Filters) The averaging of data uses both FIR and IIR filters. Consider the following averaging filters:

Filter 1: $y[n] = \dfrac{1}{N}\displaystyle\sum_{k=0}^{N-1} x[n-k]$ (N-point moving average)

Filter 2: $y[n] = \dfrac{2}{N(N+1)}\displaystyle\sum_{k=0}^{N-1}(N-k)x[n-k]$ (N-point weighted moving average)

Filter 3: $y[n] - \alpha y[n-1] = (1-\alpha)x[n], \quad \alpha = \frac{N-1}{N+1}$ (first-order exponential average)

(a) Confirm that the dc gain of each filter is unity. Which of these are FIR filters?

(b) Sketch the frequency response magnitude of each filter with $N = 4$ and $N = 9$. How will the choice of N affect the averaging?

(c) To test your filters, generate the signal $x[n] = 1-(0.6)^n$, $0 \le n \le 300$, add some noise, and apply the noisy signal to each averager and compare the results. Which filter would you recommend?

PGR2026-00033
Krisp Technologies EX1024 Page 548

Chapter 16

THE DFT AND FFT

16.0 Scope and Objectives

The Fourier series describes periodic signals by discrete spectra, whereas the DTFT describes discrete signals by periodic spectra. These results are a consequence of the fact that sampling in one domain induces periodic extension in the other. As a result, signals that are both discrete and periodic in one domain are also periodic and discrete in the other. This is the basis for the formulation of the *discrete Fourier transform* (DFT). Computationally efficient algorithms for implementing the DFT go by the generic name of *fast Fourier transforms* (FFTs). This chapter describes the DFT and its properties and develops its relationship to the Fourier series, Fourier transform, and the DTFT. It describes applications of the DFT to signal processing and concludes with a discussion of FFT algorithms for computing the DFT and its inverse.

16.1 Introduction

The N-point **discrete Fourier transform** (DFT) $X_{\text{DFT}}[k]$ of an N-sample signal $x[n]$ and the **inverse discrete Fourier transform** (IDFT), which transforms $X_{\text{DFT}}[k]$ to $x[n]$, are defined by

$$X_{\text{DFT}}[k] = \sum_{n=0}^{N-1} x[n]e^{-j2\pi nk/N}, \qquad k = 0, 1, 2, \ldots, N-1 \tag{16.1}$$

$$x[n] = \frac{1}{N}\sum_{k=0}^{N-1} X_{\text{DFT}}[k]e^{j2\pi nk/N}, \qquad n = 0, 1, 2, \ldots, N-1 \tag{16.2}$$

Each relation is a set of N equations. Each DFT sample is found as a weighted sum of all the samples in $x[n]$. One of the most important properties of the DFT and its inverse is implied periodicity. The exponential $\exp(\pm j2\pi nk/N)$ in the defining relations is periodic in both n and k with period N:

$$e^{j2\pi nk/N} = e^{j2\pi(n+N)k/N} = e^{j2\pi n(k+N)/N} \tag{16.3}$$

As a result, the DFT and its inverse are also periodic with period N, and it is sufficient to compute the results for only one period (0 to $N-1$). *Both $x[n]$ and $X_{\text{DFT}}[k]$ have a starting index of zero.*

REVIEW PANEL 16.1

The N-point DFT and N-Point IDFT Are Periodic with Period N

$$X_{\text{DFT}}[k] = \sum_{n=0}^{N-1} x[n]e^{-j2\pi nk/N},\ k = 0, 1, 2, \ldots, N-1 \qquad x[n] = \frac{1}{N}\sum_{k=0}^{N-1} X_{\text{DFT}}[k]e^{j2\pi nk/N},\ n = 0, 1, 2, \ldots, N-1$$

PGR2026-00033
Krisp Technologies EX1024 Page 549

EXAMPLE 16.1 (DFT from the Defining Relation)
Let $x[n] = \{1,\ 2,\ 1,\ 0\}$. With $N = 4$, and $e^{-j2\pi nk/N} = e^{-jnk\pi/2}$, we successively compute

$$\begin{aligned}
k = 0: \quad X_{\text{DFT}}[0] &= \sum_{n=0}^{3} x[n]e^{0} = 1 + 2 + 1 + 0 = 4 \\
k = 1: \quad X_{\text{DFT}}[1] &= \sum_{n=0}^{3} x[n]e^{-jn\pi/2} = 1 + 2e^{-j\pi/2} + e^{-j\pi} + 0 = -j2 \\
k = 2: \quad X_{\text{DFT}}[2] &= \sum_{n=0}^{3} x[n]e^{-jn\pi} = 1 + 2e^{-j\pi} + e^{-j2\pi} + 0 = 0 \\
k = 3: \quad X_{\text{DFT}}[3] &= \sum_{n=0}^{3} x[n]e^{-j3n\pi/2} = 1 + 2e^{-j3\pi/2} + e^{-j3\pi} + 0 = j2
\end{aligned}$$

The DFT is thus $X_{\text{DFT}}[k] = \{4,\ -j2,\ 0,\ j2\}$.

16.2 Properties of the DFT

The properties of the DFT are summarized in Table 16.1. They are strikingly similar to other frequency-domain transforms, but must always be used in keeping with implied periodicity (of the DFT and IDFT) in both domains.

Table 16.1 Properties of the N-Sample DFT

Property	**Signal**	**DFT**	**Remarks**
Shift	$x[n - n_0]$	$X_{\text{DFT}}[k]e^{-j2\pi kn_0/N}$	No change in magnitude
Shift	$x[n - 0.5N]$	$(-1)^k X_{\text{DFT}}[k]$	Half-period shift for even N
Modulation	$x[n]e^{j2\pi nk_0/N}$	$X_{\text{DFT}}[k - k_0]$	
Modulation	$(-1)^n x[n]$	$X_{\text{DFT}}[k - 0.5N]$	Half-period shift for even N
Folding	$x[-n]$	$X_{\text{DFT}}[-k]$	This is *circular* folding.
Product	$x[n]y[n]$	$\frac{1}{N}X_{\text{DFT}}[k] \circledast Y_{\text{DFT}}[k]$	The convolution is *periodic*.
Convolution	$x[n] \circledast y[n]$	$X_{\text{DFT}}[k]Y_{\text{DFT}}[k]$	The convolution is *periodic*.
Correlation	$x[n] \circledast\circledast y[n]$	$X_{\text{DFT}}[k]Y^*_{\text{DFT}}[k]$	The correlation is *periodic*.
Central ordinates	$x[0] = \frac{1}{N}\sum_{k=0}^{N-1} X_{\text{DFT}}[k]$		$X_{\text{DFT}}[0] = \sum_{n=0}^{N-1} x[n]$
Central ordinates	$x[\frac{N}{2}] = \frac{1}{N}\sum_{k=0}^{N-1}(-1)^k X_{\text{DFT}}[k]$ (N even)		$X_{\text{DFT}}[\frac{N}{2}] = \sum_{n=0}^{N-1}(-1)^n x[n]$ (N even)
Parseval's relation	$\frac{1}{N}\sum_{n=0}^{N-1} \lvert x[n]\rvert^2 = \sum_{k=0}^{N-1} \lvert X_{\text{DFT}}[k]\rvert^2$		

PGR2026-00033
Krisp Technologies EX1024 Page 550

16.2.1 Symmetry

In analogy with all other frequency-domain transforms, the DFT of a real sequence possesses conjugate symmetry about the origin with $X_{\text{DFT}}[-k] = X^*_{\text{DFT}}[k]$. Since the DFT is periodic, $X_{\text{DFT}}[-k] = X_{\text{DFT}}[N-k]$. This also implies conjugate symmetry about the index $k = 0.5N$, and thus

$$X_{\text{DFT}}[-k] = X^*_{\text{DFT}}[k] = X_{\text{DFT}}[N-k] \tag{16.4}$$

If N is odd, the conjugate symmetry is about the half-integer value $0.5N$. The index $k = 0.5N$ is called the **folding index**. This is illustrated in Figure 16.1.

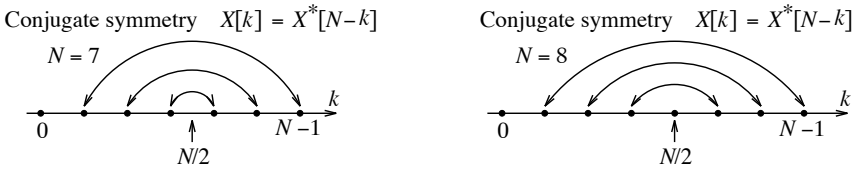


Figure 16.1 Symmetry of the DFT for real signals

Conjugate symmetry suggests that we need compute only half the DFT values to find the entire DFT sequence—another labor-saving concept! A similar result applies to the IDFT.

REVIEW PANEL 16.2

The DFT Shows Conjugate Symmetry for Real Signals

$$X_{\text{DFT}}[-k] = X^*_{\text{DFT}}[k] \qquad X_{\text{DFT}}[-k] = X_{\text{DFT}}[N-k]$$

16.2.2 Central Ordinates and Special DFT Values

The computation of the DFT at the indices $k = 0$ and (for even N) at $k = \frac{N}{2}$ can be simplified using the central ordinate theorems that arise as a direct consequence of the defining relations. In particular, we find that $X_{\text{DFT}}[0]$ equals the sum of the N signal samples $x[n]$, and $X_{\text{DFT}}[\frac{N}{2}]$ equals the sum of $(-1)^n x[n]$ (with alternating sign changes). This also implies that if $x[n]$ is real valued, so are $X_{\text{DFT}}[0]$ and $X_{\text{DFT}}[\frac{N}{2}]$. Similar results hold for the IDFT.

REVIEW PANEL 16.3

The DFT Is Easy to Compute at $k = 0$ and (for Even N) at $k = \frac{N}{2}$

$$X_{\text{DFT}}[0] = \sum_{n=0}^{N-1} x[n] \qquad X_{\text{DFT}}[\tfrac{N}{2}] = \sum_{n=0}^{N-1} (-1)^n x[n] \qquad x[0] = \frac{1}{N}\sum_{k=0}^{N-1} X_{\text{DFT}}[k] \qquad x[\tfrac{N}{2}] = \frac{1}{N}\sum_{k=0}^{N-1} (-1)^k X_{\text{DFT}}[k]$$

16.2.3 Circular Shift and Circular Symmetry

The defining relation for the DFT requires signal values for $0 \le n \le N-1$. By implied periodicity, these values correspond to one period of a periodic signal. If we wish to find the DFT of a time-shifted signal $x[n-n_0]$, its values must also be selected over $(0,\ N-1)$ from its periodic extension. This concept is called **circular shifting**. To generate $x[n-n_0]$, we delay $x[n]$ by n_0, create the periodic extension of the shifted signal, and pick N samples over $(0,\ N-1)$. This is equivalent to moving the last n_0 samples of $x[n]$ to the beginning of the sequence. Similarly, to generate $x[n+n_0]$, we move the first n_0 samples to the end

PGR2026-00033
Krisp Technologies EX1024 Page 551

of the sequence. **Circular folding** generates the signal $x[-n]$ from $x[n]$. We fold $x[n]$, create the periodic extension of the folded signal, and pick N samples of the periodic extension over $(0,\ N-1)$.

REVIEW PANEL 16.4

Generating One Period ($0 \le n \le N$) of a Circularly Shifted Periodic Signal

To generate $x[n-n_0]$: Move the last n_0 samples of $x[n]$ to the beginning.
To generate $x[n+n_0]$: Move the first n_0 samples of $x[n]$ to the end.

Even symmetry of $x[n]$ requires that $x[n] = x[-n]$. Its implied periodicity also means $x[n] = x[N-n]$, and the periodic signal $x[n]$ is said to possess **circular even symmetry**. Similarly, for **circular odd symmetry**, we have $x[n] = -x[N-n]$.

REVIEW PANEL 16.5

Circular Symmetry for Real Periodic Signals with Period N

Circular even symmetry: $x[n] = x[N-n]$ **Circular odd symmetry:** $x[n] = -x[N-n]$

16.2.4 Convolution

Convolution in one domain transforms to multiplication in the other. Due to the implied periodicity in both domains, the convolution operation describes *periodic*, not regular, convolution. This also applies to the correlation operation.

REVIEW PANEL 16.6

Multiplication in One Domain Corresponds to Discrete Periodic Convolution in the Other

Periodic Convolution

The DFT offers an indirect means of finding the periodic convolution $y[n] = x[n] \circledast h[n]$ of two sequences $x[n]$ and $h[n]$ of equal length N. We compute their N-sample DFTs $X_{\text{DFT}}[k]$ and $H_{\text{DFT}}[k]$, multiply them to obtain $Y_{\text{DFT}}[k] = X_{\text{DFT}}[k]H_{\text{DFT}}[k]$, and find the inverse of Y_{DFT} to obtain the periodic convolution $y[n]$:

$$x[n] \circledast h[n] \Longleftrightarrow X_{\text{DFT}}[k]H_{\text{DFT}}[k] \tag{16.5}$$

Periodic Correlation

Periodic correlation can be implemented using the DFT in almost exactly the same way as periodic convolution, except for an extra conjugation step prior to taking the inverse DFT. The periodic correlation of two sequences $x[n]$ and $h[n]$ of equal length N gives

$$r_{xh}[n] = x[n] \circledast\circledast\, h[n] \Longleftrightarrow X_{\text{DFT}}[k]H^*_{\text{DFT}}[k] \tag{16.6}$$

If $x[n]$ and $h[n]$ are real, the final result $r_{xh}[n]$ must also be real (to within machine roundoff).

Regular Convolution and Correlation

We can also find regular convolution (or correlation) using the DFT. For two sequences of length M and N, the regular convolution (or correlation) contains $M+N-1$ samples. We must thus pad each sequence with enough zeros, to make each sequence of length $M+N-1$, before finding the DFT.

PGR2026-00033
Krisp Technologies EX1024 Page 552

REVIEW PANEL 16.7

Regular Convolution by the DFT Requires Zero-Padding

If $x[n]$ and $h[n]$ are of length M and N, create $x_z[n]$ and $h_z[n]$, each zero-padded to length $M+N-1$. Find the DFT of the zero-padded signals, multiply the DFT sequences, and find $y[n]$ as the inverse.

16.2.5 The FFT

The DFT describes a set of N equations, each with N product terms and thus requires a total of N^2 multiplications for its computation. Computationally efficient algorithms to obtain the DFT go by the generic name FFT (fast Fourier transform) and need far fewer multiplications. In particular, **radix-2** FFT algorithms require the number of samples N to be a power of 2 and compute the DFT using only $N\log_2 N$ multiplications. We discuss such algorithms in a later section.

REVIEW PANEL 16.8

The FFT Describes Computationally Efficient Algorithms for Finding the DFT

Radix-2 FFT algorithms require the number of samples N to be a power of 2 ($N=2^m$, integer m).

EXAMPLE 16.2 (DFT Computations and Properties)

(a) Let $y[n]=\{1,\ 2,\ 3,\ 4,\ 5,\ 0,\ 0,\ 0\},\ n=0,1,2,\ldots,7$. Find one period of the circularly shifted signals $f[n]=y[n-2],\ g[n]=y[n+2]$, and the circularly folded signal $h[n]=y[-n]$ over $0\le n\le 7$.

1. To create $f[n]=y[n-2]$, we move the last two samples to the beginning. So, $f[n]=y[n-2]=\{0,\ 0,\ 1,\ 2,\ 3,\ 4,\ 5,\ 0\},\ n=0,1,\ldots,7$.
2. To create $g[n]=y[n+2]$, we move the first two samples to the end. So, $g[n]=y[n+2]=\{3,\ 4,\ 5,\ 0,\ 0,\ 0,\ 1,\ 2\},\ n=0,1,\ldots,7$.
3. To create $h[n]=y[-n]$, we fold $y[n]$ to $\{0,\ 0,\ 0,\ 5,\ 4,\ 3,\ 2,\ 1\},\ n=-7,-6,-5,\ldots,0$ and create its periodic extension by moving all samples (except $y[0]$) past $y[0]$ to get $h[n]=y[-n]=\{1,\ 0,\ 0,\ 0,\ 5,\ 4,\ 3,\ 2\},\ n=0,1,2,\ldots,7$.

(b) Let us find the DFT of $x[n]=\{1,\ 1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0\},\ n=0,1,2,\ldots,7$.

Since only $x[0]$ and $x[1]$ are nonzero, the upper index in the DFT summation will be $n=1$ and the DFT reduces to

$$X_{\text{DFT}}[k]=\sum_{n=0}^{1}x[n]e^{-j2\pi nk/8}=1+e^{-j\pi k/4},\qquad k=0,1,2,\ldots,7$$

Since $N=8$, we need compute $X_{\text{DFT}}[k]$ only for $k\le 0.5N=4$. Now, $X_{\text{DFT}}[0]=1+1=2$ and $X_{\text{DFT}}[4]=1-1=0$. For the rest ($k=1,2,3$), we compute $X_{\text{DFT}}[1]=1+e^{-j\pi/4}=1.707-j0.707$, $X_{\text{DFT}}[2]=1+e^{-j\pi/2}=1-j$, and $X_{\text{DFT}}[3]=1+e^{-j3\pi/4}=0.293-j0.707$.

By conjugate symmetry, $X_{\text{DFT}}[k]=X^*_{\text{DFT}}[N-k]=X^*_{\text{DFT}}[8-k]$. This gives

$X_{\text{DFT}}[5]=X^*_{\text{DFT}}[3]=0.293{+}j0.707,\ \ X_{\text{DFT}}[6]=X^*_{\text{DFT}}[2]=1{+}j,\ \ X_{\text{DFT}}[7]=X^*_{\text{DFT}}[1]=1.707{+}j0.707.$

Thus, $X_{\text{DFT}}[k]=\{2,\ 1.707-j0.707,\ 0.293-j0.707,\ 1-j,\ 0,\ 1+j,\ 0.293+j0.707,\ 1.707+j0.707\}$.

PGR2026-00033
Krisp Technologies EX1024 Page 553

(c) Consider the DFT pair $x[n] = \{1,\ 2,\ 1,\ 0\} \Leftrightarrow X_{\text{DFT}}[k] = \{4,\ -j2,\ 0\ , j2\}$ with $N = 4$.

1. **(Time Shift)** To find $y[n] = x[n-2]$, we move the last two samples to the beginning to get $y[n] = x[n-2] = \{1,\ 0,\ 1,\ 2\},\ n = 0, 1, 2, 3$.
   To find the DFT of $y[n] = x[n-2]$, we use the time-shift property (with $n_0 = 2$) to give $Y_{\text{DFT}}[k] = X_{\text{DFT}}[k]e^{-j2\pi kn_0/4} = X_{\text{DFT}}[k]e^{-jk\pi} = \{4,\ j2,\ 0,\ -j2\}$.
2. **(Modulation)** The sequence $Z_{\text{DFT}}[k] = X_{\text{DFT}}[k-1]$ equals $\{j2,\ 4,\ -j2,\ 0\}$. Its IDFT is $z[n] = x[n]e^{j2\pi n/4} = x[n]e^{j\pi n/2} = \{1,\ j2,\ -1,\ 0\}$.
3. **(Folding)** The sequence $g[n] = x[-n]$ is $g[n] = \{x[0], x[-1], x[-2], x[-3]\} = \{1,\ 0,\ 1,\ 2\}$.
   Its DFT equals $G_{\text{DFT}}[k] = X_{\text{DFT}}[-k] = X^*_{\text{DFT}}[k] = \{4,\ j2,\ 0,\ -j2\}$.
4. **(Conjugation)** The sequence $p[n] = x^*[n]$ is $p[n] = x^*[n] = x[n] = \{1,\ 2,\ 1,\ 0\}$. Its DFT is $P_{\text{DFT}}[k] = X^*_{\text{DFT}}[-k] = \{4,\ j2,\ 0,\ -j2\}^* = \{4,\ -j2,\ 0,\ j2\}$.
5. **(Product)** The sequence $h[n] = x[n]x[n]$ is the pointwise product. So, $h[n] = \{1,\ 4,\ 1,\ 0\}$.
   Its DFT is $H_{\text{DFT}}[k] = \frac{1}{4}X_{\text{DFT}}[k] \circledast X_{\text{DFT}}[k] = \frac{1}{4}\{4,\ -j2,\ 0,\ j2\} \circledast \{4,\ -j2,\ 0,\ j2\}$.
   Keep in mind that this is a periodic convolution.
   The result is $H_{\text{DFT}}[k] = \frac{1}{4}\{24,\ -j16,\ 0,\ j16\} = \{6,\ -j4,\ 0,\ j4\}$.
6. **(Periodic Convolution)** The periodic convolution $c[n] = x[n] \circledast x[n]$ gives
   $c[n] = \{1,\ 2,\ 1,\ 0\} \circledast \{1,\ 2,\ 1,\ 0\} = \{2,\ 4,\ 6,\ 4\}$.
   Its DFT is given by the pointwise product
   $C_{\text{DFT}}[k] = X_{\text{DFT}}[k]X_{\text{DFT}}[k] = \{16,\ -4,\ 0,\ -4\}$.
7. **(Regular Convolution)** The regular convolution $s[n] = x[n] \star x[n]$ gives
   $s[n] = \{1,\ 2,\ 1,\ 0\} \circledast \{1,\ 2,\ 1,\ 0\} = \{1,\ 4,\ 6,\ 4,\ 1,\ 0,\ 0\}$.
   Since $x[n]$ has 4 samples ($N = 4$), the DFT $S_{\text{DFT}}[k]$ of $s[n]$ is the product of the DFT of the zero-padded (to length $N + N - 1 = 7$) signal $x_z[n] = \{1,\ 2,\ 1,\ 0,\ 0,\ 0,\ 0\}$ and equals
   $\{16,\ -2.35 - j10.28,\ -2.18 + j1.05,\ 0.02 + j0.03,\ 0.02 - j0.03, -2.18 - j1.05, -2.35 + j10.28\}$.
8. **(Central Ordinates)** It is easy to check that $x[0] = \frac{1}{4}\sum X_{\text{DFT}}[k]$ and $X_{\text{DFT}}[0] = \sum x[n]$.
9. **(Parseval's Relation)** We have $\sum |x[n]|^2 = 1 + 4 + 1 + 0 = 6$.
   Since $X^2_{\text{DFT}}[k] = \{16,\ -4,\ 0,\ 4\}$, we also have $\frac{1}{4}\sum |X_{\text{DFT}}[k]|^2 = \frac{1}{4}(16 + 4 + 4) = 6$.

16.2.6 Signal Replication and Spectrum Zero Interpolation

In analogy with the DTFT and Fourier series, two useful DFT results are that replication in one domain leads to zero interpolation in the other. Formally, if $x[n] \Leftrightarrow X_{\text{DFT}}[k]$ form a DFT pair, M-fold replication of $x[n]$ to $y[n] = \{x[n],\ x[n], \ldots, x[n]\}$ leads to zero interpolation of the DFT to $Y_{\text{DFT}}[k] = MX_{\text{DFT}}[k/M]$. The multiplying factor M ensures that we satisfy Parseval's theorem and the central ordinate relations. Its dual is the result that zero interpolation of $x[n]$ to $z[n] = x[n/M]$ leads to M-fold replication of the DFT to $Z_{\text{DFT}}[k] = \{X_{\text{DFT}}[k], X_{\text{DFT}}[k], \ldots, X_{\text{DFT}}[k]\}$.

PGR2026-00033
Krisp Technologies EX1024 Page 554

REVIEW PANEL 16.9

Replication in One Domain Corresponds to Zero Interpolation in the Other

If a signal is replicated by M, its DFT is zero interpolated and scaled by M.

If $x[n] \Longleftrightarrow X_{\text{DFT}}[k]$, then $\{\underbrace{x[n],\ x[n], \ldots, x[n]}_{M\text{-fold replication}}\} \Longleftrightarrow MX_{\text{DFT}}[k/M]$.

If a signal is zero-interpolated by M, its DFT shows M-fold replication.

If $x[n] \Longleftrightarrow X_{\text{DFT}}[k]$, then $x[n/M] \Longleftrightarrow \{\underbrace{X_{\text{DFT}}[k], X_{\text{DFT}}[k], \ldots, X_{\text{DFT}}[k]}_{M\text{-fold replication}}\}$.

EXAMPLE 16.3 (Signal and Spectrum Replication)

Let $x[n] = \{2,\ 3,\ 2,\ 1\}$ and $X_{\text{DFT}}[k] = \{8,\ -j2,\ 0,\ j2\}$. Find the DFT of the 12-point signal described by $y[n] = \{x[n],\ x[n],\ x[n]\}$ and the 12-point zero-interpolated signal $h[n] = x[n/3]$.

(a) Signal replication by 3 leads to spectrum zero interpolation and multiplication by 3. Thus,

$$Y_{\text{DFT}}[k] = 3X_{\text{DFT}}[k/3] = \{24,\ 0,\ 0,\ -j6,\ 0,\ 0,\ 0,\ 0,\ 0,\ j6,\ 0,\ 0\}$$

(b) Signal zero interpolation by 3 leads to spectrum replication by 3. Thus,

$$H_{\text{DFT}}[k] = \{X_{\text{DFT}}[k],\ X_{\text{DFT}}[k],\ X_{\text{DFT}}[k]\} = \{8,\ -j2,\ 0,\ j2,\ 8,\ -j2,\ 0,\ j2,\ 8,\ -j2,\ 0,\ j2\}$$

16.2.7 Some Useful DFT Pairs

The DFT of finite sequences defined mathematically often results in very unwieldy expressions and explains the lack of many "standard" DFT pairs. However, the following DFT pairs are quite useful and easy to obtain from the defining relation and properties:

$$\{1,\ 0,\ 0, \ldots,\ 0\} \text{ (impulse)} \quad \Longleftrightarrow \quad \{1,\ 1,\ 1, \ldots,\ 1\} \text{ (constant)} \tag{16.7}$$

$$\{1,\ 1,\ 1, \ldots,\ 1\} \text{ (constant)} \quad \Longleftrightarrow \quad \{N,\ 0,\ 0, \ldots,\ 0\} \text{ (impulse)} \tag{16.8}$$

$$\alpha^n \text{ (exponential)} \quad \Longleftrightarrow \quad \frac{1-\alpha^N}{1-e^{-j2\pi k/N}} \tag{16.9}$$

$$\cos(2\pi nk_0/N) \text{ (sinusoid)} \quad \Longleftrightarrow \quad 0.5N\delta[k-k_0] + 0.5N\delta[k-(N-k_0)] \text{ (impulse pair)} \tag{16.10}$$

The first result is a direct consequence of the defining relation. For the second result, the DFT is $\sum e^{j2\pi nk/N}$, the sum of N equally spaced vectors of unit length, and equals zero (except when $k = 0$). For the third result, we use the defining relation $\sum \alpha^n e^{-j2\pi nk/N}$ and the fact that $e^{j2\pi k} = 1$ to obtain

$$X_{\text{DFT}}[k] = \sum_{k=0}^{N-1} \left(\alpha e^{-j2\pi k/N}\right)^n = \frac{1-(\alpha e^{-j2\pi k/N})^N}{1-\alpha e^{-j2\pi k/N}} = \frac{1-\alpha^N}{1-\alpha e^{-j2\pi k/N}} \tag{16.11}$$

Finally, the transform pair for the sinusoid says that for a periodic sinusoid $x[n] = \cos(2\pi nF)$ whose digital frequency is $F = k_0/N$, the DFT is a pair of impulses at $k = k_0$ and $k = N - k_0$. By Euler's relation, $x[n] = 0.5e^{j2\pi nk_0/N} + 0.5e^{-j2\pi nk_0/N}$ and, by periodicity, $0.5e^{-j2\pi nk_0/N} = 0.5e^{j2\pi n(N-k_0)/N}$. Then, with the DFT pair $1 \Leftrightarrow N\delta[k]$, and the modulation property, we get the required result.

PGR2026-00033
Krisp Technologies EX1024 Page 555

REVIEW PANEL 16.10

The N-Point DFT of a Sinusoid with Period N and $F = k_0/N$ Has Two Nonzero Samples

$$\cos(2\pi n\tfrac{k_0}{N}) \Longleftrightarrow \tfrac{N}{2}\delta[k-k_0] + \tfrac{N}{2}\delta[k-(N-k_0)] = \{0,\ldots,0,\underbrace{\tfrac{N}{2}}_{k=k_0},0,\ldots,0,\underbrace{\tfrac{N}{2}}_{k=N-k_0},0,\ldots,0\} \text{ (DFT)}$$

16.3 Connections

From a purely mathematical or computational standpoint, the DFT simply tells us how to transform a set of N numbers into another set of N numbers. Its physical significance (what the numbers mean), however, stems from its ties to the spectra of both analog and discrete signals.

The processing of analog signals using digital methods continues to gain widespread popularity. The Fourier series of periodic signals and the DTFT of discrete-time signals are duals of each other and are similar in many respects. In theory, both offer great insight into the spectral description of signals. In practice, both suffer from (similar) problems in their implementation. The finite memory limitations and finite precision of digital computers constrain us to work with a finite set of quantized numbers for describing signals in both time and frequency. This brings out two major problems inherent in the Fourier series and the DTFT as tools for digital signal processing. Both typically require an infinite number of samples (the Fourier series for its spectrum and the DTFT for its time signal). Both deal with one continuous variable (time t or digital frequency F). A numerical approximation that can be implemented using digital computers requires that we replace the continuous variable with a discrete one and limit the number of samples to a finite value in both domains. The DFT can then be viewed as a natural extension of the Fourier transform, obtained by sampling both time and frequency in turn. Sampling in one domain leads to a periodic extension in the other. A sampled representation in both domains also forces periodicity in both domains. This leads to two slightly different but functionally equivalent sets of relations, depending on the order in which we sample time and frequency, as listed in Table 16.2.

If we first sample an analog signal $x(t)$, the sampled signal has a periodic spectrum $X_p(F)$ (the DTFT), and sampling of $X_p(F)$ leads to the DFT representation. If we first sample the Fourier transform $X(f)$ in the frequency domain, the samples represent the Fourier series coefficients of a periodic time signal $x_p(t)$, and sampling of $x_p(t)$ leads to the **discrete Fourier series** (DFS) as the periodic extension of the frequency-domain samples. The DFS differs from the DFT only by a constant scale factor.

REVIEW PANEL 16.11

Sampling in One Domain Corresponds to Periodic Extension in the Other

16.3.1 Some Practical Guidelines

In general, the DFT is only an approximation to the actual (Fourier series or transform) spectrum of the underlying analog signal. The DFT spectral spacing and DFT magnitude is affected by the choice of sampling rate and how the sample values are chosen. The DFT phase is affected by the location of sampling instants. The DFT spectral spacing is affected by the sampling duration. Here are some practical guidelines on how to obtain samples of an analog signal $x(t)$ for spectrum analysis and interpret the DFT (or DFS) results.

Choice of sampling instants: The defining relation for the DFT (or DFS) mandates that samples of $x[n]$ be chosen over the range $0 \le n \le N-1$ (through periodic extension, if necessary). Otherwise, the DFT (or DFS) phase will not match the expected phase.

PGR2026-00033
Krisp Technologies EX1024 Page 556

Table 16.2 Relating Frequency-Domain Transforms

Fourier Transform	
↙	↘
Aperiodic/continuous signal: $x(t) = \int_{-\infty}^{\infty} X(f)e^{j2\pi ft}\,df$	Aperiodic/continuous spectrum: $X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}\,dt$
⇓	⇓
Sampling $x(t)$ (DTFT)	**Sampling $X(f)$ (Fourier series)**
Sampled time signal: $x[n] = \int_{1} X_p(F)e^{j2\pi nF}\,dF$ Periodic spectrum (period = 1): $X_p(F) = \sum_{n=-\infty}^{\infty} x[n]e^{-j2\pi nF}$	Sampled spectrum: $X[k] = \frac{1}{T}\int_{T} x_p(t)e^{-j2\pi kf_0t}\,dt$ Periodic time signal (period = T): $x_p(t) = \sum_{k=-\infty}^{\infty} X[k]e^{j2\pi kf_0t}$
⇓	⇓
Sampling $X_p(F)$ (DFT)	**Sampling $x_p(t)$ (DFS)**
Sampled/periodic spectrum: $X_{\text{DFT}}[k] = \sum_{n=0}^{N-1} x[n]e^{-j2\pi nk/N}$ Sampled/periodic time signal: $x[n] = \frac{1}{N}\sum_{k=0}^{N-1} X_{\text{DFT}}[k]e^{j2\pi nk/N}$	Sampled/periodic time signal: $x[n] = \sum_{k=0}^{N-1} X_{\text{DFS}}[k]e^{j2\pi nk/N}$ Sampled/periodic spectrum: $X_{\text{DFS}}[k] = \frac{1}{N}\sum_{n=0}^{N-1} x[n]e^{-j2\pi nk/N}$

Choice of samples: If a sampling instant corresponds to a jump discontinuity, the sample value should be chosen as the midpoint of the discontinuity. The reason is that the Fourier series (or transform) converges to the midpoint of any discontinuity.

Choice of frequency axis: The computation of the DFT (or DFS) is independent of the sampling frequency S or sampling interval $t_s = 1/S$. However, if an analog signal is sampled at a sampling rate S, its spectrum is periodic with period S. The DFT spectrum describes one period (N samples) of this spectrum *starting at the origin*. For sampled signals, it is useful to plot the DFT (or DFS) magnitude and phase against the analog frequency $f = kS/N$ Hz, $k = 0, 1, \ldots, N-1$ (with spacing S/N). For discrete-time signals, we can plot the DFT against the digital frequency $F = k/N$, $k = 0, 1, \ldots, N-1$ (with spacing $1/N$). These choices are illustrated in Figure 16.2.

Choice of frequency range: To compare the DFT results with conventional two-sided spectra, just remember that by periodicity, a negative frequency $-f_0$ (at the index $-k_0$) in the two-sided spectrum corresponds to the frequency $S - f_0$ (at the index $N - k_0$) in the (one-sided) DFT spectrum.

Identifying the highest frequency: *The highest frequency in the* DFT *spectrum corresponds to the folding index $k = 0.5N$ and equals $F = 0.5$ for discrete signals or $f = 0.5S$ Hz for sampled analog signals.* This highest frequency is also called the **folding frequency**. For purposes of comparison, its is sufficient to plot the DFT spectra only over $0 \le k < 0.5N$ (or $0 \le F < 0.5$ for discrete-time signals or $0 \le f < 0.5S$ Hz for sampled analog signals).

PGR2026-00033
Krisp Technologies EX1024 Page 557

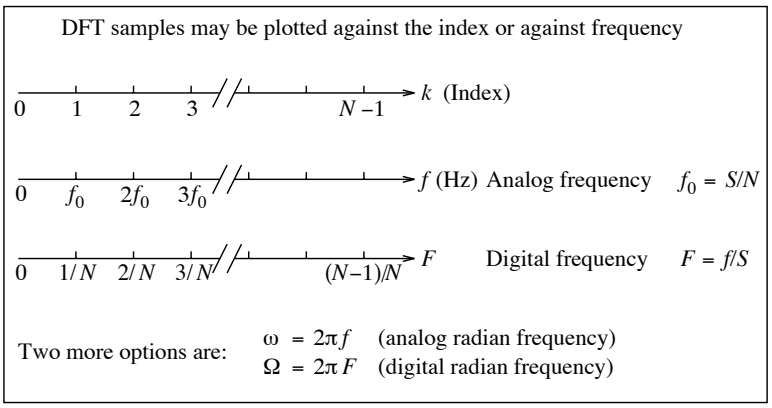


Figure 16.2 Various ways of plotting the DFT

Plotting reordered spectra: The DFT (or DFS) may also be plotted as two-sided spectra to reveal conjugate symmetry about the origin by creating its periodic extension. This is equivalent to creating a *reordered spectrum* by relocating the DFT samples at indices past the folding index $k = 0.5N$ to the left of the origin (because $X[-k] = X[N-k]$). This process is illustrated in Figure 16.3.

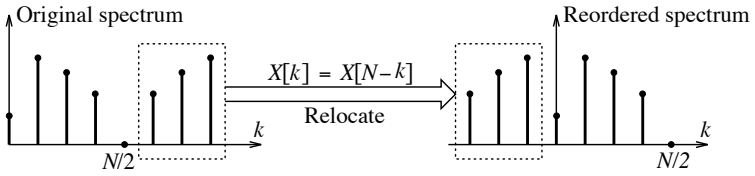


Figure 16.3 Plotting the DFT or its reordered samples

REVIEW PANEL 16.12

Practical Guidelines for Sampling a Signal and Interpreting the DFT Results

Sampling: Start at $t = 0$. Choose the midpoint value at jumps. Sample above the Nyquist rate.
Plotting: Plot DFT against index $k = 0, 1, \ldots, N-1$ or $F = \frac{k}{N}$ or $f = k\frac{S}{N}$ Hz.
Frequency spacing of DFT samples: $\Delta f = S/N$ Hz (analog) or $\Delta F = 1/N$ (digital frequency).
Highest frequency: This equals $F = 0.5$ or $f = 0.5S$ corresponding to the index $k = \frac{N}{2}$.
For long sequences: The DFT magnitude/phase are usually plotted as *continuous functions*.

16.4 Approximating the DTFT by the DFT

The DTFT relation and its inverse are

$$X_p(F) = \sum_{n=-\infty}^{\infty} x[n]e^{-j2\pi nF} \qquad x[n] = \int_1 X_p(F)e^{j2\pi nF}\,dF \tag{16.12}$$

PGR2026-00033
Krisp Technologies EX1024 Page 558

where $X_p(F)$ is periodic with unit period. If $x[n]$ is a finite N-point sequence with $n = 0, 1, \ldots, N-1$, we obtain N samples of the DTFT over one period at intervals of $1/N$ as

$$X_{\text{DFT}}[k] = \sum_{n=0}^{N-1} x[n]e^{-j2\pi nk/N}, \qquad k = 0, 1, \ldots, N-1 \tag{16.13}$$

This describes the **discrete Fourier transform** (DFT) of $x[n]$ as a sampled version of its DTFT evaluated at the frequencies $F = k/N,\ k = 0, 1, \ldots, N-1$. The DFT spectrum thus corresponds to the frequency range $0 \le F < 1$ and is plotted at the frequencies $F = k/N,\ k = 0, 1, \ldots, N-1$.

To recover the finite sequence $x[n]$ from N samples of $X_{\text{DFT}}[k]$, we use $dF \approx 1/N$ and $F \to k/N$ to approximate the integral expression in the inversion relation by

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X_{\text{DFT}}[k]e^{j2\pi nk/N}, \qquad n = 0, 1, \ldots, N-1 \tag{16.14}$$

This is the **inverse discrete Fourier transform** (IDFT). The periodicity of the IDFT implies that $x[n]$ actually corresponds to one period of a periodic signal.

If $x[n]$ is a finite N-point signal with $n = 0, 1, \ldots, N-1$, the DFT is an exact match to its DTFT $X_p(F)$ at $F = k/N,\ k = 0, 1, \ldots, N-1$, and the IDFT results in perfect recovery of $x[n]$.

If $x[n]$ is not time-limited, its N-point DFT is only an approximation to its DTFT $X_p(F)$ evaluated at $F = k/N,\ k = 0, 1, \ldots, N-1$. Due to implied periodicity, the DFT, in fact, exactly matches the DTFT of the *periodic extension* of $x[n]$ with period N at these frequencies.

If $x[n]$ is a discrete periodic signal with period N, its scaled DFT ($\frac{1}{N}X_{\text{DFT}}[k]$) is an exact match to the impulse strengths in its DTFT $X_p(F)$ at $F = k/N,\ k = 0, 1, \ldots, N-1$. In this case also, the IDFT results in perfect recovery of one period of $x[n]$ over $0 \le n \le N-1$.

REVIEW PANEL 16.13

The DFT $X_{\text{DFT}}[k]$ of an N-Sample Sequence $x[n]$ Is an Exact Match to Its Sampled DTFT

If $x[n]$ is also periodic, the scaled DFT $\frac{1}{N}X_{\text{DFT}}[k]$ is an exact match to its sampled DTFT.

EXAMPLE 16.4 (Relating the DFT and the DTFT)

(a) Let $x[n] = \{1,\ 2,\ 1,\ 0\}$. If we use the DTFT, we first find

$$X_p(F) = 1 + 2e^{-j2\pi F} + e^{-j4\pi F} + 0 = [2 + 2\cos(2\pi F)]e^{-j2\pi F}$$

With $N = 4$, we have $F = k/4,\ k = 0, 1, 2, 3$. We then obtain the DFT as

$$X_{\text{DFT}}[k] = [2 + 2\cos(2\pi k/4)]e^{-j2\pi k/4}, \quad k = 0, 1, 2, 3, \quad \text{or} \quad X_{\text{DFT}}[k] = \{4,\ -j2,\ 0,\ j2\}$$

Since $x[n]$ is a finite sequence, the DFT and DTFT show an exact match at $F = k/N,\ k = 0, 1, 2, 3$.

With $N = 4$, and $e^{-j2\pi nk/N} = e^{-jnk\pi/2}$, we compute the IDFT of $X_{\text{DFT}}[k] = \{4,\ -j2,\ 0,\ j2\}$ to give

$$n = 0: \quad x[0] = 0.25\sum_{k=0}^{3} X_{\text{DFT}}[k]e^{0} = 0.25(4 - j2 + 0 + j2) = 1$$

$$n = 1: \quad x[1] = 0.25\sum_{k=0}^{3} X_{\text{DFT}}[k]e^{-jk\pi/2} = 0.25(4 - j2e^{-j\pi/2} + 0 + j2e^{-j3\pi/2}) = 2$$

PGR2026-00033
Krisp Technologies EX1024 Page 559

$$n=2: \quad x[2] = 0.25\sum_{k=0}^{3} X_{\text{DFT}}[k]e^{-jk\pi} = 0.25(4 - j2e^{-j\pi} + 0 + j2e^{-j3\pi}) = 1$$

$$n=3: \quad x[3] = 0.25\sum_{k=0}^{3} X_{\text{DFT}}[k]e^{-j3k\pi/2} = 0.25(4 - j2e^{-j3\pi/2} + 0 + e^{-j9\pi/2}) = 0$$

The IDFT is thus $\{1,\ 2,\ 1,\ 0\}$, and recovers $x[n]$ exactly.

(b) Let $x[n] = \alpha^n u[n]$. Its DTFT is $X_p(F) = \frac{1}{1-\alpha e^{-j2\pi F}}$. Sampling $X_p(F)$ at intervals $F = k/N$ gives

$$X_p(F)\big|_{F=k/N} = \frac{1}{1 - \alpha e^{-j2\pi k/N}}$$

The N-point DFT of $x[n]$ is

$$\alpha^n \ (n = 0, 1, \ldots, N-1) \Longleftrightarrow \frac{1-\alpha^N}{1 - \alpha e^{-j2\pi k/N}}$$

Clearly, the N-sample DFT of $x[n]$ does not match the DTFT of $x[n]$ (unless $N \to \infty$).

Comment: What does match, however, is the DFT of the N-sample periodic extension $x_{\text{pe}}[n]$ and the DTFT of $x[n]$. We obtain one period of the periodic extension by wrapping around N-sample sections of $x[n] = \alpha^n$ and adding them to give

$$x_{\text{pe}}[n] = \alpha^n + \alpha^{n+N} + \alpha^{n+2N} + \cdots = \alpha^n(1 + \alpha^N + \alpha^{2N} + \cdots) = \frac{\alpha^n}{1-\alpha^N} = \frac{1}{1-\alpha^N}x[n]$$

Its DFT is thus $\dfrac{1}{1 - \alpha e^{-j2\pi k/N}}$ and matches the DTFT of $x[n]$ at $F = \frac{k}{N},\ k = 0, 1, \ldots, N-1$.

16.5 The DFT of Periodic Signals and the DFS

The Fourier series relations for a periodic signal $x_p(t)$ are

$$x_p(t) = \sum_{k=-\infty}^{\infty} X[k]e^{j2\pi k f_0 t} \qquad X[k] = \frac{1}{T}\int_T x_p(t)e^{-j2\pi k f_0 t}\,dt \tag{16.15}$$

If we acquire $x[n],\ n = 0, 1, \ldots, N-1$ as N samples of $x_p(t)$ over one period using a sampling rate of S Hz (corresponding to a sampling interval of t_s) and approximate the integral expression for $X[k]$ by a summation using $dt \to t_s,\ t \to nt_s,\ T = Nt_s$, and $f_0 = \frac{1}{T} = \frac{1}{Nt_s}$, we obtain

$$X_{\text{DFS}}[k] = \frac{1}{Nt_s}\sum_{n=0}^{N-1} x[n]e^{-j2\pi k f_0 n t_s}t_s = \frac{1}{N}\sum_{n=0}^{N-1} x[n]e^{-j2\pi nk/N}, \quad k = 0, 1, \ldots, N-1 \tag{16.16}$$

The quantity $X_{\text{DFS}}[k]$ defines the **discrete Fourier series** (DFS) as an approximation to the Fourier series coefficients of a periodic signal and equals N times the DFT.

PGR2026-00033
Krisp Technologies EX1024 Page 560

16.5.1 The Inverse DFS

To recover $x[n]$ from one period of $X_{\text{DFS}}[k]$, we use the Fourier series reconstruction relation whose summation index covers one period (from $k = 0$ to $k = N - 1$) to obtain

$$x[n] = \sum_{k=0}^{N-1} X_{\text{DFS}}[n] e^{j2\pi k f_0 n t_s} = \sum_{k=0}^{N-1} X_{\text{DFS}}[k] e^{j2\pi nk/N}, \qquad n = 0, 1, 2, \ldots, N-1 \tag{16.17}$$

This relation describes the **inverse discrete Fourier series** (IDFS). The sampling interval t_s does not enter into the computation of the DFS or its inverse. Except for a scale factor, the DFS and DFT relations are identical.

16.5.2 Understanding the DFS Results

The $X_{\text{DFS}}[k]$ describes the spectrum of a sampled signal and is thus periodic with period S (starting at the origin). Its N samples are spaced S/N Hz apart and plotted at the frequencies $f = kS/N,\ k = 0, 1, \ldots, N-1$. However, the highest frequency we can identify in the DFS spectrum is $0.5S$ (corresponding to the folding index $k = 0.5N$).

The N-sample DFS $X_{\text{DFS}}[k] = \frac{1}{N} X_{\text{DFT}}[k]$ shows an *exact* match to the Fourier series coefficients $X[k]$ of a periodic signal $x(t)$ only if all of the following conditions are satisfied:

1. The signal samples must be acquired from $x(t)$ starting at $t = 0$ (using the periodic extension of the signal, if necessary). Otherwise, the phase of the DFS coefficients will not match the phase of the corresponding Fourier series coefficients.

2. The periodic signal must contain a finite number of sinusoids (to ensure a band-limited signal with a finite highest frequency) and be sampled above the Nyquist rate. Otherwise, there will be *aliasing*, whose effects become more pronounced near the folding frequency $0.5S$. If the periodic signal is not band-limited (contains an infinite number of harmonics), we cannot sample at a rate high enough to prevent aliasing. For a pure sinusoid, the Nyquist rate corresponds to two samples per period.

3. The signal $x(t)$ must be sampled for an *integer* number of periods (to ensure a match between the periodic extension of $x(t)$ and the implied periodic extension of the sampled signal). Otherwise, the periodic extension of its samples will not match that of $x(t)$, and the DFS samples will describe the Fourier series coefficients of a different periodic signal whose harmonic frequencies do not match those of $x(t)$. This phenomenon is called **leakage** and results in nonzero spectral components at frequencies other than the harmonic frequencies of the original signal $x(t)$.

If we sample a periodic signal for an integer number of periods, the DFS (or DFT) also preserves the effects of symmetry. The DFS of an even symmetric signal is real, the DFS of an odd symmetric signal is imaginary, and the DFS of a half-wave symmetric signal is zero at even values of the index k.

REVIEW PANEL 16.14

The DFT of a Sampled Periodic Signal $x(t)$ Is Related to Its Fourier Series Coefficients

If $x(t)$ is band-limited and sampled for an integer number of periods, the DFT is an exact match to the Fourier series coefficients $X[k]$, with $X_{\text{DFT}}[k] = NX[k]$.
If $x(t)$ is not band-limited, there is *aliasing*.
If $x(t)$ is not sampled for an integer number of periods, there is *also leakage*.

PGR2026-00033
Krisp Technologies EX1024 Page 561

16.5.3 The DFS and DFT of Sinusoids

Consider the sinusoid $x(t) = \cos(2\pi f_0 t + \theta)$ whose Fourier series coefficients we know to be $0.5\angle\theta = 0.5e^{j\theta}$ at $f = f_0$ and $0.5\angle -\theta = 0.5e^{-j\theta}$ at $f = -f_0$. If $x(t)$ is sampled at the rate S, starting at $t = 0$, the sampled signal is $x[n] = \cos(2\pi nF + \theta)$, where $F = f_0/S$ is the digital frequency. As long as F is a rational fraction of the form $F = k_0/N$, we obtain N samples of $x[n]$ from k_0 full periods of $x(t)$. In this case, there is no leakage, and the N-point DFS will match the expected results and show only two nonzero DFS values, $X_{\text{DFS}}[k_0] = 0.5e^{j\theta}$ and $X_{\text{DFS}}[N - k_0] = 0.5e^{-j\theta}$. The DFT is obtained by multiplying the DFS by N.

$$\cos(2\pi n\tfrac{k_0}{N} + \theta) \Longleftrightarrow \{0, \ldots, 0, \underbrace{0.5e^{j\theta}}_{k=k_0}, 0, \ldots, 0, \underbrace{0.5e^{-j\theta}}_{k=N-k_0}, 0, \ldots, 0\} \quad \text{(DFS)} \tag{16.18}$$

For the index k_0 to lie in the range $0 \le k_0 \le N-1$, we must ensure that $0 \le F < 1$ (rather than the customary $-0.5 \le F < 0.5$). The frequency corresponding to k_0 will then be $k_0 S/N$ and will equal f_0 (if $S > 2f_0$) or its alias (if $S < 2f_0$). The nonzero DFT values will equal $X_{\text{DFT}}[k_0] = 0.5Ne^{j\theta}$ and $X_{\text{DFT}}[N - k_0] = 0.5Ne^{-j\theta}$. These results are straightforward to obtain and can be easily extended, by superposition, to the DFT of a combination of sinusoids, sampled over an integer number of periods.

REVIEW PANEL 16.15

The N-Point DFT of the Sinusoid $x[n] = \cos(2\pi n\frac{k_0}{N} + \theta)$ Contains Only Two Nonzero Samples

$$\cos(2\pi n\tfrac{k_0}{N} + \theta) \Longleftrightarrow \{0, \ldots, 0, \underbrace{\tfrac{N}{2}e^{j\theta}}_{k=k_0}, 0, \ldots, 0, \underbrace{\tfrac{N}{2}e^{-j\theta}}_{k=N-k_0}, 0, \ldots, 0\} \quad \text{(DFT)}$$

EXAMPLE 16.5 (The DFT and DFS of Sinusoids)

(a) The signal $x(t) = 4\cos(100\pi t)$ is sampled at twice the Nyquist rate for three full periods. Find and sketch its DFT.

The frequency of $x(t)$ is 50 Hz, the Nyquist rate is 100 Hz, and the sampling frequency is $S = 200$ Hz. The digital frequency is $F = 50/200 = 1/4 = 3/12 = k/N$. This means $N = 12$ for three full periods. The two nonzero DFT values will appear at $k = 3$ and $k = N - 3 = 9$. The nonzero DFT values will be $X[3] = X[9] = (0.5)(4)(N) = 24$. The signal and its DFT are sketched in Figure E16.5A.

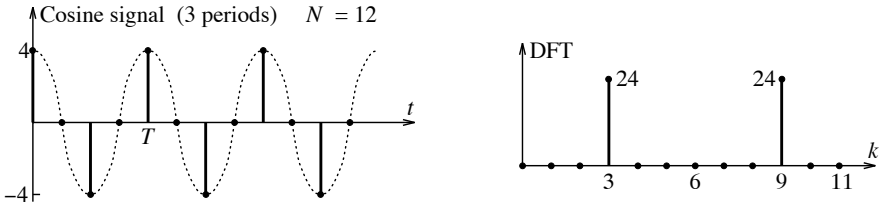


Figure E16.5A The signal and DFT for Example 16.5(a)

(b) Let $x(t) = 4\sin(72\pi t)$ be sampled at $S = 128$ Hz. Choose the minimum number of samples necessary to prevent leakage and find the DFS and DFT of the sampled signal.

The frequency of $x(t)$ is 36 Hz, so $F = 36/128 = 9/32 = k_0/N$. Thus, $N = 32$, $k_0 = 9$, and the frequency spacing is $S/N = 4$ Hz. The DFS components will appear at $k_0 = 9$ (36 Hz) and $N - k_0 = 23$ (92 Hz). The Fourier series coefficients of $x(t)$ are $-j2$ (at 36 Hz) and $j2$ (at -36 Hz).

PGR2026-00033
Krisp Technologies EX1024 Page 562

The DFS samples will be $X_{\rm DFS}[9] = -j2$, and $X_{\rm DFS}[23] = j2$. Since $X_{\rm DFT}[k] = NX_{\rm DFS}[k]$, we get $X_{\rm DFT}[9] = -j64,\ X_{\rm DFT}[23] = j64$, and thus

$$X_{\rm DFT}[k] = \{0, \ldots,\ 0,\ \underbrace{-j64}_{k=9},\ 0, \ldots,\ 0,\ \underbrace{j64}_{k=23},\ 0, \ldots,\ 0\}$$

(c) Let $x(t) = 4\sin(72\pi t) - 6\cos(12\pi t)$ be sampled at $S = 21$ Hz. Choose the minimum number of samples necessary to prevent leakage and find the DFS and DFT of the sampled signal.

Clearly, the 36-Hz term will be aliased. The digital frequencies (between 0 and 1) of the two terms are $F_1 = 36/21 = 12/7 \Rightarrow 5/7 = k_0/N$ and $F_2 = 6/21 = 2/7$. Thus, $N = 7$ and the frequency spacing is $S/N = 3$ Hz. The DFS components of the first term will be $-j2$ at $k = 5$ (15 Hz) and $j2$ at $N - k = 2$ (6 Hz). The DFS components of the second term will be -3 at $k = 2$ and -3 at $k = 5$. The DFS values will add up at the appropriate indices to give $X_{\rm DFS}[5] = -3 - j2,\ X_{\rm DFS}[2] = -3 + j2$, and

$$X_{\rm DFS}[k] = \{0,\ 0,\ \underbrace{-3+j2}_{k=2},\ 0,\ 0,\ \underbrace{-3-j2}_{k=5},\ 0\} \qquad X_{\rm DFT}[k] = NX_{\rm DFS}[k] = 7X_{\rm DFS}[k]$$

Note how the 36-Hz component was aliased to 6 Hz (the frequency of the second component).

(d) The signal $x(t) = 1 + 8\sin(80\pi t)\cos(40\pi t)$ is sampled at twice the Nyquist rate for two full periods. Is leakage present? If not, find the DFS of the sampled signal.

First, note that $x(t) = 1 + 4\sin(120\pi t) + 4\sin(40\pi t)$. The frequencies are $f_1 = 60$ Hz and $f_2 = 20$ Hz. The Nyquist rate is thus 120 Hz, and hence $S = 240$ Hz. The digital frequencies are $F_1 = 60/240 = 1/4$ and $F_2 = 20/240 = 1/12$. The fundamental frequency is $f_0 = {\rm GCD}(f_1,\ f_2) = 20$ Hz. Thus, two full periods correspond to 0.1 s or $N = 24$ samples. There is no leakage because we acquire the samples over two full periods. The index $k = 0$ corresponds to the constant (dc value). To find the indices of the other nonzero DFS samples, we compute the digital frequencies (in the form k/N) as $F_1 = 60/240 = 1/4 = 6/24$ and $F_2 = 20/240 = 1/12 = 2/24$. The nonzero DFS samples are thus $X_{\rm DFS}[0] = 1$, $X_{\rm DFS}[6] = -j2$, and $X_{\rm DFS}[2] = -j2$, and the conjugates $X_{\rm DFS}[18] = j2$ and $X_{\rm DFS}[22] = j2$. Thus,

$$X_{\rm DFS}[k] = \{0,\ 0,\ \underbrace{-j2}_{k=2},\ 0,\ 0,\ 0,\ \underbrace{-j2}_{k=6},\ 0, \ldots,\ 0,\ \underbrace{j2}_{k=18},\ 0,\ 0,\ 0,\ \underbrace{j2}_{k=22},\ 0,\ 0\}$$

16.5.4 The DFT and DFS of Sampled Periodic Signals

For periodic signals that are not band-limited, leakage can be prevented by sampling for an integer number of periods, but aliasing is unavoidable for any sampling rate. The aliasing error increases at higher frequencies. The effects of aliasing can be minimized (but not eliminated) by choosing a high enough sampling rate. A useful rule of thumb is that for the DFT to produce an error of about 5% or less up to the Mth harmonic frequency $f_M = Mf_0$, we must choose $S \geq 8f_M$ (corresponding to $N \geq 8M$ samples per period).

EXAMPLE 16.6 (DFS of Sampled Periodic Signals)
Consider a square wave $x(t)$ that equals 1 for the first half-period and -1 for the next half-period, as illustrated in Figure E16.6.

PGR2026-00033
Krisp Technologies EX1024 Page 563

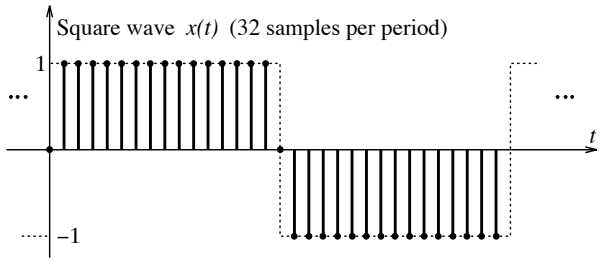


Figure E16.6 One period of the square wave periodic signal for Example 16.6

Its Fourier series coefficients are $X[k] = -j2/k\pi,\ (k \text{ odd})$. If we require the first four harmonics to be in error by no more than about 5%, we choose a sampling rate $S = 32f_0$, where f_0 is the fundamental frequency. This means that we acquire $N = 32$ samples for one period. The samples and their DFS up to $k = 8$ are listed below, along with the error in the nonzero DFS values compared with the Fourier series coefficients.

$$x[n] = \{0,\ \underbrace{1,\ 1, \ldots,\ 1,\ 1}_{15 \text{ samples}},\ 0,\ \underbrace{-1,\ -1, \ldots,\ -1,\ -1}_{15 \text{ samples}}\}$$

$$X_{\text{DFS}}[k] = \{0,\ \underbrace{-j0.6346}_{\text{off by 0.3\%}},\ 0,\ \underbrace{-j0.206}_{\text{off by 2.9\%}},\ 0,\ \underbrace{-j0.1169}_{\text{off by 8.2\%}},\ 0,\ \underbrace{-j0.0762}_{\text{off by 16.3\%}},\ 0, \ldots\}$$

Note how we picked zero as the sample value at the discontinuities. As expected, the DFS coefficients are zero for odd k (due to the half-wave symmetry in $x(t)$) and purely imaginary (due to the odd symmetry in $x(t)$). The error in the nonzero harmonics up to $k = 4$ is less than 5%.

16.5.5 The Effects of Leakage

If we do not sample $x(t)$ for an integer number of periods, the DFT will show the effects of leakage, with nonzero components appearing at frequencies other than $\pm f_0$, because the periodic extension of the signal with non-integer periods will not match the original signal, as illustrated in Figure 16.4.

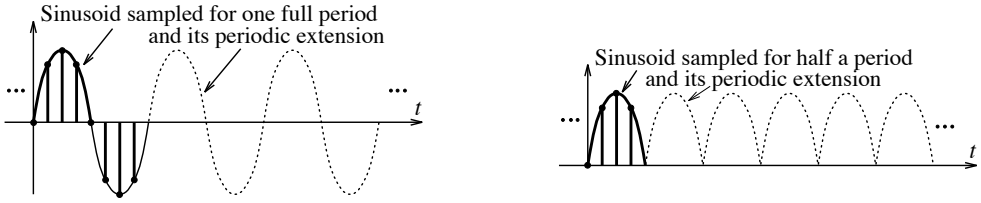


Figure 16.4 Illustrating the concept of leakage

The DFT results will also show aliasing because the periodic extension of the signal with non-integer periods will not, in general, be band-limited. As a result, we must resort to the full force of the defining relation to compute the DFT.

Suppose we sample the 1-Hz sine $x(t) = \sin(2\pi t)$ at $S = 16$ Hz. Then, $F_0 = 1/16$. If we choose $N = 8$, the DFT spectral spacing equals $S/N = 2$ Hz. In other words, there is no DFT component at 1 Hz, the frequency of the sine wave! Where should we expect to see the DFT components? If we express the digital frequency $F_0 = 1/16$ as $F_0 = k_F/N$, we obtain $k_F = NF_0 = 0.5$. Thus, F_0 corresponds to the fractional

PGR2026-00033
Krisp Technologies EX1024 Page 564

index $k_F = 0.5$, and the largest DFT components should appear at the integer index *nearest* to k_F, at $k = 0$ (or dc) and $k = 1$ (or 2 Hz). In fact, the signal and its DFS are given by

$$\begin{aligned} x[n] &= \{0,\ 0.3827,\ 0.7071,\ 0.9329,\ 1,\ 0.9329,\ 0.7071,\ 0.3827\} \\ X_{\text{DFS}}[k] &= \{\ \underbrace{0.6284}_{\text{off by 1.3\%}},\ \underbrace{-0.2207}_{\text{off by 4\%}},\ \underbrace{-0.0518}_{\text{off by 23\%}},\ \underbrace{-0.0293}_{\text{off by 61\%}},\ \underbrace{-0.0249}_{k=N/2},\ -0.0293,\ -0.0518,\ -0.2207\ \} \end{aligned}$$

As expected, the largest components appear at $k = 0$ and $k = 1$. Since $X_{\text{DFS}}[k]$ is real, the DFS results describe an even symmetric signal with nonzero average value. In fact, the periodic extension of the sampled signal over half a period actually describes a full-rectified sine with even symmetry and a fundamental frequency of 2 Hz (see Figure 16.4). The Fourier series coefficients of this full-rectified sine wave (with unit peak value) are given by

$$X[k] = \frac{2}{\pi(1 - 4k^2)}$$

We confirm that $X_{\text{DFS}}[0]$ and $X_{\text{DFS}}[1]$ show an error of less than 5%. But $X_{\text{DFS}}[2]$, $X_{\text{DFS}}[3]$, and $X_{\text{DFS}}[4]$ deviate significantly. Since the new periodic signal is no longer band-limited, the sampling rate is not high enough, and we have aliasing. The value $X_{\text{DFS}}[3]$, for example, equals the sum of the Fourier series coefficient $X[3]$ and all other Fourier series coefficients $X[11],\ X[19], \ldots$ that alias to $k = 3$. In other words,

$$X_{\text{DFS}}[3] = \sum_{m=\infty}^{\infty} X[3 + 8m] = \frac{2}{\pi} \sum_{m=\infty}^{\infty} \frac{1}{1 - 4(3 + 8m)^2}$$

Although this sum is not easily amenable to a closed-form solution, it can be computed numerically and does in fact approach $X_{\text{DFS}}[3] = -0.0293$ (for a large but finite number of coefficients).

Minimizing Leakage

Ideally, we should sample periodic signals over an integer number of periods to prevent leakage. In practice, it may not be easy to identify the period of a signal in advance. In such cases, it is best to sample over as long a signal duration as possible (to reduce the mismatch between the periodic extension of the analog and sampled signal). Sampling for a larger time (duration) not only reduces the effects of leakage but also yields a more closely spaced spectrum, and a more accurate estimate of the spectrum of the original signal. Another way to reduce leakage is to multiply the signal samples by a window function (as described later).

REVIEW PANEL 16.16

The DFT of a Sinusoid at f_0 Hz Sampled for Non-Integer Periods Shows Leakage

The largest DFT component appears at the integer index closest to $k_F = F_0 N = f_0 N/S$.

To minimize leakage: Sample for the longest duration possible or (better still) for integer periods.

EXAMPLE 16.7 (The Effects of Leakage)

The signal $x(t) = 2\cos(20\pi t) + 5\cos(100\pi t)$ is sampled at intervals of $t_s = 0.005$ s for three different durations, 0.1 s, 0.125 s, and 1.125 s. Explain the DFT spectrum for each duration.

The sampling frequency is $S = 1/t_s = 200$ Hz. The frequencies in $x(t)$ are $f_1 = 10$ Hz and $f_2 = 50$ Hz. The two-sided spectrum of $x(t)$ will show a magnitude of 1 at 10 Hz and 2.5 at 50 Hz. The fundamental frequency is 10 Hz, and the common period of $x(t)$ is 0.1 s. We have the following results with reference to Figure E16.7, which shows the DFS magnitude ($|X_{\text{DFT}}|/N$) up to the folding index (or 100 Hz).

PGR2026-00033
Krisp Technologies EX1024 Page 565

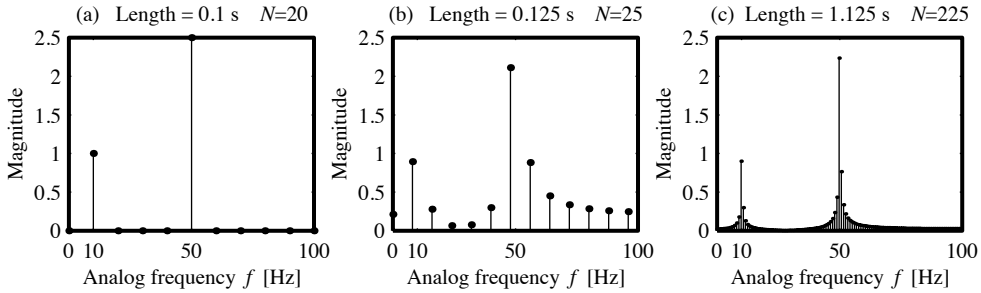


Figure E16.7 DFT results for Example 16.7

(a) The duration of 0.1 s corresponds to one full period, and $N = 20$. No leakage is present, and the DFS results reveal an *exact match* to the spectrum of $x(t)$. The nonzero components appear at the integer indices $k_1 = NF = Nf_1/S = 1$ and $k_2 = NF_2 = Nf_2/S = 5$ (corresponding to 10 Hz and 50 Hz, respectively).

(b) The duration of 0.125 s does not correspond to an integer number of periods. The number of samples over 0.125 s is $N = 25$. Leakage is present. The largest components appear at the integer indices closest to $k_1 = NF = Nf_1/S = 1.25$ (i.e., $k = 1$ or 8 Hz) and $k_2 = Nf_2/S = 6.25$ (i.e., $k = 6$ or 48 Hz).

(c) The duration of 1.125 s does not correspond to an integer number of periods. The number of samples over 1.125 s is $N = 225$. Leakage is present. The largest components appear at the integer indices closest to $k_1 = NF = Nf_1/S = 11.25$ (i.e., $k = 11$ or 9.78 Hz) and $k_2 = Nf_2/S = 56.25$ (i.e., $k = 56$ or 49.78 Hz).

Comment: The spectra reveal that the longest duration (1.125 s) also produces the smallest leakage.

16.6 The DFT of Nonperiodic Signals

The Fourier transform $X(f)$ of a nonperiodic signal $x(t)$ is continuous. To find the DFT, $x(t)$ must be sampled over a finite duration (N samples). The spectrum of a sampled signal over its principal period $(-0.5S,\ 0.5S)$ corresponds to the spectrum of the analog signal $SX(f)$, provided $x(t)$ is band-limited to $B < 0.5S$. In practice, no analog signal is truly band-limited. As a result, if $x[n]$ corresponds to N samples of the analog signal $x(t)$, obtained at the sampling rate S, the DFT $X_{\text{DFT}}[k]$ of $x[n]$ yields essentially the Fourier series of its periodic extension, and is only approximately related to $X(f)$ by

$$X_{\text{DFT}}[k] \approx SX(f)|_{f=kS/N}, \qquad 0 \le k < 0.5N \ (0 \le f < 0.5S) \tag{16.19}$$

REVIEW PANEL 16.17

The DFT of a Sampled Aperiodic Signal $x(t)$ Approximates Its Scaled Transform $SX(f)$

To find the DFT of an arbitrary signal with some confidence, we must decide on the number of samples N and the sampling rate S, based on both theoretical considerations and practical compromises. For example, one way to choose a sampling rate is based on energy considerations. We pick the sampling rate as $2B$ Hz, where the frequency range up to B Hz contains a significant fraction P of the signal energy. The number of samples should cover a large enough duration to include significant signal values.

16.6.1 Spectral Spacing and Zero-Padding

Often, the spectral spacing S/N is not small enough to make appropriate visual comparisons with the analog spectrum $X(f)$, and we need a denser or interpolated version of the DFT spectrum. To decrease the spectral

PGR2026-00033
Krisp Technologies EX1024 Page 566

spacing, we must choose a larger number of samples N. This increase in N cannot come about by increasing the sampling rate S (which would leave the spectral spacing S/N unchanged), but by increasing the duration over which we sample the signal. In other words, to reduce the frequency spacing, we must sample the signal for a longer duration at the given sampling rate. However, if the original signal is of finite duration, we can still increase N by appending zeros (zero-padding). Appending zeros does not improve accuracy because it adds no new signal information. It only decreases the spectral spacing and thus interpolates the DFT at a denser set of frequencies. To improve the accuracy of the DFT results, we must increase the number of signal samples by sampling the signal for a longer time (and not just zero-padding).

REVIEW PANEL 16.18

Signal Zero-Padding Corresponds to an Interpolated Spectrum

Zero-padding reduces the spectral spacing. It does not improve accuracy of the DFT results.
To improve accuracy, we need more signal samples (not zeros).

EXAMPLE 16.8 (DFT of Finite-Duration Signals)

(a) (Spectral Spacing) A 3-s signal is sampled at $S = 100$ Hz. The maximum spectral spacing is to be $\Delta f = 0.25$ Hz. How many samples are needed for the DFT and FFT?

If we use the DFT, the number of samples is $N = S/\Delta f = 400$. Since the 3-s signal gives only 300 signal samples, we must add 100 padded zeros. The spectral spacing is $S/N = 0.1953$ Hz as required. If we use the FFT, we need $N_{\text{FFT}} = 512$ samples (the next higher power of 2). There are now 212 padded zeros, and the spectral spacing is $S/N_{\text{FFT}} = 0.1953$ Hz (less than required).

(b) Let $x(t) = \text{tri}(t)$. Its Fourier transform is $X(f) = \text{sinc}^2(f)$. Let us choose $S = 4$ Hz and $N = 8$. To obtain samples of $x(t)$ starting at $t = 0$, we sample the periodic extension of $x(t)$ as illustrated in Figure E16.8B and obtain $t_s X_{\text{DFT}}[k]$ to give

$$\begin{aligned} x[n] &= \{1,\ 0.75,\ 0.5,\ 0.25,\ 0,\ 0.25,\ 0.5,\ 0.75\} \\ X(f) \approx t_s X_{\text{DFT}}[k] &= \{1,\ \underbrace{0.4268,}_{\text{off by 5.3\%}}\ 0,\ \underbrace{0.0732,}_{\text{off by 62.6\%}}\ \underbrace{0,}_{k=N/2}\ 0.0732,\ 0,\ 0.4268\} \end{aligned}$$

Since highest frequency present in the DFT spectrum is $0.5S = 2$ Hz, the DFT results are listed only up to $k = 4$. Since the frequency spacing is $S/N = 0.5$ Hz, we compare $t_s X_{\text{DFT}}[k]$ with $X(kS/N) = \text{sinc}^2(0.5k)$. At $k = 0$ (dc) and $k = 2$ (1 Hz), we see a perfect match. At $k = 1$ (0.5 Hz), $t_s X_{\text{DFT}}[k]$ is in error by about 5.3%, but at $k = 3$ (1.5 Hz), the error is a whopping 62.6%.

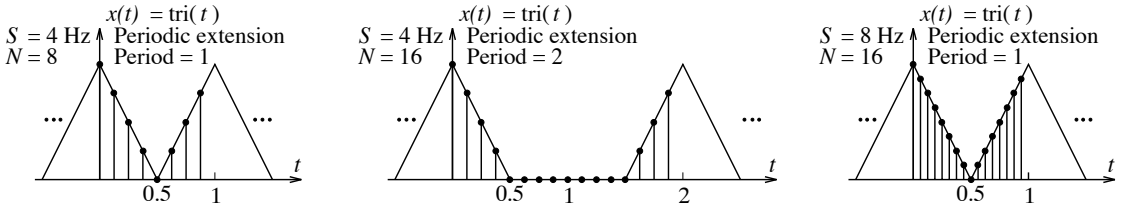


Figure E16.8B The triangular pulse signal for Example 16.8(b)

- **(Reducing Spectral Spacing)** Let us decrease the spectral spacing by zero-padding to increase the number of samples to $N = 16$. We must sample the periodic extension of the zero-padded

PGR2026-00033
Krisp Technologies EX1024 Page 567

signal, as shown in Figure E16.8B, to give

$$x[n] = \{1,\ 0.75,\ 0.5,\ 0.25,\ 0,\ \underbrace{0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0}_{8\text{ zeros}},\ 0.25,\ 0.5,\ 0.75\}$$

Note how the padded zeros appear *in the middle*. Since highest frequency present in the DFT spectrum is 2 Hz, the DFT results are listed only up to the folding index $k = 8$.

$$t_s X_{\text{DFT}}[k] = \{1,\ 0.8211,\ \underbrace{0.4268,}_{\text{off by 5.3\%}}\ 0.1012,\ 0,\ 0.0452,\ \underbrace{0.0732,}_{\text{off by 62.6\%}}\ 0.0325,\ \underbrace{0,}_{k=N/2}\ \ldots\}$$

The frequency separation is reduced to $S/N = 0.25$ Hz. Compared with $X(kf_0) = \text{sinc}^2(0.25k)$, the DFT results for $k = 2$ (0.5 Hz) and $k = 6$ (1.5 Hz) are still off by 5.3% and 62.6%, respectively. In other words, zero-padding reduces the spectral spacing, but the DFT results are no more accurate. To improve the accuracy, we must pick more signal samples.

- **(Improving Accuracy)** If we choose $S = 8$ Hz and $N = 16$, we obtain 16 samples shown in Figure E16.8B. We list the 16-sample DFT up to $k = 8$ (corresponding to the highest frequency of 4 Hz present in the DFT):

$$t_s X_{\text{DFT}}[k] = \{1,\ \underbrace{0.4105,}_{\text{off by 1.3\%}}\ 0,\ \underbrace{0.0506,}_{\text{off by 12.4\%}}\ 0,\ \underbrace{0.0226,}_{\text{off by 39.4\%}}\ 0,\ \underbrace{0.0162,}_{\text{off by 96.4\%}}\ \underbrace{0,}_{k=N/2}\ \ldots\}$$

The DFT spectral spacing is still $S/N = 0.5$ Hz. In comparison with $X(k/2) = \text{sinc}^2(0.5k)$, the error in the DFT results for the 0.5-Hz component ($k = 1$) and the 1.5-Hz component ($k = 3$) is now only about 1.3% and 12.4%, respectively. In other words, increasing the number of signal samples improves the accuracy of the DFT results. However, the error at 2.5 Hz ($k = 5$) and 3.5 Hz ($k = 7$) is 39.4% and 96.4%, respectively, and implies that the effects of aliasing are more predominant at frequencies closer to the folding frequency.

(c) Consider the signal $x(t) = e^{-t}u(t)$ whose Fourier transform is $X(f) = 1/(1 + j2\pi f)$. Since the energy E in $x(t)$ equals 1, we use Parseval's relation to estimate the bandwidth B that contains the fraction P of this energy as

$$\int_{-B}^{B} \frac{1}{1 + 4\pi^2 f^2}\, df = P \qquad\qquad B = \frac{\tan(0.5\pi P)}{2\pi}$$

1. If we choose B to contain 95% of the signal energy ($P = 0.95$), we find $B = 12.71/2\pi = 2.02$ Hz. Then, $S > 4.04$ Hz. Let us choose $S = 5$ Hz. For a spectral spacing of 1 Hz, we have $S/N = 1$ Hz and $N = 5$. So, we sample $x(t)$ at intervals of $t_s = 1/S = 0.2$ s, starting at $t = 0$, to obtain $x[n]$. Since $x(t)$ is discontinuous at $t = 0$, we pick $x[0] = 0.5$, not 1. The DFT results based on this set of choices will not be very good because with $N = 5$ we sample only a 1-s segment of $x(t)$.
2. A better choice is $N = 15$, a 3-s duration over which $x(t)$ decays to 0.05. A more practical choice is $N = 16$ (a power of 2, which allows efficient computation of the DFT, using the FFT algorithm). This gives a spectral spacing of $S/N = 5/16$ Hz. Our rule of thumb ($N > 8M$) suggests that with $N = 16$, the DFT values $X_{\text{DFT}}[1]$ and $X_{\text{DFT}}[2]$ should show an error of only about 5%. We see that $t_s X_{\text{DFT}}[k]$ does compare well with $X(f)$ (see Figure E16.8C), even though the effects of aliasing are still evident.

PGR2026-00033
Krisp Technologies EX1024 Page 568

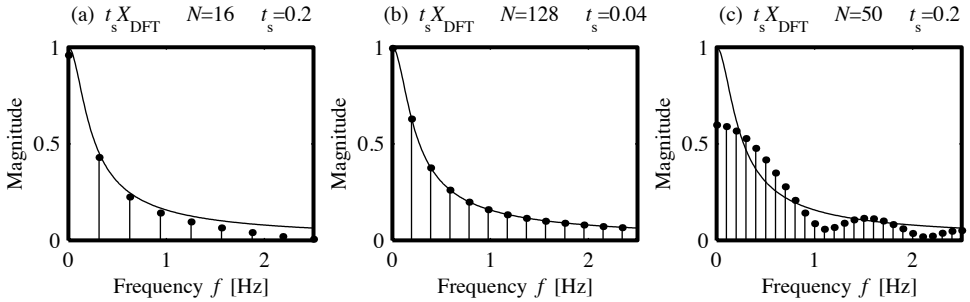


Figure E16.8C DFT results for Example 16.8(c)

3. To improve the results (and minimize aliasing), we must increase S. For example, if we require the highest frequency based on 99% of the signal energy, we obtain $B = 63.6567/2\pi = 10.13$ Hz. Based on this, let us choose $S = 25$ Hz. If we sample $x(t)$ over 5 s (by which time it decays to less that 0.01), we compute $N = (25)(5) = 125$. Choosing $N = 128$ (the next higher power of 2), we find that the 128-point DFT result $t_s X_{\text{DFT}}[k]$ is almost identical to the true spectrum $X(f)$.

4. What would happen if we choose $S = 5$ Hz and five signal samples, but reduce the spectral spacing by zero-padding to give $N = 50$? The 50-point DFT clearly shows the effects of truncation (as wiggles) and is a poor match to the true spectrum. This confirms that improved accuracy does not come by zero-padding but by including more signal samples.

16.7 Spectral Smoothing by Time Windows

Sampling an analog signal for a finite number of samples N is equivalent to multiplying the samples by a rectangular N-point window. Due to abrupt truncation of this rectangular window, the spectrum of the windowed signal shows a mainlobe and sidelobes that do not decay rapidly enough. This phenomenon is similar to the Gibbs effect, which arises during the reconstruction of periodic signals from a finite number of harmonics (an abrupt truncation of its spectrum). Just as Fourier series reconstruction uses tapered *spectral windows* to smooth the time signal, the DFT uses tapered *time-domain windows* to smooth the spectrum, but at the expense of making it broader. This is another manifestation of leakage in that the spectral energy is distributed (leaked) over a wider frequency range.

Unlike windows for Fourier series smoothing, we are not constrained by odd-length windows. As shown in Figure 16.5, an N-point DFT window is actually generated from a symmetric $(N + 1)$-point window (sampled over N intervals and symmetric about its midpoint) with its last sample discarded (in keeping with the implied periodicity of the signal and the window itself). To apply a window, we position it over the signal samples and create their pointwise product.

16.7.1 Performance Characteristics of Windows

The spectrum of all windows shows a mainlobe and decaying sidelobes, as illustrated in Figure 16.5. Measures of magnitude are often normalized by the peak magnitude P and expressed in decibels (dB) (as a gain or attenuation). These measures include the peak value P, the peak sidelobe level, the 3-dB level and the 6-dB level, and the high-frequency decay rate (in dB/decade or dB/octave). Measures of spectral width include the 3-dB width W_3, 6-dB width W_6, the width W_S to reach the peak sidelobe level (PSL), and the mainlobe width W_M. These measures are illustrated in Figure 16.6.

PGR2026-00033
Krisp Technologies EX1024 Page 569

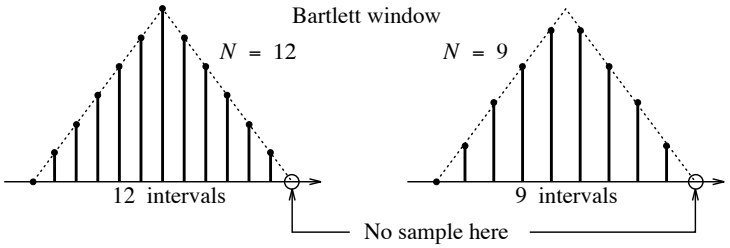


Figure 16.5 Features of a DFT window

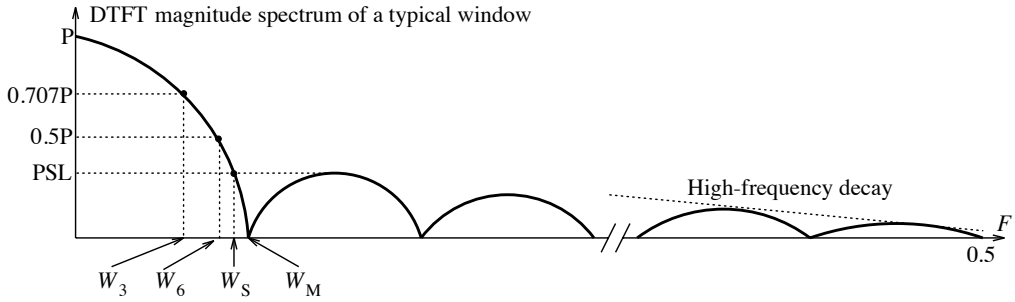


Figure 16.6 DFT spectrum of a typical window

Other measures of window performance include the **coherent gain** (CG), **equivalent noise bandwidth** (ENBW), and the **scallop loss** (SL). For an N-point window $w[n]$, these measures are defined by

$$\text{CG} = \frac{1}{N}\sum_{k=0}^{N-1}|w[k]| \qquad \text{ENBW} = \frac{N\sum_{k=0}^{N-1}|w[k]|^2}{\left|\sum_{k=0}^{N-1}w[k]\right|^2} \qquad \text{SL} = 20\log\frac{\left|\sum_{k=0}^{N-1}w[k]e^{-j\pi k/N}\right|}{\sum_{k=0}^{N-1}w[k]}\ \text{dB} \tag{16.20}$$

The reciprocal of the equivalent noise bandwidth is also called the **processing gain**. The larger the processing gain, the easier it is to reliably detect a signal in the presence of noise.

As we increase the window length, the mainlobe width of all windows decreases, but the peak sidelobe level remains more or less constant. Ideally, for a given window length, the spectrum of a window should approach an impulse with as narrow (and tall) a mainlobe as possible, and as small a peak sidelobe level as possible. The aim is to pack as much energy in a narrow mainlobe as possible and make the sidelobe level as small as possible. These are conflicting requirements in that a narrow mainlobe width also translates to a higher peak sidelobe level. Some DFT windows and their spectral characteristics are illustrated in Figure 16.7, and summarized in Table 16.3.

REVIEW PANEL 16.19

A Window Should Maximize the Mainlobe Energy and Minimize the Sidelobe Energy

A narrow mainlobe and small sidelobe are conflicting requirements for the spectrum of a window.

PGR2026-00033
Krisp Technologies EX1024 Page 570

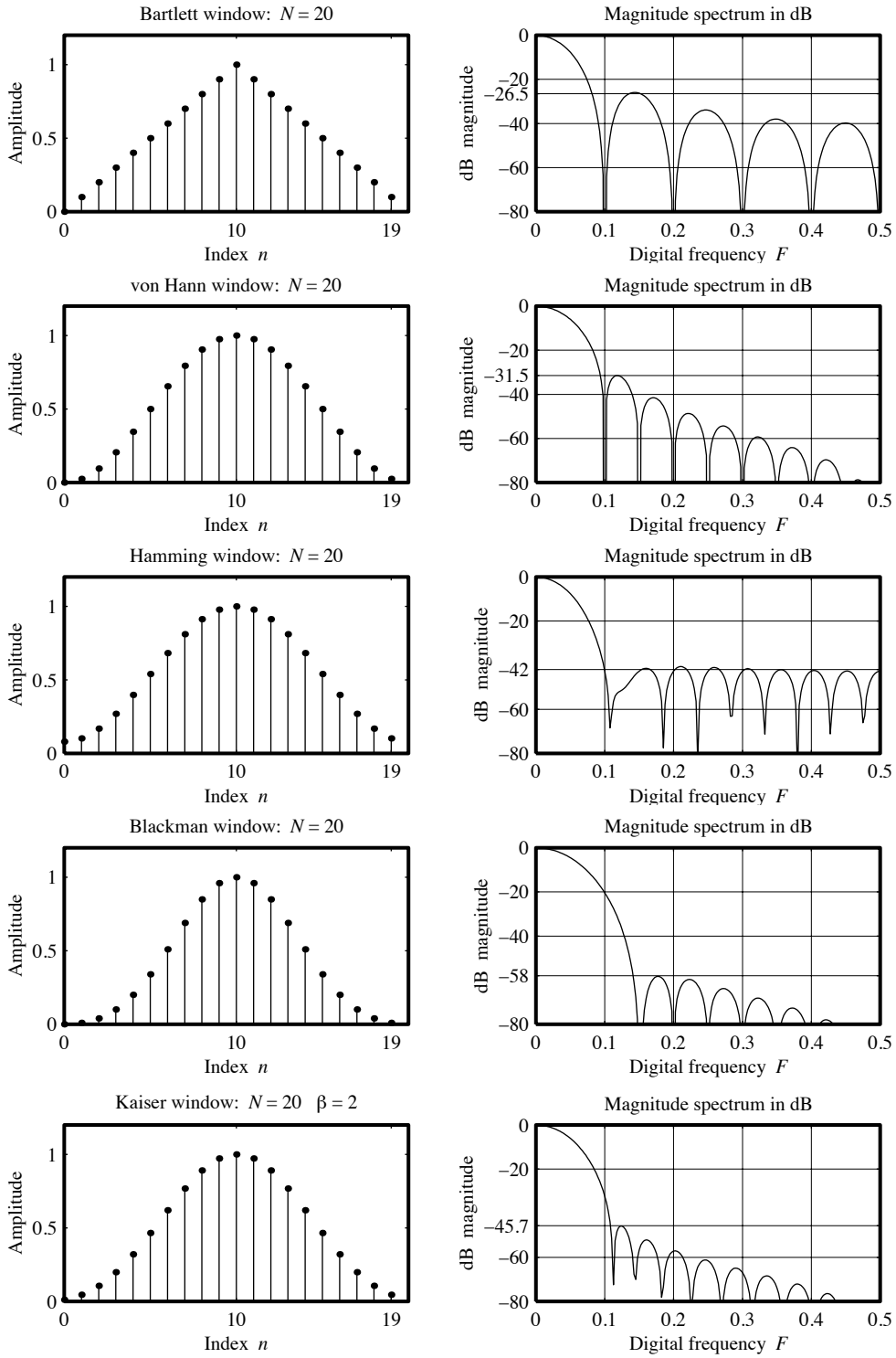


Figure 16.7 Commonly used DFT windows and their spectral characteristics

PGR2026-00033
Krisp Technologies EX1024 Page 571

Table 16.3 Some Commonly Used N-Point DFT Windows

Entry	Window	Expression for $w[n]$	$W_M = \frac{K}{N}$	Normalized Peak Sidelobe
1	Boxcar	1	$2/N$	$0.2172 \approx -13.3$ dB
2	Bartlett	$1 - \frac{2\|k\|}{N}$	$4/N$	$0.0472 \approx -26.5$ dB
3	von Hann	$0.5 + 0.5\cos(\frac{2\pi k}{N})$	$4/N$	$0.0267 \approx -31.5$ dB
4	Hamming	$0.54 + 0.46\cos(\frac{2\pi k}{N})$	$4/N$	$0.0073 \approx -42.7$ dB
5	Blackman	$0.42 + 0.5\cos(\frac{2\pi k}{N}) + 0.08\cos(\frac{4\pi k}{N})$	$6/N$	$0.0012 \approx -58.1$ dB
6	Kaiser	$\frac{I_0(\pi\beta\sqrt{1-(2k/N)^2})}{I_0(\pi\beta)}$	$\frac{2\sqrt{1+\beta^2}}{N}$	$\frac{0.22\pi\beta}{\sinh(\pi\beta)} \approx -45.7$ dB (for $\beta = 2$)

NOTES: $k = 0.5N - n$, where $n = 0, 1, \ldots, N-1$. W_M is the mainlobe width.
For the Kaiser window, $I_0(.)$ is the modified Bessel function of order zero.
For the Kaiser window, the parameter β controls the peak sidelobe level.
The von Hann window is also known as the *Hanning* window.

16.7.2 The Spectrum of Windowed Sinusoids

Consider the signal $x[n] = \cos(2\pi n F_0),\ F_0 = k_0/M$. Its DTFT is $X_p(F) = 0.5\delta(F - F_0) + 0.5\delta(F + F_0)$. If this sinusoid is windowed by an N-point window $w[n]$ whose spectrum is $W(F)$, the DTFT of the windowed signal is given by the periodic convolution

$$X_w(F) = X_p(F) \circledast W(F) = 0.5W(F - F_0) + 0.5W(F + F_0) \tag{16.21}$$

The window thus smears out the true spectrum. To obtain a windowed signal that is a replica of the spectrum of the sinusoid, we require $W(F) = \delta(F)$, which corresponds to the impractical infinite-length time window. The more the spectrum $W(F)$ of an N-point window resembles an impulse, the better the windowed spectrum matches the original.

The N-point DFT of the signal $x[n]$ may be regarded as the DTFT of the product of the infinite-length $x[n]$ and an N-point rectangular window, evaluated at $F = k/N,\ k = 0, 1, \ldots, N-1$. Since spectrum of the N-point rectangular window is $W(F) = N\frac{\text{sinc}(NF)}{\text{sinc}F}$, the DTFT spectrum of the windowed signal is

$$X_w(F) = 0.5N\,\frac{\text{sinc}[N(F - F_0)]}{\text{sinc}(F - F_0)} + 0.5N\,\frac{\text{sinc}[N(F + F_0)]}{\text{sinc}(F + F_0)} \tag{16.22}$$

The N-point DFT of the windowed sinusoid is given by $X_{\text{DFT}}[k] = X_w(F)|_{F=k/N}$. If the DFT length N equals M, the number of samples over k_0 full periods of $x[n]$, we see that $\text{sinc}[N(F - F_0)] = \text{sinc}(k - k_0)$, and this equals zero, unless $k = k_0$. Similarly, $\text{sinc}[N(F + F_0)] = \text{sinc}(k + k_0)$ is nonzero only if $k = -k_0$. The DFT thus contains only two nonzero terms and equals

$$X_{\text{DFT}}[k] = X_w(F)\Big|_{F=k/N} = 0.5N\delta[k - k_0] + 0.5N\delta[k - k_0] \quad (\text{if } N = M) \tag{16.23}$$

In other words, using an N-point rectangular window that covers an integer number of periods (M samples) of a sinusoid (i.e., with $M = N$) gives us exact results. The reason of course is that the DTFT sampling instants fall on the nulls of the sinc spectrum. If the window length N does not equal M (an integer number of periods), the sampling instants will fall between the nulls, and since the sidelobes of the sinc function are large, the DFT results will show considerable leakage. To reduce the effects of leakage, we must use windows whose spectrum shows small sidelobe levels.

PGR2026-00033
Krisp Technologies EX1024 Page 572

16.7.3 Resolution

Windows are often used to reduce the effects of leakage and improve resolution. **Frequency resolution** refers to our ability to clearly distinguish between two closely spaced sinusoids of similar amplitudes. **Dynamic-range resolution** refers to our ability to resolve large differences in signal amplitudes. The spectrum of all windows reveals a mainlobe and smaller sidelobes. It smears out the true spectrum and makes components separated by less than the mainlobe width indistinguishable. The rectangular window yields the best frequency resolution for a given length N since it has the smallest mainlobe. However, it also has the largest peak sidelobe level of any window. This leads to significant leakage and the worst dynamic-range resolution because small amplitude signals can get masked by the sidelobes of the window.

Tapered windows with less abrupt truncation show reduced sidelobe levels and lead to reduced leakage and improved dynamic-range resolution. They also show increased mainlobe widths W_M, leading to poorer frequency resolution. The choice of a window is based on a compromise between the two conflicting requirements of minimizing the mainlobe width (improving frequency resolution) and minimizing the sidelobe magnitude (improving dynamic-range resolution).

REVIEW PANEL 16.20

Good Frequency Resolution and Dynamic-Range Resolution Are Conflicting Requirements

To improve frequency resolution, use windows with narrow mainlobes.
To improve dynamic-range resolution, use windows with small sidelobes.

The mainlobe width of all windows decreases as we increase the window length. However, the peak sidelobe level remains more or less constant. To achieve a frequency resolution of Δf, the digital frequency $\Delta F = \Delta f/S$ must equal or exceed the mainlobe width W_M of the window. This yields the window length N. For a given window to achieve the same frequency resolution as the rectangular window, we require a larger window length (a smaller mainlobe width) and hence a larger signal length. The increase in signal length must come by choosing more signal samples (and not by zero-padding). To achieve a given dynamic-range resolution, however, we must select a window with small sidelobes, regardless of the window length.

REVIEW PANEL 16.21

The Smallest Frequency We Can Resolve Depends on the Mainlobe Width of the Window

To resolve frequencies separated by Δf, we require $\frac{\Delta f}{S} = W_M = \frac{K}{N}$ (window mainlobe width).
K depends on the window. To decrease Δf, increase N (more signal samples, not zero-padding).

EXAMPLE 16.9 (Frequency Resolution)

The signal $x(t) = A_1 \cos(2\pi f_0 t) + A_2 \cos[2\pi(f_0 + \Delta f)t]$, where $A_1 = A_2 = 1$, $f_0 = 30$ Hz is sampled at the rate $S = 128$ Hz. We acquire N samples, zero-pad them to length N_{FFT}, and obtain the N_{FFT}-point FFT.

1. What is the smallest Δf that can be resolved for

$N = 256$, $N_{\text{FFT}} = 2048$ using a rectangular and von Hann (Hanning) window.
$N = 512$, $N_{\text{FFT}} = 2048$ using a rectangular and von Hann window.
$N = 256$, $N_{\text{FFT}} = 4096$ using a rectangular and von Hann window.

2. How do the results change if $A_2 = 0.05$?

(a) (Frequency Resolution) Since $\Delta F = \Delta f/S = W_M$, we have $\Delta f = SW_M$. We compute:

Rectangular window: $\Delta f = SW_M = 2S/N = 1$ Hz for $N = 256$ and 0.5 Hz for $N = 512$

von Hann window: $\Delta f = SW_M = 4S/N = 2$ Hz for $N = 256$ and 1 Hz for $N = 512$

PGR2026-00033
Krisp Technologies EX1024 Page 573

Note that N_{FFT} governs only the FFT spacing S/N_{FFT}, whereas N governs only the frequency resolution S/N (which does not depend on the zero-padded length). Figure E16.9A shows the FFT spectra, plotted as continuous curves, over a selected frequency range. We make the following remarks:

1. For a given signal length N, the rectangular window resolves a smaller Δf but also has the largest sidelobes (panels a and b). This means that the effects of leakage are more severe for a rectangular window than for any other.
2. We can resolve a smaller Δf by increasing the signal length N *alone* (panel c). To resolve the same Δf with a von Hann window, we must *double* the signal length N (panel d). This means that we can improve resolution only by increasing the number of signal samples (adding more signal information). How many more signal samples we require will depend on the desired resolution and the type of window used.
3. We cannot resolve a smaller Δf by increasing the zero-padded length N_{FFT} *alone* (panels e and f). This means that increasing the number of samples by zero-padding cannot improve resolution. Zero-padding simply interpolates the DFT at a denser set of frequencies. It simply cannot improve the accuracy of the DFT results because adding more zeros does not add more signal information.

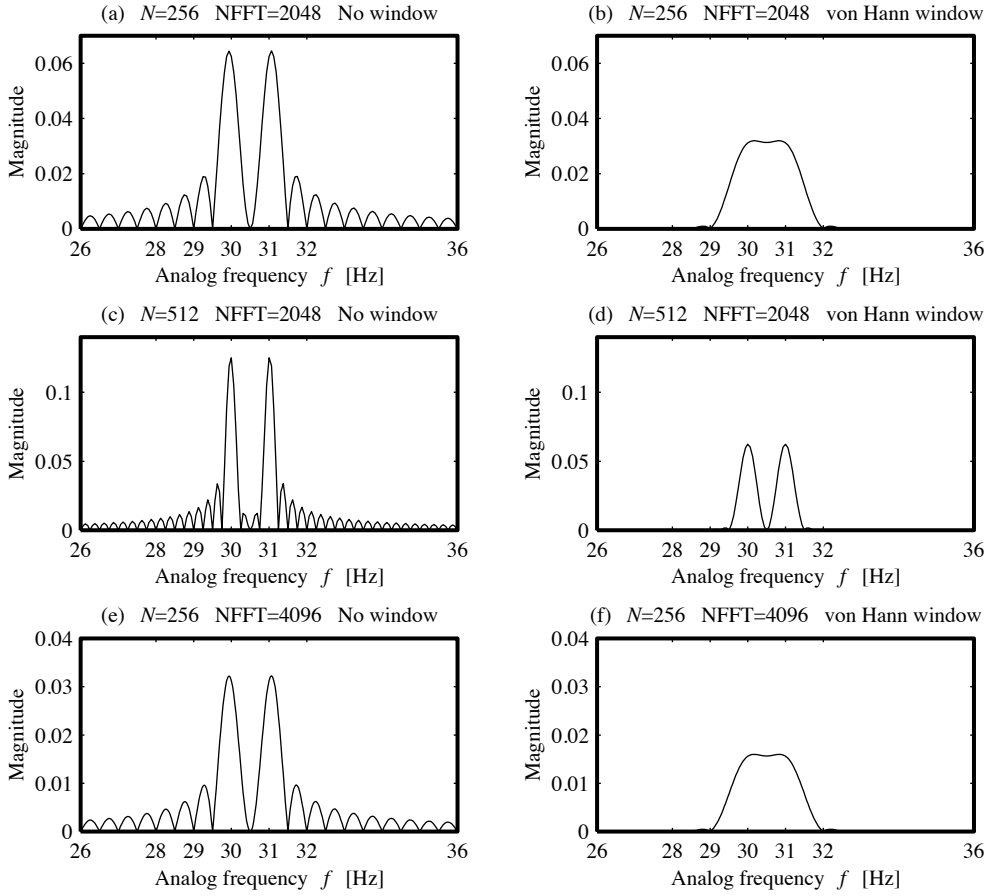


Figure E16.9A DFT spectra for Example 16.9(a)

PGR2026-00033
Krisp Technologies EX1024 Page 574

(b) (Dynamic-Range Resolution) If $A_2 = 0.05$ (26 dB below A_1), the large sidelobes of the rectangular window (13 dB below the peak) will mask the second peak at 31 Hz, even if we increase N and N_{FFT}. This is illustrated in Figure E16.9B(a) (where the peak magnitude is normalized to unity, or 0 dB) for $N = 512$ and $N_{\text{FFT}} = 4096$. For the same values of N and N_{FFT}, however, the smaller sidelobes of the von Hann window (31.5 dB below the peak) do allow us to resolve two distinct peaks in the windowed spectrum, as shown in Figure E16.9B(b).

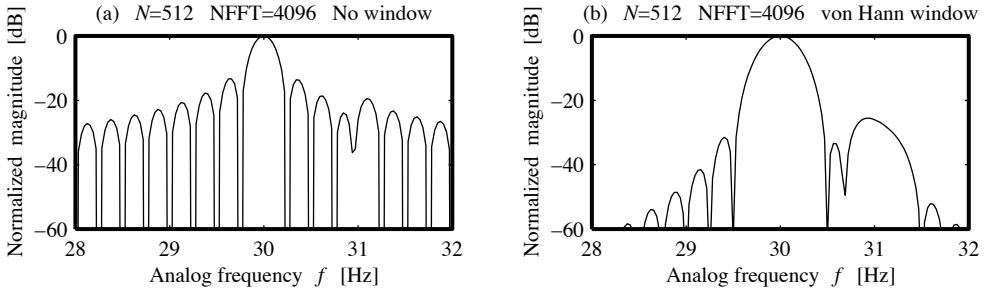


Figure E16.9B DFT spectra for Example 16.9(b)

16.7.4 Detecting Hidden Periodicity Using the DFT

Given an analog signal $x(t)$ known to contain periodic components, how do we estimate their frequencies and magnitude? There are several ways. Most rely on statistical estimates, especially if the signal $x(t)$ is also corrupted by noise. Here is a simpler, perhaps more intuitive—but by no means the best—approach based on the effects of aliasing. The location and magnitude of the components in the DFT spectrum can change with the sampling rate if this rate is below the Nyquist rate. Due to aliasing, the spectrum may not drop to zero at $0.5S$ and may even show increased magnitudes as we move toward the folding frequency. If we try to minimize the effects of noise by using a lowpass filter, we must ensure that its cutoff frequency exceeds the frequency of all the components of interest present in $x(t)$. We have no a priori way of doing this. A better way, if the data can be acquired repeatedly, is to use the average of many runs. Averaging minimizes noise while preserving the integrity of the signal.

A crude estimate of the sampling frequency may be obtained by observing the most rapidly varying portions of the signal. Failing this, we choose an arbitrary but small sampling frequency, sample $x(t)$, and observe the DFT spectrum. We repeat the process with increasing sampling rates and observe how the DFT spectrum changes, and when the spectrum shows little change in the *location* of its spectral components, we have the right spectrum and the right sampling frequency. This trial-and-error method is illustrated in Figure 16.8 and actually depends on aliasing for its success.

If $x(t)$ is a sinusoid, its magnitude A is computed from $X_{\text{DFT}}[k_0] = 0.5NA$, where the index k_0 corresponds to the peak in the N-point DFT. However, if other nonperiodic signals are also present in $x(t)$, this may not yield a correct result. A better estimate of the magnitude of the sinusoid may be obtained by comparing two DFT results of different lengths—say, $N_1 = N$ and $N_2 = 2N$. The N_1-point DFT at the index k_1 of the peak will equal $0.5N_1A$ plus a contribution due to the nonperiodic signals. Similarly, the N_2-point DFT will show a peak at k_2 (where $k_2 = 2k_1$ if $N_2 = 2N_1$), and its value will equal $0.5N_2A$ plus a contribution due to the nonperiodic signals. If the nonperiodic components do not affect the spectrum significantly, the difference in these two values will cancel out the contribution due to the nonperiodic components, and yield an estimate for the magnitude of the sinusoid from

$$X_{\text{DFT2}}[k_2] - X_{\text{DFT1}}[k_1] = 0.5N_2A - 0.5N_1A \tag{16.24}$$

PGR2026-00033
Krisp Technologies EX1024 Page 575

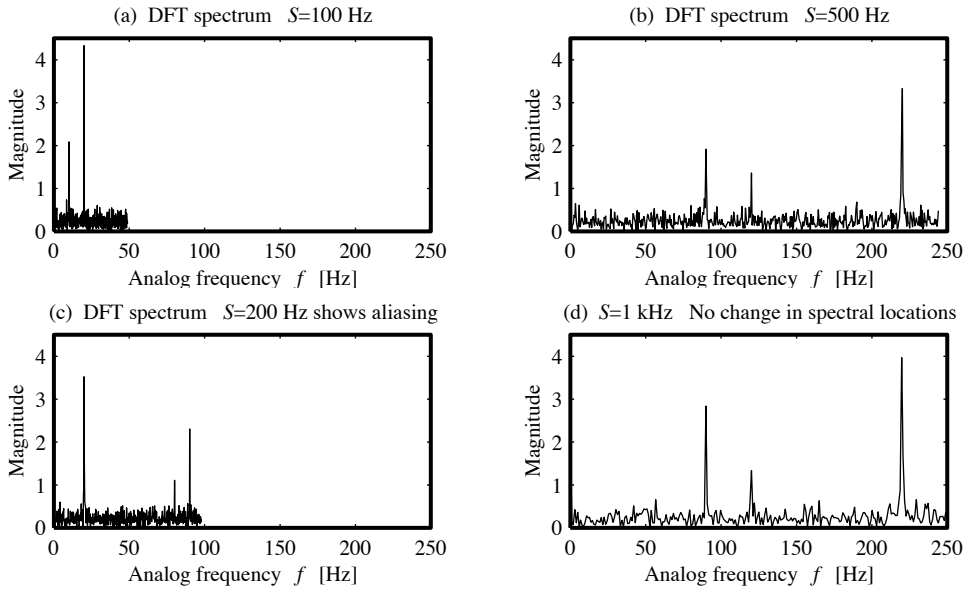


Figure 16.8 Trial-and-error method for obtaining the DFT spectrum

EXAMPLE 16.10 (Detecting Hidden Periodicity)
A signal $x(t)$ is known to contain a sinusoidal component. The 80-point DFT result and a comparison of the 80-point and 160-point DFT are shown in Figure E16.10. Estimate the frequency and magnitude of the sinusoid and its DFT index. The sampling rate is $S = 10$ Hz.

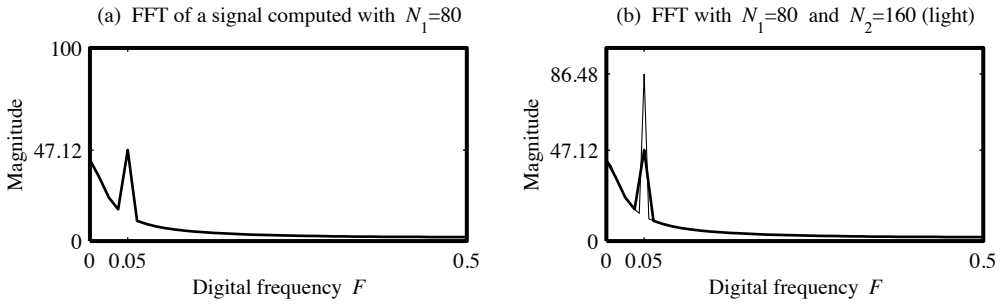


Figure E16.10 DFT spectra for Example 16.10

The comparison of the two DFT results suggests a peak at $F = 0.05$ and the presence of a sinusoid. Since the sampling rate is $S = 10$ Hz, the frequency of the sinusoid is $f = FS = 0.5$ Hz. Let $N_1 = 80$ and $N_2 = 160$. The peak in the N_1-point DFT occurs at the index $k_1 = 4$ because $F = 0.05 = k_1/N_1 = 4/80$. Similarly, the peak in the N_2-point DFT occurs at the index $k_2 = 8$ because $F = 0.05 = k_2/N_2 = 8/160$. Since the two spectra do not differ much, except near the peak, the difference in the peak values allows us to compute the peak value A of the sinusoid from

$$X_{\text{DFT2}}[k_2] - X_{\text{DFT1}}[k_1] = 86.48 - 47.12 = 0.5N_2A - 0.5N_1A = 40A$$

Thus, $A = 0.984$ and implies the presence of the 0.5-Hz sinusoidal component $0.984\cos(\pi t + \theta)$.

PGR2026-00033
Krisp Technologies EX1024 Page 576

Comment: The DFT results shown in Figure E16.10 are actually for the signal $x(t) = \cos(\pi t) + e^{-t}$, sampled at $S = 10$ Hz. The sinusoidal component has unit peak value, and the DFT estimate ($A = 0.984$) differs from this value by less than 2%. Choosing larger DFT lengths would improve the accuracy of the estimate. However, the 80-point DFT alone yields the estimate $A = 47.12/40 = 1.178$ (an 18% difference), whereas the 160-point DFT alone yields $A = 86.48/80 = 1.081$ (an 8% difference).

16.8 Applications in Signal Processing

The applications of the DFT and FFT span a wide variety of disciplines. Here, we briefly describe some applications directly related to digital signal processing.

16.8.1 Convolution of Long Sequences

A situation that often arises in practice is the processing of a long stream of incoming data by a filter whose impulse response is much shorter than that of the incoming data. The convolution of a short sequence $h[n]$ of length N (such as an averaging filter) with a very long sequence $x[n]$ of length $L \gg N$ (such as an incoming stream of data) can involve large amounts of computation and memory. There are two preferred alternatives, both of which are based on sectioning the long sequence $x[n]$ into shorter ones. The DFT offers a useful means of finding such a convolution. It even allows on-line implementation if we can tolerate a small processing delay.

The Overlap-Add Method

Suppose $h[n]$ is of length N, and the length of $x[n]$ is $L = mN$ (if not, we can always zero-pad it to this length). We partition $x[n]$ into m segments $x_0[n], x_1[n], \ldots, x_{m-1}[n]$, each of length N. We find the regular convolution of each section with $h[n]$ to give the partial results $y_0[n], y_1[n], \ldots, y_{m-1}[n]$. Using superposition, the total convolution is the sum of their shifted (by multiples of N) versions

$$y[n] = y_0[n] + y_1[n-N] + y_2[n-2N] + \cdots + y_{m-1}[n-(m-1)N] \tag{16.25}$$

Since each regular convolution contains $2N-1$ samples, we zero-pad $h[n]$ and each section $x_k[n]$ with $N-1$ zeros before finding $y_k[n]$ using the FFT. Splitting $x[n]$ into equal-length segments is not a strict requirement. We may use sections of different lengths, provided we keep track of how much each partial convolution must be shifted before adding the results.

The Overlap-Save Method

The regular convolution of sequences of length L and N has $L+N-1$ samples. If $L > N$ and we zero-pad the second sequence to length L, their periodic convolution has $2L-1$ samples. Its first $N-1$ samples are contaminated by wraparound, and the rest correspond to the regular convolution. To understand this, let $L = 16$ and $N = 7$. If we pad N by nine zeros, their regular convolution has 31 (or $2L-1$) samples with nine trailing zeros ($L - N = 9$). For periodic convolution, 15 samples ($L - 1 = 15$) are wrapped around. Since the last nine (or $L - N$) are zeros, only the first six samples ($L - N - (N-1) = N - 1 = 6$) of the periodic convolution are contaminated by wraparound. This idea is the basis for the overlap-save method. First, we add $N-1$ *leading* zeros to the longer sequence $x[n]$ and section it into k *overlapping* (by $N-1$) segments of length M. Typically, we choose $M \approx 2N$. Next, we zero-pad $h[n]$ (with *trailing* zeros) to length M, and find the *periodic* convolution of $h[n]$ with each section of $x[n]$. Finally, we discard the first $N-1$ (contaminated) samples from each convolution and glue (concatenate) the results to give the required convolution.

PGR2026-00033
Krisp Technologies EX1024 Page 577

In either method, the FFT of the shorter sequence need be found only once, stored, and reused for all subsequent partial convolutions. Both methods allow on-line implementation if we can tolerate a small processing delay that equals the time required for each section of the long sequence to arrive at the processor (assuming the time taken for finding the partial convolutions is less than this processing delay). The correlation of two sequences may also be found in exactly the same manner, using either method, provided we use a folded version of one sequence.

EXAMPLE 16.11 (Overlap-Add and Overlap-Save Methods of Convolution)
Let $x[n] = \{1, 2, 3, 3, 4, 5\}$ and $h[n] = \{1, 1, 1\}$. Here $L = 6$ and $N = 3$.

(a) To find their convolution using the overlap-add method, we section $x[n]$ into two sequences given by $x_0[n] = \{1, 2, 3\}$ and $x_1[n] = \{3, 4, 5\}$, and obtain the two convolution results:

$$y_0[n] = x_0[n] \star h[n] = \{1, 3, 6, 5, 3\} \qquad y_1[n] = x_1[n] \star h[n] = \{3, 7, 12, 9, 5\}$$

Shifting and superposition results in the required convolution $y[n]$ as

$$y[n] = y_0[n] + y_1[n-3] = \left\{ \begin{array}{r} 1, 3, 6, 5, 3 \\ \quad 3, 7, 12, 9, 5 \end{array} \right\} = \{1, 3, 6, 8, 10, 12, 9, 5\}$$

This result can be confirmed using any of the convolution algorithms described in Chapter 7.

(b) To find their convolution using the overlap-add method, we start by creating the zero-padded sequence $x[n] = \{0, 0, 1, 2, 3, 3, 4, 5\}$. If we choose $M = 5$, we get three overlapping sections of $x[n]$ (we need to zero-pad the last one) described by

$$x_0[n] = \{0, 0, 1, 2, 3\} \qquad x_1[n] = \{2, 3, 3, 4, 5\} \qquad x_2[n] = \{4, 5, 0, 0, 0\}$$

The zero-padded $h[n]$ becomes $h[n] = \{1, 1, 1, 0, 0\}$. Periodic convolution gives

$$\begin{aligned} x_0[n] \circledast h[n] &= \{5, 3, 1, 3, 6\} \\ x_1[n] \circledast h[n] &= \{11, 10, 8, 10, 12\} \\ x_2[n] \circledast h[n] &= \{4, 9, 9, 5, 0\} \end{aligned}$$

We discard the first two samples from each convolution and glue the results to obtain

$$y[n] = x[n] \star h[n] = \{1, 3, 6, 8, 10, 12, 9, 5, 0\}$$

Note that the last sample (due to the zero-padding) is redundant, and may be discarded.

16.8.2 Deconvolution

Given a signal $y[n]$ that represents the output of some system with impulse response $h[n]$, how do we recover the input $x[n]$ where $y[n] = x[n] \star h[n]$? One method is to undo the effects of convolution using deconvolution. The time-domain approach to deconvolution was studied in Chapters 6 and 7. Here, we examine a frequency-domain alternative based on the DFT (or FFT).

The idea is to transform the convolution relation using the FFT to obtain $Y_{\text{FFT}}[k] = X_{\text{FFT}}[k]H_{\text{FFT}}[k]$, compute $X_{\text{FFT}}[k] = Y_{\text{FFT}}[k]/H_{\text{FFT}}[k]$ by pointwise division, and then find $x[n]$ as the IFFT of $X_{\text{FFT}}[k]$. This process does work in many cases, but it has two disadvantages. First, it fails if $H_{\text{FFT}}[k]$ equals zero at some index because we get division by zero. Second, it is quite sensitive to noise in the input $x[n]$ and to the accuracy with which $y[n]$ is known.

PGR2026-00033
Krisp Technologies EX1024 Page 578

16.8.3 Band-Limited Signal Interpolation

Interpolation of $x[n]$ by M to a new signal $x_I[n]$ is equivalent to a sampling rate increase by M. If the signal has been sampled above the Nyquist rate, signal interpolation should add no new information to the spectrum. The idea of zero-padding forms the basis for an interpolation method using the DFT in the sense that the (MN)-sample DFT of the interpolated signal should contain N samples corresponding to the DFT of $x[n]$, while the rest should be zero. Thus, if we find the DFT of $x[n]$, zero-pad it (by inserting zeros about the folding index) to increase its length to MN, and find the inverse DFT of the zero-padded sequence, we should obtain the interpolated signal $x_I[n]$. This approach works well for band-limited signals (such as pure sinusoids sampled above the Nyquist rate over an integer number of periods). To implement this process, we split the N-point DFT $X_{\text{DFT}}[k]$ of $x[n]$ about the folding index $N/2$. If N is even, the folding index falls on the sample value $X[N/2]$, and it must also be split in half. We then insert enough zeros in the middle to create a padded sequence $X_{\text{zp}}[k]$ with MN samples. It has the form

$$X_{\text{zp}}[k] = \begin{cases} \{X[0], \ldots,\ X[N/2-1],\quad \underbrace{0, \ldots,\ 0,}_{(M-1)N \text{ zeros}} \quad X[N/2+1], \ldots,\ X[N-1]\} & (\text{odd } N) \\ \{X[0], \ldots,\ 0.5X[N/2],\quad \underbrace{0, \ldots,\ 0,}_{N(M-1)-1 \text{ zeros}} \quad 0.5X[N/2], \ldots,\ X[N-1]\} & (\text{even } N) \end{cases} \tag{16.26}$$

The inverse DFT of $X_{\text{zp}}[k]$ will include the factor $1/MN$ and its machine computation may show (small) imaginary parts. We therefore retain only its real part, and divide by M, to obtain the interpolated signal $x_I[n]$, which contains $M-1$ interpolated values between each sample of $x[n]$:

$$x_I[n] = \frac{1}{M}\text{Re}\Big[\text{IDFT}\{X_{\text{zp}}[k]\}\Big] \tag{16.27}$$

This method is entirely equivalent to creating a zero-interpolated signal (which produces spectrum replication), and filtering the replicated spectrum (by zeroing out the spurious images). For periodic band-limited signals sampled above the Nyquist rate for an integer number of periods, the interpolation is exact. For all others, imperfections show up as a poor match, especially near the ends, since we are actually interpolating to zero outside the signal duration.

EXAMPLE 16.12 (Signal Interpolation Using the FFT)

(a) For a sinusoid sampled over one period with four samples, we obtain the signal $x[n] = \{0, 1, 0, -1\}$. Its DFT is $X_{\text{DFT}}[k] = \{0, -j2, 0, j2\}$. To interpolate this by $M = 8$, we generate the 32-sample zero-padded sequence $Z_T = \{0, -j2, 0,\ (27 \text{ zeros}),\ 0, j2\}$. The interpolated sequence (the IDFT of Z_T) shows an exact match to the sinusoid, as illustrated in Figure E16.12)(a).

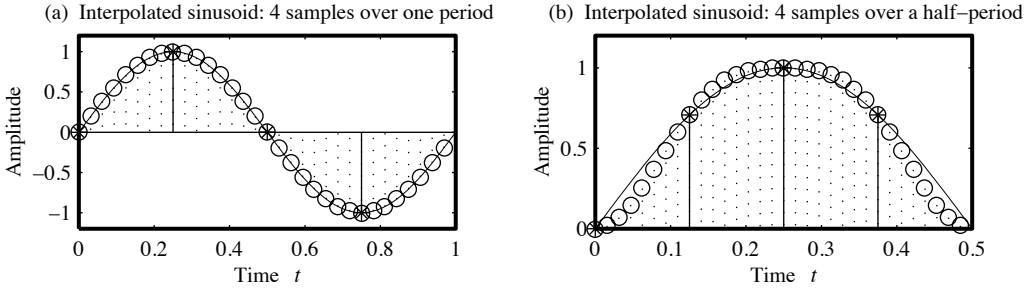


Figure E16.12 Interpolated sinusoids for Example 16.12

PGR2026-00033
Krisp Technologies EX1024 Page 579

(b) For a sinusoid sampled over a half-period with four samples, interpolation does not yield exact results, as shown in Figure E16.12(b). Since we are actually sampling one period of a full-rectified sine (the periodic extension), the signal is not band-limited, and the chosen sampling frequency is too low. This shows up as a poor match, especially near the ends of the sequence.

16.8.4 The Discrete Hilbert Transform

The Hilbert transform was discussed in Chapter 10. The **discrete Hilbert transform** of a sequence may be obtained by using the FFT. The Hilbert transform of $x[n]$ involves the convolution of $x[n]$ with the impulse response $h[n]$ of the Hilbert transformer whose spectrum is $H(F) = -j\,\mathrm{sgn}(F)$. The easiest way to perform this convolution is by FFT methods. We find the N-point FFT of $x[n]$ and multiply by the *periodic extension* (from $k = 0$ to $k = N - 1$) of N samples of $\mathrm{sgn}(F)$, which may be written as

$$\mathrm{sgn}[k] = \{0,\ \underbrace{1, 1, \ldots, 1,}_{\frac{N}{2}-2 \text{ samples}}\ 0,\ \underbrace{-1, -1, \ldots, -1}_{\frac{N}{2}-2 \text{ samples}}\} \tag{16.28}$$

The inverse FFT of the product, multiplied by the omitted factor $-j$, yields the Hilbert transform.

16.9 Spectrum Estimation

The **power spectral density** (PSD) $R_{xx}(f)$ of an analog power signal or random signal $x(t)$ is the Fourier transform of its autocorrelation function $r_{xx}(t)$ and is a real, non-negative, even function with $R_{xx}(0)$ equal to the average power in $x(t)$. If a signal $x(t)$ is sampled and available only over a finite duration, the best we can do is *estimate* the PSD of the underlying signal $x(t)$ from the given finite record. This is because the spectrum of finite sequences suffers from leakage and poor resolution. The PSD estimate of a noisy analog signal $x(t)$ from a finite number of its samples is based on two fundamentally different approaches. The first, **non-parametric spectrum estimation**, makes no assumptions about the data. The second, **parametric spectrum estimation**, models the data as the output of a digital filter excited by a noise input with a constant power spectral density, estimates the filter coefficients, and in so doing arrives at an estimate of the true PSD. Both rely on statistical measures to establish the quality of the estimate.

16.9.1 The Periodogram Estimate

The simplest non-parametric estimate is the **periodogram** $P[k]$ that is based on the DFT of an N-sample sequence $x[n]$. It is defined as

$$P[k] = \frac{1}{N}|X_{\mathrm{DFT}}[k]|^2 \tag{16.29}$$

Although $P[k]$ provides a good estimate for deterministic, band-limited, power signals sampled above the Nyquist rate, it yields poor estimates for noisy signals because the quality of the estimate does not improve with increasing record length N (even though the spectral spacing decreases).

EXAMPLE 16.13 (The Concept of the Periodogram)
The sequence $x[n] = \{0,\ 1,\ 0,\ -1\}$ is obtained by sampling a sinusoid for one period at twice the Nyquist rate. Find its periodogram estimate.
The DFT of $x[n]$ is $X_{\mathrm{DFT}}[k] = \{0,\ -j2,\ 0,\ j2\}$. Thus, $P[k] = \frac{1}{N}|X_{\mathrm{DFT}}[k]|^2 = \{0,\ 1,\ 0,\ 1\}$.
This is usually plotted as a *bar graph* with each sample occupying a bin width $\Delta F = 1/N = 0.25$.
The total power thus equals $\Delta F \sum P[k] = 0.5$, and matches the true power in the sinusoid.

PGR2026-00033
Krisp Technologies EX1024 Page 580

16.9.2 PSD Estimation by the Welch Method

The **Welch method** is based on the concept of averaging the results of overlapping periodogram estimates and finds widespread use in spectrum estimation. The N-sample signal $x[n]$ is sectioned into K overlapping M-sample sections $x_M[n]$, each with an overlap of D samples. A 50%– 75% overlap is common practice. Each section is windowed by an M-point window $w[n]$ to control leakage. The PSD of each windowed section is found using the periodogram $\frac{1}{M}|X_M[k]|^2$, and the K periodograms are averaged to obtain an M-point **averaged periodogram**. It is customary to normalize (divide) this result by the power in the window signal ($\sum |W_{\text{DFT}}[k]|^2$ or $\frac{1}{M}\sum |w[n]|^2$). Sections of larger length result in a smaller spectral spacing, whereas more sections result in a smoother estimate (at the risk of masking sharp details). The number of sections chosen is a trade-off between decreasing the spectral spacing and smoothing the estimate. Averaging the results of the (presumably uncorrelated) segments reduces the statistical fluctuations (but at the expense of a larger spectral spacing). In the related **Bartlett method**, the segments are neither overlapped nor windowed. Figure 16.9(a) shows the Welch PSD of a 400-point chirp signal, whose digital frequency varies from $F = 0.2$ to $F = 0.4$, using a 45% overlap and a 64-point von Hann window (shown dark), and a rectangular window (no window). Note how the windowing results in significant smoothing.

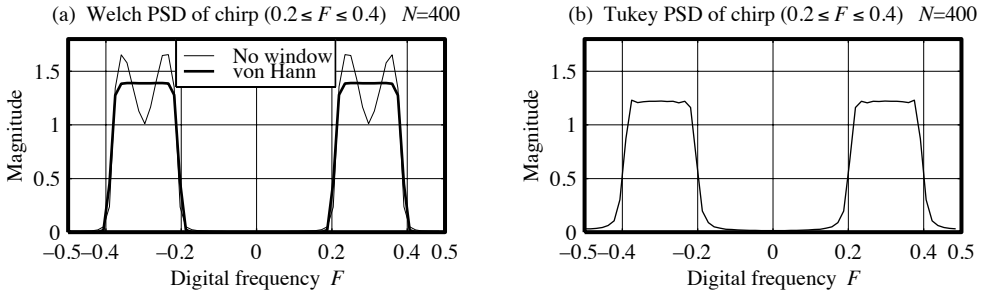


Figure 16.9 Welch and Tukey PSD of a chirp signal

16.9.3 PSD Estimation by the Blackman-Tukey Method

The **Blackman-Tukey method** relies on finding the PSD from the windowed autocorrelation, which is assumed to be zero past the window length. The N-sample signal $x[n]$ is zero-padded to get the $2N$-sample signal $y[n]$. The *linear* autocorrelation of $x[n]$ equals the *periodic* autocorrelation of $y[n]$ and is evaluated by finding the FFT $Y_{\text{FFT}}[k]$ and taking the IFFT of $Y_{\text{FFT}}[k]Y^*_{\text{FFT}}[k]$ to obtain the $2N$-sample autocorrelation estimate $r_{xx}[n]$. This autocorrelation is then windowed by an M-point window (to smooth the spectrum and reduce the effects of poor autocorrelation estimates due to the finite data length). The FFT of the M-sample windowed autocorrelation yields the **smoothed periodogram**. Using a smaller M (narrower windows) for the autocorrelation provides greater smoothing but may also mask any peaks or obscure the sharp details. Typical values of the window length M range from $M = 0.1N$ to $M = 0.5N$. Only windows whose transform is entirely positive should be used. Of the few that meet this constraint, the most commonly used is the Bartlett (triangular) window. Figure 16.9(b) shows the Tukey PSD of a 1000-point chirp signal, whose digital frequency varies from $F = 0.2$ to $F = 0.4$, using a 64-point Bartlett window.

The Welch method is more useful for detecting closely spaced peaks of similar magnitudes (*frequency resolution*). The Blackman-Tukey method is better for detecting well-separated peaks with different magnitudes (*dynamic-range resolution*). Neither is very effective for short data lengths.

PGR2026-00033
Krisp Technologies EX1024 Page 581

16.9.4 Non-Parametric System Identification

The Fourier transform $R_{yx}(f)$ of the cross-correlation $r_{yx}(t) = y(t) \star\star x(t) = y(t) \star x(-t)$ of two random signals $x(t)$ and $y(t)$ is called the **cross-spectral density**. For a system with impulse response $h(t)$, the input $x(t)$ yields the output $y(t) = x(t) \star h(t)$, and we have

$$r_{yx}(t) = y(t) \star x(-t) = h(t) \star x(t) \star x(-t) = h(t) \star r_{xx}(t) \tag{16.30}$$

The Fourier transform of both sides yields

$$R_{yx}(f) = H(f)R_{xx}(f) \qquad\qquad H(f) = \frac{R_{yx}(f)}{R_{xx}(f)} \tag{16.31}$$

This relation allows us to identify an unknown transfer function $H(f)$ as the ratio of the cross-spectral density $R_{yx}(f)$ and the PSD $R_{xx}(f)$ of the input $x(t)$. If $x(t)$ is a noise signal with a constant power spectral density, its PSD is a constant of the form $R_{xx}(f) = K$. Then, $H(f) = R_{yx}(f)/K$ is directly proportional to the cross-spectral density.

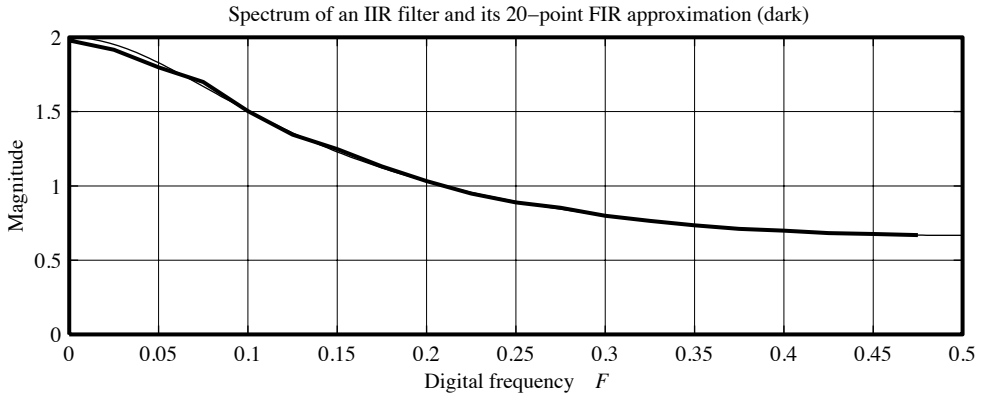


Figure 16.10 Spectrum of an IIR filter and its non-parametric FIR filter estimate

This approach is termed **non-parametric** because it presupposes no model for the system. In practice, we use the FFT to approximate $R_{xx}(f)$ and $R_{yx}(f)$ (using the Welch method, for example) by the finite N-sample sequences $R_{xx}[k]$ and $R_{yx}[k]$. As a consequence, the transfer function $H[k] = R_{yx}[k]/R_{xx}[k]$ also has N samples and describes an FIR filter. Its inverse FFT yields the N-sample impulse response $h[n]$. Figure 16.10 shows the spectrum of an IIR filter defined by $y[n] - 0.5y[n-1] = x[n]$ (or $h[n] = (0.5)^n u[n]$), and its 20-point FIR estimate obtained by using a 400-sample noise sequence and the Welch method.

16.9.5 Time-Frequency Plots

In practical situations, we are often faced with the task of finding how the spectrum of signals varies with time. A simple approach is to section the signal into overlapping segments, window each section to reduce leakage, and find the PSD (using the Welch method, for example). The PSD for each section is then staggered and stacked to generate what is called a **waterfall plot** or **time-frequency plot**. Figure 16.11 shows such a waterfall plot for the sum of a single frequency sinusoid and a chirp signal whose frequency varies linearly with time. It provides a much better visual indication of how the spectrum evolves with time.

The fundamental restriction in obtaining the time-frequency information, especially for short time records, is that *we cannot localize both time and frequency to arbitrary precision.* Recent approaches are based on

PGR2026-00033
Krisp Technologies EX1024 Page 582

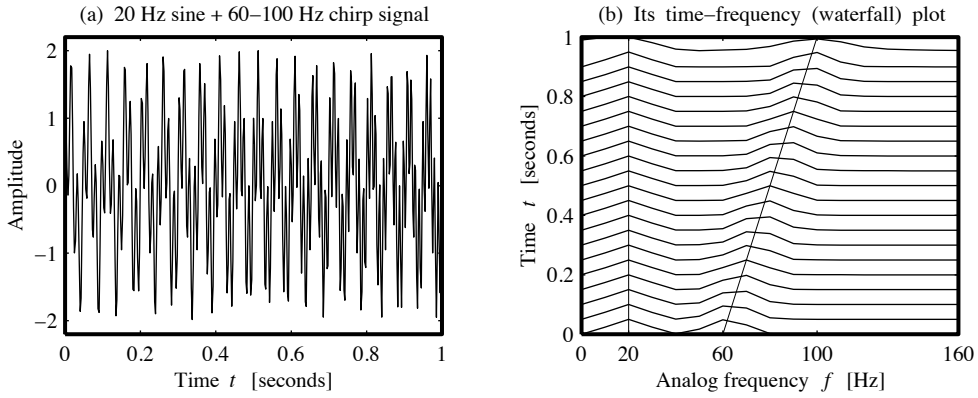


Figure 16.11 Time-frequency plot for the sum of a sinusoid and a chirp signal

expressing signals in terms of **wavelets** (much like the Fourier series). Wavelets are functions that show the "best" possible localization characteristics in both the time-domain and the frequency-domain, and are an area of intense ongoing research.

16.10 Matrix Formulation of the DFT and IDFT

If we let $W_N = \exp(-j2\pi/N)$, the defining relations for the DFT and IDFT may be written as

$$X_{\text{DFT}}[k] = \sum_{n=0}^{N-1} x[n]W_N^{nk}, \qquad k = 0, 1, \ldots, N-1 \tag{16.32}$$

$$x[n] = \frac{1}{N}\sum_{k=0}^{N-1} X_{\text{DFT}}[k][W_N^{nk}]^*, \qquad n = 0, 1, \ldots, N-1 \tag{16.33}$$

The first set of N DFT equations in N unknowns may be expressed in matrix form as

$$\mathbf{X} = \mathbf{W}_N\mathbf{x} \tag{16.34}$$

Here, $\mathbf{X}$ and $\mathbf{x}$ are $(N \times 1)$ matrices, and $\mathbf{W}_N$ is an $(N \times N)$ square matrix called the **DFT matrix**. The full matrix form is described by

$$\begin{bmatrix} X[0] \\ X[1] \\ X[2] \\ \vdots \\ X[N-1] \end{bmatrix} = \begin{bmatrix} W_N^0 & W_N^0 & W_N^0 & \ldots & W^0 \\ W_N^0 & W_N^1 & W_N^2 & \ldots & W_N^{N-1} \\ W_N^0 & W_N^2 & W_N^4 & \ldots & W_N^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ W_N^0 & W_N^{N-1} & W_N^{2(N-1)} & \ldots & W_N^{(N-1)(N-1)} \end{bmatrix} \begin{bmatrix} x[0] \\ x[1] \\ x[2] \\ \vdots \\ x[N-1] \end{bmatrix} \tag{16.35}$$

The exponents t in the elements W_N^t of $\mathbf{W}_N$ are called **twiddle factors**.

PGR2026-00033
Krisp Technologies EX1024 Page 583

EXAMPLE 16.14 (The DFT from the Matrix Formulation)
Let $x[n] = \{1,\ 2,\ 1,\ 0\}$. Then, with $N = 4$ and $W_N = \exp(-j2\pi/4) = -j$, the DFT may be obtained by solving the matrix product:

$$\begin{bmatrix} X[0] \\ X[1] \\ X[2] \\ X[3] \end{bmatrix} = \begin{bmatrix} W_N^0 & w_N^0 & W_N^0 & W_N^0 \\ W_N^0 & w_N^1 & W_N^2 & W_N^3 \\ W_N^0 & w_N^2 & W_N^4 & W_N^6 \\ W_N^0 & w_N^3 & W_N^6 & W_N^9 \end{bmatrix} \begin{bmatrix} x[0] \\ x[1] \\ x[2] \\ x[3] \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -j & -1 & j \\ 1 & -1 & 1 & -1 \\ 1 & j & -1 & -j \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ -j2 \\ 0 \\ j2 \end{bmatrix}$$

The result is $X_{\text{DFT}}[k] = \{4,\ -j2,\ 0,\ j2\}$.

16.10.1 The IDFT from the Matrix Form

The matrix $\mathbf{x}$ may be expressed in terms of the inverse of $\mathbf{W}_N$ as

$$\mathbf{x} = \mathbf{W}_N^{-1}\mathbf{X} \tag{16.36}$$

The matrix $\mathbf{W}_N^{-1}$ is called the **IDFT matrix**. We may also obtain $\mathbf{x}$ directly from the IDFT relation in matrix form, where the change of index from n to k, and the change in the sign of the exponent in $\exp(j2\pi nk/N)$, lead to a **conjugate transpose** of $\mathbf{W}_N$. We then have

$$\mathbf{x} = \frac{1}{N}[\mathbf{W}_N^*]^T\mathbf{X} \tag{16.37}$$

Comparison of the two forms suggests that

$$\mathbf{W}_N^{-1} = \frac{1}{N}[\mathbf{W}_N^*]^T \tag{16.38}$$

This very important result shows that $\mathbf{W}_N^{-1}$ requires only conjugation and transposition of $\mathbf{W}_N$, an obvious computational advantage.

The elements of the DFT and IDFT matrices satisfy $A_{ij} = A^{(i-1)(j-1)}$. Such matrices are known as **Vandermonde matrices**. They are notoriously ill conditioned insofar as their numerical inversion. This is not the case for $\mathbf{W}_N$, however. The product of the DFT matrix $\mathbf{W}_N$ with its conjugate transpose matrix equals the identity matrix $\mathbf{I}$. Matrices that satisfy such a property are called **unitary**. For this reason, the DFT and IDFT, which are based on unitary operators, are also called **unitary transforms**.

16.10.2 Using the DFT to Find the IDFT

Both the DFT and IDFT are matrix operations, and there is an inherent symmetry in the DFT and IDFT relations. In fact, we can obtain the IDFT by finding the DFT of the conjugate sequence and then conjugating the results and dividing by N. Mathematically,

$$x[n] = \text{IDFT}\{X_{\text{DFT}}[k]\} = \frac{1}{N}\Big(\text{DFT}\{X_{\text{DFT}}^*[k]\}\Big)^* \tag{16.39}$$

This result invokes the conjugate symmetry and duality of the DFT and IDFT, and suggests that the DFT algorithm itself can also be used to find the IDFT. In practice, this is indeed what is done.

PGR2026-00033
Krisp Technologies EX1024 Page 584

EXAMPLE 16.15 (Using the DFT to Find the IDFT)
Let us find the IDFT of $X_{\text{DFT}}[k] = \{4,\ -j2,\ 0,\ j2\}$ using the DFT. First, we conjugate the sequence to get $X^*_{\text{DFT}}[k] = \{4,\ j2,\ 0,\ -j2\}$. Next, we find the DFT of $X^*_{\text{DFT}}[k]$, using the 4×4 DFT matrix of the previous example, to give

$$\text{DFT}\{X^*_{\text{DFT}}[k]\} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -j & -1 & j \\ 1 & -1 & 1 & -1 \\ 1 & j & -1 & -j \end{bmatrix} \begin{bmatrix} 4 \\ -j2 \\ 0 \\ j2 \end{bmatrix} = \begin{bmatrix} 4 \\ 8 \\ 4 \\ 0 \end{bmatrix}$$

Finally, we conjugate this result and divide by $N = 4$ to get the IDFT of $X_{\text{DFT}}[k]$ as

$$\text{IDFT}\ \{X_{\text{DFT}}[k]\} = \tfrac{1}{4}\{4,\ 8,\ 4,\ 0\} = \{1,\ 2,\ 1,\ 0\}$$

16.11 The FFT

The importance of the DFT stems from the fact that it is amenable to fast and efficient computation using algorithms called **fast Fourier transform**, or FFT, algorithms. Fast algorithms reduce the problem of calculating an N-point DFT to that of calculating many smaller-size DFTs. The key ideas in optimizing the computation are based on the following ideas.

Symmetry and Periodicity

All FFT algorithms take advantage of the symmetry and periodicity of the exponential $W_N = \exp(-j2\pi n/N)$, as listed in Table 16.4. The last entry in this table, for example, suggests that $W_{N/2} = \exp(j2\pi n/0.5N)$ is periodic with period $\frac{N}{2}$.

Table 16.4 Symmetry and Periodicity of $W_N = \exp(-j2\pi/N)$

Entry	Exponential Form	Symbolic Form
1	$e^{-j2\pi n/N} = e^{-j2\pi(n+N)/N}$	$W_N^{n+N} = W_N^n$
2	$e^{-j2\pi(n+N/2)/N} = -e^{-j2\pi n/N}$	$W_N^{n+N/2} = -W_N^n$
3	$e^{-j2\pi K} = e^{-j2\pi NK/N} = 1$	$W_N^{NK} = 1$
4	$e^{-j2(2\pi/N)} = e^{-j2\pi/(N/2)}$	$W_N^2 = W_{N/2}$

Choice of Signal Length

We choose the signal length N as a number that is the product of many smaller numbers r_k such that $N = r_1 r_2 \ldots r_m$. A more useful choice results when the factors are equal, such that $N = r^m$. The factor r is called the **radix**. By far the most practically implemented choice for r is 2, such that $N = 2^m$, and leads to the **radix-2** FFT algorithms.

Index Separation and Storage

The computation is carried out separately on even-indexed and odd-indexed samples to reduce the computational effort. All algorithms allocate storage for computed results. The less the storage required, the more efficient is the algorithm. Many FFT algorithms reduce storage requirements by performing computations *in place* by storing results in the same memory locations that previously held the data.

PGR2026-00033
Krisp Technologies EX1024 Page 585

16.11.1 Some Fundamental Results

We begin by considering two trivial, but extremely important, results.

- **1-point transform:** The DFT of a single number A is the number A itself.
- **2-point transform:** The DFT of a 2-point sequence is easily found to be

$$X_{\text{DFT}}[0] = x[0] + x[1] \quad \text{and} \quad X_{\text{DFT}}[1] = x[0] - x[1] \tag{16.40}$$

The single most important result in the development of a radix-2 FFT algorithm is that an N-sample DFT can be written as the sum of two $\frac{N}{2}$-sample DFTs formed from the even- and odd-indexed samples of the original sequence. Here is the development:

$$\begin{aligned} X_{\text{DFT}}[k] &= \sum_{n=0}^{N-1} x[n]W_N^{nk} = \sum_{n=0}^{N/2-1} x[2n]W_N^{2nk} + \sum_{n=0}^{N/2-1} x[2n+1]W_N^{(2n+1)k} \\ X_{\text{DFT}}[k] &= \sum_{n=0}^{N/2-1} x[2n]W_N^{2nk} + W_N^k \sum_{n=0}^{N/2-1} x[2n+1]W_N^{2nk} \\ &= \sum_{n=0}^{N/2-1} x[2n]W_{N/2}^{nk} + W_N \sum_{n=0}^{N/2-1} x[2n+1]W_{N/2}^{nk} \end{aligned}$$

If $X^e[k]$ and $X^o[k]$ denote the DFT of the even- and odd-indexed sequences of length $N/2$, we can rewrite this result as

$$X_{\text{DFT}}[k] = X^e[k] + W_N^k X^o[k], \qquad k = 0, 1, 2, \ldots, N-1 \tag{16.41}$$

Note carefully that the index k in this expression varies from 0 to $N-1$ and that $X^e[k]$ and $X^o[k]$ are both periodic in k with period $N/2$; we thus have two periods of each to yield $X_{\text{DFT}}[k]$. Due to periodicity, we can split $X_{\text{DFT}}[k]$ and compute the first half and next half of the values as

$$\begin{aligned} X_{\text{DFT}}[k] &= X^e[k] + W_N^k X^o[k], \qquad k = 0, 1, 2, \ldots, \tfrac{1}{2}N - 1 \\ X_{\text{DFT}}[k + \tfrac{N}{2}] &= X^e[k + \tfrac{1}{2}N] + W_N^{k+\frac{N}{2}} X^o[k + \tfrac{N}{2}] \\ &= X^e[k] - W_N^k X^o[k], \qquad k = 0, 1, 2, \ldots, \tfrac{N}{2} - 1 \end{aligned}$$

This result is known as the **Danielson-Lanczos lemma**. Its *signal-flow graph* is shown in Figure 16.12 and is called a **butterfly** due to its characteristic shape.

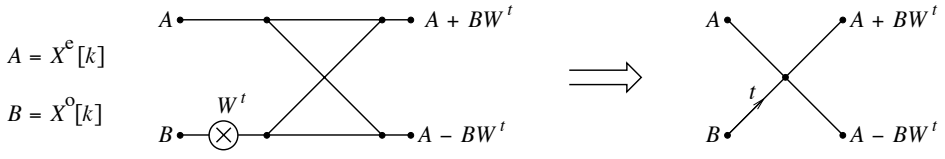


Figure 16.12 A typical butterfly

The inputs X^e and X^o are transformed into $X^e + W_N^k X^o$ and $X^e - W_N^k X^o$. A butterfly operates on one pair of samples and involves *two complex additions and one complex multiplication.* For N samples, there are $N/2$ butterflies in all. Starting with N samples, this lemma reduces the computational complexity by

PGR2026-00033
Krisp Technologies EX1024 Page 586

evaluating the DFT of two $\frac{N}{2}$-point sequences. The DFT of each of these can once again be reduced to the computation of sequences of length $N/4$ to yield

$$X^e[k] = X^{ee}[k] + W_{N/2}^k X^{eo}[k] \qquad X^o[k] = X^{oe}[k] + W_{N/2}^k X^{oo}[k] \tag{16.42}$$

Since $W_{N/2}^k = W_N^{2k}$, we can rewrite this expression as

$$X^e[k] = X^{ee}[k] + W_N^{2k} X^{eo}[k] \qquad X^o[k] = X^{oe}[k] + W_N^{2k} X^{oo}[k] \tag{16.43}$$

Carrying this process to its logical extreme, if we choose the number N of samples as $N = 2^m$, we can reduce the computation of an N-point DFT to the computation of 1-point DFTs in m stages. And the 1-point DFT is just the sample value itself (repeated with period 1). This process is called *decimation*. The FFT results so obtained are actually in **bit-reversed order**. If we let $e = 0$ and $o = 1$ and then reverse the order, we have the sample number in binary representation. The reason is that splitting the sequence into even and odd indices is equivalent to testing each index for the least significant bit (0 for even, 1 for odd). We describe two common in-place FFT algorithms based on decimation. A summary appears in Table 16.5.

Table 16.5 FFT Algorithms for Computing the DFT

Entry	**Characteristic**	**Decimation in Frequency**	**Decimation in Time**
1	Number of samples	$N = 2^m$	$N = 2^m$
2	Input sequence	Natural order	Bit-reversed order
3	DFT result	Bit-reversed order	Natural order
4	Computations	In place	In place
5	Number of stages	$m = \log_2 N$	$m = \log_2 N$
6	Multiplications	$\frac{N}{2}\log_2 M$ (complex)	$\frac{N}{2}N\log_2 M$(complex)
7	Additions	$N\log_2 M$ (complex)	$N\log_2 M$ (complex)
	Structure of the ith Stage		
8	No. of butterflies	$\frac{N}{2}$	$\frac{N}{2}$
9	Butterfly input	A (top) and B (bottom)	A (top) and B (bottom)
10	Butterfly output	$(A + B)$ and $(A - B)W_N^t$	$(A + BW_N^t)$ and $(A - BW_N^t)$
11	Twiddle factors t	$2^{i-1}Q,\ Q = 0, 1, \ldots, P - 1$	$2^{m-i}Q,\ Q = 0, 1, \ldots, P - 1$
12	Values of P	$P = 2^{m-i}$	$P = 2^{i-1}$

16.11.2 The Decimation-in-Frequency FFT Algorithm

The decimation-in-frequency (DIF) FFT algorithm starts by reducing the single N-point transform at each successive stage to two $\frac{N}{2}$-point transforms, then four $\frac{N}{4}$-point transforms, and so on, until we arrive at N 1-point transforms that correspond to the actual DFT. With the input sequence in natural order, computations can be done in place, but the DFT result is in bit-reversed order and must be reordered.

The algorithm slices the input sequence $x[n]$ into two halves and leads to

$$X_{\text{DFT}}[k] = \sum_{n=0}^{N-1} x[n]W_N^{nk} = \sum_{n=0}^{N/2-1} x[n]W_N^{nk} + \sum_{n=N/2}^{N/2-1} x[n]W_N^{nk} = \sum_{n=0}^{N/2-1} x[n]W_N^{nk} + \sum_{n=0}^{N/2-1} x[n+N/2]W_N^{(n+\frac{N}{2})k}$$

PGR2026-00033
Krisp Technologies EX1024 Page 587

This may be rewritten as

$$X_{\text{DFT}}[k] = \sum_{n=0}^{N/2-1} x[n]W_N^{nk} + W_N^{Nk/2} \sum_{n=0}^{N/2-1} x[n+\tfrac{N}{2}]W_N^{nk} = \sum_{n=0}^{N/2-1} x[n]W_N^{nk} + (-1)^k \sum_{n=0}^{N/2-1} x[n+\tfrac{N}{2}]W_N^{nk}$$

Separating even and odd indices, and letting $x[n] = x^a$ and $x[n+N/2] = x^b$,

$$X_{\text{DFT}}[2k] = \sum_{n=0}^{N/2-1} [x^a + x^b]W_N^{2nk}, \quad k = 0, 1, 2, \ldots, \tfrac{N}{2} - 1 \tag{16.44}$$

$$X_{\text{DFT}}[2k+1] = \sum_{n=0}^{N/2-1} [x^a - x^b]W_N^{(2k+1)n} = \sum_{n=0}^{N/2-1} [x^a - x^b]W_N^n W_N^{2nk}, \quad k = 0, 1, \ldots, \tfrac{N}{2} - 1 \tag{16.45}$$

Since $W_N^{2nk} = W_{N/2}^{nk}$, the even-indexed and odd-indexed terms describe a $\frac{N}{2}$-point DFT. The computations result in a butterfly structure with inputs $x[n]$ and $x[n+\frac{N}{2}]$, and outputs $X_{\text{DFT}}[2k] = \{x[n] + x[n+\frac{N}{2}]\}$ and $X_{\text{DFT}}[2k+1] = \{x[n] - x[n+\frac{N}{2}]\}W_N^n$. Its butterfly structure is shown in Figure 16.13.

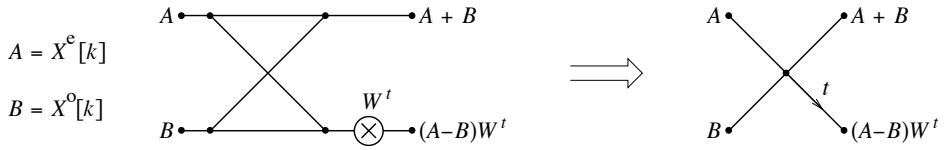


Figure 16.13 A typical butterfly for the decimation-in-frequency FFT algorithm

The factors W^t, called **twiddle factors**, appear only in the lower corners of the butterfly wings at each stage. Their exponents t have a definite order, described as follows for an $N = 2^m$-point FFT algorithm with m stages:

1. Number P of distinct twiddle factors W^t at ith stage: $P = 2^{m-i}$.
2. Values of t in the twiddle factors W^t: $t = 2^{i-1}Q$ with $Q = 0, 1, 2, \ldots, P-1$.

The DIF algorithm is illustrated in Figure 16.14 for $N = 2$, $N = 4$, and $N = 8$.

EXAMPLE 16.16 (A 4-Point Decimation-in-Frequency FFT Algorithm)
For a 4-point DFT, we use the above equations to obtain

$$X_{\text{DFT}}[2k] = \sum_{n=0}^{1} \{x[n] + x[n+2]\}W_4^{2nk}$$

$$X_{\text{DFT}}[2k+1] = \sum_{n=0}^{1} \{x[n] - x[n+2]\}W_4^n W_4^{2nk}, \qquad k = 0, 1$$

Since $W_4^0 = 1$ and $W_4^2 = -1$, we arrive at the following result:

$$X_{\text{DFT}}[0] = x[0] + x[2] + x[1] + x[3] \qquad X_{\text{DFT}}[2] = x[0] + x[2] - \{x[1] + x[3]\}$$
$$X_{\text{DFT}}[1] = x[0] - x[2] + W_4\{x[1] - x[3]\} \qquad X_{\text{DFT}}[3] = x[0] - x[2] - W_4\{x[1] - x[3]\}$$

We do not reorder the input sequence before using it.

PGR2026-00033
Krisp Technologies EX1024 Page 588

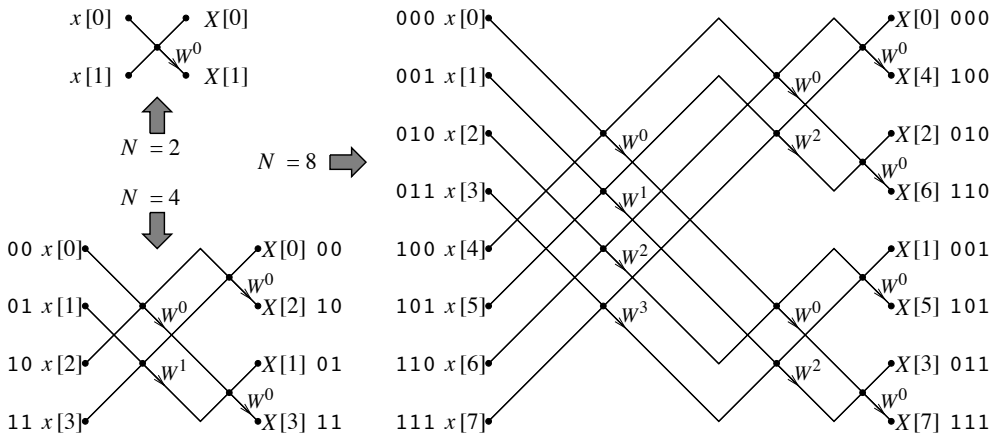


Figure 16.14 The decimation-in-frequency FFT algorithm for $N = 2,\ 4,\ 8$

16.11.3 The Decimation-in-Time FFT Algorithm

In the decimation-in-time (DIT) FFT algorithm, we start with N 1-point transforms, combine adjacent pairs at each successive stage into 2-point transforms, then 4-point transforms, and so on, until we get a single N-point DFT result. With the input sequence in bit-reversed order, the computations can be done in place, and the DFT is obtained in natural order. Thus, for a 4-point input, the binary indices $\{00, 01, 10, 11\}$ reverse to $\{00, 10, 01, 11\}$, and we use the bit-reversed order $\{x[0], x[2], x[1], x[3]\}$.

For an 8-point input sequence, $\{000, 001, 010, 011, 100, 101, 110, 111\}$, the reversed sequence corresponds to $\{000, 100, 010, 110, 001, 101, 011, 111\}$ or $\{x[0], x[4], x[6], x[2], x[1], x[5], x[3], x[7]\}$, and we use this sequence to perform the computations.

At a typical stage, we obtain

$$X_{\text{DFT}}[k] = X^e[k] + W_N^k X^o[k] \qquad X_{\text{DFT}}[k + \tfrac{N}{2}] = X^e[k] - W_N^k X^o[k] \tag{16.46}$$

Its butterfly structure is shown in Figure 16.15.

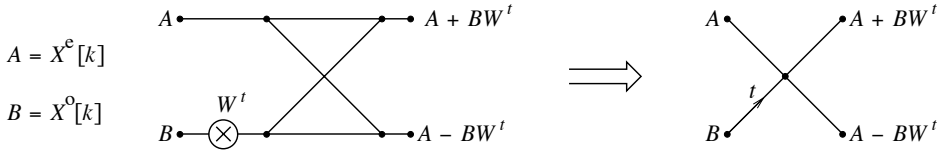


Figure 16.15 A typical butterfly for the decimation-in-time FFT algorithm

As with the decimation-in-frequency algorithm, the twiddle factors W^t at each stage appear only in the bottom wing of each butterfly. The exponents t also have a definite (and almost similar) order described by

1. Number P of distinct twiddle factors W^t at ith stage: $P = 2^{i-1}$.
2. Values of t in the twiddle factors W^t: $t = 2^{m-i}Q$ with $Q = 0, 1, 2, \ldots, P-1$.

The DIT algorithm is illustrated in Figure 16.16 for $N = 2,\ N = 4$, and $N = 8$.

PGR2026-00033
Krisp Technologies EX1024 Page 589

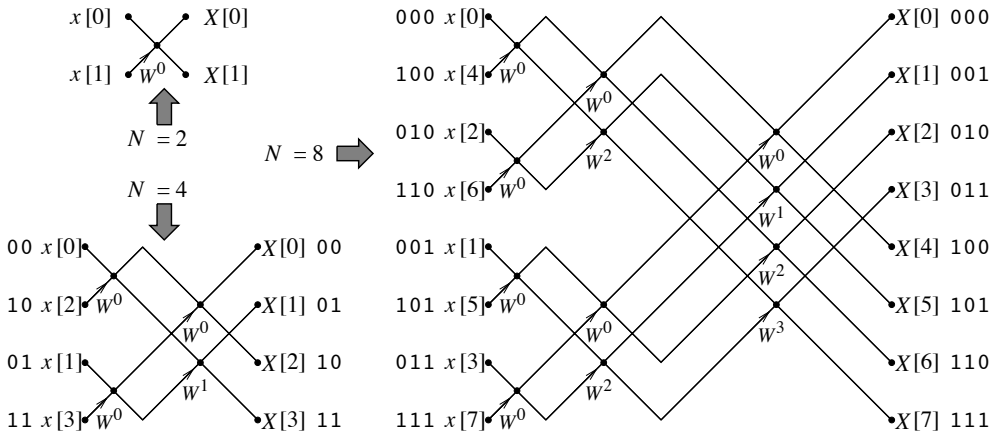


Figure 16.16 The decimation-in-time FFT algorithm for $N = 2,\ 4,\ 8$

In both the DIF algorithm and DIT algorithm, it is possible to use a sequence in natural order and get DFT results in natural order. This, however, requires more storage, since the computations cannot now be done in place.

EXAMPLE 16.17 (A 4-Point Decimation-in-Time FFT Algorithm)
For a 4-point DFT, with $W_4 = e^{-j\pi/2} = -j$, we have

$$X_{\text{DFT}}[k] = \sum_{n=0}^{3} x[n]W_4^{nk}, \qquad k = 0, 1, 2, 3$$

We group by even-indexed and odd-indexed samples of $x[n]$ to obtain

$$X_{\text{DFT}}[k] = X^e[k] + W_4^k X^o[k] \qquad \begin{cases} X^e[k] = x[0] + x[2]W_4^{2k}, \\ X^o[k] = x[1] + x[3]W_4^{2k}, \end{cases} \qquad k = 0, 1, 2, 3$$

Using periodicity, we simplify this result to

$$\begin{array}{l} X_{\text{DFT}}[k] = X^e[k] + W_4^k X^o[k] \\ X_{\text{DFT}}[k + \frac{1}{2}N] = X^e[k] - W_4^k X^o[k] \end{array} \qquad \begin{cases} X^e[k] = x[0] + x[2]W_4^{2k}, \\ X^o[k] = x[1] + x[3]W_4^{2k}, \end{cases} \qquad k = 0, 1$$

These equations yield $X_{\text{DFT}}[0]$ through $X_{\text{DFT}}[3]$ as

$$\begin{aligned} X_{\text{DFT}}[0] &= X^e[0] + W_4^0 X^o[0] = x[0] + x[2]W_4^0 + W_4^0\{x[1] + x[3]W_4^0\} \\ X_{\text{DFT}}[1] &= X^e[1] + W_4^1 X^o[1] = x[0] + x[2]W_4^2 + W_4^1\{x[1] + x[3]W_4^2\} \\ X_{\text{DFT}}[2] &= X^e[0] - W_4^0 X^o[0] = x[0] + x[2]W_4^0 - W_4^0\{x[1] + x[3]W_4^0\} \\ X_{\text{DFT}}[3] &= X^e[1] - W_4^1 X^o[1] = x[0] + x[2]W_4^2 - W_4^1\{x[1] + x[3]W_4^2\} \end{aligned}$$

Upon simplification, using $W_4^0 = 1$, $W_4^1 = W_4 = 1$, and $W_4^2 = -1$, we obtain

$$\begin{array}{ll} X_{\text{DFT}}[0] = x[0] + x[2] + x[1] + x[3] & X_{\text{DFT}}[1] = x[0] - x[2] + W_4\{x[1] - x[3]\} \\ X_{\text{DFT}}[2] = x[0] + x[2] - \{x[1] + x[3]\} & X_{\text{DFT}}[3] = x[0] - x[2] - W_4\{x[1] - x[3]\} \end{array}$$

PGR2026-00033
Krisp Technologies EX1024 Page 590

16.11.4 Computational Cost

Both the DIF and DIT N-point FFT algorithms involve $m = \log_2 N$ stages and $0.5N$ butterflies per stage. The FFT computation thus requires $0.5mN = 0.5N \log_2 N$ complex multiplications and $mN = N \log_2 N$ complex additions. Table 16.6 shows how the FFT stacks up against direct DFT evaluation for $N = 2^m$, counting all operations. Note, however, that multiplications take the bulk of computing time.

Table 16.6 Computational Cost of the DFT and FFT

Feature	N**-Point DFT**	N**-Point FFT**
Algorithm	Solution of N equations in N unknowns	$0.5N$ butterflies/stage, m stages Total butterflies $= 0.5mN$
Multiplications	N per equation	1 per butterfly
Additions	$N-1$ per equation	2 per butterfly
Total multiplications	N^2	$0.5mN = 0.5N \log_2 N$
Total additions	$N(N-1)$	$mN = N \log_2 N$

The difference between $0.5 \log_2 N$ and N^2 may not seem like much for small N. For example, with $N = 16$, $0.5N \log_2 N = 64$ and $N^2 = 256$. For large N, however, the difference is phenomenal. For $N = 1024 = 2^{10}$, $0.5N \log_2 N \approx 5000$ and $N^2 \approx 10^6$. This is like waiting 1 minute for the FFT result and (more than) 3 hours for the identical direct DFT result. Note that $N \log_2 N$ is nearly linear with N for large N, whereas N^2 shows a much faster, quadratic growth.

In the DFT relations, since there are N factors that equal W^0 or 1, we require only $N^2 - N$ complex multiplications. In the FFT algorithms, the number of factors that equal 1 doubles (or halves) at each stage and is given by $1+2+2^2+\cdots+2^{m-1} = 2^m - 1 = N-1$. We thus actually require only $0.5N \log_2 N - (N-1)$ complex multiplications for the FFT. The DFT requires N^2 values of W^k, but the FFT requires at most N such values at each stage. Due to the periodicity of W^k, only about $\frac{3}{4}N$ of these values are distinct. Once computed, they can be stored and used again. However, this hardly affects the comparison for large N. Since computers use real arithmetic, the number of real operations may be found by noting that one complex addition involves two real additions, and one complex multiplication involves four real multiplications and three real additions (because $(A + jB)(C + jD) = AC - BD + jBC + jAD$).

Speed of Fast Convolution

A direct computation of the convolution of two N-sample signals requires N^2 complex multiplications. The FFT method works with sequences of length $2N$. The number of complex multiplications involved is $2(N \log_2 2N)$ to find the FFT of the two sequences, $2N$ to form the product sequence, and $N \log_2 N$ to find the IFFT sequence that gives the convolution. It thus requires $3N \log_2 2N + 2N$ complex multiplications. If $N = 2^m$, the FFT approach becomes computationally superior only for $m > 5$ ($N > 32$) or so.

16.12 Why Equal Lengths for the DFT and IDFT?

The DTFT of an N-sample discrete time signal $x[n]$ is given by

$$X_p(F) = \sum_{n=0}^{N-1} x[n] e^{-j2\pi nF} \tag{16.47}$$

PGR2026-00033
Krisp Technologies EX1024 Page 591

If we sample F at M intervals over one period, the frequency interval F_0 equals $1/M$ and $F \to kF_0 = \frac{k}{M},\ k = 0, 1, \ldots, M-1$, and we get

$$X_{\text{DFT}}[k] = \sum_{n=0}^{N-1} x[n] e^{-j2\pi nk/M}, \qquad k = 0, 1, \ldots, M-1 \tag{16.48}$$

With $W_M = e^{-j2\pi/M}$, this equation can be written as

$$X_{\text{DFT}}[k] = \sum_{n=0}^{N-1} x[n] W_M^{nk}, \qquad k = 0, 1, \ldots, M-1 \tag{16.49}$$

This is a set of *M equations in N unknowns* and describes *the M-point DFT of the N-sample sequence* $x[n]$. It may be written in matrix form as

$$\mathbf{X} = \mathbf{W}_M \mathbf{x} \tag{16.50}$$

Here, $\mathbf{X}$ is an $M \times 1$ matrix, $\mathbf{x}$ is an $N \times 1$ matrix, and $\mathbf{W}_M$ is an $(M \times N)$ matrix. In full form,

$$\begin{bmatrix} X[0] \\ X[1] \\ X[2] \\ \vdots \\ X[M-1] \end{bmatrix} = \begin{bmatrix} W_M^0 & W_M^0 & W_M^0 & \ldots & W_M^0 \\ W_M^0 & W_M^1 & W_M^2 & \ldots & W_M^{N-1} \\ W_M^0 & W_M^2 & W_M^4 & \ldots & W_M^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ W_M^0 & W_M^{M-1} & W_M^{2(M-1)} & \ldots & W_M^{(N-1)(M-1)} \end{bmatrix} \begin{bmatrix} x[0] \\ x[1] \\ x[2] \\ \vdots \\ x[N-1] \end{bmatrix} \tag{16.51}$$

EXAMPLE 16.18 (A 4-Point DFT from a 3-Point Sequence)
Let $x[n] = \{1,\ 2,\ 1\}$. We have $N = 3$. The DTFT of this signal is

$$X_p(F) = 1 + 2e^{-j2\pi F} + e^{-j4\pi F} = [2 + 2\cos(2\pi F)]e^{-j2\pi F}$$

If we pick $M = 4$, we have $F = k/4$, $k = 0, 1, 2, 3$, and obtain the DFT as

$$X_{\text{DFT}}[k] = [2 + 2\cos(2\pi k/4)]e^{-j2\pi k/4}, \quad k = 0, 1, 2, 3 \quad \text{or} \quad X_{\text{DFT}}[k] = \{4,\ -j2,\ 0,\ j2\}$$

Using matrix notation with $W_M = e^{-j2\pi/4} = -j$, we can also find $X_{\text{DFT}}[k]$ as

$$\begin{bmatrix} X[0] \\ X[1] \\ X[2] \\ X[3] \end{bmatrix} = \begin{bmatrix} W_M^0 & W_M^0 & W_M^0 \\ W_M^0 & W_M^1 & W_M^2 \\ W_M^0 & W_M^2 & W_M^4 \\ W_M^0 & W_M^3 & W_M^6 \end{bmatrix} \begin{bmatrix} x[0] \\ x[1] \\ x[2] \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -j & -1 \\ 1 & -1 & 1 \\ 1 & j & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 \\ -j2 \\ 0 \\ j2 \end{bmatrix}$$

Thus, $X_{\text{DFT}}[k] = \{4,\ -j2,\ 0,\ j2\}$, as before.

16.12.1 The Inverse DFT

How do we obtain the N-sample sequence $x[n]$ from the M-sample DFT? It would seem that we require the product of $\mathbf{X}$, an $M \times 1$ matrix with an $N \times M$ matrix to give $\mathbf{x}$ as an $N \times 1$ matrix. But what is this $(M \times N)$ matrix, and how is it related to the $M \times N$ matrix $\mathbf{W}_M$? To find out, recall that

$$x[n] = \int_1 X_p(F) e^{j2\pi nF}\, dF \tag{16.52}$$

PGR2026-00033
Krisp Technologies EX1024 Page 592

Converting this to discrete form with $F = kF_0 = \frac{k}{M}$ results in periodicity of $x[n]$ with period 1, and we obtain N samples of $x[n]$ using

$$x[n] = \frac{1}{M} \sum_{k=0}^{M-1} X_{\text{DFT}}[k] e^{j2\pi nk/M}, \qquad n = 0, 1, \ldots, N-1 \tag{16.53}$$

For $N < M$, one period of $x[n]$ is a zero-padded M-sample sequence. For $N > M$, however, one period of $x[n]$ is the *periodic extension* of the N-sample sequence with period M.

The sign of the exponent and the interchange of the indices n and k allow us to set up the matrix formulation for obtaining $x[n]$ using an $N \times M$ inversion-matrix $\mathbf{W}_I$ that just equals $\frac{1}{M}$ times $[\mathbf{W}_M^*]^T$, the **conjugate transpose** of the $M \times N$ DFT matrix $\mathbf{W}_M$. Its product with the $N \times 1$ matrix corresponding to $X[k]$ yields the $M \times 1$ matrix for $x[n]$. We thus have the forward and inverse matrix relations:

$$\mathbf{X} = \mathbf{W}_M \mathbf{x} \quad \text{(DFT)} \qquad\qquad \mathbf{x} = \mathbf{W}_I \mathbf{X} = \frac{1}{M} [\mathbf{W}_M^*]^T \mathbf{X} \quad \text{(IDFT)} \tag{16.54}$$

These results are valid for any choice of M and N. An interesting result is that the product of $\mathbf{W}_M$ with $\mathbf{W}_I$ is the $N \times N$ identity matrix.

EXAMPLE 16.19 (A 3-Point IDFT from a 4-Point DFT)
Let $X_{\text{DFT}}[k] = \{4,\ -j2,\ 0,\ j2\}$ and $M = 4$.

The IDFT matrix equals

$$\mathbf{W}_I = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 \\ 1 & j & -1 \\ 1 & -1 & 1 \\ 1 & -j & -1 \end{bmatrix}^{*T} = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -j & -1 & j \\ 1 & -1 & 1 & -1 \end{bmatrix}$$

We then get the IDFT as

$$\mathbf{x} = \mathbf{W}_I \mathbf{X} = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -j & -1 & j \\ 1 & -1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 4 \\ -j2 \\ 0 \\ j2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

The important thing to realize is that $x[n]$ is actually periodic with $M = 4$, and one period of $x[n]$ is the zero-padded sequence $\{1,\ 2,\ 1,\ 0\}$.

16.12.2 How Unequal Lengths Affect the DFT Results

Even though the M-point IDFT of an N-point sequence is valid for any M, the choice of M affects the nature of $x[n]$ through the IDFT and its inherent periodic extension.

1. If $M = N$, the IDFT is periodic, with period M, and its one period equals the N-sample $x[n]$. Both the DFT matrix and IDFT matrix are square ($M \times M$), and allow a simple inversion relation to go back and forth between the two.

2. If $M > N$, the IDFT is periodic with period M. Its one period is the original N-sample $x[n]$ with $M - N$ padded zeros. The choice $M > N$ is equivalent to using a zero-padded version of $x[n]$ with a total of M samples and $M \times M$ square matrices for both the DFT matrix and IDFT matrix.

PGR2026-00033
Krisp Technologies EX1024 Page 593

3. If $M < N$, the IDFT is periodic with period $M < N$. Its one period is the periodic extension of the N-sample $x[n]$ with period M. It thus yields a signal that corresponds to $x[n]$ *wrapped around* after M samples and does not recover the original $x[n]$.

EXAMPLE 16.20 (The Importance of Periodic Extension)
Let $x[n] = \{1,\ 2,\ 1\}$. We have $N = 3$. The DTFT of this signal is

$$X_p(F) = 1 + 2e^{-j2\pi F} + e^{-j4\pi F} = [2 + 2\cos(2\pi F)]e^{-j2\pi F}$$

We sample $X_p(F)$ at M intervals and find the IDFT as $y[n]$. What do we get?

(a) For $M = 3$, we should get $y[n] = \{1,\ 2,\ 1\} = x[n]$. Let us find out.

With $M = 3$, we have $F = k/3$ for $k = 0, 1, 2$, and $X_{\text{DFT}}[k]$ becomes

$$X_{\text{DFT}}[k] = [2 + 2\cos(2k\pi/3)]e^{-j2k\pi/3} = \left\{4,\ -\tfrac{1}{2} - j\sqrt{\tfrac{3}{4}},\ -\tfrac{1}{2} + j\sqrt{\tfrac{3}{4}}\right\}$$

We also find $\mathbf{W}_I$ and the IDFT $\mathbf{x} = \mathbf{W}_I\mathbf{X}$ as

$$\mathbf{W}_I = \frac{1}{3}\begin{bmatrix} W_M^0 & W_M^0 & W_M^0 \\ W_M^0 & W_M^1 & W_M^2 \\ W_M^0 & W_M^2 & W_M^4 \end{bmatrix}^{*T} = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & -\frac{1}{2} + j\sqrt{\frac{3}{4}} & -\frac{1}{2} - j\sqrt{\frac{3}{4}} \\ 1 & -\frac{1}{2} - j\sqrt{\frac{3}{4}} & -\frac{1}{2} + j\sqrt{\frac{3}{4}} \end{bmatrix}$$

$$\mathbf{x} = \mathbf{W}_I\mathbf{X} = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & -\frac{1}{2} + j\sqrt{\frac{3}{4}} & -\frac{1}{2} - j\sqrt{\frac{3}{4}} \\ 1 & -\frac{1}{2} - j\sqrt{\frac{3}{4}} & -\frac{1}{2} + j\sqrt{\frac{3}{4}} \end{bmatrix}\begin{bmatrix} 4 \\ -\frac{1}{2} - j\sqrt{\frac{3}{4}} \\ -\frac{1}{2} + j\sqrt{\frac{3}{4}} \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

This result is periodic with $M = 3$, and one period of this equals $x[n]$.

(b) For $M = 4$, we should get a new sequence $y[n] = \{1,\ 2,\ 1,\ 0\}$ that corresponds to a zero-padded version of $x[n]$, and we do (the details were worked out in Example 16.19).

(c) For $M = 2$, we should get a new sequence $z[n] = \{2,\ 2\}$ that corresponds to the periodic extension of $x[n]$ with period 2.

With $M = 2$ and $k = 0, 1$, we have $Z_{\text{DFT}}[k] = [2 + 2\cos(\pi k)]e^{-jk\pi} = \{4,\ 0\}$.
Since $e^{-j2\pi/M} = e^{-j\pi} = -1$, we can find the IDFT $z[n]$ directly from the definition as

$$z[0] = 0.5\{X_{\text{DFT}}[0] + X_{\text{DFT}}[1]\} = 2 \qquad z[1] = 0.5\{X_{\text{DFT}}[0] - X_{\text{DFT}}[1]\} = 2$$

The sequence $z[n] = \{2,\ 2\}$ is periodic with $M = 2$. As expected, this equals one period of the periodic extension of $x[n] = \{1,\ 2,\ 1\}$ (with wraparound past two samples).

PGR2026-00033
Krisp Technologies EX1024 Page 594

CHAPTER 16 PROBLEMS

DRILL AND REINFORCEMENT

16.1 (DFT from Definition) Compute the DFT and DFS of the following signals.

(a) $x[n] = \{1, 2, 1, 2\}$ **(b)** $x[n] = \{2, 1, 3, 0, 4\}$
(c) $x[n] = \{2, 2, 2, 2\}$ **(d)** $x[n] = \{1, 0, 0, 0, 0, 0, 0, 0\}$

16.2 (IDFT from Definition) Compute the IDFT of the following.

(a) $X_{\text{DFT}}[k] = \{2, -j, 0, j\}$ **(b)** $X_{\text{DFT}}[k] = \{4, -1, 1, 1, -1\}$
(c) $X_{\text{DFT}}[k] = \{1, 2, 1, 2\}$ **(d)** $X_{\text{DFT}}[k] = \{1, 0, 0, j, 0, -j, 0, 0\}$

16.3 (Symmetry) For the DFT of each real sequence, compute the boxed quantities.

(a) $X_{\text{DFT}}[k] = \left\{0,\ \boxed{X_1},\ 2+j,\ -1,\ \boxed{X_4},\ j\right\}$

(b) $X_{\text{DFT}}[k] = \left\{1,\ 2,\ \boxed{X_2},\ \boxed{X_3},\ 0,\ 1-j,\ -2,\ \boxed{X_7}\right\}$

16.4 (Properties) The DFT of $x[n]$ is $X_{\text{DFT}}[k] = \{1, 2, 3, 4\}$. Find the DFT of each of the following sequences, using properties of the DFT.

(a) $y[n] = x[n-2]$ **(b)** $f[n] = x[n+6]$ **(c)** $g[n] = x[n+1]$
(d) $h[n] = e^{jn\pi/2}x[n]$ **(e)** $p[n] = x[n] \circledast x[n]$ **(f)** $q[n] = x^2[n]$
(g) $r[n] = x[-n]$ **(h)** $s[n] = x^*[n]$ **(i)** $v[n] = x^2[-n]$

16.5 (Replication and Zero Interpolation) The DFT of $x[n]$ is $X_{\text{DFT}}[k] = \{1, 2, 3, 4, 5\}$.

(a) What is the DFT of the replicated signal $y[n] = \{x[n], x[n]\}$?
(b) What is the DFT of the replicated signal $f[n] = \{x[n], x[n], x[n]\}$?
(c) What is the DFT of the zero-interpolated signal $g[n] = x[n/2]$?
(d) What is the DFT of the zero-interpolated signal $h[n] = x[n/3]$?

16.6 (DFT of Pure Sinusoids) Determine the DFS and DFT of $x(t) = \sin(2\pi f_0 t + \frac{\pi}{3})$ (without doing any DFT computations) if we sample this signal, starting at $t = 0$, and acquire the following:

(a) 4 samples over 1 period **(b)** 8 samples over 2 periods
(c) 8 samples over 1 period **(d)** 18 samples over 3 periods
(e) 8 samples over 5 periods **(f)** 16 samples over 10 periods

16.7 (DFT of Sinusoids) The following signals are sampled starting at $t = 0$. Find their DFS and DFT and identify the indices of the nonzero DFT components.

(a) $x(t) = \cos(4\pi t)$ sampled at 25 Hz, using the minimum number of samples to prevent leakage
(b) $x(t) = \cos(20\pi t) + 2\sin(40\pi t)$ sampled at 25 Hz with $N = 15$
(c) $x(t) = \sin(10\pi t) + 2\sin(40\pi t)$ sampled at 25 Hz for 1 s
(d) $x(t) = \sin(40\pi t) + 2\sin(60\pi t)$ sampled at intervals of 0.004 s for four periods

16.8 (Aliasing and Leakage) The following signals are sampled and the DFT of the sampled signal obtained. Which cases will show aliasing, or leakage, or both? In which cases can the effects of leakage and/or aliasing be avoided, and how?

PGR2026-00033
Krisp Technologies EX1024 Page 595

(a) A 100-Hz sinusoid for 2 periods at 400 Hz
(b) A 100-Hz sinusoid for 4 periods at 70 Hz
(c) A 100-Hz sinusoid at 400 Hz, using 10 samples
(d) A 100-Hz sinusoid for 2.5 periods at 70 Hz
(e) A sum of sinusoids at 100 Hz and 150 Hz for 100 ms at 450 Hz
(f) A sum of sinusoids at 100 Hz and 150 Hz for 1 period at 250 Hz
(g) A sum of sinusoids at 100 Hz and 150 Hz for 0.5 period at 400 Hz
(h) A square wave with period $T = 10$ ms for 2 periods at 400 Hz
(i) A square wave with period $T = 10$ ms for 1.5 periods at 400 Hz

16.9 (Spectral Spacing) What is the spectral spacing in the 500-point DFT of a sampled signal obtained by sampling an analog signal at 1 kHz?

16.10 (Spectral Spacing) We wish to sample a signal of 1-s duration, band-limited to 50 Hz, and compute the DFT of the sampled signal.
(a) Using the minimum sampling rate that avoids aliasing, what is the spectral spacing Δf, and how many samples are acquired?
(b) How many padding zeros are needed to reduce the spacing to $0.5\Delta f$, using the minimum sampling rate to avoid aliasing if we use the DFT?
(c) How many padding zeros are needed to reduce the spacing to $0.5\Delta f$, using the minimum sampling rate to avoid aliasing if we use a radix-2 FFT?

16.11 (Convolution and the DFT) Consider two sequences $x[n]$ and $h[n]$ of length 12 samples and 20 samples, respectively.
(a) How many padding zeros are needed for $x[n]$ and $h[n]$ in order to find their regular convolution $y[n]$, using the DFT?
(b) If we pad $x[n]$ with eight zeros and find the periodic convolution $y_p[n]$ of the resulting 20-point sequence with $h[n]$, for what indices will the samples of $y[n]$ and $y_p[n]$ be identical?

16.12 (FFT) Write out the DFT sequence that corresponds to the following bit-reversed sequences obtained using the DIF FFT algorithm.

(a) $\{1, 2, 3, 4\}$ **(b)** $\{0, -1, 2, -3, 4, -5, 6, -7\}$

16.13 (FFT) Set up a flowchart showing all twiddle factors and values at intermediate nodes to compute the DFT of $x[n] = \{1, 2, 2, 2, 1, 0, 0, 0\}$, using

(a) The 8-point DIF algorithm. **(b)** The 8-point DIT algorithm.

16.14 (Spectral Spacing and the FFT) We wish to sample a signal of 1-s duration, and band-limited to 100 Hz, in order to compute its spectrum. The spectral spacing should not exceed 0.5 Hz. Find the minimum number N of samples needed and the actual spectral spacing Δf if we use

(a) The DFT. **(b)** The radix-2 FFT.

16.15 (Convolution) Find the linear convolution of $x[n] = \{\overset{\Downarrow}{1}, 2, 1\}$ and $h[n] = \{\overset{\Downarrow}{1}, 2, 3\}$, using
(a) The time-domain convolution operation.
(b) The DFT and zero-padding.
(c) The radix-2 FFT and zero-padding.

PGR2026-00033
Krisp Technologies EX1024 Page 596

16.16 (Convolution) Find the *periodic* convolution of $x[n] = \{\overset{\Downarrow}{1}, 2, 1\}$ and $h[n] = \{\overset{\Downarrow}{1}, 2, 3\}$, using
- **(a)** The time-domain convolution operation.
- **(b)** The DFT operation. Is this result identical to that of part (a)?
- **(c)** The radix-2 FFT and zero-padding. Is this result identical to that of part (a)? Should it be?

16.17 (Correlation) Find the *periodic* correlation r_{xh} of $x[n] = \{\overset{\Downarrow}{1}, 2, 1\}$ and $h[n] = \{\overset{\Downarrow}{1}, 2, 3\}$, using
- **(a)** The time-domain correlation operation.
- **(b)** The DFT.

16.18 (Convolution of Long Sequences) Let $x[n] = \{\overset{\Downarrow}{1}, 2, 1\}$ and $h[n] = \{\overset{\Downarrow}{1}, 2, 1, 3, 2, 2, 3, 0, 1, 0, 2, 2\}$.
- **(a)** Find their convolution using the overlap-add method.
- **(b)** Find their convolution using the overlap-save method.
- **(c)** Are the results identical to the time-domain convolution of $x[n]$ and $h[n]$?

REVIEW AND EXPLORATION

16.19 (Properties) Let $X_{\text{DFT}}[k]$ be the N-point DFT of a real signal. How many DFT samples will always be real, and what will be their index k? [Hint: Use the concept of conjugate symmetry and consider the cases for odd N and even N separately.]

16.20 (Properties) Let $X_{\text{DFT}}[k]$ be the N-point DFT of a (possibly complex) signal $x[n]$. What can you say about the symmetry of $x[n]$ for the following $X_{\text{DFT}}[k]$?
- **(a)** Conjugate symmetric
- **(b)** Real but with no symmetry
- **(c)** Real and even symmetric
- **(d)** Real and odd symmetric
- **(e)** Imaginary and odd symmetric

16.21 (Properties) For each DFT pair shown, compute the values of the boxed quantities, using properties such as conjugate symmetry and Parseval's theorem.
- **(a)** $\left\{\boxed{x_0},\ 3,\ -4,\ 0,\ 2\right\} \Longleftrightarrow \left\{5,\ \boxed{X_1},\ -1.28 - j4.39,\ \boxed{X_3},\ 8.78 - j1.4\right\}$
- **(b)** $\left\{\boxed{x_0},\ 3,\ -4,\ 2,\ 0,\ 1\right\} \Longleftrightarrow \left\{4,\ \boxed{X_1},\ 4 - j5.2,\ \boxed{X_3},\ \boxed{X_4},\ 4 - j1.73\right\}$

16.22 (Properties) Let $x[n] = \{\overset{\Downarrow}{1},\ -2,\ 3,\ -4,\ 5,\ -6\}$. Without evaluating its DFT $X[k]$, compute the following:

(a) $X[0]$ **(b)** $\displaystyle\sum_{k=0}^{5} X[k]$ **(c)** $X[3]$ **(d)** $\displaystyle\sum_{k=0}^{5} |X[k]|^2$ **(e)** $\displaystyle\sum_{k=0}^{5} (-1)^k X[k]$

16.23 (DFT Computation) Find the N-point DFT of each of the following signals.

(a) $x[n] = \delta[n]$	**(b)** $x[n] = \delta[n - K],\ K < N$
(c) $x[n] = \delta[n - 0.5N]$ (N even)	**(d)** $x[n] = \delta[n - 0.5(N-1)]$ (N odd)
(e) $x[n] = 1$	**(f)** $x[n] = \delta[n - 0.5(N-1)] + \delta[n - 0.5(N+1)]$ (N odd)
(g) $x[n] = (-1)^n$ (N even)	**(h)** $x[n] = e^{j4n\pi/N}$
(i) $x[n] = \cos(4n\pi/N)$	**(j)** $x[n] = \cos(4n\pi/N + 0.25\pi)$

PGR2026-00033
Krisp Technologies EX1024 Page 597

16.24 (Properties) The DFT of a signal $x[n]$ is $X_{\text{DFT}}[k]$. If we use its conjugate $Y_{\text{DFT}}[k] = X^*_{\text{DFT}}[k]$ and obtain its DFT as $y[n]$, how is $y[n]$ related to $x[n]$?

16.25 (Properties) Let $X[k] = \{\overset{\Downarrow}{1},\ -2,\ 1-j,\ j2,\ 0, \ldots\}$ be the 8-point DFT of a real signal $x[n]$.

(a) Determine $X[k]$ in its entirety.
(b) What is the DFT $Y[k]$ of the signal $y[n] = (-1)^n x[n]$?
(c) What is the DFT $G[k]$ of the zero-interpolated signal $g[n] = x[n/2]$?
(d) What is the DFT $H[k]$ of $h[n] = \{x[n],\ x[n],\ x[n]\}$ obtained by threefold replication of $x[n]$?

16.26 (Spectral Spacing) We wish to sample the signal $x(t) = \cos(50\pi t) + \sin(200\pi t)$ at 800 Hz and compute the N-point DFT of the sampled signal $x[n]$.

(a) Let $N = 100$. At what indices would you expect to see the spectral peaks? Will the peaks occur at the frequencies of $x(t)$?
(b) Let $N = 128$. At what indices would you expect to see the spectral peaks? Will the peaks occur at the frequencies of $x(t)$?

16.27 (Spectral Spacing) We wish to sample the signal $x(t) = \cos(50\pi t) + \sin(80\pi t)$ at 100 Hz and compute the N-point DFT of the sampled signal $x[n]$.

(a) Let $N = 100$. At what indices would you expect to see the spectral peaks? Will the peaks occur at the frequencies of $x(t)$?
(b) Let $N = 128$. At what indices would you expect to see the spectral peaks? Will the peaks occur at the frequencies of $x(t)$?

16.28 (Spectral Spacing) We wish to identify the 21-Hz component from the N-sample DFT of a signal. The sampling rate is 100 Hz, and only 128 signal samples are available.

(a) If $N = 128$, will there be a DFT component at 21 Hz? If not, what is the frequency closest to 21 Hz that can be identified? What DFT index does this correspond to?
(b) Assuming that all signal samples must be used and zero-padding is allowed, what is the smallest value of N that will result in a DFT component at 21 Hz? How many padding zeros will be required? At what DFT index will the 21-Hz component appear?

16.29 (Sampling Frequency) For each of the following signals, estimate the sampling frequency and sampling duration by arbitrarily choosing the bandwidth as the frequency where $|H(f)|$ is 5% of its maximum and the signal duration as the time at which $x(t)$ is 1% of its maximum.

(a) $x(t) = e^{-t}u(t)$ **(b)** $x(t) = te^{-t}u(t)$ **(c)** $x(t) = \text{tri}(t)$

16.30 (Sampling Frequency and Spectral Spacing) It is required to sample the signal $x(t) = e^{-t}u(t)$ and compute the DFT of the sampled signal. The signal is sampled for a duration D that contains 95% of the signal energy. How many samples are acquired if the sampling rate S is chosen to ensure that

(a) The aliasing level at $f = 0.5S$ due to the first replica is less than 1% of the peak level?
(b) The energy in the aliased signal past $f = 0.5S$ is less than 1% of the total signal energy?

16.31 (Sampling Rate and the DFT) A periodic square wave $x(t)$ with a duty ratio of 0.5 and period $T = 2$ is sampled for one full period to obtain N samples. The N-point DFT of the samples corresponds to the signal $y(t) = A + B\sin(\pi t)$.

PGR2026-00033
Krisp Technologies EX1024 Page 598

(a) What are the possible values of N for which you could obtain such a result? For each such choice, compute the values of A and B.
(b) What are the possible values of N for which you could obtain such a result using the radix-2 FFT? For each such choice, compute the values of A and B.
(c) Is it possible for $y(t)$ to be identical to $x(t)$ for any choice of sampling rate?

16.32 (Sampling Rate and the DFT) A periodic signal $x(t)$ with period $T = 2$ is sampled for one full period to obtain N samples. The signal reconstructed from the N-point DFS of the samples is $y(t)$.
(a) Will the DFS show the effects of leakage?
(b) Let $N = 8$. How many harmonics of $x(t)$ can be identified in the DFS? What constraints on $x(t)$ will ensure that $y(t) = x(t)$?
(c) Let $N = 12$. How many harmonics of $x(t)$ can be identified in the DFS? What constraints on $x(t)$ will ensure that $y(t) = x(t)$?

16.33 (DFT from Definition) Use the defining relation to compute the N-point DFT of the following:
(a) $x[n] = \delta[n], \quad 0 \leq n \leq N-1$
(b) $x[n] = \alpha^n, \quad 0 \leq n \leq N-1$
(c) $x[n] = e^{j\pi n/N}, \quad 0 \leq n \leq N-1$

16.34 (DFT Concepts) The signal $x(t) = \cos(150\pi t) + \cos(180\pi t)$ is to be sampled and analyzed using the DFT.
(a) What is the minimum sampling $S_{\min}$ rate to prevent aliasing?
(b) With the sampling rate chosen as $S = 2S_{\min}$, what is the minimum number of samples $N_{\min}$ required to prevent leakage?
(c) With the sampling rate chosen as $S = 2S_{\min}$, and $N = 3N_{\min}$, what is the DFT of the sampled signal?
(d) With $S = 160$ Hz and $N = 256$, at what DFT indices would you expect to see spectral peaks? Will leakage be present? Will aliasing occur?

16.35 (DFT Concepts) The signal $x(t) = \cos(50\pi t) + \cos(80\pi t)$ is sampled at $S = 200$ Hz.
(a) What is the minimum number of samples required to prevent leakage?
(b) Find the DFT of the sampled signal if $x(t)$ is sampled for 1 s.
(c) What are the DFT indices of the spectral peaks if $N = 128$?

16.36 (DFT Concepts) Let $x[n] = \cos(2n\pi F_0)$.
(a) Use the defining relation to compute its N-point DFT.
(b) For what values of F_0 would you expect aliasing and/or leakage?
(c) Use the above results to compute the DFT if $N = 8$ and $F_0 = 0.25$.
(d) Use the above results to compute the DFT if $N = 8$ and $F_0 = 1.25$.
(e) Use the above results to compute the DFT if $N = 9$ and $F_0 = 0.2$.

16.37 (Resolution) A signal is sampled at 5 kHz to acquire N samples of $x[n]$.
(a) Let $N = 1000$. What is the frequency resolution in the DFT of the sampled signal $x[n]$?
(b) Let $N = 1000$. What is the frequency resolution in the DFT of the von Hann windowed signal?
(c) Find the smallest value of N to ensure a frequency resolution of 2 Hz with no window.
(d) Find the smallest value of N to ensure a frequency resolution of 2 Hz with a von Hann window.

PGR2026-00033
Krisp Technologies EX1024 Page 599

COMPUTATION AND DESIGN

dfftgui A GUI for Visualization of the FFT of Sinusoids

The graphical user interface `dfftgui` allows you to visualize the FFT of sampled sinusoids with or without windowing.
You can select the sinusoids, the sampling rate, and the window.
The interface displays the sampled signal and its spectrum.
To explore this routine, type `dfftgui` at the MATLAB prompt.

tfestgui A GUI for Transfer Function Estimation

The graphical user interface `tfestgui` allows you to approximate a given transfer function by an FIR filter and visualize the response of each.
The transfer function estimate is based on the Welch PSD estimate to a white-noise input.
To explore this routine, type `tfestgui` at the MATLAB prompt.

chirpgui A GUI for Visualization of a Chirp Signal and Its Spectrum

The graphical user interface `chirpgui` allows you to visualize a chirp signal, its PSD estimate, and its time-frequency spectrum, which shows how its frequency varies with time.
You can select the chirp signal parameters and the PSD estimation method.
To explore this routine, type `chirpgui` at the MATLAB prompt.

16.38 (DFT Properties) Consider the signal $x[n] = n+1,\ 0 \le n \le 7$. Use MATLAB to compute its DFT. Confirm the following properties by computing the DFT.

(a) The DFT of $y[n] = x[-n]$ to confirm the (circular) folding property
(b) The DFT of $f[n] = x[n-2]$ to confirm the (circular) shift property
(c) The DFT of $g[n] = x[n/2]$ to confirm the zero-interpolation property
(d) The DFT of $h[n] = \{x[n],\ x[n]\}$ to confirm the signal-replication property
(e) The DFT of $p[n] = x[n]\cos(0.5n\pi)$ to confirm the modulation property
(f) the DFT of $r[n] = x^2[n]$ to confirm the multiplication property
(g) The DFT of $s[n] = x[n] \circledast x[n]$ to confirm the periodic convolution property

16.39 (IDFT from DFT) Consider the signal $x[n] = (1+j)n,\ 0 \le n \le 9$.

(a) Find its DFT $X[k]$. Find the DFT of the sequence $0.1X[k]$. Does this appear to be related to the signal $x[n]$?
(b) Find the DFT of the sequence $0.1X^*[k]$. Does this appear to be related to $x[n]$?
(c) Use your results to explain how you might find the IDFT of a sequence by using the forward DFT algorithm alone.

16.40 (Resolution) We wish to compute the radix-2 FFT of the signal samples acquired from the signal $x(t) = A\cos(2\pi f_0 t) + B\cos[2\pi(f_0 + \Delta f)t]$, where $f_0 = 100$ Hz. The sampling frequency is $S = 480$ Hz.

(a) Let $A = B = 1$. What is the smallest number of signal samples $N_{\min}$ required for a frequency resolution of $\Delta f = 2$ Hz if no window is used? How does this change if we wish to use a von Hann window? What about a Blackman window? Plot the FFT magnitude for each case to confirm your expectations.

PGR2026-00033
Krisp Technologies EX1024 Page 600

(b) Let $A = 1$ and $B = 0.02$. Argue that we cannot obtain a frequency resolution of $\Delta f = 2$ Hz if no window is used. Plot the FFT magnitude for various lengths to justify your argument. Of the Bartlett, Hamming, von Hann, and Blackman windows, which ones can we use to obtain a resolution of $\Delta f = 2$ Hz? Which one will require the minimum number of samples, and why? Plot the FFT magnitude for each applicable window to confirm your expectations.

16.41 (Convolution) Consider the sequences $x[n] = \{1, 2, 1, 2, 1\}$ and $h[n] = \{1, 2, 3, 3, 5\}$.

(a) Find their regular convolution using three methods: the convolution operation; zero-padding and the DFT; and zero-padding to length 16 and the DFT. How are the results of each operation related? What is the effect of zero-padding?

(b) Find their periodic convolution using three methods: regular convolution and wraparound; the DFT; and zero-padding to length 16 and the DFT. How are the results of each operation related? What is the effect of zero-padding?

16.42 (Convolution) Consider the signals $x[n] = 4(0.5)^n,\ 0 \le n \le 4$, and $h[n] = n,\ 0 \le n \le 10$.

(a) Find their regular convolution $y[n] = x[n] \star h[n]$, using the MATLAB routine `conv`.

(b) Use the FFT and IFFT to obtain the regular convolution, assuming the *minimum* length N (that each sequence must be zero-padded to) for correct results.

(c) How do the results change if each sequence is zero-padded to length $N + 2$?

(d) How do the results change if each sequence is zero-padded to length $N - 2$?

16.43 (FFT of Noisy Data) Sample the sinusoid $x(t) = \cos(2\pi f_0 t)$ with $f_0 = 8$ Hz at $S = 64$ Hz for 4 s to obtain a 256-point sampled signal $x[n]$. Also generate 256 samples of a uniformly distributed noise sequence $s[n]$ with zero mean.

(a) Display the first 32 samples of $x[n]$. Can you identify the period from the plot? Compute and plot the DFT of $x[n]$. Does the spectrum match your expectations?

(b) Generate the noisy signal $y[n] = x[n] + s[n]$ and display the first 32 samples of $y[n]$. Do you detect any periodicity in the data? Compute and plot the DFT of the noisy signal $y[n]$. Can you identify the frequency and magnitude of the periodic component from the spectrum? Do they match your expectations?

(c) Generate the noisy signal $z[n] = x[n]s[n]$ (by elementwise multiplication) and display the first 32 samples of $z[n]$. Do you detect any periodicity in the data? Compute and plot the DFT of the noisy signal $z[n]$. Can you identify the frequency the periodic component from the spectrum?

16.44 (Filtering a Noisy ECG Signal: I) During recording, an electrocardiogram (ECG) signal, sampled at 300 Hz, gets contaminated by 60-Hz hum. Two beats of the original and contaminated signal (600 samples) are provided on disk as `ecgo.mat` and `ecg.mat`. Load these signals into MATLAB (for example, use the command `load ecgo`). In an effort to remove the 60-Hz hum, use the DFT as a filter to implement the following steps.

(a) Compute (but do not plot) the 600-point DFT of the contaminated ECG signal.

(b) By hand, compute the DFT indices that correspond to the 60-Hz signal.

(c) Zero out the DFT components corresponding the 60-Hz signal.

(d) Take the IDFT to obtain the filtered ECG and display the original and filtered signal.

(e) Display the DFT of the original and filtered ECG signal and comment on the differences.

(f) Is the DFT effective in removing the 60-Hz interference?

PGR2026-00033
Krisp Technologies EX1024 Page 601

16.45 (Filtering a Noisy ECG Signal: II) Continuing with Problem 16.44, load the original and contaminated ECG signal sampled at 300 Hz with 600 samples provided on disk as `ecgo.mat` and `ecg.mat` (for example, use the command `load ecgo`). Truncate each signal to 512 samples. In an effort to remove the 60-Hz hum, use the DFT as a filter to implement the following steps.

(a) Compute (but do not plot) the 512-point DFT of the contaminated ECG signal.
(b) Compute the DFT indices *closest to* 60 Hz and zero out the DFT at these indices.
(c) Take the IDFT to obtain the filtered ECG and display the original and filtered signal.
(d) Display the DFT of the original and filtered ECG signal and comment on the differences.
(e) The DFT is not effective in removing the 60-Hz interference. Why?
(f) From the DFT plots, suggest and implement a method for improving the results (by zeroing out a larger portion of the DFT around 60 Hz, for example).

16.46 (Decoding a Mystery Message) During transmission, a message signal gets contaminated by a low-frequency signal and high-frequency noise. The message can be decoded only by displaying it in the time domain. The contaminated signal is provided on disk as `mystery1.mat`. Load this signal into MATLAB (use the command `load mystery1`). In an effort to decode the message, use the DFT as a filter to implement the following steps and determine what the decoded message says.

(a) Display the contaminated signal. Can you "read" the message?
(b) Take the DFT of the signal to identify the range of the message spectrum.
(c) Zero out the DFT component corresponding to the low-frequency signal.
(d) Zero out the DFT components corresponding to the high-frequency noise.
(e) Take the IDFT to obtain the filtered signal and display it to decode the message.

16.47 (Spectrum Estimation) The FFT is extensively used in estimating the spectrum of various signals, detecting periodic components buried in noise, or detecting long-term trends. The monthly rainfall data, for example, tends to show periodicity (an annual cycle). However, long-term trends may also be present due to factors such as deforestation and soil erosion that tend to reduce rainfall amounts over time. Such long-term trends are often masked by the periodicity in the data and can be observed only if the periodic components are first removed (filtered).

(a) Generate a signal $x[n] = 0.01n + \sin(n\pi/6),\ 0 \le n \le 500$, and add some random noise to simulate monthly rainfall data. Can you observe any periodicity or long-term trend from a plot of the data?
(b) Find the FFT of the rainfall data. Can you identify the periodic component from the FFT magnitude spectrum?
(c) Design a notch filter to remove the periodic component from the rainfall data. You may identify the frequency to be removed from $x[n]$ (if you have not been able to identify it from the FFT). Filter the rainfall data through your filter and plot the filtered data. Do you observe any periodicity in the filtered data? Can you detect any long-term trends from the plot?
(d) To detect the long-term trend, pass the filtered data through a moving average filter. Experiment with different lengths. Does averaging of the filtered data reveal the long-term trend? Explain how you might go about quantifying the trend.

16.48 (The FFT as a Filter) We wish to filter out the 60-Hz interference from the signal

$$x(t) = \cos(100\pi t) + \cos(120\pi t)$$

by sampling $x(t)$ at $S = 500$ Hz and passing the sampled signal $x[n]$ through a lowpass filter with a cutoff frequency of $f_C = 55$ Hz. The N-point FFT of the filtered signal may be obtained by simply

PGR2026-00033
Krisp Technologies EX1024 Page 602

zeroing out the FFT $X[k]$ of the sampled signal between the indices $M = \text{int}(Nf_C/S)$ and $N - M$ (corresponding to the frequencies between f_C and $S - f_C$). This is entirely equivalent to multiplying $X[k]$ by a *filter function* $H[k]$ of the form

$$H[k] = \{\ 1,\ \ (M \text{ ones}),\ (N - 2M - 1 \text{ zeros}),\ (M \text{ ones})\ \}$$

The FFT of the filtered signal equals $Y[k] = H[k]X[k]$, and the filtered signal $y[n]$ is obtained by computing its IFFT.

(a) Start with the smallest value of N you need to resolve the two frequencies and successively generate the sampled signal $x[n]$, its FFT $X[k]$, the filter function $H[k]$, the FFT $Y[k]$ of the filtered signal as $Y[k] = H[k]X[k]$, and its IFFT $y[n]$. Plot $X[k]$ and $Y[k]$ on the same plot and $x[n]$ and $y[n]$ on the same plot. Is the 60-Hz signal completely blocked? Is the filtering effective?

(b) Double the value of N several times and, for each case, repeat the computations and plots of part (a). Is there a noticeable improvement in the filtering?

(c) The filter described by the N-sample $h[n]$ is not very useful because its true frequency response $H(F)$ matches the N-point FFT only at N points and varies considerably in between. To see this, superimpose the DTFT $H(F)$ of the N-sample $h[n]$ (using the value of N from part (a)) over $0 \le F \le 1$ (using $4N$ points) and the N-point FFT $X[k]$. How does $H(F)$ differ from $H[k]$? What is the reason for this difference? Can a larger N reduce the differences? If not, how can the differences be minimized? (This forms the subject of frequency sampling filters that we discuss in Chapter 20.)

16.49 (Band-Limited Interpolation) Consider the band-limited signal $x(t) = \sin(200\pi t)$.

(a) Sample $x(t)$ at $S = 400$ Hz to obtain the sampled signal $x[n]$ with $N = 4$ samples. Compute and plot the DFT $X[k]$ of $x[n]$. Can you identify the frequency and magnitude from the spectrum?

(b) Zero-interpolate the signal $G[k] = 8X[k]$ by inserting 28 zeros about the middle (about index $k = 0.5N$) to obtain the interpolated 32-point spectrum $Y[k]$. Compute the IDFT of $Y[k]$ to obtain the signal $y[n]$. Sample $x(t)$ again, but at $8S$ Hz, to obtain the sampled signal $y_1[n]$. Plot $e[n] = y[n] - y_1[n]$. Are $y[n]$ and $y_1[n]$ identical (to within machine roundoff)? Should they be?

(c) Sample $x(t)$ at $S = 800$ Hz to obtain the sampled signal $x[n]$ with $N = 4$ samples. Compute its DFT $X[k]$ and zero-interpolate the signal $G[k] = 8X[k]$ by inserting 28 zeros about the middle (about index $k = 0.5N$) to obtain the interpolated 32-point spectrum $Y[k]$. Compute the IDFT of $Y[k]$ to obtain the signal $y[n]$. Sample $x(t)$ again, but at $8S$ Hz, to obtain the sampled signal $y_1[n]$. Plot $e[n] = y[n] - y_1[n]$. Are $y[n]$ and $y_1[n]$ identical? Should they be?

(d) Explain the differences in the results of parts (a) and (b).

16.50 (Decimation) To decimate a signal $x[n]$ by N, we use a lowpass filter (to band-limit the signal to $F = 0.5/N$), followed by a down-sampler (that retains only every Nth sample). In this problem, ignore the lowpass filter.

(a) Generate the test signal $x[n] = \cos(0.2n\pi) + \cos(0.3n\pi),\ 0 \le n \le 59$. Plot its DFT. Can you identify the frequencies present?

(b) Decimate $x[n]$ by $N = 2$ to obtain the signal $x_2[n]$. Is the signal $x[n]$ sufficiently band-limited in this case? Plot the DFT of $x_2[n]$. Can you identify the frequencies present? Do the results match your expectations? Would you be able to recover $x[n]$ from band-limited interpolation (by $N = 2$) of $x_2[n]$?

PGR2026-00033
Krisp Technologies EX1024 Page 603

(c) Decimate $x[n]$ by $N = 4$ to obtain the signal $x_4[n]$. Is the signal $x[n]$ sufficiently band-limited in this case? Plot the DFT of $x_4[n]$. Can you identify the frequencies present? If not, explain how the result differs from that of part (b). Would you be able to recover $x[n]$ from band-limited interpolation (by $N = 4$) of $x_4[n]$?

16.51 (DFT of Large Data Sets) The DFT of a large N-point data set may be obtained from the DFT of smaller subsets of the data. In particular, if $N = RC$, we arrange the data as an $R \times C$ matrix (by filling along columns), find the DFT of each row, multiply each result at the location $(r,\ c)$ by $W_{rc} = e^{-j2\pi rc/N}$, where $r = 0, 1, \ldots, R-1$ and $c = 0, 1, \ldots, C-1$, find the DFT of each column, and reshape the result (by rows) into the required N-point DFT. Let $x[n] = n + 1,\ 0 \le n \le 11$.

(a) Find the DFT of $x[n]$ by using this method with $R = 3,\ C = 4$.
(b) Find the DFT of $x[n]$ by using this method with $R = 4,\ C = 3$.
(c) Find the DFT of $x[n]$ using the MATLAB command `fft`.
(d) Do all methods yield identical results? Can you justify your answer?

16.52 (Time-Frequency Plots) This problem deals with time-frequency plots of a sum of sinusoids.

(a) Generate 600 samples of the signal $x[n] = \cos(0.1n\pi) + \cos(0.4n\pi) + \cos(0.7n\pi)$, the sum of three pure cosines at $F = 0.05,\ 0.2,\ 0.35$. Use the MATLAB command `fft` to plot its DFT magnitude. Use the ADSP routine `timefreq` to display its time-frequency plot. What do the plots reveal?
(b) Generate 200 samples each of the three signals $y_1[n] = \cos(0.1n\pi)$, $y_2[n] = \cos(0.4n\pi)$, and $y_3[n] = \cos(0.7n\pi)$. Concatenate them to form the 600-sample signal $y[n] = \{y_1[n],\ y_2[n],\ y_3[n]\}$. Plot its DFT magnitude and display its time-frequency plot. What do the plots reveal?
(c) Compare the DFT magnitude and time-frequency plots of $x[n]$ and $y[n]$ How do they differ?

16.53 (Deconvolution) The FFT is a useful tool for deconvolution. Given an input signal $x[n]$ and the system response $y[n]$, the system impulse response $h[n]$ may be found from the IDFT of the ratio $H_{\text{DFT}}[k] = Y_{\text{DFT}}[k]/X_{\text{DFT}}[k]$. Let $x[n] = \{1,\ 2,\ 3,\ 4\}$ and $h[n] = \{1,\ 2,\ 3\}$.

(a) Obtain the convolution $y[n] = x[n] \star h[n]$. Now zero-pad $x[n]$ to the length of $y[n]$ and find the DFT of the two sequences and their ratio $H_{\text{DFT}}[k]$. Does the IDFT of $H_{\text{DFT}}[k]$ equal $h[n]$ (to within machine roundoff)? Should it?
(b) Repeat part (a) with $x[n] = \{1,\ 2,\ 3,\ 4\}$ and $h[n] = \{1,\ 2,\ -3\}$. Does the method work for this choice? Does the IDFT of $H_{\text{DFT}}[k]$ equal $h[n]$ (to within machine roundoff)?
(c) Repeat part (a) with $x[n] = \{1,\ 2,\ -3\}$ and $h[n] = \{1,\ 2,\ 3,\ 4\}$. Show that the method does not work because the division yields infinite or indeterminate results (such as $\frac{1}{0}$ or $\frac{0}{0}$). Does the method work if you replace zeros by very small quantities (for example, 10^{-10})? Should it?

16.54 (FFT of Two Real Signals at Once) Show that it is possible to find the FFT of two real sequences $x[n]$ and $y[n]$ from a single FFT operation on the complex signal $g[n] = x[n] + jy[n]$ as

$$X_{\text{DFT}}[k] = 0.5(G^*_{\text{DFT}}[N-k] + G_{\text{DFT}}[k]) \qquad Y_{\text{DFT}}[k] = j0.5(G^*_{\text{DFT}}[N-k] - G_{\text{DFT}}[k])$$

Use this result to find the FFT of $x[n] = \{1, 2, 3, 4\}$ and $y[n] = \{5, 6, 7, 8\}$ and compare the results with their FFT computed individually.

16.55 (Quantization Error) Quantization leads to noisy spectra. Its effects can be studied only in statistical terms. Let $x(t) = \cos(20\pi t)$ be sampled at 50 Hz to obtain the 256-point sampled signal $x[n]$.

PGR2026-00033
Krisp Technologies EX1024 Page 604

(a) Plot the linear and decibel magnitude of the DFT of $x[n]$.

(b) Quantize $x[n]$ by rounding to B bits to generate the quantized signal $y[n]$. Plot the linear and decibel magnitude of the DFT of $y[n]$. Compare the DFT spectra of $x[n]$ and $y[n]$ for $B = 8,\ 4,\ 2$, and 1. What is the effect of decreasing the number of bits on the DFT spectrum of $y[n]$?

(c) Repeat parts (a) and (b), using quantization by truncation. How do the spectra differ in this case?

(d) Repeat parts (a)–(c) after windowing $x[n]$ by a von Hann window. What is the effect of windowing?

16.56 (Sampling Jitter) During the sampling operation, the phase noise on the sampling clock can result in **jitter**, or random variations in the time of occurrence of the true sampling instant. Jitter leads to a noisy spectral and its effects can be studied only in statistical terms. Consider the analog signal $x(t) = \cos(2\pi f_0 t)$ sampled at a rate S that equals three times the Nyquist rate.

(a) Generate a time array t_n of 256 samples at intervals of $t_s = 1/S$. Generate the sampled signal $x[n]$ from values of $x(t)$ at the time instants in t_n. Plot the DFT magnitude of $x[n]$.

(b) Add some uniformly distributed random noise with a mean of zero and a noise amplitude of At_s to t_n to form the new time array t_{nn}. Generate the sampled signal $y[n]$ from values of $x(t)$ at the time instants in t_{nn}. Plot the DFT magnitude of $y[n]$ and compare with the DFT magnitude of $x[n]$ for $A = 0.01,\ 0.1,\ 1,\ 10$. What is the effect of increasing the noise amplitude on the DFT spectrum of $y[n]$? What is the largest value of A for which you can still identify the signal frequency from the DFT of $y[n]$?

(c) Repeat parts (a) and (b) after windowing $x[n]$ and $y[n]$ by a von Hann window. What is the effect of windowing?

PGR2026-00033
Krisp Technologies EX1024 Page 605

Chapter 17

THE z-TRANSFORM

17.0 Scope and Objectives

The z-transform plays the same role for discrete-time signals and systems as does the Laplace transform for continuous-time signals and systems. This chapter deals with the z-transform, its operational properties, and some of its applications in systems analysis and signal processing. Our description of the z-transform closely parallels that of Laplace transforms, the better to highlight the similarities between the two. The genesis of the z-transform may be viewed at several levels. We develop it as an independent transformation method in order to keep the discussion self-contained and explore the connections between the various transform methods toward the end of this chapter.

17.1 The Two-Sided z-Transform

The **two-sided z-transform** $X(z)$ of a discrete signal $x[n]$ is defined as

$$X(z) = \sum_{k=-\infty}^{\infty} x[k]z^{-k} \quad \text{(two-sided } z\text{-transform)} \tag{17.1}$$

The relation between $x[n]$ and $X(z)$ is denoted symbolically by

$$x[n] \Longleftrightarrow X(z) \tag{17.2}$$

Here, $x[n]$ and $X(z)$ form a transform pair, and the double arrow implies a one-to-one correspondence between the two.

17.1.1 What the z-Transform Reveals

The complex quantity z generalizes the concept of digital frequency F or Ω to the complex domain and is usually described in polar form as

$$z = |r|e^{j2\pi F} = |r|e^{j\Omega} \tag{17.3}$$

Values of z can be plotted on an Argand diagram called the ***z*-plane**.

The defining relation is a power series (Laurent series) in z. The term for each index k is the product of the sample value $x[k]$ and z^{-k}.

For the sequence $x[n] = \{-7,\ 3,\ \overset{\Downarrow}{1},\ 4,\ -8,\ 5\}$, for example, the z-transform may be written as

$$X(z) = -7z^2 + 3z^1 + z^0 + 4z^{-1} - 8z^{-2} + 5z^{-3}$$

PGR2026-00033
Krisp Technologies EX1024 Page 606

Comparing $x[n]$ and $X(z)$, we observe that the quantity z^{-1} plays the role of a unit delay operator. The sample location $n = 2$ in $x[n]$, for example, corresponds to the term with z^{-2} in $X(z)$. In concept, then, it is not hard to go back and forth between a sequence and its z-transform if all we are given is a finite number of samples.

REVIEW PANEL 17.1

The Two-Sided z-Transform of Finite Sequences Is a Power Series in z

Since the defining relation for $X(z)$ describes a power series, it may not converge for all z. The values of z for which it does converge define the **region of convergence** (ROC) for $X(z)$. Two completely different sequences may produce the same *two-sided* z-transform $X(z)$, but with different regions of convergence. It is important (unlike Laplace transforms) that we specify the ROC associated with each $X(z)$, especially when dealing with the two-sided z-transform.

17.1.2 Some z-Transform Pairs Using the Defining Relation

Table 17.1 lists the z-transforms of some useful signals. We provide some examples using the defining relation to find z-transforms. For finite-length sequences, the z-transform may be written as a polynomial in z. For sequences with a large number of terms, the polynomial form can get to be unwieldy unless we can find closed-form solutions.

EXAMPLE 17.1 (The z-Transform from the Defining Relation)

(a) Let $x[n] = \delta[n]$. Its z-transform is $X(z) = 1$. The ROC is the entire z-plane.

(b) Let $x[n] = 2\delta[n+1] + \delta[n] - 5\delta[n-1] + 4\delta[n-2]$. This describes the sequence $x[n] = \{2, \overset{\Downarrow}{1}, -5, 4\}$. Its z-transform is evaluated as $X(z) = 2z + 1 - 5z^{-1} + 4z^{-2}$. No simplifications are possible. The ROC is the entire z-plane, except $z = 0$ and $z = \infty$ (or $0 < |z| < \infty$).

(c) Let $x[n] = u[n] - u[n-N]$. This represents a sequence of N samples, and its z-transform may be written as

$$X(z) = 1 + z^{-1} + z^{-2} + \cdots + z^{-(N-1)}$$

A closed-form solution may be found using the defining relation as follows:

$$X(z) = \sum_{k=0}^{N-1} z^{-k} = \frac{1 - z^{-N}}{1 - z^{-1}}, \quad z \neq 1 \qquad \text{ROC: } z \neq 0$$

The ROC is the entire z-plane, except $z = 0$ (or $|z| > 0$). Note that if $z = 1$, we get $X(z) = N$.

(d) Let $x[n] = u[n]$. We evaluate its z-transform using the defining relation as follows:

$$X(z) = \sum_{k=0}^{\infty} z^{-k} = \sum_{k=0}^{\infty} (z^{-1})^k = \frac{1}{1 - z^{-1}} = \frac{z}{z-1}, \qquad \text{ROC: } |z| > 1$$

The geometric series converges only for $|z^{-1}| < 1$ or $|z| > 1$, which defines the ROC for this $X(z)$.

PGR2026-00033
Krisp Technologies EX1024 Page 607

Table 17.1 A Short Table of z-Transform Pairs

Entry	Signal	z-Transform	ROC
		Finite Sequences	
1	$\delta[n]$	1	all z
2	$u[n] - u[n-N]$	$\dfrac{1 - z^{-N}}{1 - z^{-1}}$	$z \neq 0$
		Causal Signals	
3	$u[n]$	$\dfrac{z}{z-1}$	$\lvert z\rvert > 1$
4	$\alpha^n u[n]$	$\dfrac{z}{z-\alpha}$	$\lvert z\rvert > \lvert\alpha\rvert$
5	$(-\alpha)^n u[n]$	$\dfrac{z}{z+\alpha}$	$\lvert z\rvert > \lvert\alpha\rvert$
6	$nu[n]$	$\dfrac{z}{(z-1)^2}$	$\lvert z\rvert > 1$
7	$n\alpha^n u[n]$	$\dfrac{z\alpha}{(z-\alpha)^2}$	$\lvert z\rvert > \lvert\alpha\rvert$
8	$\cos(n\Omega)u[n]$	$\dfrac{z^2 - z\cos\Omega}{z^2 - 2z\cos\Omega + 1}$	$\lvert z\rvert > 1$
9	$\sin(n\Omega)u[n]$	$\dfrac{z\sin\Omega}{z^2 - 2z\cos\Omega + 1}$	$\lvert z\rvert > 1$
10	$\alpha^n\cos(n\Omega)u[n]$	$\dfrac{z^2 - \alpha z\cos\Omega}{z^2 - 2\alpha z\cos\Omega + \alpha^2}$	$\lvert z\rvert > \lvert\alpha\rvert$
11	$\alpha^n\sin(n\Omega)u[n]$	$\dfrac{\alpha z\sin\Omega}{z^2 - 2\alpha z\cos\Omega + \alpha^2}$	$\lvert z\rvert > \lvert\alpha\rvert$
		Anti-Causal Signals	
12	$-u[-n-1]$	$\dfrac{z}{z-1}$	$\lvert z\rvert < 1$
13	$-nu[-n-1]$	$\dfrac{z}{(z-1)^2}$	$\lvert z\rvert < 1$
14	$-\alpha^n u[-n-1]$	$\dfrac{z}{z-\alpha}$	$\lvert z\rvert < \lvert\alpha\rvert$
15	$-n\alpha^n u[-n-1]$	$\dfrac{z\alpha}{(z-\alpha)^2}$	$\lvert z\rvert < \lvert\alpha\rvert$

PGR2026-00033
Krisp Technologies EX1024 Page 608

(e) Let $x[n] = \alpha^n u[n]$. Using the defining relation, its z-transform and ROC are

$$X(z) = \sum_{k=0}^{\infty} \alpha^k z^{-k} = \sum_{k=0}^{\infty} \left(\frac{\alpha}{z}\right)^k = \frac{1}{1-(\alpha/z)} = \frac{z}{z-\alpha}, \qquad \text{ROC: } |z| > |\alpha|$$

17.1.3 The Region of Convergence

For a finite sequence $x[n]$, the z-transform $X(z)$ is a polynomial in z or z^{-1} and converges (is finite) for all z, except $z = 0$, if $X(z)$ contains terms of the form z^{-k} (or $x[n]$ is nonzero for $n > 0$), and/or $z = \infty$ if $X(z)$ contains terms of the form z^k (or $x[n]$ is nonzero for $n < 0$). Thus, the ROC for finite sequences is the entire z-plane, except perhaps for $z = 0$ and/or $z = \infty$, as applicable.

In general, if $X(z)$ is a rational function in z, as is often the case, its ROC actually depends on the one- or two-sidedness of $x[n]$, as illustrated in Figure 17.1.

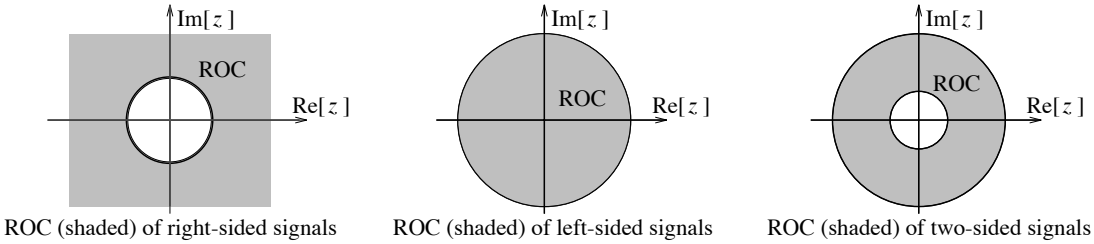


Figure 17.1 The ROC (shown shaded) of the z-transform for various sequences

The ROC *excludes all pole locations* (denominator roots) where $X(z)$ becomes infinite. As a result, the ROC of right-sided signals is $|z| > |p|_{\max}$ and lies exterior to a circle of radius $|p|_{\max}$, the magnitude of the largest pole. The ROC of causal signals, with $x[n] = 0,\ n < 0$, excludes the origin and is given by $0 > |z| > p|_{\max}$. Similarly, the ROC of a left-sided signal $x[n]$ is $|z| < |p|_{\min}$ and lies interior to a circle of radius $|p|_{\min}$, the smallest pole magnitude of $X(z)$. Finally, the ROC of a two-sided signal $x[n]$ is $|p|_{\min} < |z| < |p|_{\max}$, an annulus whose radii correspond to the smallest and largest pole locations in $X(z)$. We use *inequalities* of the form $|z| < |\alpha|$ (and not $|z| \leq |\alpha|$), for example, because $X(z)$ may not converge at the boundary $|z| = |\alpha|$.

REVIEW PANEL 17.2

The ROC of the z-Transform $X(z)$ Determines the Nature of the Signal $x[n]$

Finite-length $x[n]$:	ROC of $X(z)$ is all the z-plane, except perhaps for $z = 0$ and/or $z = \infty$.
Right-sided $x[n]$:	ROC of $X(z)$ is outside a circle whose radius is the largest pole magnitude.
Left-sided $x[n]$:	ROC of $X(z)$ is inside a circle whose radius is the smallest pole magnitude.
Two-sided $x[n]$:	ROC of $X(z)$ is an annulus bounded by the largest and smallest pole radius.

Why We Must Specify the ROC

Consider the signal $y[n] = -\alpha^n u[-n-1] = -1,\ n = -1, -2, \ldots$. The two-sided z-transform of $y[n]$, using a change of variables, can be written as

$$Y(z) = \sum_{k=-\infty}^{-1} -\alpha^k z^{-k} = \sum_{m=1}^{\infty} -\left(\frac{z}{\alpha}\right)^m = -\frac{z/\alpha}{1-(z/\alpha)} = \frac{z}{z-\alpha}, \qquad \text{ROC: } |z| > |\alpha|$$

PGR2026-00033
Krisp Technologies EX1024 Page 609

The ROC of $Y(z)$ is $|z| < |\alpha|$. The z-transform of $\alpha^n u[n]$ is also $z/(z-\alpha)$, but with an ROC of $|z| > |\alpha|$. Do you see the problem? We cannot uniquely identify a signal from its transform alone, unless we also specify the ROC. In this book, we will assume a right-sided signal if no ROC is specified.

REVIEW PANEL 17.3

We Assume Right-Sided Signals if No ROC Is Specified.

EXAMPLE 17.2 (Identifying the ROC)

(a) Let $x[n] = \{4,\ -3,\ \underset{\Downarrow}{2},\ 6\}$. The ROC of $X(z)$ is $0 < |z| < \infty$ and *excludes* $z = 0$ and $z = \infty$ because $x[n]$ is nonzero for $n < 0$ and $n > 0$.

(b) Let $X(z) = \dfrac{z}{z-2} + \dfrac{z}{z+3}$.

Its ROC depends on the nature of $x[n]$.

If $x[n]$ is assumed right-sided, the ROC is $|z| > 3$ (because $|p|_{\max} = 3$).

If $x[n]$ is assumed left-sided, the ROC is $|z| < 2$ (because $|p|_{\min} = 2$).

If $x[n]$ is assumed two-sided, the ROC is $2 < |z| < 3$.

The region $|z| < 2$ and $|z| > 3$ does not correspond to a valid region of convergence because we must find a region that is common to both terms.

17.2 Properties of the Two-Sided z-Transform

The z-transform is a linear operation and obeys superposition. The properties of the z-transform, listed in Table 17.2, are based on the linear nature of the z-transform operation.

Time Shift: To prove the time-shift property of the two-sided z-transform, we use a change of variables. We start with the pair $x[n] \Leftrightarrow X(z)$. If $y[n] = x[n-N]$, its z-transform is

$$Y(z) = \sum_{k=-\infty}^{\infty} x[k-N]z^{-k} \tag{17.4}$$

With the change of variable $m = k - N$, the new summation index m still ranges from $-\infty$ to ∞ (since N is finite), and we obtain

$$Y(z) = \sum_{m=-\infty}^{\infty} x[m]z^{-(m+N)} = z^{-N} \sum_{m=-\infty}^{\infty} x[m]z^{-m} = z^{-N}X(z) \tag{17.5}$$

The factor z^{-N} with $X(z)$ induces a right shift of N in $x[n]$.

REVIEW PANEL 17.4

Time-Shift Property for the Two-Sided z-Transform: $x[n-N] \iff z^{-N}X(z)$

PGR2026-00033
Krisp Technologies EX1024 Page 610

Table 17.2 Properties of the Two-Sided z-Transform

Entry	Property	Signal	z-Transform
1	Shifting	$x[n-N]$	$z^{-N}X(z)$
2	Reflection	$x[-n]$	$X\left(\frac{1}{z}\right)$
3	Anti-causal	$x[-n]u[-n-1]$	$X\left(\frac{1}{z}\right) - x[0]$ (for causal $x[n]$)
4	Scaling	$\alpha^n x[n]$	$X\left(\frac{z}{\alpha}\right)$
5	Times-n	$nx[n]$	$-z\,\frac{dX(z)}{dz}$
6	Times-cos	$\cos(n\Omega)x[n]$	$0.5\left[X(ze^{j\Omega}) + X(ze^{-j\Omega})\right]$
7	Times-sin	$\sin(n\Omega)x[n]$	$j0.5\left[X(ze^{j\Omega}) - X(ze^{-j\Omega})\right]$
8	Convolution	$x[n] \star h[n]$	$X(z)H(z)$

Times-n: The times-n property is established by taking derivatives, to yield

$$X(z) = \sum_{k=-\infty}^{\infty} x[k]z^{-k} \qquad \frac{dX(z)}{dz} = \sum_{k=-\infty}^{\infty} \frac{d}{dz}\left[x[k]z^{-k}\right] = \sum_{k=-\infty}^{\infty} -kx[k]z^{-(k+1)} \tag{17.6}$$

Multiplying both sides by $-z$, we obtain

$$-z\,\frac{dX(z)}{dz} = \sum_{k=-\infty}^{\infty} kx[k]z^{-k} \tag{17.7}$$

This represents the transform of $nx[n]$.

REVIEW PANEL 17.5

The Times-n Property: $nx[n] \iff -z\,\frac{dX(z)}{dz}$

Scaling: The scaling property follows from the transform of $y[n] = \alpha^n x[n]$, to yield

$$Y(z) = \sum_{k=-\infty}^{\infty} \alpha^k x[k]z^{-k} = \sum_{k=-\infty}^{\infty} x[k]\left(\frac{z}{\alpha}\right)^{-k} = X\left(\frac{z}{\alpha}\right) \tag{17.8}$$

If $x[n]$ is multiplied by $e^{jn\Omega}$ or $\left(e^{j\Omega}\right)^n$, we then obtain the pair $e^{jn\Omega}x[n] \Leftrightarrow X(ze^{-j\Omega})$. An extension of this result, using Euler's relation, leads to the times-cos and times-sin properties:

$$\cos(n\Omega)x[n] = 0.5x[n]\left(e^{jn\Omega} + e^{-jn\Omega}\right) \iff 0.5\left[X\left(ze^{j\Omega}\right) + X\left(ze^{-j\Omega}\right)\right] \tag{17.9}$$

$$\sin(n\Omega)x[n] = -j0.5x[n]\left(e^{jn\Omega} - e^{-jn\Omega}\right) \iff j0.5\left[X\left(ze^{j\Omega}\right) - X\left(ze^{-j\Omega}\right)\right] \tag{17.10}$$

In particular, if $\alpha = -1$, we obtain the useful result $(-1)^n x[n] \Leftrightarrow X(-z)$.

PGR2026-00033
Krisp Technologies EX1024 Page 611

REVIEW PANEL 17.6

The Scaling Property: $\alpha^n x[n] \iff X(z/\alpha)$ and $(-1)^n x[n] \iff X(-z)$

Convolution: The convolution property is based on the fact that multiplication in the time domain corresponds to convolution in any transformed domain. The z-transforms of sequences are polynomials, and multiplication of two polynomials corresponds to the convolution of their coefficient sequences. This property finds extensive use in the analysis of systems in the transformed domain.

REVIEW PANEL 17.7

The Convolution Property: $x[n] \star h[n] \iff X(z)H(z)$

Folding: The folding property comes about if we use $\alpha = -1$ in the scaling property (or $k \to -k$ in the defining relation). With $x[n] \Leftrightarrow X(z)$ and $y[n] = x[-n]$, we get

$$Y(z) = \sum_{k=-\infty}^{\infty} x[-k]z^{-k} = \sum_{k=-\infty}^{\infty} x[k]z^{k} = \sum_{k=-\infty}^{\infty} x[k](1/z)^{-k} = X(1/z) \tag{17.11}$$

If the ROC of $x[n]$ is $|z| > |\alpha|$, the ROC of the folded signal $x[-n]$ becomes $|1/z| > |\alpha|$ or $|z| < 1/|\alpha|$.

REVIEW PANEL 17.8

The Folding Property of the Two-Sided z-Transform

$x[-n] \iff X(1/z)$ (the ROC changes from $|z| > |\alpha|$ to $|z| < 1/|\alpha|$)

The Folding Property and Symmetric Signals

The folding property is useful in checking for signal symmetry from its z-transform. For an even symmetric signal, we have $x[n] = x[-n]$, and thus $X(z) = X(1/z)$. For an odd symmetric signal, we have $x[n] = x[-n]$, and thus $X(z) = -X(1/z)$.

REVIEW PANEL 17.9

A Property of the z-Transform of Symmetric Sequences

Even symmetry: $x[n]=x[-n] \quad X(z)=X(1/z)$ **Odd symmetry:** $x[n]=-x[-n] \quad X(z)=-X(1/z)$

The Folding Property and Anti-Causal Signals

The folding property is also useful in finding the transform of anti-causal signals. From the causal signal $x[n]u[n] \Leftrightarrow X(z)$ (with ROC $|z| > |\alpha|$), we find the transform of $x[-n]u[-n]$ as $X(1/z)$ (whose ROC is $|z| < 1/|\alpha|$). The anti-causal signal $y[n] = x[-n]u[-n-1]$ (which excludes the sample at $n = 0$) can then be written as $y[n] = x[-n]u[-n] - x[0]\delta[n]$, as illustrated in Figure 17.2.

The transform of $x[-n]u[-n-1]$ then gives

$$x[-n]u[-n-1] \Leftrightarrow X(1/z) - x[0], \quad |z| < 1/|\alpha|, \qquad \text{where} \qquad x[n]u[n] \Leftrightarrow X(z), \quad |z| > |\alpha| \tag{17.12}$$

REVIEW PANEL 17.10

How to Find the z-Transform of a Left-Sided Signal from Its Right-Sided Version

If $x[n]u[n] \iff X(z), \quad |z| > |\alpha|, \quad$ then $\quad x[-n]u[-n-1] \iff X(1/z) - x[0], \quad |z| < 1/|\alpha|.$

PGR2026-00033
Krisp Technologies EX1024 Page 612

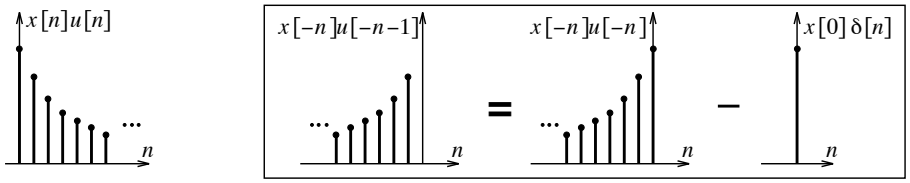


Figure 17.2 Finding the z-transform of an anti-causal signal from a causal version

EXAMPLE 17.3 (z-Transforms Using Properties)

(a) Using the times-n property, the z-transform of $y[n] = nu[n]$ is

$$Y(z) = -z\frac{d}{dz}\left[\frac{z}{z-1}\right] = -z\left[\frac{-z}{(z-1)^2} + \frac{1}{z-1}\right] = \frac{z}{(z-1)^2}$$

(b) With $x[n] = \alpha^n nu[n]$, we use scaling to obtain the z-transform:

$$X(z) = \frac{z/\alpha}{[(z/\alpha)-1]^2} = \frac{z\alpha}{(z-\alpha)^2}$$

(c) We find the transform of the N-sample exponential pulse $x[n] = \alpha^n(u[n] - u[n-N])$. We let $y[n] = u[n] - u[n-N]$. Its z-transform is

$$Y(z) = \frac{1-z^{-N}}{1-z^{-1}}, \quad |z| \neq 1$$

Then, the z-transform of $x[n] = \alpha^n y[n]$ becomes

$$X[z] = \frac{1-(z/\alpha)^{-N}}{1-(z/\alpha)^{-1}}, \quad z \neq \alpha$$

(d) The z-transforms of $x[n] = \cos(n\Omega)u[n]$ and $y[n] = \sin(n\Omega)u[n]$ are found using the times-cos and times-sin properties:

$$X(z) = 0.5\left[\frac{ze^{j\Omega}}{ze^{j\Omega}-1} + \frac{ze^{-j\Omega}}{ze^{-j\Omega}-1}\right] = \frac{z^2 - z\cos\Omega}{z^2 - 2z\cos\Omega + 1}$$

$$Y(z) = j0.5\left[\frac{ze^{j\Omega}}{ze^{j\Omega}-1} - \frac{ze^{-j\Omega}}{ze^{-j\Omega}-1}\right] = \frac{z\sin\Omega}{z^2 - 2z\cos\Omega + 1}$$

(e) The z-transforms of $f[n] = \alpha^n\cos(n\Omega)u[n]$ and $g[n] = \alpha^n\sin(n\Omega)u[n]$ follow from the results of part (d) and the scaling property:

$$F(z) = \frac{(z/\alpha)^2 - (z/\alpha)\cos\Omega}{(z/\alpha)^2 - 2(z/\alpha)\cos\Omega + 1} = \frac{z^2 - \alpha z\cos\Omega}{z^2 - 2\alpha z\cos\Omega + \alpha^2}$$

$$G(z) = \frac{(z/\alpha)\sin\Omega}{(z/\alpha)^2 - 2(z/\alpha)\cos\Omega + 1} = \frac{\alpha z\sin\Omega}{z^2 - 2\alpha z\cos\Omega + \alpha^2}$$

PGR2026-00033
Krisp Technologies EX1024 Page 613

(f) We use the folding property to find the transform of $x[n] = \alpha^{-n}u[-n-1]$. We start with the transform pair $y[n] = \alpha^n u[n] \Leftrightarrow z/(z-\alpha)$, ROC: $|z| > |\alpha|$. With $y[0] = 1$ and $x[n] = y[-n] - \delta[n]$, we find

$$\alpha^{-n}u[-n-1] \Longleftrightarrow \frac{1/z}{1/z-\alpha} - 1 = \frac{\alpha z}{1-\alpha z}, \qquad \text{ROC: } |z| < \frac{1}{|\alpha|}$$

If we replace α by $1/\alpha$, and change the sign of the result, we get

$$-\alpha^n u[-n-1] \Longleftrightarrow \frac{z}{z-\alpha}, \qquad \text{ROC: } |z| < |\alpha|$$

(g) We use the folding property to find the transform of $x[n] = \alpha^{|n|}$, $|\alpha| < 1$ (a two-sided decaying exponential). We write this as $x[n] = \alpha^n u[n] + \alpha^{-n}u[n] - \delta[n]$ (a one-sided decaying exponential and its folded version, less the extra sample included at the origin), as illustrated in Figure E17.3G.

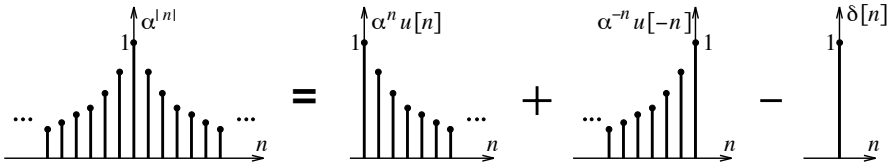


Figure E17.3G The signal for Example 17.3(g)

Its z-transform then becomes

$$X(z) = \frac{z}{z-\alpha} + \frac{1/z}{(1/z)-\alpha} - 1 = \frac{z}{z-\alpha} - \frac{z}{z-(1/\alpha)}, \qquad \text{ROC: } |\alpha| < |z| < \frac{1}{|\alpha|}$$

Note that the ROC is an annulus that corresponds to a two-sided sequence, and describes a valid region only if $|\alpha| < 1$.

17.3 Poles, Zeros, and the z-Plane

The z-transform of many signals is a *rational* function of the form

$$X(z) = \frac{N(z)}{D(z)} = \frac{B_0 + B_1 z^{-1} + B_2 z^{-2} + \cdots + B_M z^{-M}}{1 + A_1 z^{-1} + A_2 z^{-2} + \cdots + A_N z^{-N}} \tag{17.13}$$

Here, the coefficient A_0 of the leading term in the denominator has been normalized to unity.

Denoting the roots of $N(z)$ by z_i, $i = 1, 2, \ldots, M$ and the roots of $D(z)$ by p_k, $k = 1, 2, \ldots, N$, we may also express $X(z)$ in factored form as

$$X(z) = K\frac{N(z)}{D(z)} = K\frac{(z-z_1)(z-z_2)\cdots(z-z_M)}{(z-p_1)(z-p_2)\cdots(z-p_N)} \tag{17.14}$$

Assuming that common factors have been canceled, the p roots of $N(z)$ and the q roots of $D(z)$ are termed the **zeros** and the **poles** of the transfer function, respectively.

PGR2026-00033
Krisp Technologies EX1024 Page 614

17.3.1 Pole-Zero Plots

A plot of the poles (denoted by $\times$) and zeros (denoted by circles) of a rational function $X(z)$ in the z-plane constitutes a **pole-zero plot**, and provides a visual picture of the root locations. For multiple roots, we indicate their multiplicity next to the root location on the plot.

Clearly, we can find $X(z)$ directly from a pole-zero plot of the root locations but only to within the multiplicative gain factor K. If the value of K is also shown on the plot, $X(z)$ is known in its entirety.

REVIEW PANEL 17.11

Poles and Zeros of a Rational Function $X(z) = N(z)/D(z)$

Zeros: Roots of the numerator $N(z)$ **Poles:** Roots of the denominator $D(z)$

EXAMPLE 17.4 (Pole-Zero Plots)

(a) Let $H(z) = \dfrac{2z(z+1)}{(z-\frac{1}{3})(z^2+\frac{1}{4})(z^2+4z+5)}$.

The numerator degree is 2. The two zeros are $z = 0$ and $z = -1$.

The denominator degree is 5. The five finite poles are at $z = -\frac{1}{3},\ z = \pm j\frac{1}{2}$, and $z = -2 \pm j$.

The gain factor is $K = 2$. The pole-zero plot is shown in Figure E17.4(a).

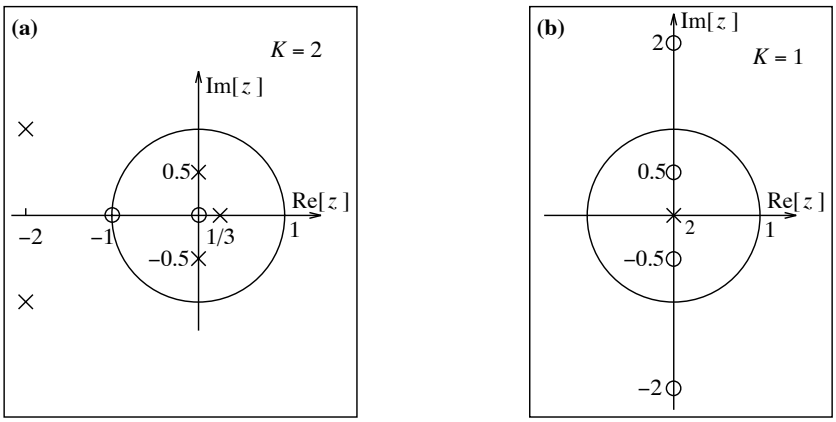


Figure E17.4 Pole-zero plots for Example 17.4(a and b)

(b) What is the z-transform corresponding to the pole-zero pattern of Figure E17.4(b)? Does it represent a symmetric signal?

If we let $X(z) = KN(z)/D(z)$, the four zeros correspond to the numerator $N(z)$ given by

$$N(z) = (z - j0.5)(z + j2)(z + j0.5)(z - j2) = z^4 + 4.25z^2 + 1$$

The two poles at the origin correspond to the denominator $D(z) = z^2$. With $K = 1$, the z-transform is given by

$$X(z) = K\frac{N(z)}{D(z)} = \frac{z^4 + 4.25z^2 + 1}{z^2} = z^2 + 4.25 + z^{-2}$$

PGR2026-00033
Krisp Technologies EX1024 Page 615

Checking for symmetry, we find that $X(z) = X(1/z)$, and thus $x[n]$ is even symmetric. In fact, $x[n] = \delta[n+2] + 4.25\delta[n] + \delta[n-2] = \{1,\ \underset{\Downarrow}{4.25},\ 1\}$. We also note that each zero is paired with its reciprocal ($j0.5$ with $-j2$, and $-j0.5$ with $j2$), a characteristic of symmetric sequences.

17.4 The Transfer Function

The response $y[n]$ of a system with impulse response $h[n]$, to an arbitrary input $x[n]$, is given by the convolution $y[n] = x[n] \star h[n]$. Since the convolution operation transforms to a product, we have

$$Y(z) = X(z)H(z) \qquad \text{or} \qquad H(z) = \frac{Y(z)}{X(z)} \tag{17.15}$$

The time-domain and z-domain equivalence of these operations is illustrated in Figure 17.3.

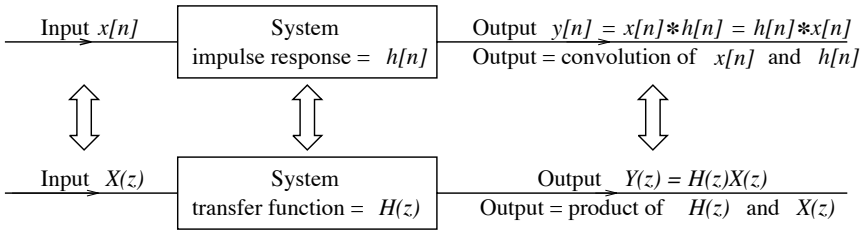


Figure 17.3 System description in the time domain and z-domain

The transfer function is defined only for *relaxed LTI systems*, either as the ratio of the output $Y(z)$ and input $X(z)$, or as the z-transform of the system impulse response $h[n]$.

A relaxed LTI system is also described by the difference equation:

$$y[n] + A_1 y[n-1] + \cdots + A_N y[n-N] = B_0 x[n] + B_1 x[n-1] + \cdots + B_M x[n-M] \tag{17.16}$$

Its z-transform results in the transfer function:

$$H(z) = \frac{Y(z)}{X(z)} = \frac{B_0 + B_1 z^{-1} + \cdots + B_M z^{-M}}{1 + A_1 z^{-1} + \cdots + A_N z^{-N}} \tag{17.17}$$

The transfer function is thus a ratio of polynomials in z. It also allows us to retrieve either the system difference equation or the impulse response.

REVIEW PANEL 17.12

The Transfer Function of Relaxed LTI Systems Is a Rational Function of z

$$H(z) = \frac{Y(z)\ \text{(transformed output)}}{X(z)\ \text{(transformed input)}} \qquad H(z) = \mathcal{Z}\{h[n]\} \quad (z\text{-transform of impulse response})$$

PGR2026-00033
Krisp Technologies EX1024 Page 616

17.4.1 System Representation in Various Forms

A relaxed LTI system may be described in various forms: by its difference equation, its impulse response, or its transfer function. Given one form, it is fairly easy to obtain the others using time-domain methods and/or frequency-domain transformations. For example, given the impulse response $h[n]$, we can find the transfer function from its z-transform $H(z)$. If we express the transfer function as $H(z) = Y(z)/X(z)$, cross-multiplication and inverse transformation lead to the system difference equation.

REVIEW PANEL 17.13

Three Representations of a Relaxed LTI System

Difference equation: from $H(z) = \frac{Y(z)}{X(z)}$ by cross-multiplication and inverse z-transform (IZT)

Transfer function: $H(z) = \frac{Y(z)}{X(z)}$ **Impulse response:** $h[n] = \text{IZT}\{H(z)\}$

EXAMPLE 17.5 (System Representation in Various Forms)

(a) Let $y[n] = 0.8y[n-1] + 2x[n]$. We obtain its transfer function and impulse response as follows:

$$Y(z) = 0.8z^{-1}Y(z) + 2X(z) \qquad H(z) = \frac{Y(z)}{X(z)} = \frac{2}{1 - 0.8z^{-1}} = \frac{2z}{z - 0.8} \qquad h[n] = 2(0.8)^n u[n]$$

(b) Let $h[n] = \delta[n] - 0.4(0.5)^n u[n]$. We obtain its transfer function as

$$H(z) = 1 - \frac{0.4z}{z - 0.5} = \frac{0.6z - 0.5}{z - 0.5}$$

We also obtain the difference equation by expressing the transfer function as

$$H(z) = \frac{Y(z)}{X(z)} = \frac{0.6z - 0.5}{z - 0.5} \quad \text{or} \quad \frac{0.6 - 5z^{-1}}{1 - 0.5z^{-1}}$$

and using cross-multiplication to give

$$(z - 0.5)Y(z) = (0.6z - 0.5)X(z) \quad \text{or} \quad (1 - 0.5z^{-1})Y(z) = (0.6 - 0.5z^{-1})X(z)$$

The difference equation may then be found using forward differences or backward differences as

$$y[n+1] - 0.5y[n] = 0.6x[n+1] - 0.5x[n] \quad \text{or} \quad y[n] - 0.5y[n-1] = 0.6x[n] - 0.5x[n-1]$$

The impulse response, the transfer function, and the difference equation describe the same system.

17.4.2 Poles and Zeros of the Transfer Function

The transfer function of an LTI system may be expressed in factored form as

$$H(z) = \frac{B_0 + B_1 z^{-1} + \cdots + B_M z^{-M}}{1 + A_1 z^{-1} + \cdots + A_N z^{-N}} = K\frac{(z - z_1)\cdots(z - z_M)}{(z - p_1)\cdots(z - p_N)} \tag{17.18}$$

The poles of $H(z)$ are called **natural modes** or **natural frequencies**. The poles of $Y(z) = H(z)X(z)$ determine the form of the system response. Clearly, the natural frequencies in $H(z)$ will always appear in the system response, unless they are canceled by any corresponding zeros in $X(z)$. The zeros of $H(z)$ may be regarded as the (complex) frequencies that are blocked by the system. For causal systems, the number of zeros cannot exceed the number of poles. In other words, the degree of the numerator polynomial in $H(z)$ cannot exceed the degree of the denominator polynomial, and the transfer function must be *proper*.

PGR2026-00033
Krisp Technologies EX1024 Page 617

17.4.3 Stability of LTI Systems

In the time domain, bounded-input, bounded-output (BIBO) stability of an LTI system requires an absolutely summable impulse response $h[n]$. For a causal system, this is equivalent to requiring the poles of the transfer function $H(z)$ to lie entirely *within* the unit circle in the z-plane. This equivalence stems from the following observations:

Poles outside the unit circle ($|z| > 1$) lead to exponential growth even if the input is bounded.
Example: $H(z) = z/(z-3)$ results in the growing exponential $(3)^n u[n]$.

Multiple poles on the unit circle always result in polynomial growth.
Example: $H(z) = 1/[z(z-1)^2]$ produces a ramp function in $h[n]$.

Simple (non-repeated) poles on the unit circle can also lead to an unbounded response.
Example: A simple pole at $z = 1$ leads to $H(z)$ with a factor $z/(z-1)$. If $X(z)$ also contains a pole at $z = 1$, the response $Y(z)$ will contain the term $z/(z-1)^2$ and exhibit polynomial growth.

None of these types of time-domain terms is absolutely summable, and their presence leads to system instability. Formally, for BIBO stability, all the poles of $H(z)$ must *lie inside (and exclude)* the unit circle $|z| = 1$. This is both a necessary and sufficient condition for the stability of causal systems.

If a system has simple (non-repeated) poles on the unit circle, it is sometimes called **marginally stable**. If a system has all its poles and zeros inside the unit circle, it is called a **minimum-phase system**. We study such systems later, in this and the next chapter.

Stability of anti-causal systems: In analogy with the discussion for causal systems, the stability of anti-causal systems requires that all the poles of $H(z)$ *lie outside, and exclude*, the unit circle $|z| = 1$.

17.4.4 Stability and the ROC

The ROC of a stable system (be it causal, anti-causal, or two-sided) always includes the unit circle. The various situations are illustrated in Figure 17.4.

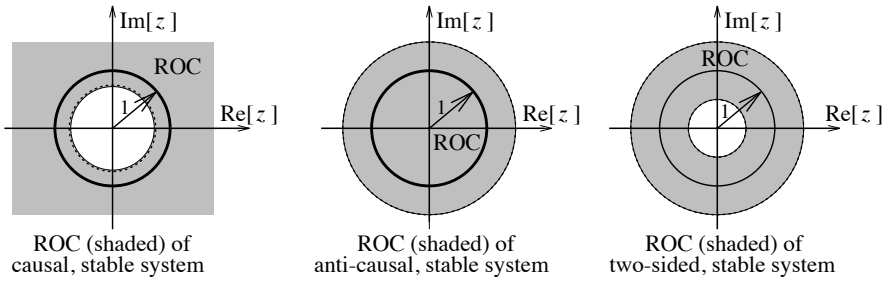


Figure 17.4 The ROC of stable systems (shown shaded) always includes the unit circle

In the time domain, a causal system requires a causal impulse response $h[n]$ with $h[n] = 0, n < 0$. In the z-domain, this is equivalent to a transfer function $H(z)$ that is proper and whose ROC lies outside a circle of finite radius. For stability, the poles of $H(z)$ must lie inside the unit circle. Thus, *for a system to be both causal and stable, the* ROC *must include the unit circle*.

The stability of an anti-causal system requires all the poles to lie outside the unit circle, and an ROC that lies inside a circle of finite radius. Thus, *for a system to be both anti-causal and stable, the* ROC *must include the unit circle*. Similarly, the ROC of a stable, two-sided system must be an annulus that includes the unit circle, and all its poles must lie *outside* outside this annulus.

PGR2026-00033
Krisp Technologies EX1024 Page 618

REVIEW PANEL 17.14

The ROC of Stable LTI Systems Always Includes the Unit Circle

Stable, causal system: All the poles must lie inside the unit circle.
Stable, anti-causal system: All the poles must lie outside the unit circle.
Stability from impulse response: $h[n]$ must be *absolutely summable* ($\sum |h[k]| < \infty$).

EXAMPLE 17.6 (Stability of a Recursive Filter)

(a) Let $H(z) = \dfrac{z}{z-\alpha}$.

If the ROC is $|z| > |\alpha|$, its impulse response is $h[n] = \alpha^n u[n]$, and the system is causal.

For stability, we require $|\alpha| < 1$ (for the ROC to include the unit circle).

(b) Let $H(z) = \dfrac{z}{z-\alpha}$, as before.

If the ROC is $|z| < |\alpha|$, its impulse response is $h[n] = -\alpha^n u[-n-1]$, and the system is anti-causal.

For stability, we require $|\alpha| > 1$ (for the ROC to include the unit circle).

17.5 The Inverse z-Transform

The formal inversion relation that yields $x[n]$ from $X(z)$ actually involves complex integration and is described by

$$x[n] = \frac{1}{j2\pi} \oint_{\Gamma} X(z) z^{n-1}\, dz \tag{17.19}$$

Here, Γ describes a clockwise contour of integration (such as the unit circle) that encloses the origin. Evaluation of this integral requires knowledge of complex variable theory. In this text, we pursue simpler alternatives, which include long division and partial fraction expansion.

17.5.1 Inverse z-Transform of Finite Sequences

For finite-length sequences, $X(z)$ has a polynomial form that immediately reveals the required sequence $x[n]$. The ROC can also be discerned from the polynomial form of $X(z)$.

EXAMPLE 17.7 (Inverse Transform of Sequences)

(a) Let $X(z) = 3z^{-1} + 5z^{-3} + 2z^{-4}$. This transform corresponds to a causal sequence. We recognize $x[n]$ as a sum of shifted impulses given by

$$x[n] = 3\delta[n-1] + 5\delta[n-3] + 2\delta[n-4]$$

This sequence can also be written as $x[n] = \{\overset{\Downarrow}{0}, 3, 0, 5, 2\}$.

(b) Let $X(z) = 2z^2 - 5z + 5z^{-1} - 2z^{-2}$. This transform corresponds to a noncausal sequence. Its inverse transform is written, by inspection, as

$$x[n] = \{2, -5, \overset{\Downarrow}{0}, 5, -2\}$$

Comment: Since $X(z) = -X(1/z)$, $x[n]$ should possess odd symmetry. It does.

17.5.2 Inverse z-Transform by Long Division

A second method requires $X(z)$ as a rational function (a ratio of polynomials) along with its ROC. For a right-sided signal (whose ROC is $|z| > |\alpha|$), we arrange the numerator and denominator in *ascending powers of* z, and use long division to obtain a power series in decreasing powers of z, whose inverse transform corresponds to the right-sided sequence.

For a left-sided signal (whose ROC is $|z| < |\alpha|$), we arrange the numerator and denominator in *descending powers of* z, and use long division to obtain a power series in increasing powers of z, whose inverse transform corresponds to the left-sided sequence.

This approach, however, becomes cumbersome if more than just the first few terms of $x[n]$ are required. It is not often that the first few terms of the resulting sequence allow its general nature or form to be discerned. If we regard the rational z-transform as a transfer function $H(z) = P(z)/Q(z)$, the method of long division is simply equivalent to finding the first few terms of its impulse response recursively from its difference equation.

REVIEW PANEL 17.15

Finding Inverse Transforms of $X(z) = N(z)/D(z)$ by Long Division

Right-sided: Put $N(z)$, $D(z)$ in *ascending powers of* z. Obtain a power series in decreasing powers of z.
Left-sided: Put $N(z)$, $D(z)$ in *descending powers of* z. Obtain a power series in increasing powers of z.

EXAMPLE 17.8 (Inverse Transforms by Long Division)

(a) We find the right-sided inverse of $H(z) = \dfrac{z-4}{1-z+z^2}$.

We arrange the polynomials in descending powers of z and use long division to get

$$\begin{array}{r} z^{-1} - 3z^{-2} - 4z^{-3} \cdots \\ z^2 - z + 1 \,\overline{)\, z - 4} \\ \underline{z - 1 + z^{-1}} \\ -3 - z^{-1} \\ \underline{-3 + 3z^{-1} - 3z^{-2}} \\ -4z^{-1} + 3z^{-2} \\ \underline{-4z^{-1} + 4z^{-2} - 4z^{-3}} \\ -z^{-2} + 4z^{-3} \\ \cdots \end{array}$$

This leads to $H(z) = z^{-1} - 3z^{-2} - 4z^{-3}$. The sequence $h[n]$ can be written as

$$h[n] = \delta(n-1) - 3\delta[n-2] - 4\delta[n-3] \quad \text{or} \quad h[n] = \{\overset{\Downarrow}{0}, 1, -3, -4, \ldots\}$$

PGR2026-00033
Krisp Technologies EX1024 Page 620

(b) We could also have found the inverse by setting up the the difference equation corresponding to $H(z) = Y(z)/X(z)$, to give

$$y[n] - y[n-1] + y[n-2] = x[n-1] - 4x[n-2]$$

With $x[n] = \delta[n]$, its impulse response $h[n]$ is

$$h[n] - h[n-1] + h[n-2] = \delta[n] - 4\delta[n-2]$$

With $h[-1] = h[-2] = 0$, (a relaxed system), we recursively obtain the first few values of $h[n]$ as

$$\begin{aligned}
n = 0: \quad h[0] &= h[-1] - h[-2] + \delta[-1] - 4\delta[-2] = 0 - 0 + 0 - 0 = 0 \\
n = 1: \quad h[1] &= h[0] - h[-1] + \delta[0] - 4\delta[-1] = 0 - 0 + 1 - 0 = 1 \\
n = 2: \quad h[2] &= h[1] - h[0] + \delta[1] - 4\delta[0] = 1 - 0 + 0 - 4 = -3 \\
n = 3: \quad h[3] &= h[2] - h[1] + \delta[2] - 4\delta[1] = -3 - 1 + 0 + 0 = -4
\end{aligned}$$

These are identical to the values obtained using long division in part (a).

(c) We find the left-sided inverse of $H(z) = \dfrac{z-4}{1-z+z^2}$.

We arrange the polynomials in descending powers of z and use long division to obtain

$$\begin{array}{r}
\multicolumn{1}{r}{-4 - 3z + z^2 \cdots} \\
1 - z + z^2 \;\overline{)\,-4 + z} \\
\multicolumn{1}{r}{-4 + 4z - 4z^2} \\
\cline{2-2}
\multicolumn{1}{r}{-3z + 4z^2} \\
\multicolumn{1}{r}{-3z + 3z^2 - 3z^3} \\
\cline{2-2}
\multicolumn{1}{r}{z^2 + 3z^3} \\
\multicolumn{1}{r}{z^2 - z^3 + z^4} \\
\cline{2-2}
\multicolumn{1}{r}{4z^3 - z^4} \\
\multicolumn{1}{r}{\cdots}
\end{array}$$

Thus, $H(z) = -4 - 3z + z^2$. The sequence $h[n]$ can then be written as

$$h[n] = -4\delta[n] - 3\delta[n+1] + \delta[n+2] \quad \text{or} \quad h[n] = \{\ldots, 1, -3, \overset{\Downarrow}{-4}\}$$

(d) We could also have found the inverse by setting up the the difference equation in the form

$$h[n-2] = h[n-1] + \delta[n-1] - 4\delta[n-2]$$

With $h[1] = h[2] = 0$, we can generate $h[0], h[-1], h[-2], \ldots$, recursively, to obtain the same result as in part (c).

PGR2026-00033
Krisp Technologies EX1024 Page 621

17.5.3 Inverse z-Transform from Partial Fractions

A much more useful method for inversion of the z-transform relies on its *partial fraction expansion* into terms whose inverse transform can be identified using a table of transform pairs. This is analogous to finding inverse Laplace transforms, but with one major difference. Since the z-transform of standard sequences in Table 17.1 involves the factor z in the numerator, it is more convenient to perform the partial fraction expansion for $Y(z) = X(z)/z$ rather than for $X(z)$. We then multiply through by z to obtain terms describing $X(z)$ in a form ready for inversion. This also implies that, for partial fraction expansion, it is $Y(z) = X(z)/z$ and not $X(z)$ that must be a proper rational function. The form of the expansion depends on the nature of the poles (denominator roots) of $Y(z)$. The constants in the partial fraction expansion are often called **residues**.

Distinct Linear Factors

If $Y(z)$ contains only distinct poles, we express it in the form:

$$Y(z) = \frac{P(z)}{(z+p_1)(z+p_2)\cdots(z+p_N)} = \frac{K_1}{z+p_1} + \frac{K_2}{z+p_2} + \cdots + \frac{K_N}{z+p_N} \tag{17.20}$$

To find the mth coefficient K_m, we multiply both sides by $(z+p_m)$ to get

$$(z+p_m)Y(z) = K_1\frac{(z+p_m)}{(z+p_1)} + \cdots + K_m + \cdots + K_N\frac{(z+p_m)}{(z+p_N)} \tag{17.21}$$

With both sides evaluated at $z = -p_m$, we obtain K_m as

$$K_m = (z+p_m)Y(z)\Big|_{z=-p_m} \tag{17.22}$$

In general, $Y(z)$ will contain terms with real constants and terms with complex conjugate residues, and may be written as

$$Y(z) = \frac{K_1}{z+p_1} + \frac{K_2}{z+p_2} + \cdots + \frac{A_1}{z+r_1} + \frac{A_1^*}{z+r_1^*} + \frac{A_2}{z+r_2} + \frac{A_2^*}{z+r_2^*} + \cdots \tag{17.23}$$

For a real root, the residue (coefficient) will also be real. For each pair of complex conjugate roots, the residues will also be complex conjugates, and we thus need compute only one of these.

Repeated Factors

If the denominator of $Y(z)$ contains the repeated term $(z+r)^k$, the partial fraction expansion corresponding to the repeated terms has the form

$$Y(z) = (\text{other terms}) + \frac{A_0}{(z+r)^k} + \frac{A_1}{(z+r)^{k-1}} + \cdots + \frac{A_{k-1}}{z+r} \tag{17.24}$$

Observe that the constants A_j ascend in index j from 0 to $k-1$, whereas the denominators $(z+r)^n$ descend in power n from k to 1. Their evaluation requires $(z+r)^kY(z)$ and its derivatives. We successively find

$$\begin{aligned} A_0 &= (z+r)^kY(z)\Big|_{z=-r} & A_2 &= \frac{1}{2!}\frac{d^2}{dz^2}[(z+r)^kY(z)]\Big|_{z=-r} \\ A_1 &= \frac{d}{dz}[(z+r)^kY(z)]\Big|_{z=-r} & A_n &= \frac{1}{n!}\frac{d^n}{dz^n}[(z+r)^kY(z)]\Big|_{z=-r} \end{aligned} \tag{17.25}$$

Even though this process allows us to find the coefficients independently of each other, the algebra in finding the derivatives can become tedious if the multiplicity k of the roots exceeds 2 or 3. Table 17.3 lists some transform pairs useful for inversion by partial fractions.

PGR2026-00033
Krisp Technologies EX1024 Page 622

Table 17.3 Inverse z-Transform of Partial Fraction Expansion (PFE) Terms

Entry	**PFE Term** $X(z)$	**Causal Signal** $x[n],\ n \geq 0$
Note: For anti-causal sequences, we get the signal $-x[n]u[-n-1]$ where $x[n]$ is as listed.		
1	$\dfrac{z}{z-\alpha}$	α^n
2	$\dfrac{z}{(z-\alpha)^2}$	$n\alpha^{(n-1)}$
3	$\dfrac{z}{(z-\alpha)^{N+1}} \quad (N>1)$	$\dfrac{n(n-1)\cdots(n-N+1)}{N!}\alpha^{(n-N)}$
4	$\dfrac{z(C+jD)}{z-\alpha e^{j\Omega}} + \dfrac{z(C-jD)}{z-\alpha e^{-j\Omega}}$	$2\alpha^n[C\cos(n\Omega) - D\sin(n\Omega)]$
5	$\dfrac{zK\angle\phi}{z-\alpha e^{j\Omega}} + \dfrac{zK\angle-\phi}{z-\alpha e^{-j\Omega}}$	$2K\alpha^n\cos(n\Omega+\phi)$
6	$\dfrac{z(C+jD)}{(z-\alpha e^{j\Omega})^2} + \dfrac{z(C-jD)}{(z-\alpha e^{-j\Omega})^2}$	$2n\alpha^{n-1}\Big(C\cos[(n-1)\Omega] - D\sin[(n-1)\Omega]\Big)$
7	$\dfrac{zK\angle\phi}{(z-\alpha e^{j\Omega})^2} + \dfrac{zK\angle-\phi}{(z-\alpha e^{-j\Omega})^2}$	$2Kn\alpha^{n-1}\cos[(n-1)\Omega+\phi]$
8	$\dfrac{zK\angle\phi}{(z-\alpha e^{j\Omega})^{N+1}} + \dfrac{zK\angle-\phi}{(z-\alpha e^{-j\Omega})^{N+1}}$	$2K\dfrac{n(n-1)\cdots(n-N+1)}{N!}\alpha^{(n-N)}\cos[(n-N)\Omega+\phi]$

REVIEW PANEL 17.16

Partial Fraction Expansion of $Y(z) = X(z)/z$ Depends on Its Poles (Denominator Roots)

Distinct roots: $Y(z) = \displaystyle\prod_{m=1}^{N}\frac{P(z)}{z+p_m} = \sum_{m=1}^{N}\frac{K_m}{z+p_m}$, where $K_m = (z+p_m)Y(z)\big|_{z=-p_m}$

Repeated: $\dfrac{1}{(z+r)^k}\displaystyle\prod_{m=1}^{N}\frac{P(z)}{z+p_m} = \sum_{m=1}^{N}\frac{K_m}{z+p_m} + \sum_{n=0}^{k-1}\frac{A_n}{(z+r)^{k-n}}$, where $A_n = \dfrac{1}{n!}\dfrac{d^n}{dz^n}[(z+r)^kY(z)]\Big|_{z=-r}$

EXAMPLE 17.9 (Inverse Transform of Right-Sided Signals)

(a) (Non-Repeated Roots) We find the inverse of $X(z) = \dfrac{1}{(z-0.25)(z-0.5)}$.

We first form $Y(z) = \dfrac{X(z)}{z}$, and expand $Y(z)$ into partial fractions, to obtain

$$Y(z) = \frac{X(z)}{z} = \frac{1}{z(z-0.25)(z-0.5)} = \frac{8}{z} - \frac{16}{z-0.24} + \frac{8}{z-0.5}$$

Multiplying through by z, we get

$$X(z) = 8 - \frac{16z}{z-0.25} + \frac{8z}{z-0.5} \qquad x[n] = 8\delta[n] - 16(0.25)^n u[n] + 8(0.5)^n u[n]$$

PGR2026-00033
Krisp Technologies EX1024 Page 623

Its first few samples, $x[0] = 0,\ x[1] = 0,\ x[2] = 1$, and $x[3] = 0.75$ can be checked by long division.

Comment: An alternate approach (not recommended) is to expand $X(z)$ itself as

$$X(z) = \frac{-4}{z - 0.25} + \frac{4}{z - 0.5} \qquad x[n] = -4(0.25)^{n-1}u[n-1] + 4(0.5)^{n-1}u[n-1]$$

Its inverse requires the shifting property. This form is functionally equivalent to the previous case. For example, we find that $x[0] = 0,\ x[1] = 0,\ x[2] = 1$, and $x[3] = 0.25$, as before.

(b) (Repeated Roots) We find the inverse of $X(z) = \dfrac{z}{(z-1)^2(z-2)}$.

We obtain $Y(z) = \dfrac{X(z)}{z}$, and set up its partial fraction expansion as

$$Y(z) = \frac{X(z)}{z} = \frac{1}{(z-1)^2(z-2)} = \frac{A}{z-2} + \frac{K_0}{(z-1)^2} + \frac{K_1}{z-1}$$

The constants in the partial fraction expansion are

$$A = \left.\frac{1}{(z-1)^2}\right|_{z=2} = 1 \quad K_0 = \left.\frac{1}{z-2}\right|_{z=1} = -1 \quad K_1 = \left.\frac{d}{dz}\left[\frac{1}{z-2}\right]\right|_{z=1} = -1$$

Substituting into $Y(z)$ and multiplying through by z, we get

$$X(z) = \frac{z}{z-2} - \frac{z}{(z-1)^2} - \frac{z}{z-1} \qquad x[n] = (2)^n u[n] - nu[n] - u[n] = (2^n - n - 1)u[n]$$

The first few values $x[0] = 0,\ x[1] = 0,\ x[2] = 1,\ x[3] = 4$, and $x[4] = 11$ can be easily checked by long division.

(c) (Complex Roots) We find the inverse of $X(z) = \dfrac{z^2 - 3z}{(z-2)(z^2 - 2z + 2)}$.

We set up the partial fraction expansion for $Y(z) = \dfrac{X(z)}{z}$ as

$$Y(z) = \frac{X(z)}{z} = \frac{z-3}{(z-2)(z^2-2z+2)} = \frac{A}{z-2} + \frac{K}{z-1-j} + \frac{K^*}{z-1+j}$$

We evaluate the constants A and K, to give

$$A = \left.\frac{z-3}{z^2-2z+2}\right|_{z=2} = -0.5 \qquad K = \left.\frac{z-3}{(z-2)(z-1+j)}\right|_{z=1+j} = 0.25 - j0.75$$

Substituting into the expression for $Y(z)$, and multiplying through by z, we get

$$X(z) = \frac{-0.5}{z-2} + \frac{(0.25 - j0.75)z}{z-1-j} + \frac{(0.25 + j0.75)z}{z-1+j}$$

PGR2026-00033
Krisp Technologies EX1024 Page 624

We note that $(z-1-j) = (z-\sqrt{2}e^{j\pi/4})$. From Table 17.3 (entry 4), the constant associated with this factor is $K = C + jD = 0.25 - j0.75$. Thus, $C = 0.25$, $D = -0.75$, and

$$x[n] = -0.5(2)^n u[n] + 2(\sqrt{2})^n[0.25\cos(\tfrac{n\pi}{4}) + 0.75\sin(\tfrac{n\pi}{4})]u[n]$$

Alternatively, with $0.25 - j0.25 = 0.7906\angle -71.56^\circ$, Table 17.3 (entry 5) also gives

$$x[n] = -0.5(2)^n u[n] + 2(0.7906)(\sqrt{2})^n \cos(\tfrac{n\pi}{4} - 71.56^\circ)$$

The two forms, of course, are identical. Which one we pick is a matter of personal preference.

(d) (Inverse Transform of Quadratic Forms) We find the inverse of $X(z) = \dfrac{z}{z^2+4}$.

One way is partial fractions. Here, we start with $B\alpha^n \sin(n\Omega)u[n] \Leftrightarrow \dfrac{Bz\alpha\sin\Omega}{z^2 - 2z\alpha\cos\Omega + \alpha^2}$.

Comparing denominators, $\alpha^2 = 4$, and thus $\alpha = \pm 2$. Let us pick $\alpha = 2$.

We also have $2\alpha\cos\Omega = 0$. Thus, $\Omega = \pi/2$.

Finally, comparing numerators, $Bz\alpha\sin\Omega = z$ or $B = 0.5$. Thus,

$$x[n] = 0.5(2)^n \sin(n\pi/2)u[n]$$

Comment: Had we picked $\alpha = -2$, we would still have come out with the same answer.

(e) (Inverse Transform of Quadratic Forms) Let $X(z) = \dfrac{z^2+z}{z^2-2z+4}$. We start with

$$A\alpha^n\cos(n\Omega)u[n] \Longleftrightarrow \frac{A(z^2 - z\alpha\cos\Omega)}{z^2 - 2z\alpha\cos\Omega + \alpha^2} \qquad B\alpha^n\cos(n\Omega)u[n] \Longleftrightarrow \frac{Bz\alpha\sin\Omega}{z^2 - 2z\alpha\cos\Omega + \alpha^2}$$

Comparing denominators, we find $\alpha = 2$. Then, $2\alpha\cos\Omega = 2$, and thus $\cos\Omega = 0.5$ or $\Omega = \pi/3$.

Now, $A(z^2 - z\alpha\cos\Omega) = A(z^2 - z)$ and $Bz\alpha\sin\Omega = Bz\sqrt{3}$. We express the numerator of $X(z)$ as a sum of these forms to get $z^2 + z = (z^2 - z) + 2z = (z^2 - z) + (2/\sqrt{3})(z\sqrt{3})$ (with $A = 1$ and $B = 2/\sqrt{3} = 1.1547$). Thus,

$$x[n] = (2)^n\cos(\tfrac{n\pi}{3})u[n] + 1.1547(2)^n\sin(\tfrac{n\pi}{3})u[n]$$

17.5.4 The ROC and Inversion

We have so far been assuming right-sided sequences when no ROC is given. Only when the ROC is specified do we obtain a unique sequence from $X(z)$. Sometimes, the ROC may be specified indirectly by requiring the system to be stable, for example. Since the ROC of a stable system includes the unit circle, this gives us a clue to the type of inverse we require.

PGR2026-00033
Krisp Technologies EX1024 Page 625

EXAMPLE 17.10 (Inversion and the ROC)

(a) Find all possible inverse transforms of $X(z) = \dfrac{z}{(z-0.25)(z-0.5)}$.

The partial fraction expansion of $Y(z) = \dfrac{X(z)}{z}$ leads to $X(z) = \dfrac{-4z}{z-0.25} + \dfrac{4z}{z-0.5}$.

1. If the ROC is $|z| > 0.5$, $x[n]$ is causal and stable, and we obtain

$$x[n] = -4(0.25)^n u[n] + 4(0.5)^n u[n]$$

2. If the ROC is $|z| < 0.25$, $x[n]$ is anti-causal and unstable, and we obtain

$$x[n] = 4(0.25)^n u[-n-1] - 4(0.5)^n u[-n-1]$$

3. If the ROC is $0.25 < |z| < 0.5$, $x[n]$ is two-sided and unstable. This ROC is valid only if $\frac{-4z}{z-0.25}$ describes a causal sequence (ROC $|z| > 0.25$), and $\frac{4z}{z-0.5}$ describes an anti-causal sequence (ROC $|z| < 0.5$). With this in mind, we obtain

$$x[n] = \underbrace{-4(0.25)^n u[n]}_{\text{ROC: } |z|>0.25} \quad \underbrace{-4(0.5)^n u[-n-1]}_{\text{ROC: } |z|<0.5}$$

(b) Find the unique inverse transforms of the following, assuming each system is stable:

$$H_1(z) = \frac{z}{(z-0.4)(z+0.6)} \qquad H_2(z) = \frac{2.5z}{(z-0.5)(z+2)} \qquad H_3(z) = \frac{z}{(z-2)(z+3)}$$

Partial fraction expansion leads to

$$H_1(z) = \frac{z}{z-0.4} - \frac{z}{z+0.6} \qquad H_2(z) = \frac{z}{z-0.5} - \frac{z}{z+2} \qquad H_3(z) = \frac{z}{z-2} - \frac{z}{z+3}$$

To find the appropriate inverse, the key is to recognize that the ROC must include the unit circle. Looking at the pole locations, we see that

$H_1(z)$ is stable if its ROC is $|z| > 0.6$.

Its inverse is causal, with $h_1[n] = (0.4)^n u[n] - (-0.6)^n u[n]$.

$H_2(z)$ is stable if its ROC is $0.5 < |z| < 2$.

Its inverse is two-sided, with $h_2[n] = (0.5)^n u[n] + (-2)^n u[-n-1]$.

$H_3(z)$ is stable if its ROC is $|z| < 2$.

Its inverse is anti-causal, with $h_3[n] = -(2)^n u[-n-1] + (-3)^n u[-n-1]$.

PGR2026-00033
Krisp Technologies EX1024 Page 626

17.6 The One-Sided z-Transform

The **one-sided z-transform** is particularly useful in the analysis of causal LTI systems. It is defined by

$$X(z) = \sum_{k=0}^{\infty} x[k]z^{-k} \quad \text{(one-sided } z\text{-transform)} \tag{17.26}$$

The lower limit of zero in the summation implies that the one-sided z-transform of an arbitrary signal $x[n]$ and its causal version $x[n]u[n]$ are identical. Most of the properties of the two-sided z-transform also apply to the one-sided version.

REVIEW PANEL 17.17

The Scaling, Times-n, and Convolution Properties of the One-Sided z-Transform

$$\alpha^n x[n] \Longleftrightarrow X(z/\alpha) \qquad nx[n] \Longleftrightarrow -z\frac{dX(z)}{dz} \qquad x[n] \star h[n] \Longleftrightarrow X(z)H(z)$$

However, the shifting property of the two-sided z-transform must be modified for use with right-sided (or causal) signals that are nonzero for $n < 0$. We also develop new properties, such as the initial value theorem and the final value theorem, that are unique to the one-sided z-transform. These properties are summarized in Table 17.4.

Table 17.4 Properties Unique to the One-Sided z-Transform

Property	**Signal**	**One-Sided z-Transform**
Right shift	$x[n-1]$	$z^{-1}X(z) + x[-1]$
	$x[n-2]$	$z^{-2}X(z) + z^{-1}x[-1] + x[-2]$
	$x[n-N]$	$z^{-N}X(z) + z^{-(N-1)}x[-1] + z^{-(N-2)}x[-2] + \cdots + x[-N]$
Left shift	$x[n+1]$	$zX(z) - zx[0]$
	$x[n+2]$	$z^2X(z) - z^2x[0] - zx[1]$
	$x[n+N]$	$z^NX(z) - z^Nx[0] - z^{N-1}x[1] - \cdots - zx[N-1]$
Switched periodic	$x_p[n]u[n]$	$\dfrac{X_1(z)}{1-z^{-N}}$ ($x_1[n]$ is the first period of $x_p[n]$)
Initial value theorem:	$x[0] = \lim_{z\to\infty} X(z)$	
Final value theorem:	$\lim_{n\to\infty} x[n] = \lim_{z\to 1}(z-1)X(z)$	

EXAMPLE 17.11 (Properties of the One-Sided z-Transform)

(a) Find the z-transform of $x[n] = n(4)^{0.5n}u[n]$.

We rewrite this as $x[n] = n(2)^n u[n]$ to get $X(z) = \dfrac{2z}{(z-2)^2}$.

PGR2026-00033
Krisp Technologies EX1024 Page 627

(b) Find the z-transform of $x[n] = (2)^{n+1}u[n-1]$.

We rewrite this as $x[n] = (2)^2(2)^{n-1}u[n-1]$ to get $X(z) = \dfrac{z^{-1}(4z)}{z-2} = \dfrac{4}{z-2}$.

(c) Let $x[n] \Longleftrightarrow \dfrac{4z}{(z+0.5)^2} = X(z)$, with ROC: $|z| > 0.5$. Find the z-transform of the signals $h[n] = nx[n]$ and $y[n] = x[n] \star x[n]$.

By the times-n property, we have $H(z) = -zX'(z)$, which gives

$$H(z) = -z\left[\frac{-8z}{(z+0.5)^3} + \frac{4}{(z+0.5)^2}\right] = \frac{4z^2 - 2z}{(z+0.5)^3}$$

By the convolution property, $Y(z) = X^2(z) = \dfrac{16z^2}{(z+0.5)^4}$.

(d) Let $(4)^n u[n] \Longrightarrow X(z)$. Find the signal corresponding to $F(z) = X^2(z)$ and $G(z) = X(2z)$.

By the convolution property, $f[n] = (4)^n u[n] \star (4)^n u[n] = (n+1)(4)^n u[n]$.

By the scaling property, $G(z) = X(2z) = X(z/0.5)$ corresponds to the signal $g[n] = (0.5)^n x[n]$. Thus, we have $g[n] = (2)^n u[n]$.

17.6.1 The Right-Shift Property of the One-Sided z-Transform

The one-sided z-transform of a sequence $x[n]$ and its causal version $x[n]u[n]$ are identical. A right shift of $x[n]$ brings samples for $n < 0$ into the range $n \geq 0$, as illustrated in Figure 17.5, and leads to the z-transforms

$$x[n-1] \Longleftrightarrow z^{-1}X(z) + x[-1] \qquad x[n-2] \Longleftrightarrow z^{-2}X(z) + z^{-1}x[-1] + x[-2] \tag{17.27}$$

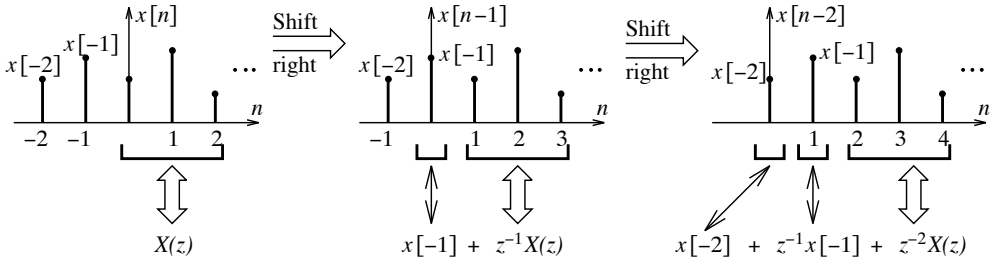


Figure 17.5 Illustrating the right-shift property of the one-sided z-transform

These results generalize to

$$x[n-N] \Longleftrightarrow z^{-N}X(z) + z^{-(N-1)}x[-1] + z^{-(N-2)}x[-2] + \cdots + x[-N] \tag{17.28}$$

For causal signals (for which $x[n] = 0,\ n < 0$), this result reduces to $x[n-N] \Leftrightarrow z^{-N}X(z)$.

PGR2026-00033
Krisp Technologies EX1024 Page 628

REVIEW PANEL 17.18

The Right-Shift Property of the One-Sided z-Transform

$$x[n-1] \Longleftrightarrow z^{-1}X(z) + x[-1] \qquad x[n-2] \Longleftrightarrow z^{-2}X(z) + z^{-1}x[-1] + x[-2]$$

17.6.2 The Left-Shift Property of the One-Sided z-Transform

A left shift of $x[n]u[n]$ moves samples for $n \geq 0$ into the range $n < 0$, and these samples no longer contribute to the z-transform of the causal portion, as illustrated in Figure 17.6.

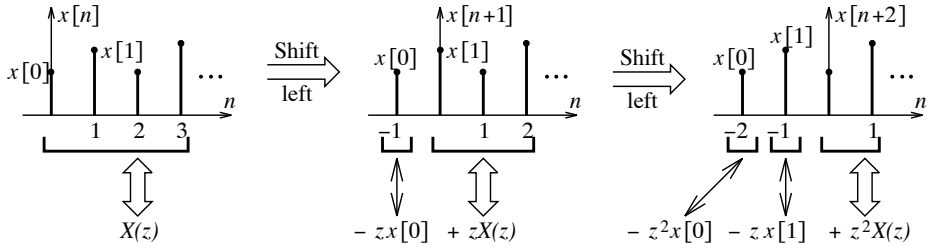


Figure 17.6 Illustrating the left-shift property of the one-sided z-transform

This leads to the z-transforms

$$x[n+1] \Longleftrightarrow zX(z) - zx[0] \qquad x[n+2] \Longleftrightarrow z^2X(z) - z^2x[0] - zx[1] \tag{17.29}$$

By successively shifting $x[n]u[n]$ to the left, we obtain the general relation

$$X[n+N] \Longleftrightarrow z^NX(z) - z^Nx[0] - z^{N-1}x[1] - \cdots - zx[N-1] \tag{17.30}$$

The right-shift and left-shift properties of the one-sided z-transform form the basis for finding the response of causal LTI systems with nonzero initial conditions.

REVIEW PANEL 17.19

The Left-Shift Property of the One-Sided z-Transform

$$x[n+1] \Longleftrightarrow zX(z) - zx[0] \qquad x[n+2] \Longleftrightarrow z^2X(z) - z^2x[0] - zx[1]$$

EXAMPLE 17.12 (Using the Shift Properties)

(a) Using the right-shift property and superposition, we obtain the z-transform of the first difference of $x[n]$ as

$$y[n] = x[n] - x[n-1] \Longleftrightarrow X(z) - z^{-1}X(z) = (1 - z^{-1})X(z)$$

(b) Consider the signal $x[n] = \alpha^n$. Its one-sided z-transform is identical to that of $\alpha^n u[n]$ and equals $X(z) = z/(z-\alpha)$. If $y[n] = x[n-1]$, the right-shift property, with $N = 1$, yields

$$Y(z) = z^{-1}X(z) + x[-1] = \frac{1}{z-\alpha} + \alpha^{-1}$$

The additional term α^{-1} arises because $x[n]$ is not causal.

PGR2026-00033
Krisp Technologies EX1024 Page 629

(c) (The Left-Shift Property)

Consider the shifted step $u[n+1]$. Its one-sided z-transform should be identical to that of $u[n]$ since $u[n]$ and $u[n+1]$ are identical for $n \geq 0$.

With $u[n] \Leftrightarrow z/(z-1)$ and $u[0] = 1$, the left-shift property gives

$$u[n+1] \Longleftrightarrow zU(z) - zu[0] = \frac{z^2}{z-1} - z = \frac{z}{z-1}$$

(d) With $y[n] = \alpha^n u[n] \Leftrightarrow z/(z-\alpha)$ and $y[0] = 1$, the left-shift property gives

$$y[n+1] = \alpha^{n+1} u[n+1] \Longleftrightarrow z\left[\frac{z}{z-\alpha}\right] - z = \frac{\alpha z}{z-\alpha}$$

17.6.3 The Initial Value Theorem and Final Value Theorem

The initial value theorem and final value theorem apply only to the one-sided z-transform and the proper part $X(z)$ of a rational z-transform.

REVIEW PANEL 17.20

The Initial Value Theorem and Final Value Theorem for the One-Sided z-Transform

Initial value: $x[0] = \lim_{z\to\infty} X(z)$ **Final value:** $x[\infty] = \lim_{z\to 1}(z-1)X(z)$

Final value theorem is meaningful only if: Poles of $(z-1)X(z)$ are inside unit circle.

With $X(z)$ described by $x[0] + x[1]z^{-1} + x[2]z^{-2} + \cdots$, it should be obvious that only $x[0]$ survives as $z \to \infty$ and the initial value equals $x[0] = \lim_{z\to\infty} X(z)$.

To find the final value, we evaluate $(z-1)X(z)$ at $z = 1$. It yields meaningful results only when the poles of $(z-1)X(z)$ have magnitudes smaller than unity (lie within the unit circle in the z-plane). As a result:

1. $x[\infty] = 0$ if all poles of $X(z)$ lie within the unit circle (since $x[n]$ will then contain only exponentially damped terms).
2. $x[\infty]$ is constant if there is a single pole at $z = 1$ (since $x[n]$ will then include a step).
3. $x[\infty]$ is indeterminate if there are complex conjugate poles on the unit circle (since $x[n]$ will then include sinusoids). The final value theorem can yield absurd results if used in this case.

EXAMPLE 17.13 (Initial and Final Value Theorems)

Let $X(z) = \dfrac{z(z-2)}{(z-1)(z-0.5)}$. We then find

$$\text{The initial value: } x[0] = \lim_{z\to\infty} X(z) = \lim_{z\to\infty} \frac{1-2z^{-1}}{(1-z^{-1})(1-0.5z^{-1})} = 1$$

$$\text{The final value: } \lim_{n\to\infty} x[n] = \lim_{z\to 1}(z-1)X(z) = \lim_{z\to 1} \frac{z(z-2)}{z-0.5} = -2$$

PGR2026-00033
Krisp Technologies EX1024 Page 630

17.6.4 The z-Transform of Switched Periodic Signals

Consider a causal signal $x[n] = x_p[n]u[n]$ where $x_p[n]$ is periodic with period N. If $x_1[n]$ describes the first period of $x[n]$ and has the z-transform $X_1(z)$, then the z-transform of $x[n]$ can be found as the superposition of the z-transform of the shifted versions of $x_1[n]$:

$$X(z) = X_1(z) + z^{-N}X_1(z) + z^{-2N}X_1(z) + \cdots = X_1(z)[1 + z^{-N} + z^{-2N} + \cdots] \tag{17.31}$$

Expressing the geometric series in closed form, we obtain

$$X(z) = \frac{1}{1 - z^{-N}}X_1(z) = \frac{z^N}{z^N - 1}X_1(z) \tag{17.32}$$

REVIEW PANEL 17.21

The z-Transform of a Switched Periodic Signal $x_p[n]u[n]$ with Period N

$X(z) = \dfrac{X_1(z)}{1 - z^{-N}}$ ($X_1(z)$ is z-transform of first period $x_1[n]$)

EXAMPLE 17.14 (z-Transform of Switched Periodic Signals)

(a) Find the z-transform of a periodic signal whose first period is $x_1[n] = \{\overset{\Downarrow}{0},\ 1,\ -2\}$.

The period of $x[n]$ is $N = 3$. We then find the z-transform of $x[n]$ as

$$X(z) = \frac{X_1(z)}{1 - z^{-N}} = \frac{z^{-1} - 2z^{-2}}{1 - z^{-3}}$$

(b) Find the z-transform of $x[n] = \sin(0.5n\pi)u[n]$.

The digital frequency of $x[n]$ is $F = \frac{1}{4}$. So $N = 4$. The first period of $x[n]$ is $x_1[n] = \{\overset{\Downarrow}{0},\ 1,\ 0,\ -1\}$. The z-transform of $x[n]$ is thus

$$X(z) = \frac{X_1(z)}{1 - z^{-N}} = \frac{z^{-1} - z^{-3}}{1 - z^{-4}} = \frac{z^{-1}}{1 + z^{-2}}$$

(c) Find the causal signal corresponding to $X(z) = \dfrac{2 + z^{-1}}{1 - z^{-3}}$.

Comparing with the z-transform for a switched periodic signal, we recognize $N = 3$ and $X_1(z) = 2 - z^{-1}$. Thus, the first period of $x[n]$ is $\{\overset{\Downarrow}{2},\ 1,\ 0\}$.

(d) Find the causal signal corresponding to $X(z) = \dfrac{z^{-1}}{1 + z^{-3}}$.

We first rewrite $X(z)$ as

$$X(z) = \frac{(z^{-1})(1 - z^{-3})}{(1 + z^{-3})(1 - z^{-3})} = \frac{z^{-1} - z^{-4}}{1 - z^{-6}}$$

Comparing with the z-transform for a switched periodic signal, we recognize the period as $N = 6$, and $X_1(z) = z^{-1} - z^{-4}$. Thus, the first period of $x[n]$ is $\{\overset{\Downarrow}{0},\ 1,\ 0,\ 0,\ -1,\ 0\}$.

PGR2026-00033
Krisp Technologies EX1024 Page 631

17.7 The z-Transform and System Analysis

The one-sided z-transform serves as a useful tool for analyzing LTI systems described by difference equations or transfer functions. The key is of course that the solution methods are much simpler in the transformed domain because convolution transforms to a multiplication. Naturally, the time-domain response requires an inverse transformation, a penalty exacted by all methods in the transformed domain.

17.7.1 Systems Described by Difference Equations

For a system described by a difference equation, the solution is based on transformation of the difference equation using the shift property and incorporating the effect of initial conditions (if present), and subsequent inverse transformation using partial fractions to obtain the time-domain response. The response may be separated into its zero-state component (due only to the input), and zero-input component (due to the initial conditions), in the z-domain itself.

REVIEW PANEL 17.22

Solving Difference Equations Using the z-Transform

Relaxed system: Transform the difference equation, then find $Y(z)$ and its inverse.
Not relaxed: Transform using the shift property and initial conditions. Find $Y(z)$ and its inverse.

EXAMPLE 17.15 (Solution of Difference Equations)

(a) Solve the difference equation $y[n] - 0.5y[n-1] = 2(0.25)^n u[n]$ with $y[-1] = -2$.

Transformation using the right-shift property yields

$$Y(z) - 0.5\{z^{-1}Y(z) + y[-1]\} = \frac{2z}{z-0.25} \qquad Y(z) = \frac{z(z+0.25)}{(z-0.25)(z-0.5)}$$

We use partial fractions to get $\dfrac{Y(z)}{z} = \dfrac{z+0.25}{(z-0.25)(z-0.5)} = \dfrac{-2}{z-0.25} + \dfrac{3}{z-0.5}$.

Multiplying through by z and taking inverse transforms, we obtain

$$Y(z) = \frac{-2z}{z-0.25} + \frac{3z}{z-0.5} \qquad y[n] = [-2(0.25)^n + 3(0.5)^n]u[n]$$

(b) Let $y[n+1] - 0.5y[n] = 2(0.25)^{n+1}u[n+1]$ with $y[-1] = -2$.

We transform the difference equation using the *left-shift property*. The solution will require $y[0]$. By recursion, with $n = -1$, we obtain $y[0] - 0.5y[-1] = 2$ or $y[0] = 2 + 0.5y[-1] = 2 - 1 = 1$.

Let $x[n] = (0.25)^n u[n]$. Then, by the left-shift property $x[n+1] \Leftrightarrow zX(z) - zx[0]$ (with $x[0] = 1$),

$$(0.25)^{n+1}u[n+1] \Longleftrightarrow z\left[\frac{z}{z-0.25}\right] - z = \frac{0.25z}{z-0.25}$$

We now transform the difference equation using the left-shift property:

$$zY(z) - zy[0] - 0.5Y(z) = \frac{0.5z}{z-0.25} \qquad Y(z) = \frac{z(z+0.25)}{(z-0.25)(z-0.5)}$$

This is identical to the result of part (a), and thus $y[n] = -2(0.25)^n + 3(0.5)^n$, as before.

Comment: By time invariance, this represents the same system as in part (a).

PGR2026-00033
Krisp Technologies EX1024 Page 632

(c) (Zero-Input and Zero-State Response) Let $y[n] - 0.5y[n-1] = 2(0.25)^n u[n]$, with $y[-1] = -2$. Upon transformation using the right-shift property, we obtain

$$Y(z) - 0.5\{z^{-1}Y(z) + y[-1]\} = \frac{2z}{z-0.25} \qquad (1 - 0.5z^{-1})Y(z) = \frac{2z}{z-0.25} - 1$$

1. **Zero-state response:** For the zero-state response, we assume zero initial conditions to obtain

$$(1 - 0.5z^{-1})Y_{\text{zs}}(z) = \frac{2z}{z-0.25} \qquad Y_{\text{zs}}(z) = \frac{2z^2}{(z-0.25)(z-0.5)}$$

Upon partial fraction expansion, we obtain

$$\frac{Y_{\text{zs}}(z)}{z} = \frac{2z}{(z-0.25)(z-0.5)} = -\frac{2}{z-0.25} + \frac{4}{z-0.5}$$

Multiplying through by z and inverse transforming the result, we get

$$Y_{\text{zs}}(z) = \frac{-2z}{z-0.25} + \frac{4z}{z-0.5} \qquad y_{\text{zs}}[n] = -2(0.25)^n u[n] + 4(0.25)^n u[n]$$

2. **Zero-input response:** For the zero-input response, we assume zero input (the right-hand side) and use the right-shift property to get

$$Y_{\text{zi}}(z) - 0.5\{z^{-1}Y_{\text{zi}}(z) + y[-1]\} = 0 \qquad Y_{\text{zi}}(z) = \frac{-z}{z-0.5}$$

This is easily inverted to give $y_{\text{zi}}[n] = -(\frac{1}{2})^n u[n]$.

3. **Total response:** We find the total response as

$$y[n] = y_{\text{zs}}[n] + y_{\text{zi}}[n] = -2(0.25)^n u[n] + 3(0.5)^n u[n]$$

17.7.2 Systems Described by the Transfer Function

The response $Y(z)$ of a relaxed LTI system equals the product $X(z)H(z)$ of the transformed input and the transfer function. It is often much easier to work with the transfer function description of a linear system. If we let $H(z) = N(z)/D(z)$, the zero-state response $Y(z)$ of a relaxed system to an input $X(z)$ may be expressed as $Y(z) = X(z)H(z) = X(z)N(z)/D(z)$. If the system is not relaxed, the initial conditions result in an additional contribution, the zero-input response $Y_{\text{zi}}(z)$, which may be written as $Y_{\text{zi}}(z) = N_{\text{zi}}(z)/D(z)$. To evaluate $Y_{\text{zi}}(z)$, we first set up the system difference equation, and then use the shift property to transform it in the presence of initial conditions.

REVIEW PANEL 17.23

System Analysis Using the Transfer Function

Zero-state response: Evaluate $Y(z) = X(z)H(z)$ and take inverse transform.
Zero-input response: Find difference equation. Transform this using the shift property and initial conditions. Find the response in the z-domain and take inverse transform.

PGR2026-00033
Krisp Technologies EX1024 Page 633

EXAMPLE 17.16 (System Response from the Transfer Function)

(a) (A Relaxed System) Let $H(z) = \dfrac{3z}{z-0.4}$. To find the zero-state response of this system to $x[n] = (0.4)^n u[n]$, we first transform the input to $X(z) = \dfrac{z}{z-0.4}$. Then,

$$Y(z) = H(z)X(z) = \frac{3z^2}{(z-0.4)^2} \qquad y[n] = 3(n+1)(0.4)^n u[n]$$

(b) (Step Response) Let $H(z) = \dfrac{4z}{z-0.5}$.

To find its step response, we let $x[n] = u[n]$. Then $X(z) = \dfrac{z}{z-1}$, and the output equals

$$Y(z) = H(z)X(z) = \left(\frac{4z}{z-0.5}\right)\left(\frac{z}{z-1}\right) = \frac{4z^2}{(z-1)(z-0.5)}$$

Using partial fraction expansion of $Y(z)/z$, we obtain

$$\frac{Y(z)}{z} = \frac{4z}{(z-1)(z-0.5)} = \frac{8}{z-0.5} - \frac{4}{z-1}$$

Thus,

$$Y(z) = \frac{8z}{z-0.5} - \frac{4z}{z-1} \qquad y[n] = 8u[n] - (0.5)^n u[n]$$

The first term in $y[n]$ is the *steady-state* response, which can be found much more easily as described shortly in Section 17.8.1.

(c) (A Second-Order System) Let $H(z) = \dfrac{z^2}{z^2 - \frac{1}{6}z - \frac{1}{6}}$. Let the input be $x[n] = 4u[n]$ and the initial conditions be $y[-1] = 0,\ y[-2] = 12$.

1. **Zero-state and zero-input response:** The zero-state response is found directly from $H(z)$ as

$$Y_{\text{zs}}(z) = X(z)H(z) = \frac{4z^3}{(z^2 - \frac{1}{6}z - \frac{1}{6})(z-1)} = \frac{4z^3}{(z-\frac{1}{2})(z+\frac{1}{3})(z-1)}$$

Partial fractions of $Y_{\text{zs}}(z)/z$ and inverse transformation give

$$Y_{\text{zs}}(z) = \frac{-2.4z}{z-\frac{1}{2}} + \frac{0.4z}{z+\frac{1}{3}} + \frac{6z}{z-1} \qquad y_{\text{zs}}[n] = -2.4(\tfrac{1}{2})^n u[n] + 0.4(-\tfrac{1}{3})^n u[n] + 6u[n]$$

To find the zero-input response, we first set up the difference equation. We start with $H(z) = \dfrac{Y(z)}{X(z)} = \dfrac{z^2}{z^2 - \frac{1}{6}z - \frac{1}{6}}$, or $(z^2 - \frac{1}{6}z - \frac{1}{6})Y(z) = z^2 X(z)$. This gives

$$[1 - \tfrac{1}{6}z^{-1} - \tfrac{1}{6}z^{-2}]Y(z) = X(z) \qquad y[n] - \tfrac{1}{6}y[n-1] - \tfrac{1}{6}y[n-2] = x[n]$$

PGR2026-00033
Krisp Technologies EX1024 Page 634

We now assume $x[n] = 0$ (zero input) and transform this equation, using the *right-shift* property, to obtain the zero-input response from

$$Y_{\text{zi}}(z) - \tfrac{1}{6}\{z^{-1}Y_{\text{zi}}(z) + y[-1]\} - \tfrac{1}{6}\{z^{-2}Y_{\text{zi}}(z) + z^{-1}y[-1] + y[-2]\} = 0$$

With $y[-1] = 0$ and $y[-2] = 12$, this simplifies to

$$Y_{\text{zi}}(z) = \frac{2z^2}{z^2 - \frac{1}{6}z - \frac{1}{6}} = \frac{2z^2}{(z - \frac{1}{2})(z + \frac{1}{3})}$$

Partial fraction expansion of $Y_{\text{zi}}(z)/z$ and inverse transformation lead to

$$Y_{\text{zi}}(z) = \frac{0.75z}{z - \frac{1}{2}} + \frac{0.8z}{z + \frac{1}{3}} \qquad y_{\text{zi}}[n] = 1.2(\tfrac{1}{2})^n u[n] + 0.8(-\tfrac{1}{3})^n u[n]$$

Finally, we find the total response as

$$y[n] = y_{\text{zs}}[n] + y_{\text{zi}}[n] = -1.2(\tfrac{1}{2})^n u[n] + 1.2(-\tfrac{1}{3})^n u[n] + 6u[n]$$

2. **Natural and forced response:** By inspection, the natural and forced components of $y[n]$ are

$$y_N[n] = -1.2(\tfrac{1}{2})^n u[n] + 1.2(-\tfrac{1}{3})^n u[n] \qquad y_F[n] = 6u[n]$$

Comment: Alternatively, we could transform the system difference equation to obtain

$$Y(z) - \tfrac{1}{6}\{z^{-1}Y(z) + y[-1]\} - \tfrac{1}{6}\{z^{-2}Y(z) + z^{-1}y[-1] + y[-2]\} = \frac{4z}{z - 1}$$

This simplifies to

$$Y(z) = \frac{z^2(6z - 2)}{(z - 1)(z^2 - \frac{1}{6}z - \frac{1}{6})} = \frac{-1.2z}{z - \frac{1}{2}} + \frac{1.2z}{z + \frac{1}{3}} + \frac{6z}{z - 1}$$

The steady-state response corresponds to terms of the form $z/(z - 1)$ (step functions). For this example, $Y_F(z) = \frac{6z}{z-1}$ and $y_F[n] = 6u[n]$.
Since the poles of $(z - 1)Y(z)$ lie within the unit circle, $y_F[n]$ can also be found by the final value theorem:

$$y_F[n] = \lim_{z \to 1}(z - 1)Y(z) = \lim_{z \to 1} \frac{z^2(6z - 2)}{z^2 - \frac{1}{6}z - \frac{1}{6}} = 6$$

17.8 Frequency Response

If the transfer function $H(z)$ of a linear system is evaluated for values of $z = e^{j2\pi F} = e^{j\Omega}$ (on the unit circle), we obtain the **steady-state transfer function** or **frequency response**, $H_p(F)$, of the system. This is just the DTFT of the impulse response $h[n]$. The quantity $\Omega = 2\pi F$ describes the angular orientation. Figure 17.7 shows how the values of z, Ω, and F are related on the unit circle.

In analogy with the Laplace transform, $H_p(F)$ describes the response of a system to discrete-time harmonics. Since $H_p(F)$ is complex, it can be written as

$$H_p(F) = |H_p(F)|\angle\phi(F) = |H_p(F)|e^{j\phi(F)} \tag{17.33}$$

PGR2026-00033
Krisp Technologies EX1024 Page 635

Recall that $H_p(F)$ is periodic with period 1 and, for real $h[n]$, shows conjugate symmetry over the principal period $(-0.5,\ 0.5)$. This implies that $H_p(-F) = H_p^*(F)$, $|H_p(F)|$ is even symmetric and $\angle\phi(F)$ is odd symmetric. Since $H_p(F)$ and $H(z)$ are related by $z = e^{j2\pi F} = e^{j\Omega}$, the frequency response $H_p(F)$ at $F = 0$ (or $\Omega = 0$) is equivalent to evaluating $H(z)$ at $z = 1$, and the response $H_p(F)$ at $F = 0.5$ (or $\Omega = \pi$) is equivalent to evaluating $H(z)$ at $z = -1$.

$$H_p(F)\big|_{F=0} = H_p(\Omega)\big|_{\Omega=0} = H(z)\big|_{z=1} \qquad H_p(F)\big|_{F=0.5} = H_p(\Omega)\big|_{\Omega=\pi} = H(z)\big|_{z=-1} \tag{17.34}$$

The frequency response is useful primarily for stable systems for which the natural response does, indeed, decay with time.

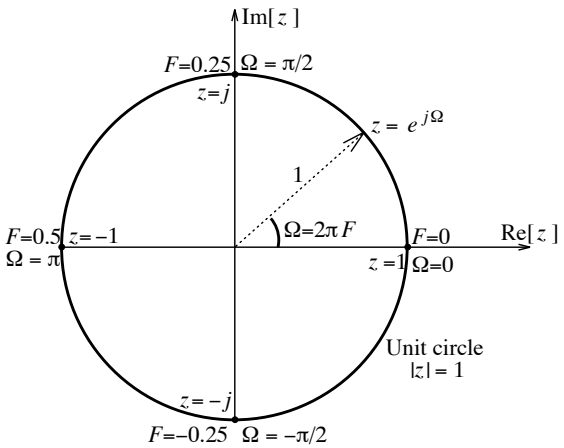


Figure 17.7 Relating the variables z, F, and Ω through the unit circle

REVIEW PANEL 17.24

The Frequency Response Is the z-Transform Evaluated on the Unit Circle

It corresponds to the DTFT: $H_p(F) = H(z)\big|_{z=e^{j2\pi F}}$ or $H_p(\Omega) = H(z)\big|_{z=e^{j\Omega}}$

EXAMPLE 17.17 (Frequency Response of a Recursive Filter)

Let $y[n] = \alpha y[n-1] + x[n],\ 0 < \alpha < 1$. To find its frequency response, we transform the difference equation and find $H(z)$ as

$$H(z) = \frac{Y(z)}{X(z)} = \frac{1}{1 - \alpha z^{-1}}$$

Next, we let $z = e^{j2\pi F}$ and evaluate $H_p(F)$ as

$$H_p(F) = \frac{1}{1 - \alpha e^{-j2\pi F}} = \frac{1}{1 - \alpha\cos(2\pi F) + j\alpha\sin(2\pi F)}$$

The magnitude and phase of $H_p(F)$ then equal

$$|H_p(F)| = \left[\frac{1}{1 - 2\alpha\cos(2\pi F) + \alpha^2}\right]^{1/2} \qquad \phi(F) = \tan^{-1}\left[\frac{1 - \alpha\cos(2\pi F)}{\alpha\sin(2\pi F)}\right]$$

PGR2026-00033
Krisp Technologies EX1024 Page 636

Typical plots of the magnitude and phase are shown in Figure E17.17 over the principal period $(-0.5,\ 0.5)$. Note the conjugate symmetry (even symmetry of $|H_p(F)|$ and odd symmetry of $\phi(F)$).

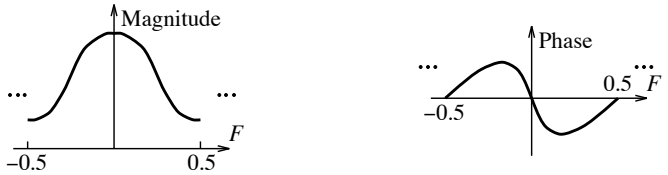


Figure E17.17 Magnitude and phase of $H_p(F)$ for Example 17.17

17.8.1 The Steady-State Response to Harmonic Inputs

The everlasting discrete-time harmonic $e^{j2\pi nF_0}$ is an eigensignal of discrete-time LTI systems. The response is also a harmonic $H_0e^{j(2\pi nF_0+\phi_0)}$ at the input frequency F_0 whose magnitude H_0 and phase ϕ_0 correspond to the system function $H_p(F_0) = H_0\angle\phi_0$ evaluated at the input frequency $F = F_0$. Euler's relation and superposition generalizes this result to sinusoidal inputs. To find the steady-state response $y_{ss}[n]$ of an LTI system to the sinusoid $x[n] = A\cos(2\pi nF_0 + \theta)$, we evaluate $H_p(F_0) = H_0\angle\phi_0$ (at the input frequency F_0) and obtain $y_{ss}[n] = AH_0\cos(2\pi nF_0 + \theta + \phi_0)$. If the input consists of harmonics at different frequencies, we use superposition of the individual *time-domain* responses.

We can also find the steady-state component from the total response (using z-transforms), but this defeats the whole purpose if all we are after is the steady-state response.

REVIEW PANEL 17.25

Finding the Steady-State Response of an LTI System to a Sinusoidal Input

Input: $x[n] = A\cos(2\pi nF_0 + \theta)$ **Transfer function:** Evaluate $H(z)$ at $z = e^{j2\pi F_0}$ as $H_0\angle\phi_0$.
Steady-state output: $y[n] = AH_0\cos(2\pi nF_0 + \theta + \phi_0)$

EXAMPLE 17.18 (Steady-State Response of a Recursive Filter)

(a) Let $y[n] = 0.5y[n-1] + x[n]$. We find its steady-state response to $x[n] = 10\cos(0.5n\pi + 60^\circ)$.

First, we find $H_p(F)$ from $H(z)$ as

$$H(z) = \frac{1}{1-0.5z^{-1}} \qquad H_p(F) = H(z)\Big|_{z=e^{j\Omega}} = \frac{1}{1-0.5e^{-j2\pi F}}$$

Next, we first evaluate $H_p(F)$ at the input frequency $F = 0.25$

$$H_p(0.25) = \frac{1}{1-0.5e^{-j2\pi/4}} = \frac{1}{1+j0.5} = 0.89443\angle -26.6^\circ$$

Then, $y_{ss}[n] = 10(0.89443)\cos(0.5n\pi + 60^\circ - 26.6^\circ) = 8.9443\cos(0.5n\pi + 33.4^\circ)$.

(b) Let $H(z) = \dfrac{2z-1}{z^2+0.5z+0.5}$. We find its steady-state response to $x[n] = 6u[n]$.

Since the input is a constant for $n \geq 0$, the input frequency $F = 0$ corresponds to $z = e^{j2\pi F} = 1$.

We thus require $H(z)\big|_{z=1} = 0.5$. Then, $y_{ss}[n] = (6)(0.5) = 3$.

Of course, in either part, the total response would require more laborious methods, but the steady-state component would turn out to be just what we found here.

PGR2026-00033
Krisp Technologies EX1024 Page 637

17.9 Connections

The z-transform may be viewed as a generalization of the DTFT to complex frequencies. The DTFT describes a signal as a sum of weighted harmonics, or complex exponentials. However, it cannot handle exponentially growing signals, and it cannot handle initial conditions in system analysis. The z-transform overcomes these shortcomings by using exponentially weighted harmonics in its definition.

$$X_p(\Omega) = \sum_{k=-\infty}^{\infty} x[n]e^{-jn\Omega} \qquad X(z) = \sum_{k=-\infty}^{\infty} x[n]\left(re^{j\Omega}\right)^{-k} = \sum_{k=-\infty}^{\infty} x[n]z^n$$

A lower limit of zero defines the *one-sided Laplace transform* and applies to causal signals, permitting the analysis of systems with arbitrary initial conditions to arbitrary inputs.

Many properties of the z-transform and the DTFT show correspondence. However, the shifting property for the one-sided z-transform must be modified to account for initial values. Due to the presence of the convergence factor r^{-k}, the z-transform $X(z)$ of $x[n]$ no longer displays the conjugate symmetry present in its DTFT. Since $z = re^{j2\pi F} = re^{j\Omega}$ is complex, the z-transform may be plotted as a surface in the z-plane. The DTFT (with $z = e^{j2\pi F} = e^{j\Omega}$) is just the cross section of this surface along the unit circle ($|z| = 1$).

17.9.1 Relating the z-Transform and DTFT

For absolutely summable signals, the DTFT is simply the one-sided z-transform with $z = e^{j2\pi F}$. The DTFT of signals that are not absolutely summable almost invariably contains impulses. However, for such signals, the z-transform equals just the non-impulsive portion of the DTFT, with $z = e^{j2\pi F}$. In other words, for absolutely summable signals we can always find their z-transform from the DTFT, but we cannot always find their DTFT from the z-transform.

REVIEW PANEL 17.26

Relating the z-Transform and the DTFT

From $X(z)$ to DTFT: If $x[n]$ is *absolutely summable*, simply replace z by $e^{j2\pi F}$ (or $e^{j\Omega}$).
From DTFT to $X(z)$: Delete impulsive terms in DTFT and replace $e^{j2\pi F}$ (or $e^{j\Omega}$) by z.

EXAMPLE 17.19 (The z-Transform and DTFT)

(a) The signal $x[n] = \alpha^n u[n],\ |\alpha| < 1$, is absolutely summable. Its DTFT equals $X_p(F) = \dfrac{1}{1-\alpha e^{-j2\pi F}}$. We can find the z-transform of $x[n]$ from its DTFT as $X(z) = \dfrac{1}{1-\alpha z^{-1}} = \dfrac{z}{z-\alpha}$. We can also find the DTFT from the z-transform by reversing the steps.

(b) The signal $x[n] = u[n]$ is not absolutely summable. Its DTFT is $X_p(F) = \dfrac{1}{1-e^{-j2\pi F}} + 0.5\delta(F)$. We can find the z-transform of $u[n]$ as the impulsive part in the DTFT, with $e^{j2\pi F} = z$, to give $X(z) = \dfrac{1}{1-z^{-1}} = \dfrac{z}{z-1}$. However, we cannot recover the DTFT from its z-transform in this case.

PGR2026-00033
Krisp Technologies EX1024 Page 638

CHAPTER 17 PROBLEMS

DRILL AND REINFORCEMENT

17.1 (The z-Transform of Sequences) Use the defining relation to find the z-transform and its region of convergence for the following:

(a) $x[n] = \{1, 2, \overset{\Downarrow}{3}, 2, 1\}$ **(b)** $x[n] = \{-1, 2, \overset{\Downarrow}{0}, -2, 1\}$
(c) $x[n] = \{\overset{\Downarrow}{1}, 1, 1, 1\}$ **(d)** $x[n] = \{1, 1, -1, \overset{\Downarrow}{-1}\}$

17.2 (z-Transforms) Find the z-transforms and specify the ROC for the following:

(a) $x[n] = (2)^{n+2}u[n]$ **(b)** $x[n] = n(2)^{0.2n}u[n]$
(c) $x[n] = (2)^{n+2}u[n-1]$ **(d)** $x[n] = n(2)^{n+2}u[n-1]$
(e) $x[n] = (n+1)(2)^n u[n]$ **(f)** $x[n] = (n-1)(2)^{n+2}u[n]$

17.3 (z-Transforms) Find the z-transforms and specify the ROC for the following:

(a) $x[n] = \cos(\frac{n\pi}{4} - \frac{\pi}{4})u[n]$ **(b)** $x[n] = (0.5)^n \cos(\frac{n\pi}{4})u[n]$
(c) $x[n] = (0.5)^n \cos(\frac{n\pi}{4} - \frac{\pi}{4})u[n]$ **(d)** $x[n] = (\frac{1}{3})^n(u[n] - u[n-4])$
(e) $x[n] = n(0.5)^n \cos(\frac{n\pi}{4})u[n]$ **(f)** $x[n] = [(0.5)^n - (-0.5)^n]nu[n]$

17.4 (Two-Sided z-Transform) Find the z-transform $X(z)$ and its ROC for the following:

(a) $x[n] = u[-n-1]$ **(b)** $x[n] = (0.5)^{-n}u[-n-1]$
(c) $x[n] = (0.5)^{|n|}$ **(d)** $x[n] = u[-n-1] + (\frac{1}{3})^n u[n]$
(e) $x[n] = (0.5)^{-n}u[-n-1] + (\frac{1}{3})^n u[n]$ **(f)** $x[n] = (0.5)^{|n|} + (-0.5)^{|n|}$

17.5 (Properties) Let $x[n] = nu[n]$. Find $X(z)$, using the following:

(a) The defining relation for the z-transform
(b) The times-n property
(c) The convolution result $u[n] \star u[n] = (n+1)u[n+1]$ and the shifting property
(d) The convolution result $u[n] \star u[n] = (n+1)u[n]$ and superposition

17.6 (Properties) The z-transform of $x[n]$ is $X(z) = \dfrac{4z}{(z+0.5)^2}$, $|z| > 0.5$. Find the z-transform of the following using properties and specify the region of convergence.

(a) $y[n] = x[n-2]$ **(b)** $d[n] = (2)^n x[n]$ **(c)** $f[n] = nx[n]$
(d) $g[n] = (2)^n nx[n]$ **(e)** $h[n] = n^2 x[n]$ **(f)** $p[n] = [n-2]x[n]$
(g) $q[n] = x[-n]$ **(h)** $r[n] = x[n] - x[n-1]$ **(i)** $s[n] = x[n] \star x[n]$

17.7 (Properties) The z-transform of $x[n] = (2)^n u[n]$ is $X(z)$. Use properties to find the time signal corresponding to the following:

(a) $Y(z) = X(2z)$ **(b)** $F(z) = X(1/z)$ **(c)** $G(z) = zX'(z)$
(d) $H(z) = \dfrac{zX(z)}{z-1}$ **(e)** $D(z) = \dfrac{zX(2z)}{z-1}$ **(f)** $P(z) = z^{-1}X(z)$
(g) $Q(z) = z^{-2}X(2z)$ **(h)** $R(z) = X^2(z)$ **(i)** $S(z) = X(-z)$

PGR2026-00033
Krisp Technologies EX1024 Page 639

17.8 (Inverse Transforms of Polynomials) Find the inverse z-transform $x[n]$ for the following:

(a) $X(z) = 2 - z^{-1} + 3z^{-3}$ **(b)** $X(z) = (2 + z^{-1})^3$
(c) $X(z) = (z - z^{-1})^2$ **(d)** $X(z) = (z - z^{-1})^2(2 + z)$

17.9 (Inverse Transforms by Long Division) Assume that $x[n]$ represents a right-sided signal. Determine the ROC of the following z-transforms and compute the values of $x[n]$ for $n = 0, 1, 2, 3$.

(a) $X(z) = \dfrac{(z+1)^2}{z^2+1}$ **(b)** $X(z) = \dfrac{z+1}{z^2+2}$
(c) $X(z) = \dfrac{1}{z^2-0.25}$ **(d)** $X(z) = \dfrac{1-z^{-2}}{2+z^{-1}}$

17.10 (Inverse Transforms by Partial Fractions) Assume that $x[n]$ represents a right-sided signal. Determine the ROC of the following z-transforms and compute $x[n]$, using partial fractions.

(a) $X(z) = \dfrac{z}{(z+1)(z+2)}$ **(b)** $X(z) = \dfrac{16}{(z-2)(z+2)}$
(c) $X(z) = \dfrac{3z^2}{(z^2-1.5z+0.5)(z-0.25)}$ **(d)** $X(z) = \dfrac{3z^3}{(z^2-1.5z+0.5)(z-0.25)}$
(e) $X(z) = \dfrac{3z^4}{(z^2-1.5z+0.5)(z-0.25)}$ **(f)** $X(z) = \dfrac{4z}{(z+1)^2(z+3)}$

17.11 (Inverse Transforms by Partial Fractions) Assume that $x[n]$ represents a right-sided signal. Determine the ROC of the following z-transforms and compute $x[n]$, using partial fractions.

(a) $X(z) = \dfrac{z}{(z^2+z+0.25)(z+1)}$ **(b)** $X(z) = \dfrac{z}{(z^2+z+0.25)(z+0.5)}$
(c) $X(z) = \dfrac{1}{(z^2+z+0.25)(z+1)}$ **(d)** $X(z) = \dfrac{z}{(z^2+z+0.5)(z+1)}$
(e) $X(z) = \dfrac{z^3}{(z^2-z+0.5)(z-1)}$ **(f)** $X(z) = \dfrac{z^2}{(z^2+z+0.5)(z+1)}$
(g) $X(z) = \dfrac{2z}{(z^2-0.25)^2}$ **(h)** $X(z) = \dfrac{2}{(z^2-0.25)^2}$
(i) $X(z) = \dfrac{z}{(z^2+0.25)^2}$ **(j)** $X(z) = \dfrac{z^2}{(z^2+0.25)^2}$

17.12 (Inverse Transforms by Long Division) Assume that $x[n]$ represents a left-sided signal. Determine the ROC of the following z-transforms and compute the values of $x[n]$ for $n = -1, -2, -3$.

(a) $X(z) = \dfrac{z^2+4z}{z^2-z+2}$ **(b)** $X(z) = \dfrac{z}{(z+1)^2}$
(c) $X(z) = \dfrac{z^2}{z^3+z-1}$ **(d)** $X(z) = \dfrac{z^3+1}{z^2+1}$

17.13 (The ROC and Inverse Transforms) Let $X(z) = \dfrac{z^2+5z}{z^2-2z-3}$. Which of the following describe a valid ROC for $X(z)$? For each valid ROC, find $x[n]$, using partial fractions.

(a) $|z| < 1$ **(b)** $|z| > 3$ **(c)** $1 < |z| < 3$ **(d)** $|z| < 1$ and $z > 3$

PGR2026-00033
Krisp Technologies EX1024 Page 640

17.14 (Convolution) Compute $y[n]$, using the z-transform. Then verify your results by finding $y[n]$, using time-domain convolution.

(a) $y[n] = \{-\overset{\Downarrow}{1}, 2, 0, 3\} \star \{2, 0, \overset{\Downarrow}{3}\}$ **(b)** $y[n] = \{-1, 2, \overset{\Downarrow}{0}, -2, 1\} \star \{-1, 2, \overset{\Downarrow}{0}, -2, 1\}$

(c) $y[n] = (2)^n u[n] \star (2)^n u[n]$ **(d)** $(2)^n u[n] \star (3)^n u[n]$

17.15 (Initial Value and Final Value Theorems) Assume that $X(z)$ corresponds to a right-sided signal $x[n]$ and find the initial and final values of $x[n]$.

(a) $X(z) = \dfrac{2}{z^2 + \frac{1}{6}z - \frac{1}{6}}$ **(b)** $X(z) = \dfrac{2z^2}{z^2 + z + 0.25}$

(c) $X(z) = \dfrac{2z}{z^2 + z - 1}$ **(d)** $X(z) = \dfrac{2z^2 + 0.25}{(z-1)(z+0.25)}$

(e) $X(z) = \dfrac{z + 0.25}{z^2 + 0.25}$ **(f)** $X(z) = \dfrac{2z + 1}{z^2 - 0.5z - 0.5}$

17.16 (System Representation) Find the transfer function and difference equation for the following causal systems. Investigate their stability, using each system representation.

(a) $h[n] = (2)^n u[n]$ **(b)** $h[n] = [1 - (\frac{1}{3})^n]u[n]$

(c) $h[n] = n(\frac{1}{3})^n u[n]$ **(d)** $h[n] = 0.5\delta[n]$

(e) $h[n] = \delta[n] - (-\frac{1}{3})^n u[n]$ **(f)** $h[n] = [(2)^n - (3)^n]u[n]$

17.17 (System Representation) Find the transfer function and impulse response of the following causal systems. Investigate the stability of each system.

(a) $y[n] + 3y[n-1] + 2y[n-2] = 2x[n] + 3x[n-1]$

(b) $y[n] + 4y[n-1] + 4y[n-2] = 2x[n] + 3x[n-1]$

(c) $y[n] = 0.2x[n]$

(d) $y[n] = x[n] + x[n-1] + x[n-2]$

17.18 (System Representation) Set up the system difference equations and impulse response of the following causal systems. Investigate the stability of each system.

(a) $H(z) = \dfrac{3}{z+2}$ **(b)** $H(z) = \dfrac{1 + 2z + z^2}{(1+z^2)(4+z^2)}$

(c) $H(z) = \dfrac{2}{1+z} - \dfrac{1}{2+z}$ **(d)** $H(z) = \dfrac{2z}{1+z} - \dfrac{1}{2+z}$

17.19 (Zero-State Response) Find the zero-state response of the following systems, using the z-transform.

(a) $y[n] - 0.5y[n-1] = 2u[n]$ **(b)** $y[n] - 0.4y[n-1] = (0.5)^n$

(c) $y[n] - 0.4y[n-1] = (0.4)^n$ **(d)** $y[n] - 0.5y[n-1] = \cos(n\pi/2)$

17.20 (System Response) Consider the system $y[n] - 0.5y[n-1] = x[n]$. Find its zero-state response to the following inputs, using the z-transform.

(a) $x[n] = u[n]$ **(b)** $x[n] = (0.5)^n u[n]$ **(c)** $x[n] = \cos(n\pi/2)u[n]$

(d) $x[n] = (-1)^n u[n]$ **(e)** $x[n] = (j)^n u[n]$ **(f)** $x[n] = (\sqrt{j})^n u[n] + (\sqrt{j})^{-n} u[n]$

17.21 (Zero-State Response) Find the zero-state response of the following systems, using the z-transform.

(a) $y[n] - 1.1y[n-1] + 0.3y[n-2] = 2u[n]$ **(b)** $y[n] - 0.9y[n-1] + 0.2y[n-2] = (0.5)^n$

(c) $y[n] + 0.7y[n-1] + 0.1y[n-2] = (0.5)^n$ **(d)** $y[n] - 0.25y[n-2] = \cos(n\pi/2)$

PGR2026-00033
Krisp Technologies EX1024 Page 641

17.22 (System Response) Let $y[n] - 0.5y[n-1] = x[n]$, with $y[-1] = -1$. Find the response $y[n]$ of this system for the following inputs, using the z-transform.

(a) $x[n] = 2u[n]$ **(b)** $x[n] = (0.25)^n u[n]$ **(c)** $x[n] = n(0.25)^n u[n]$
(d) $x[n] = (0.5)^n u[n]$ **(e)** $x[n] = n(0.5)^n$ **(f)** $x[n] = (0.5)^n \cos(0.5n\pi)$

17.23 (System Response) Find the response $y[n]$ of the following systems, using the z-transform.

(a) $y[n] + 0.1y[n-1] - 0.3y[n-2] = 2u[n]$ $y[-1] = 0$ $y[-2] = 0$
(b) $y[n] - 0.9y[n-1] + 0.2y[n-2] = (0.5)^n$ $y[-1] = 1$ $y[-2] = -4$
(c) $y[n] + 0.7y[n-1] + 0.1y[n-2] = (0.5)^n$ $y[-1] = 0$ $y[-2] = 3$
(d) $y[n] - 0.25y[n-2] = (0.4)^n$ $y[-1] = 0$ $y[-2] = 3$
(e) $y[n] - 0.25y[n-2] = (0.5)^n$ $y[-1] = 0$ $y[-2] = 0$

17.24 (System Response) For each system, evaluate the response $y[n]$, using the z-transform.

(a) $y[n] - 0.4y[n-1] = x[n]$ $x[n] = (0.5)^n u[n]$ $y[-1] = 0$
(b) $y[n] - 0.4y[n-1] = 2x[n] + x[n-1]$ $x[n] = (0.5)^n u[n]$ $y[-1] = 0$
(c) $y[n] - 0.4y[n-1] = 2x[n] + x[n-1]$ $x[n] = (0.5)^n u[n]$ $y[-1] = 5$
(d) $y[n] + 0.5y[n-1] = x[n] - x[n-1]$ $x[n] = (0.5)^n u[n]$ $y[-1] = 2$
(e) $y[n] + 0.5y[n-1] = x[n] - x[n-1]$ $x[n] = (-0.5)^n u[n]$ $y[-1] = 0$

17.25 (System Response) Find the response $y[n]$ of the following systems, using the z-transform.

(a) $y[n] - 0.4y[n-1] = 2(0.5)^{n-1}u[n-1]$ $y[-1] = 2$
(b) $y[n] - 0.4y[n-1] = (0.4)^n u[n] + 2(0.5)^{n-1}u[n-1]$ $y[-1] = 2.5$
(c) $y[n] - 0.4y[n-1] = n(0.5)^n u[n] + 2(0.5)^{n-1}u[n-1]$ $y[-1] = 2.5$

17.26 (System Response) The transfer function of a system is $H(z) = \dfrac{2z(z-1)}{4 + 4z + z^2}$. Find its response $y[n]$ for the following inputs.

(a) $x[n] = \delta[n]$ **(b)** $x[n] = 2\delta[n] + \delta[n+1]$ **(c)** $x[n] = u[n]$
(d) $x[n] = (2)^n u[n]$ **(e)** $x[n] = nu[n]$ **(f)** $x[n] = \cos(\frac{n\pi}{2})u[n]$

17.27 (System Analysis) Find the impulse response $h[n]$ and the step response $s[n]$ of the causal digital filters described by

(a) $H(z) = \dfrac{4z}{z - 0.5}$ **(b)** $y[n] + 0.5y[n-1] = 6x[n]$

17.28 (System Analysis) Find the zero-state response, zero-input response, and total response for each of the following systems, using the z-transform.

(a) $y[n] - \frac{1}{4}y[n-1] = (\frac{1}{3})^n u[n]$ $y[-1] = 8$
(b) $y[n] + 1.5y[n-1] + 0.5y[n-2] = (-0.5)^n u[n]$ $y[-1] = 2$ $y[-2] = -4$
(c) $y[n] + y[n-1] + 0.25y[n-2] = 4(0.5)^n u[n]$ $y[-1] = 6$ $y[-2] = -12$
(d) $y[n] - y[n-1] + 0.5y[n-2] = (0.5)^n u[n]$ $y[-1] = -1$ $y[-2] = -2$

17.29 (Steady-State Response) The transfer function of a system is $H(z) = \dfrac{2z(z-1)}{z^2 + 0.25}$. Find its *steady-state* response for the following inputs.

(a) $x[n] = 4u[n]$ **(b)** $x[n] = 4\cos(\frac{n\pi}{2} + \frac{\pi}{4})u[n]$
(c) $x[n] = \cos(\frac{n\pi}{2}) + \sin(\frac{n\pi}{2})$ **(d)** $x[n] = 4\cos(\frac{n\pi}{4}) + 4\sin(\frac{n\pi}{2})$

PGR2026-00033
Krisp Technologies EX1024 Page 642

17.30 (Response of Digital Filters) Consider the averaging filter $y[n] = 0.5x[n] + x[n-1] + 0.5x[n-2]$.

(a) Find its impulse response $h[n]$, its transfer function $H(z)$, and its frequency response $H(F)$.
(b) Find its response $y[n]$ to the input $x[n] = \{\overset{\Downarrow}{2}, 4, 6, 8\}$.
(c) Find its response $y[n]$ to the input $x[n] = \cos(\frac{n\pi}{3})$.
(d) Find its response $y[n]$ to the input $x[n] = \cos(\frac{n\pi}{3}) + \sin(\frac{2n\pi}{3}) + \cos(\frac{n\pi}{2})$.

17.31 (Frequency Response) For each filter, sketch the magnitude spectrum and identify the filter type.
(a) $h[n] = \delta[n] - \delta[n-2]$ **(b)** $y[n] - 0.25y[n-1] = x[n] - x[n-1]$
(c) $H(z) = \dfrac{z-2}{z-0.5}$ **(d)** $y[n] - y[n-1] + 0.25y[n-2] = x[n] + x[n-1]$

17.32 (Transfer Function) The input to a digital filter is $x[n] = \{\overset{\Downarrow}{1},\ 0.5\}$, and the response is described by $y[n] = \delta[n+1] - 2\delta[n] - \delta[n-1]$.
(a) What is the filter transfer function $H(z)$?
(b) Does $H(z)$ describe an IIR filter or FIR filter?
(c) Is the filter stable? Is it causal?

17.33 (System Response) A system is described by $H(z) = \dfrac{z}{(z-0.5)(z+2)}$. Find the ROC and impulse response $h[n]$ of this system and state whether the system is stable or unstable if
(a) $h[n]$ is assumed to be causal.
(b) $h[n]$ is assumed to be anti-causal.
(c) $h[n]$ is assumed to be two-sided.

17.34 (System Response) Consider the 2-point averager with $y[n] = 0.5x[n] + 0.5x[n-1]$.
(a) Sketch its frequency response and identify the filter type.
(b) Find its response $y[n]$ to the input $x[n] = \cos(n\pi/2)$.
(c) Find its response $y[n]$ to the input $x[n] = \delta[n]$.
(d) Find its response $y[n]$ to the input $x[n] = 1$.
(e) Find its response $y[n]$ to the input $x[n] = 3 + 2\delta[n] - 4\cos(n\pi/2)$.

17.35 (System Response) Consider the 3-point averager with $h[n] = \frac{1}{3}\{1,\ \overset{\Downarrow}{1},\ 1\}$.
(a) Sketch its frequency response and identify the filter type.
(b) Find its response $y[n]$ to the input $x[n] = \cos(n\pi/3)$.
(c) Find its response $y[n]$ to the input $x[n] = \delta[n]$.
(d) Find its response $y[n]$ to the input $x[n] = (-1)^n$.
(e) Find its response $y[n]$ to the input $x[n] = 3 + 3\delta[n] - 6\cos(n\pi/3)$.

17.36 (System Response) Consider the tapered 3-point averager with $h[n] = \{\overset{\Downarrow}{0.5},\ 1,\ 0.5\}$.
(a) Sketch its frequency response and identify the filter type.
(b) Find its response $y[n]$ to the input $x[n] = \cos(n\pi/2)$.
(c) Find its response $y[n]$ to the input $x[n] = \delta[n-1]$.
(d) Find its response $y[n]$ to the input $x[n] = 1 + (-1)^n$.
(e) Find its response $y[n]$ to the input $x[n] = 3 + 2\delta[n] - 4\cos(2n\pi/2)$.

PGR2026-00033
Krisp Technologies EX1024 Page 643

17.37 (System Response) Consider the 2-point differencer with $h[n] = \delta[n] - \delta[n-1]$.

(a) Sketch its frequency response and identify the filter type.
(b) Find its response $y[n]$ to the input $x[n] = \cos(n\pi/2)$.
(c) Find its response $y[n]$ to the input $x[n] = u[n]$.
(d) Find its response $y[n]$ to the input $x[n] = (-1)^n$.
(e) Find its response $y[n]$ to the input $x[n] = 3 + 2u[n] - 4\cos(n\pi/2)$.

17.38 (Frequency Response) Consider the filter described by $y[n] + 0.5y[n-1] = 0.5x[n] + x[n-1]$.

(a) Sketch its frequency response and identify the filter type.
(b) Find its response $y[n]$ to the input $x[n] = \cos(n\pi/2)$.
(c) Find its response $y[n]$ to the input $x[n] = \delta[n]$.
(d) Find its response $y[n]$ to the input $x[n] = (-1)^n$.
(e) Find its response $y[n]$ to the input $x[n] = 4 + 2\delta[n] - 4\cos(n\pi/2)$.

REVIEW AND EXPLORATION

17.39 (The z-Transform of Sequences) Find the z-transform and its ROC for the following:

(a) $x[n] = u[n+2] - u[n-2]$
(b) $x[n] = (0.5)^n(u[n+2] - u[n-2])$
(c) $x[n] = (0.5)^{|n|}(u[n+2] - u[n-2])$

17.40 (z-Transforms) Find the z-transforms and specify the ROC for the following:

(a) $x[n] = (0.5)^{2n}u[n]$ **(b)** $x[n] = n(0.5)^{2n}u[n]$ **(c)** $(0.5)^{-n}u[n]$
(d) $(0.5)^n u[-n]$ **(e)** $(0.5)^n u[-n-1]$ **(f)** $(0.5)^{-n}u[-n-1]$

17.41 (The ROC) The transfer function of a system is $H(z)$. What can you say about the ROC of $H(z)$ for the following cases?

(a) $h[n]$ is a causal signal.
(b) The system is stable.
(c) The system is stable, and $h[n]$ is a causal signal.

17.42 (Poles, Zeros, and the ROC) The transfer function of a system is $H(z)$. What can you say about the poles and zeros of $H(z)$ for the following cases?

(a) The system is stable.
(b) The system is causal and stable.
(c) The system is an FIR filter with real coefficients.
(d) The system is a linear-phase FIR filter with real coefficients.
(e) The system is a causal, linear-phase FIR filter with real coefficients.

17.43 (z-Transforms and ROC) Consider the signal $x[n] = \alpha^n u[n] + \beta^n u[-n-1]$. Find its z-transform $X(z)$. Will $X(z)$ represent a valid transform for the following cases?

(a) $\alpha > \beta$ **(b)** $\alpha < \beta$ **(c)** $\alpha = \beta$

PGR2026-00033
Krisp Technologies EX1024 Page 644

17.44 (z-Transforms) Find the z-transforms (if they exist) and specify their ROC.

(a) $x[n] = (2)^{-n}u[n] + 2^n u[-n]$ **(b)** $x[n] = (0.25)^n u[n] + 3^n u[-n]$
(c) $x[n] = (0.5)^n u[n] + 2^n u[-n-1]$ **(d)** $x[n] = (2)^n u[n] + (0.5)^n u[-n-1]$
(e) $x[n] = \cos(0.5n\pi)u[n]$ **(f)** $x[n] = \cos(0.5n\pi + 0.25\pi)u[n]$
(g) $x[n] = e^{jn\pi}u[n]$ **(h)** $x[n] = e^{jn\pi/2}u[n]$
(i) $x[n] = e^{jn\pi/4}u[n]$ **(j)** $x[n] = (\sqrt{j})^n u[n] + (\sqrt{j})^{-n}u[n]$

17.45 (z-Transforms and ROC) The causal signal $x[n] = \alpha^n u[n]$ has the transform $X(z)$ whose ROC is $|z| > \alpha$. Find the ROC of the z-transform of the following:

(a) $y[n] = x[n-5]$
(b) $p[n] = x[n+5]$
(c) $g[n] = x[-n]$
(d) $h[n] = (-1)^n x[n]$
(e) $r[n] = \alpha^n x[n]$

17.46 (z-Transforms and ROC) The anti-causal signal $x[n] = -\alpha^n u[-n-1]$ has the transform $X(z)$ whose ROC is $|z| < \alpha$. Find the ROC of the z-transform of the following:

(a) $y[n] = x[n-5]$
(b) $p[n] = x[n+5]$
(c) $g[n] = x[-n]$
(d) $h[n] = (-1)^n x[n]$
(e) $r[n] = \alpha^n x[n]$

17.47 (z-Transforms) Find the z-transform $X(z)$ of $x[n] = \alpha^{|n|}$ and specify the region of convergence of $X(z)$. Consider the cases $|\alpha| < 1$ and $|\alpha| > 1$ separately.

17.48 (Properties) The z-transform of a signal $x[n]$ is $X(z) = \dfrac{4z}{(z+0.5)^2}$, $|z| > 0.5$. Find the z-transform and its ROC for the following.

(a) $y[n] = (-1)^n x[n]$ **(b)** $f[n] = x[2n]$
(c) $g[n] = (j)^n x[n]$ **(d)** $h[n] = x[n+1] + x[n-1]$

17.49 (Properties) The z-transform of the signal $x[n] = (2)^n u[n]$ is $X(z)$. Use properties to find the time signal corresponding to the following.

(a) $F(z) = X(-z)$ **(b)** $G(z) = X(1/z)$ **(c)** $H(z) = zX'(-z)$

17.50 (Properties) The z-transform of a signal $x[n]$ is $X(z) = \dfrac{z}{z-0.4}$.

(a) Without computing inverse transforms, find the z-transform $X_e(z)$ and the ROC for the even part of $x[n]$. Confirm your answer by computing $x[n]$ and $x_e[n]$ and finding its z-transform.

(b) Can you find $X_e(z)$ if $X(z) = \dfrac{z}{z-4}$? Explain.

17.51 (Properties) Find the z-transform of $x[n] = \text{rect}(n/2N) = u[n+N] - u[n-N-1]$. Use this result to evaluate the z-transform of $y[n] = \text{tri}(n/N)$.

PGR2026-00033
Krisp Technologies EX1024 Page 645

17.52 (Pole-Zero Patterns and Symmetry) Plot the pole-zero patterns for each $X(z)$. Which of these describe symmetric time sequences?

(a) $X(z) = \dfrac{z^2 + z - 1}{z}$ **(b)** $X(z) = \dfrac{z^4 + 2z^3 + 3z^2 + 2z + 1}{z^2}$

(c) $X(z) = \dfrac{z^4 - z^3 + z - 1}{z^2}$ **(d)** $X(z) = \dfrac{(z^2 - 1)(z^2 + 1)}{z^2}$

17.53 (Switched Periodic Signals) Find the z-transform of each switched periodic signal.

(a) $x[n] = \{\overset{\Downarrow}{2},\ 1,\ 3,\ 0, \ldots\}, \quad N = 4$ **(b)** $x[n] = \cos(n\pi/2)u[n]$

(c) $x[n] = \{\overset{\Downarrow}{0},\ 1,\ 1,\ 0,\ 0, \ldots\}, \quad N = 5$ **(d)** $x[n] = \cos(0.5n\pi + 0.25\pi)u[n]$

17.54 (Inverse Transforms) For each $X(z)$, find the signal $x[n]$ for each valid ROC.

(a) $X(z) = \dfrac{z}{(z + 0.4)(z - 0.6)}$ **(b)** $X(z) = \dfrac{3z^2}{z^2 - 1.5z + 0.5}$

17.55 (Poles and Zeros) Make a rough sketch of the pole and zero locations of the z-transform of each of the signals shown in Figure P17.55.

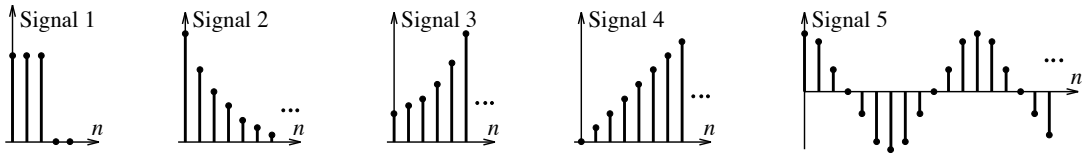


Figure P17.55 Figure for Problem 17.55

17.56 (Poles and Zeros) Find the transfer function corresponding to each pole-zero pattern shown in Figure P17.56 and identify the filter type.

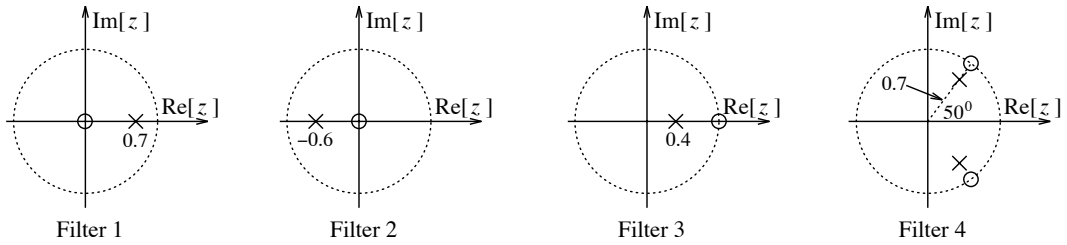


Figure P17.56 Figure for Problem 17.56

17.57 (Causality and Stability) How can you identify whether a system is a causal and/or stable system from the following information?

(a) Its impulse response $h[n]$

(b) Its transfer function $H(z)$ and its region of convergence

(c) Its system difference equation

(d) Its pole-zero plot

17.58 (Inverse Transforms) Consider the *stable* system described by $y[n] + \alpha y[n-1] = x[n] + x[n-1]$.

(a) Find its causal impulse response $h[n]$ and specify the range of α and the ROC of $H(z)$.

(b) Find its anti-causal impulse response $h[n]$ and specify the range of α and the ROC of $H(z)$.

PGR2026-00033
Krisp Technologies EX1024 Page 646

17.59 (Inverse Transforms) Each $X(z)$ below represents the z-transform of a switched periodic signal $x_p[n]u[n]$. Find one period $x_1[n]$ of each signal and verify your results using inverse transformation by long division.

(a) $X(z) = \dfrac{1}{1-z^{-3}}$ **(b)** $X(z) = \dfrac{1}{1+z^{-3}}$

(c) $X(z) = \dfrac{1+2z^{-1}}{1-z^{-4}}$ **(d)** $X(z) = \dfrac{3+2z^{-1}}{1+z^{-4}}$

17.60 (Inverse Transforms) Let $H(z) = (z-0.5)(2z+4)(1-z^{-2})$.

(a) Find its inverse transform $h[n]$.

(b) Does $h[n]$ describe a symmetric sequence?

(c) Does $h[n]$ describe a linear phase sequence?

17.61 (Inverse Transforms) Let $H(z) = \dfrac{z}{(z-0.5)(z+2)}$.

(a) Find its impulse response $h[n]$ if it is known that this represents a stable system. Is this system causal?

(b) Find its impulse response $h[n]$ if it is known that this represents a causal system. Is this system stable?

17.62 (Inverse Transforms) Let $H(z) = \dfrac{z}{(z-0.5)(z+2)}$. Establish the ROC of $H(z)$, find its impulse response $h[n]$, and investigate its stability for the following:

(a) A causal $h[n]$ **(b)** An anti-causal $h[n]$ **(c)** A two-sided $h[n]$

17.63 (Periodic Signal Generators) Find the transfer function $H(z)$ of a filter whose impulse response is a periodic sequence with first period $x[n] = \{\overset{\Downarrow}{1},\ 2,\ 3,\ 4,\ 6,\ 7,\ 8\}$. Find the difference equation and sketch a realization for this filter.

17.64 (Periodic Signal Generators) Find the transfer function $H(z)$ and the difference equation of a filter whose impulse response is $h[n] = \cos(2n\pi F_0)$ and sketch its realization. What is the unit step response of this filter for $F_0 = 0.25$?

17.65 (Steady-State Response) The filter $H(z) = A\dfrac{z-\alpha}{z-0.5\alpha}$ is designed to have a steady-state response of unity if the input is $u[n]$ and a steady-state response of zero if the input is $\cos(n\pi)$. What are the values of A and α?

17.66 (Steady-State Response) The filter $H(z) = A\dfrac{z-\alpha}{z-0.5\alpha}$ is designed to have a steady-state response of zero if the input is $u[n]$ and a steady-state response of unity if the input is $\cos(n\pi)$. What are the values of A and α?

17.67 (System Response) Find the response of the following filters to the unit step $x[n] = u[n]$, and to the alternating unit step $x[n] = (-1)^n u[n]$.

(a) $h[n] = \delta[n] - \delta[n-1]$ (differencing operation)

(b) $h[n] = \{\overset{\Downarrow}{0.5},\ 0.5\}$ (2-point average)

(c) $h[n] = \frac{1}{N}\displaystyle\sum_{k=0}^{N-1} \delta[n-k],\ N = 3$ (moving average)

PGR2026-00033
Krisp Technologies EX1024 Page 647

(d) $h[n] = \frac{2}{N(N+1)} \sum_{k=0}^{N-1} (N-k)\delta[n-k],\ N = 3$ (weighted moving average)

(e) $y[n] - \alpha y[n-1] = (1-\alpha)x[n],\ \ \alpha = \frac{N-1}{N+1},\ N = 3$ (exponential average)

17.68 (Steady-State Response) Consider the following DSP system:

$$x(t) \longrightarrow \boxed{\text{sampler}} \longrightarrow \boxed{\text{digital filter } H(z)} \longrightarrow \boxed{\text{ideal LPF}} \longrightarrow y(t)$$

The input is $x(t) = 2 + \cos(10\pi t) + \cos(20\pi t)$. The sampler is ideal and operates at a sampling rate of S Hz. The digital filter is described by $H(z) = 0.1S\dfrac{z-1}{z-0.5}$. The ideal lowpass filter has a cutoff frequency of $0.5S$ Hz.

(a) What is the smallest value of S that will prevent aliasing?

(b) Let $S = 40$ Hz and $H(z) = 1 + z^{-2} + z^{-4}$. What is the *steady-state* output $y(t)$?

(c) Let $S = 40$ Hz and $H(z) = \dfrac{z^2+1}{z^4+0.5}$. What is the *steady-state* output $y(t)$?

17.69 (System Analysis) The impulse response of a system is $h[n] = \delta[n] - \alpha\delta[n-1]$. Determine α and make a sketch of the pole-zero plot for this system to act as

(a) A lowpass filter. **(b)** A highpass filter. **(c)** An allpass filter.

17.70 (System Analysis) The impulse response of a system is $h[n] = \alpha^n u[n]$. Determine α and make a sketch of the pole-zero plot for this system to act as

(a) A stable lowpass filter. **(b)** A stable highpass filter. **(c)** An allpass filter.

17.71 (System Analysis) Consider a system whose impulse response is $h[n] = (0.5)^n u[n]$. Find its response to the following inputs.

(a) $x[n] = \delta[n]$ **(b)** $x[n] = u[n]$
(c) $x[n] = (0.25)^n u[n]$ **(d)** $x[n] = (0.5)^n[n]$
(e) $x[n] = \cos(n\pi)$ **(f)** $x[n] = \cos(n\pi)u[n]$
(g) $x[n] = \cos(0.5n\pi)$ **(h)** $x[n] = \cos(0.5n\pi)u[n]$

17.72 (System Analysis) Consider a system whose impulse response is $h[n] = n(0.5)^n u[n]$. What input $x[n]$ will result in each of the following outputs?

(a) $y[n] = \cos(0.5n\pi)$
(b) $y[n] = 2 + \cos(0.5n\pi)$
(c) $y[n] = \cos^2(0.25n\pi)$

17.73 (System Response) Consider the system $y[n] - 0.25y[n-2] = x[n]$. Find its response $y[n]$, using z-transforms, for the following inputs.

(a) $x[n] = 2\delta[n-1] + u[n]$ **(b)** $x[n] = 2 + \cos(0.5n\pi)$

17.74 (Poles and Zeros) It is known that the transfer function $H(z)$ of a filter has two poles at $z = 0$, two zeros at $z = -1$, and a dc gain of 8.

(a) Find the filter transfer function $H(z)$ and impulse response $h[n]$.

(b) Is this an IIR or FIR filter?

PGR2026-00033
Krisp Technologies EX1024 Page 648

(c) Is this a causal or noncausal filter?
(d) Is this a linear-phase filter? If so, what is the symmetry in $h[n]$?
(e) Repeat parts (a)–(d) if another zero is added at $z = -1$.

17.75 (System Response) The signal $x[n] = (0.5)^n u[n]$ is applied to a digital filter, and the response is $y[n]$. Find the filter transfer function and state whether it is an IIR or FIR filter and whether it is a linear-phase filter if the system output $y[n]$ is the following:

(a) $y[n] = \delta[n] + 0.5\delta[n-1]$
(b) $y[n] = \delta[n] - 2\delta[n-1]$
(c) $y[n] = (-0.5)^n u[n]$

17.76 (Frequency Response) Sketch the pole-zero plot and frequency response of the following systems and describe the function of each system.

(a) $y[n] = 0.5x[n] + 0.5x[n-1]$ **(b)** $y[n] = 0.5x[n] - 0.5x[n-1]$

(c) $h[n] = \frac{1}{3}\{1, \overset{\Downarrow}{1}, 1\}$ **(d)** $h[n] = \frac{1}{3}\{1, \overset{\Downarrow}{-1}, 1\}$

(e) $h[n] = \{\overset{\Downarrow}{0.5}, 1, 0.5\}$ **(f)** $h[n] = \{\overset{\Downarrow}{0.5}, -1, 0.5\}$

(g) $H(z) = \dfrac{z - 0.5}{z + 0.5}$ **(h)** $H(z) = \dfrac{z - 2}{z + 0.5}$

(i) $H(z) = \dfrac{z + 2}{z + 0.5}$ **(j)** $H(z) = \dfrac{z^2 + 2z + 3}{3z^2 + 2z + 1}$

17.77 (Interconnected Systems) Consider two systems whose impulse response is $h_1[n] = \delta[n] + \alpha\delta[n-1]$ and $h_2[n] = (0.5)^n u[n]$. Find the overall system transfer function and the response $y[n]$ of the overall system to the input $x[n] = (0.5)^n u[n]$, and to the input $x[n] = \cos(n\pi)$ if

(a) The two systems are connected in parallel with $\alpha = 0.5$.
(b) The two systems are connected in parallel with $\alpha = -0.5$.
(c) The two systems are connected in cascade with $\alpha = 0.5$.
(d) The two systems are connected in cascade with $\alpha = -0.5$.

17.78 (Interconnected Systems) The transfer function $H(z)$ of the cascade of two systems $H_1(z)$ and $H_2(z)$ is known to be $H(z) = \dfrac{z^2 + 0.25}{z^2 - 0.25}$. It is also known that the unit step response of the first system is $[2 - (0.5)^n]u[n]$. Determine $H_1(z)$ and $H_2(z)$.

17.79 (Frequency Response) Consider the filter realization of Figure P17.79. Find the transfer function $H(z)$ of the overall system if the impulse response of the filter is given by

(a) $h[n] = \delta[n] - \delta[n-1]$. **(a)** $h[n] = 0.5\delta[n] + 0.5\delta[n-1]$.

Find the difference equation relating $y[n]$ and $x[n]$ from $H(z)$ and investigate the stability of the overall system.

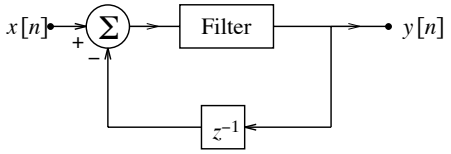


Figure P17.79 Filter realization for Problem 17.79

PGR2026-00033
Krisp Technologies EX1024 Page 649

17.80 (Systems in Cascade) Consider the following system:

$$x[n] \longrightarrow \boxed{H_1(z)} \longrightarrow \boxed{H_2(z)} \longrightarrow \boxed{H_3(z)} \longrightarrow y[n]$$

It is known that $h_1[n] = 0.5(0.4)^n u[n]$, $H_2(z) = \dfrac{A(z+\alpha)}{z+\beta}$, and $h_3[n] = \delta[n] + 0.5\delta[n-1]$. Choose A, α, and β such that the overall system represents an identity system.

17.81 (Recursive and Non-Recursive Filters) Consider two filters described by

$$(1)\ h[n] = \{\overset{\Downarrow}{1}, 1, 1\} \qquad (2)\ y[n] - y[n-1] = x[n] - x[n-3]$$

(a) Find the transfer function of each filter.
(b) Find the response of each filter to the input $x[n] = \cos(n\pi)$.
(c) Are the two filters related in any way?

COMPUTATION AND DESIGN

dtfgui A GUI for Discrete-Time System Visualization

The graphical user interface `dtfgui` allows you to visualize the impulse response, the pole-zero plot, and the frequency response of a discrete-time system.
You can select the transfer function and plot results using various options.
To explore this routine, type `dtfgui` at the MATLAB prompt.

17.82 (System Response in Symbolic Form) The ADSP routine `sysresp2` returns the system response in symbolic form. See Chapter 21 for examples of its usage. Obtain the response of the following filters and plot the response for $0 \le n \le 30$.

(a) The step response of $y[n] - 0.5y[n] = x[n]$
(b) The impulse response of $y[n] - 0.5y[n] = x[n]$
(c) The zero-state response of $y[n] - 0.5y[n] = (0.5)^n u[n]$
(d) The complete response of $y[n] - 0.5y[n] = (0.5)^n u[n],\ y[-1] = -4$
(e) The complete response of $y[n] + y[n-1] + 0.5y[n-2] = (0.5)^n u[n],\ y[-1] = -4,\ y[-2] = 3$

17.83 (Steady-State Response in Symbolic Form) The ADSP routine `ssresp` yields a symbolic expression for the steady-state response to sinusoidal inputs (see Chapter 21 for examples of its usage). Find the steady-state response to the input $x[n] = 2\cos(0.2n\pi - \frac{\pi}{3})$ for each of the following systems and plot the results over $0 \le n \le 50$.

(a) $y[n] - 0.5y[n] = x[n]$
(b) $y[n] + y[n-1] + 0.5y[n-2] = 3x[n]$

PGR2026-00033
Krisp Technologies EX1024 Page 650

Chapter 18

APPLICATIONS OF THE z-TRANSFORM

18.0 Scope and Objectives

This chapter deals with applications of the z-transform. It starts with aspects of digital filter realization. It provides a graphical interpretation of the frequency response and introduces the concept of filter design by pole-zero placement. It examines several useful filters, such as minimum-phase filters, allpass filters, and comb filters, and concludes with some applications in audio signal processing and communications.

18.1 Transfer Function Realization

The realization of digital filters described by transfer functions parallels the realization based on difference equations. The nonrecursive and recursive filters described by

$$H_N(z) = B_0 + B_1 z^{-1} + \cdots + B_M z^{-M} \quad y[n] = B_0 x[n] + B_1 x[n-1] + \cdots + B_M x[n-M] \tag{18.1}$$

$$H_R(z) = \frac{1}{1 + A_1 z^{-1} + \cdots + A_N z^{-N}} \quad y[n] = -A_1 y[n-1] - \cdots - A_N y[n-N] + x[n] \tag{18.2}$$

can be realized using the **feed-forward** (nonrecursive) structure and **feedback** (recursive) structure, as shown in Figure 18.1.

Now, consider the general difference equation

$$y[n] = -A_1 y[n-1] - \cdots - A_N y[n-N] + B_0 x[n] + B_1 x[n-1] + \cdots + B_N x[n-N] \tag{18.3}$$

We choose $M = N$ with no loss of generality, since some of the coefficients B_k may always be set to zero. The transfer function (with $M = N$) then becomes

$$H(z) = \frac{B_0 + B_1 z^{-1} + \cdots + B_N z^{-N}}{1 + A_1 z^{-1} + A_2 z^{-2} + \cdots + A_N z^{-N}} = H_N(z) H_R(z) \tag{18.4}$$

The transfer function $H(z) = H_N(z)H_R(z)$ is the product of the transfer functions of a recursive and a nonrecursive system. Its realization is thus a cascade of the realizations for the recursive and nonrecursive portions, as shown in Figure 18.2(a). This form describes a **direct form I** realization. It uses $2N$ delay elements to realize an Nth-order difference equation and is therefore not very efficient.

Since LTI systems can be cascaded in any order, we can switch the recursive and nonrecursive parts to get the structure of Figure 18.2(b). This structure suggests that each pair of feed-forward and feedback signals can be obtained from a single delay element instead of two. This allows us to use only N delay elements

PGR2026-00033
Krisp Technologies EX1024 Page 651

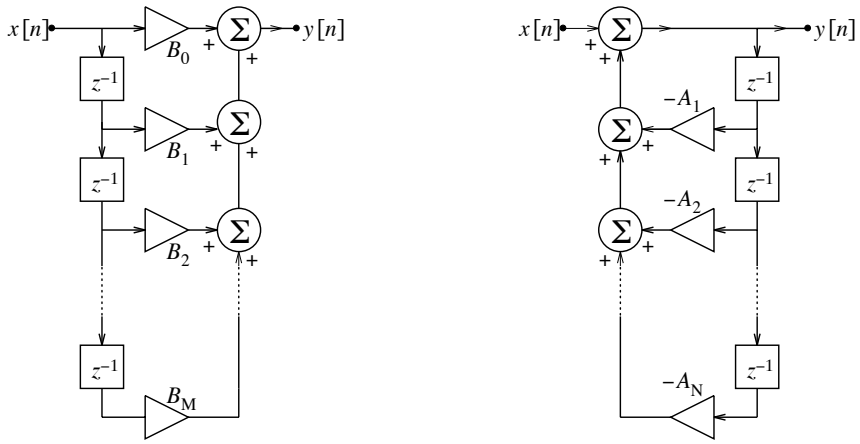


Figure 18.1 Realization of a nonrecursive (left) and recursive (right) digital filter

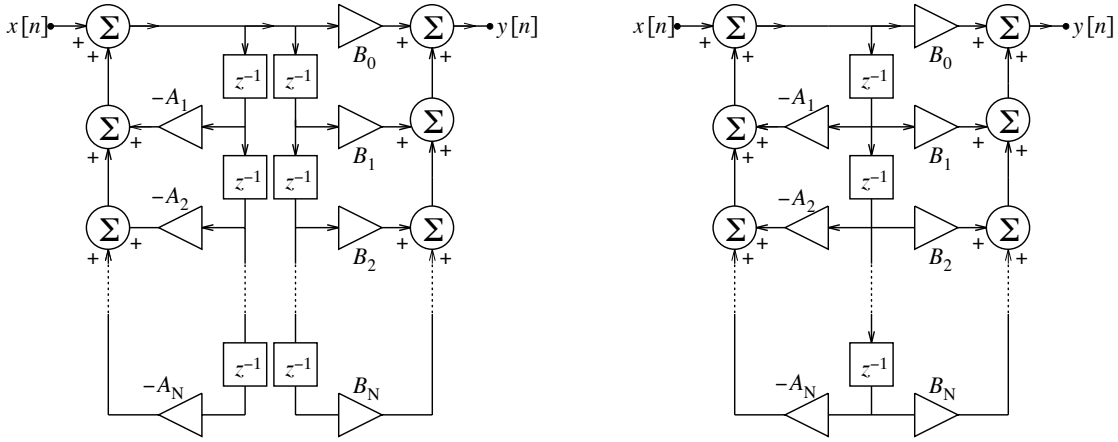


Figure 18.2 Direct form I (left) and canonical, or direct form II (right), realization of a digital filter

and results in the **direct form II**, or **canonic**, realization. The term *canonic* implies a realization with the minimum number of delay elements.

If M and N are not equal, some of the coefficients (A_k or B_k) will equal zero and will result in missing signal paths corresponding to these coefficients in the filter realization.

REVIEW PANEL 18.1

Digital Filter Realization

FIR: No feedback paths **IIR:** Both feed-forward and feedback paths

18.1.1 Transposed Realization

The direct form II also yields a transposed realization if we turn the realization around (and reverse the input and input), replace summing junctions by nodes (and vice versa), and reverse the direction of signal flow. Such a realization is developed in Figure 18.3.

PGR2026-00033
Krisp Technologies EX1024 Page 652

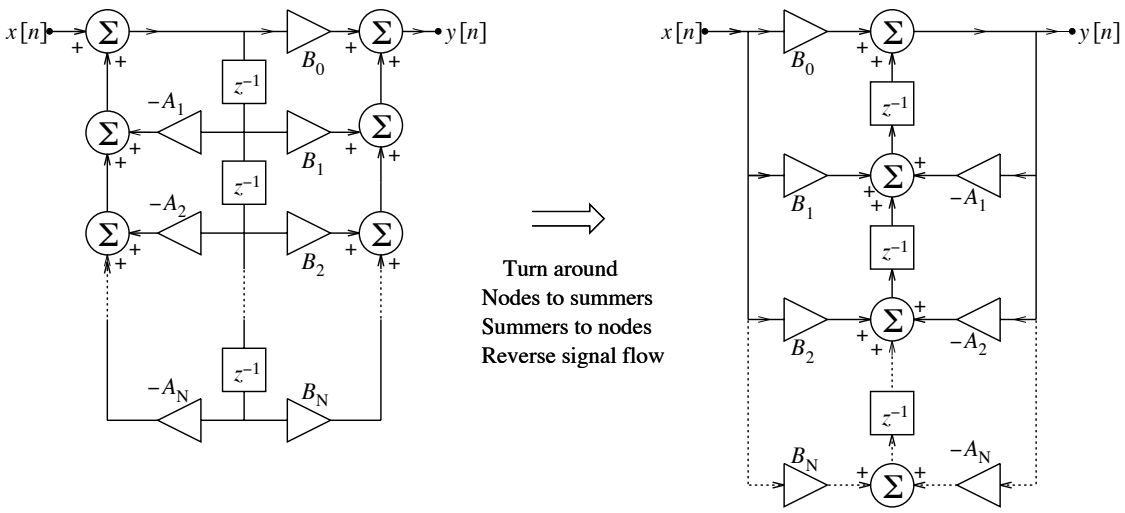


Figure 18.3 Direct form II (left) and transposed (right) realization of a digital filter

EXAMPLE 18.1 (Direct Form II and Transposed Realizations)
Consider a system described by $2y[n] - y[n-2] - 4y[n-3] = 3x[n-2]$. Its transfer function is

$$H(z) = \frac{3z^{-2}}{2 - z^{-2} - 4z^{-3}} = \frac{1.5z}{z^3 - 0.5z - 2}$$

This is a third-order system. To sketch its direct form II and transposed realizations, we compare $H(z)$ with the generic third-order transfer function to get

$$H(z) = \frac{B_0 z^3 + B_1 z^2 + B_2 z + B_3}{z^3 + A_1 z^2 + A_2 z + A_3}$$

The nonzero constants are $B_2 = 1.5$, $A_2 = -0.5$, and $A_3 = -2$. Using these, we obtain the direct form II and transposed realizations shown in Figure E18.1.

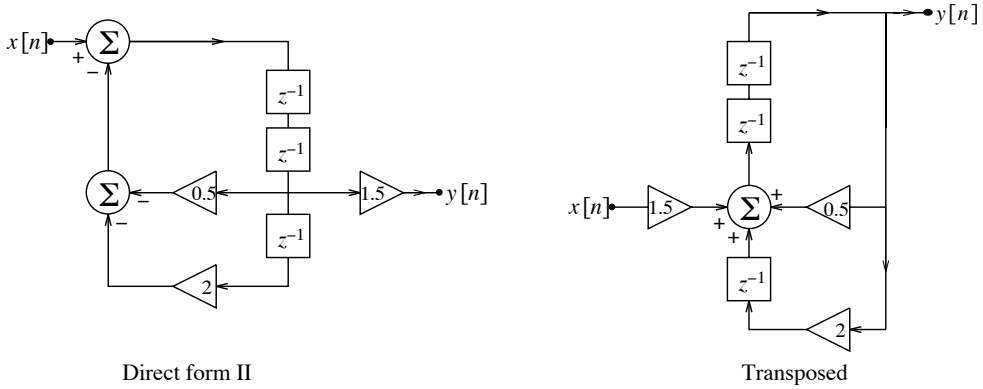


Figure E18.1 Direct form II (left) and transposed (right) realization of the system for Example 18.1

PGR2026-00033
Krisp Technologies EX1024 Page 653

18.2 Interconnected Systems

The z-transform is well suited to the study of interconnected LTI systems. Figure 18.4 shows the interconnection of two *relaxed* systems in cascade and in parallel.

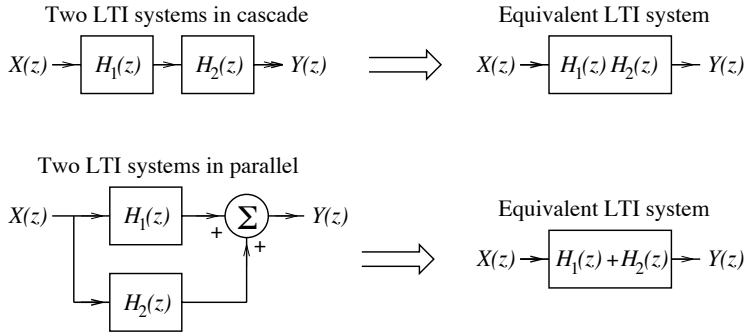


Figure 18.4 Cascade and parallel systems and their equivalents

The overall transfer function of a **cascaded system** is the product of the individual transfer functions. For n systems in cascade, the overall impulse response $h_C[n]$ is the convolution of the individual impulse responses $h_1[n]$, $h_2[n], \ldots$. Since the convolution operation transforms to a product, we have

$$H_C(z) = H_1(z)H_2(z)\cdots H_n(z) \qquad \text{(for } n \text{ systems in cascade)} \tag{18.5}$$

We can also factor a given transfer function $H(z)$ into the product of first-order and second-order transfer functions and realize $H(z)$ in cascaded form.

For systems in **parallel**, the overall transfer function is the sum of the individual transfer functions. For n systems in parallel,

$$H_P(z) = H_1(z) + H_2(z) + \cdots + H_n(z) \qquad \text{(for } n \text{ systems in parallel)} \tag{18.6}$$

We can also use partial fractions to express a given transfer function $H(z)$ as the sum of first-order and/or second-order subsystems, and realize $H(z)$ as a parallel combination.

REVIEW PANEL 18.2

Overall Impulse Response and Transfer Function of Systems in Cascade and Parallel

Cascade: Convolve individual impulse responses. Multiply individual transfer functions.
Parallel: Add individual impulse responses. Add individual transfer functions.

EXAMPLE 18.2 (Systems in Cascade and Parallel)

(a) Two digital filters are described by $h_1[n] = \alpha^n u[n]$ and $h_2[n] = (-\alpha)^n u[n]$. The transfer function of their cascade is $H_C(z)$ and of their parallel combination is $H_P(z)$. How are $H_C(z)$ and $H_P(z)$ related?

The transfer functions of the two filters are $H_1(z) = \dfrac{z}{z-\alpha}$ and $H_2(z) = \dfrac{z}{z+\alpha}$. Thus,

$$H_C(z) = H_1(z)H_2(z) = \frac{z^2}{z^2-\alpha^2} \qquad H_P(z) = H_1(z) + H_2(z) = \frac{2z^2}{z^2-\alpha^2}$$

So, $H_P(z) = 2H_C(z)$.

PGR2026-00033
Krisp Technologies EX1024 Page 654

(b) Find a cascaded realization for $H(z) = \dfrac{z^2(6z-2)}{(z-1)(z^2-\frac{1}{6}z-\frac{1}{6})}$.

This system may be realized as a cascade $H(z) = H_1(z)H_2(z)$, as shown in Figure E18.2B, where

$$H_1(z) = \frac{z^2}{z^2-\frac{1}{6}z-\frac{1}{6}} \qquad H_2(z) = \frac{6z-2}{z-1}$$

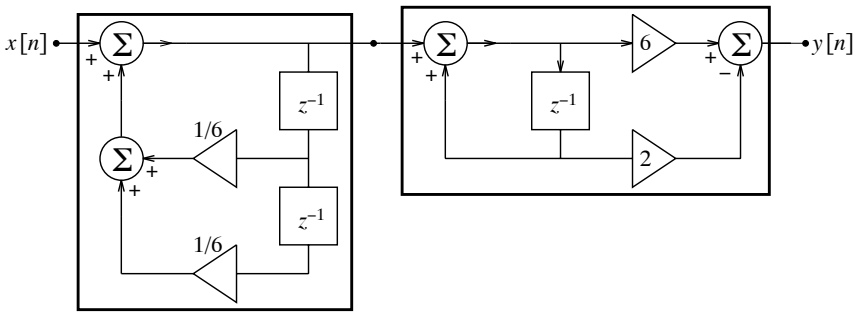


Figure E18.2B Cascade realization of the system for Example 18.2(b)

(c) Find a parallel realization for $H(z) = \dfrac{z^2}{(z-1)(z-0.5)}$

Using partial fractions, we find $H(z) = \dfrac{2z}{z-1} - \dfrac{z}{z-0.5} = H_1(z) - H_2(z)$.

The two subsystems $H_1(z)$ and $H_2(z)$ may now be used to obtain the parallel realization, as shown in Figure E18.2C.

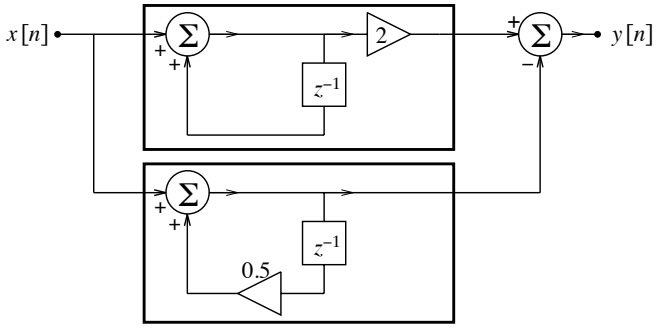


Figure E18.2C Parallel realization of the system for Example 18.2(c)

(d) Is the cascade or parallel combination of two *linear-phase* filters also linear phase? Explain.

Linear-phase filters are described by symmetric impulse response sequences.

The impulse response of their cascade is also symmetric because it is the convolution of two symmetric sequences. So, the cascade of two linear-phase filters is *always* linear phase.

The impulse response of their parallel combination is the sum of their impulse responses. Since the sum of symmetric sequences is not always symmetric (unless both are odd symmetric or both are even symmetric), the parallel combination of two linear-phase filters is not always linear phase.

PGR2026-00033
Krisp Technologies EX1024 Page 655

18.2.1 Inverse Systems

The **inverse system** corresponding to a transfer function $H(z)$ is denoted by $H^{-1}(z)$, and defined as

$$H^{-1}(z) = H_I(z) = \frac{1}{H(z)} \tag{18.7}$$

The cascade of a system and its inverse has a transfer function of unity:

$$H_C(z) = H(z)H^{-1}(z) = 1 \qquad h_C[n] = \delta[n] \tag{18.8}$$

This cascaded system is called an **identity system**, and its impulse response equals $h_C[n] = \delta[n]$. The inverse system can be used to undo the effect of the original system. We can also describe $h_C[n]$ by the convolution $h[n] \star h_I[n]$. It is far easier to find the inverse of a system in the transformed domain.

REVIEW PANEL 18.3

Relating the Impulse Response and Transfer Function of a System and Its Inverse

If the system is described by $H(z)$ and $h[n]$ and its inverse by $H_I(z)$ and $h_I[n]$, then

$$H(z)H_I(z) = 1 \qquad h[n] \star h_I[n] = \delta[n]$$

EXAMPLE 18.3 (Inverse Systems)

Consider a system with the difference equation $y[n] = \alpha y[n-1] + x[n]$.

To find the inverse system, we evaluate $H(z)$ and take its reciprocal. Thus,

$$H(z) = \frac{1}{1-\alpha z^{-1}} \qquad H_I(z) = \frac{1}{H(z)} = 1 - \alpha z^{-1}$$

The difference equation of the inverse system is $y[n] = x[n] - \alpha x[n-1]$.

18.3 Minimum-Phase Systems

Consider a system described in factored form by

$$H(z) = K\frac{(z-z_1)(z-z_2)(z-z_3)\cdots(z-z_m)}{(z-p_1)(z-p_2)(z-p_3)\cdots(z-p_n)} \tag{18.9}$$

If we replace K by $-K$, or a factor $(z-\alpha)$ by $(\frac{1}{z}-\alpha)$ or by $\frac{1-\alpha z}{z}$ or by $(1-\alpha z)$, the magnitude of $|H_p(F)|$ remains unchanged, and only the phase is affected. If $K > 0$ and all the poles and zeros of $H(z)$ lie inside the unit circle, $H(z)$ is stable and defines a **minimum-phase system**. It shows the *smallest group delay* (and the *smallest deviation from zero phase*) at every frequency among all systems with the same magnitude response. A stable system is called **mixed phase** if some of its zeros lie outside the unit circle and **maximum phase** if all its zeros lie outside the unit circle. *Of all stable systems with the same magnitude response, there is only one minimum-phase system.*

REVIEW PANEL 18.4

All Poles and Zeros of a Minimum-Phase System Lie Inside the Unit Circle

PGR2026-00033
Krisp Technologies EX1024 Page 656

EXAMPLE 18.4 (The Minimum-Phase Concept)
Consider the transfer function of the following systems:

$$H_1(z) = \frac{(z-\frac{1}{2})(z-\frac{1}{4})}{(z-\frac{1}{3})(z-\frac{1}{5})} \qquad H_2(z) = \frac{(1-\frac{1}{2}z)(z-\frac{1}{4})}{(z-\frac{1}{3})(z-\frac{1}{5})} \qquad H_3(z) = \frac{(1-\frac{1}{2}z)(1-\frac{1}{4}z)}{(z-\frac{1}{3})(z-\frac{1}{5})}$$

Each system has poles inside the unit circle and is thus stable. All systems have the same magnitude. Their phase and delay are different, as shown Figure E18.4.

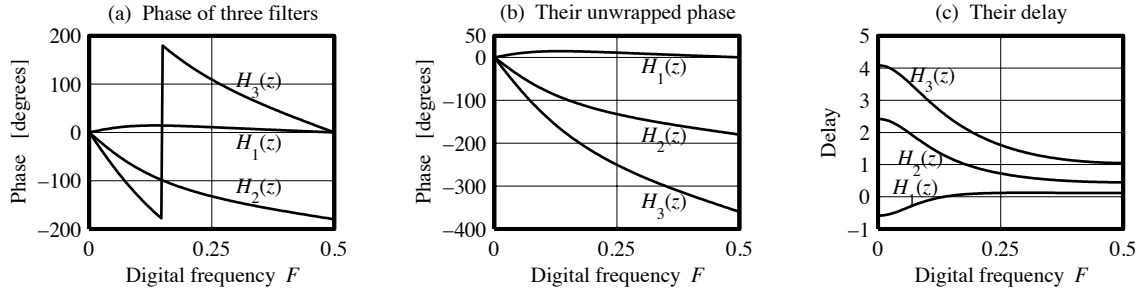


Figure E18.4 Response of the systems for Example 18.4

The phase response confirms that $H_1(z)$ is a minimum-phase system with no zeros outside the unit circle, $H_2(z)$ is a mixed-phase system with one zero outside the unit circle, and $H_3(z)$ is a maximum-phase system with all its zeros outside the unit circle.

18.3.1 Minimum-Phase Systems from the Magnitude Spectrum

The design of many digital systems is often based on a specified magnitude response $|H_p(F)|$. The phase response is then selected to ensure a causal, stable system. The transfer function of such a system is unique and may be found by writing its magnitude squared function $|H_p(F)|^2$ as

$$|H_p(F)|^2 = H_p(F)H_p(-F) = H(z)H(1/z)|_{z\to\exp(j2\pi F)} \tag{18.10}$$

From $|H_p(F)|^2$, we can reconstruct $H_T(z) = H(z)H(1/z)$, which displays conjugate reciprocal symmetry. For every root r_k, there is a root at $1/r_k^*$. We thus select only the roots lying inside the unit circle to extract the minimum-phase transfer function $H(z)$. The following example illustrates the process.

EXAMPLE 18.5 (Finding the Minimum-Phase System)
Find the minimum-phase transfer function $H(z)$ corresponding to $|H_p\Omega)|^2 = \dfrac{5+4\cos\Omega}{17+8\cos\Omega}$.

We use Euler's relation to give $|H_p(\Omega)|^2 = \dfrac{5+2e^{j\Omega}+2e^{-j\Omega}}{17+4e^{j\Omega}+4e^{-j\Omega}}$. Upon substituting $e^{j\Omega}\to z$, we obtain $H_T(z)$ as

$$H_T(z) = H(z)H(1/z) = \frac{5+2z+2/z}{17+4z+4/z} = \frac{2z^2+5z+2}{4z^2+17z+4} = \frac{(2z+1)(z+2)}{(4z+1)(z+4)}$$

To extract $H(z)$, we pick the roots of $H_T(z)$ that correspond to $|z|<1$. This yields

$$H(z) = \frac{(2z+1)}{(4z+1)} = 0.5\frac{(z+0.5)}{(z+0.25)}$$

We find that $|H(z)|_{z=1} = |H_p(\Omega)|_{\Omega=0} = 0.6$, implying identical dc gains.

PGR2026-00033
Krisp Technologies EX1024 Page 657

18.3.2 Linear-Phase Sequences

Linear phase plays an important role in the design of digital filters. Sequences that are symmetric about the origin (or shifted versions of such sequences) are endowed with linear phase and constant delay. The pole and zero locations of such sequences cannot be arbitrary. The poles must lie at the origin for a sequence $x[n]$ to be of finite length. Since signal symmetry (about the origin) requires $x[n] = \pm x[-n]$ and thus $X(z) = \pm X(1/z)$, the zeros must occur in reciprocal pairs (and conjugate pairs if complex) and exhibit what is called **conjugate reciprocal symmetry**. This is illustrated in Figure 18.5. Sequences that are symmetric about their midpoint also display conjugate reciprocal symmetry.

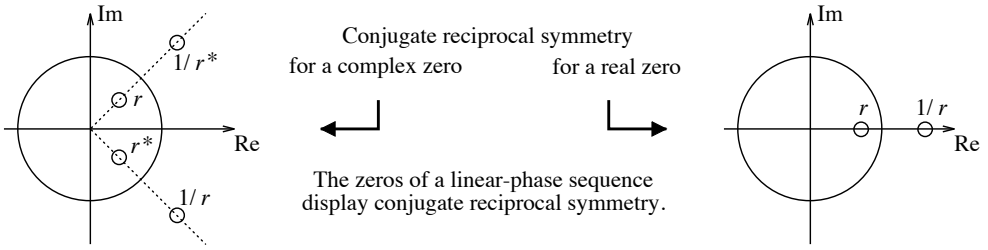


Figure 18.5 Conjugate reciprocal symmetry

Real zeros at $z = 1$ or $z = -1$ need not be paired (they form their own reciprocals), but all other real zeros must be paired with their reciprocals. Complex zeros on the unit circle must be paired with their conjugates (which also form their reciprocals), whereas complex zeros anywhere else must occur in conjugate reciprocal quadruples. For a linear-phase sequence to be odd symmetric, there must be *an odd number of zeros at* $z = 1$.

REVIEW PANEL 18.5

Characteristics of a Linear-Phase Sequence

Time domain: A linear-phase signal $x[n]$ is symmetric about its midpoint.
Pole-zero plot: All poles are at $z = 0$ (for finite length). Zeros at $z = \pm 1$ can occur singly.
Other zeros occur in (reciprocal) pairs or conjugate reciprocal quadruples.
Symmetry about $n = 0$: $X(z) = X(\frac{1}{z})$ (even $x[n]$) $\quad X(z) = -X(\frac{1}{z})$ (odd $x[n]$)
About midpoint: Odd symmetry if an odd number of zeros at $z = 1$ (even symmetry otherwise)

EXAMPLE 18.6 (Linear-Phase Sequences)

(a) Does $X(z) = 1 + 2z^{-1} + 2z^{-2} + z^{-3}$ describe a linear-phase sequence?

We see that $x[n] = \{\overset{\Downarrow}{1},\ 2,\ 2,\ 1\}$ is symmetric about its midpoint $n = 1.5$.

Since $X(z) = (z^3+2z^2+2z+1)/z^3$, its poles are at $z = 0$. Its zeros are at $z = -1$ and $z = -0.5 \pm j0.866$. The zeros show conjugate reciprocal symmetry because the zeros at $z = -0.5 \pm j0.866$ (on the unit circle) are their own reciprocals, as is the zero at $z = -1$. Thus, $X(z)$ corresponds to a linear-phase signal $x[n]$.

Finally, since $X(z) \neq \pm X(1/z)$, the signal $x[n]$ is not symmetric about the origin $n = 0$.

PGR2026-00033
Krisp Technologies EX1024 Page 658

(b) Let $x[n] = \delta[n+2] + 4.25\delta[n] + \delta[n-2]$. Sketch the pole-zero plot of $X(z)$.

We find $X(z) = z^2 + 4.25 + z^{-2}$. Since $X(z) = X(1/z)$, $x[n]$ is even symmetric about $n = 0$.

In factored form, $X(z) = \dfrac{z^4 + 4.25z^2 + 1}{z^2} = \dfrac{(z + j0.5)(z - j0.5)(z + j2)(z - j2)}{z^2}$.

Its four zeros are at $j0.5$, $-j2$, $-j0.5$, and $j2$. The pole-zero plot is shown in Figure E18.6 (left panel). Note the conjugate reciprocal symmetry of the zeros.

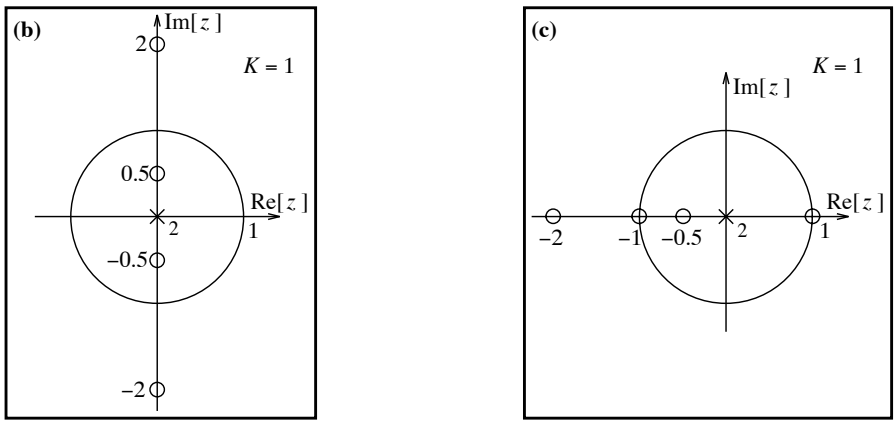


Figure E18.6 Pole-zero plots of the sequences for Example 18.6(b and c)

(c) Sketch the pole-zero plot for $X(z) = z^2 + 2.5z - 2.5z^{-1} - z^{-2}$. Is $x[n]$ a linear-phase sequence?

Since $X(z) = -X(1/z)$, $x[n]$ is odd symmetric. In fact, $x[n] = \{1,\ 2.5,\ \underset{\Downarrow}{0},\ -2.5,\ -1\}$.

In factored form, $X(z) = \dfrac{(z-1)(z+1)(z+0.5)(z+2)}{z^2}$.

We see a pair of reciprocal zeros at -0.5 and -2. The zeros at $z = -1$ and $z = 1$ are their own reciprocals and are not paired. The pole-zero plot is shown in Figure E18.6 (right panel).

18.4 The Frequency Response: A Graphical Interpretation

The factored form or pole-zero plot of $H(z)$ is quite useful if we want a qualitative picture of its frequency response $H(F)$ or $H(\Omega)$. Consider the stable transfer function $H(z)$ given in factored form by

$$H(z) = \frac{8(z-1)}{(z - 0.6 - j0.6)(z - 0.6 + j0.6)} \tag{18.11}$$

Its frequency response $H(\Omega)$ and magnitude $|H(\Omega)|$ at $\Omega = \Omega_0$ are given by

$$H(\Omega_0) = \frac{8(e^{j\Omega_0} - 1)}{(e^{j\Omega_0} - 0.6 - j0.6)(e^{j\Omega_0} - 0.6 + j0.6]} = \frac{8\mathbf{N}_1}{\mathbf{D}_1\mathbf{D}_2} \tag{18.12}$$

$$|H(\Omega_0)| = 8\frac{|\mathbf{N}_1|}{|\mathbf{D}_1||\mathbf{D}_2|} = (\text{gain factor})\frac{\textit{product of}\ \text{distances from zeros}}{\textit{product of}\ \text{distances from poles}} \tag{18.13}$$

PGR2026-00033
Krisp Technologies EX1024 Page 659

Analytically, the magnitude $|H(\Omega_0)|$ is the ratio of the magnitudes of each term. Graphically, the complex terms may be viewed in the z-plane as vectors $\mathbf{N}_1$, $\mathbf{D}_1$, and $\mathbf{D}_2$, directed from each pole or zero location to the location $\Omega = \Omega_0$ on the unit circle corresponding to $z = e^{j\Omega_0}$. The gain factor times the ratio of the vector magnitudes (the product of distances from the zeros divided by the product of distances from the poles) yields $|H(\Omega_0)|$, the magnitude at $\Omega = \Omega_0$. The difference in the angles yields the phase at $\Omega = \Omega_0$. The vectors and the corresponding magnitude spectrum are sketched for several values of Ω in Figure 18.6.

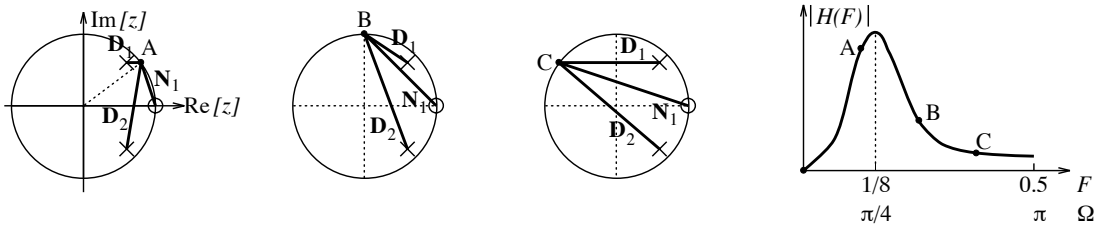


Figure 18.6 Graphical interpretation of the frequency response

A graphical evaluation can yield exact results but is much more suited to obtaining a qualitative estimate of the magnitude response. We observe how the vector ratio $\mathbf{N}_1/\mathbf{D}_1\mathbf{D}_2$ influences the magnitude as Ω is increased from $\Omega = 0$ to $\Omega = \pi$. For our example, at $\Omega = 0$, the vector $\mathbf{N}_1$ is zero, and the magnitude is zero. For $0 < \Omega < \pi/4$ (point A), both $|\mathbf{N}_1|$ and $|\mathbf{D}_2|$ increase, but $|\mathbf{D}_1|$ decreases. Overall, the response is small but increasing. At $\Omega = \pi/4$, the vector $\mathbf{D}_1$ attains its smallest length, and we obtain a peak in the response. For $\frac{\pi}{4} < \Omega < \pi$ (points B and C), $|\mathbf{N}_1|$ and $|\mathbf{D}_1|$ are of nearly equal length, while $|\mathbf{D}_2|$ is increasing. The magnitude is thus decreasing. The form of this response is typical of a bandpass filter.

18.4.1 The Rubber Sheet Analogy

If we imagine the z-plane as a rubber sheet, tacked down at the zeros of $H(z)$ and poked up to an infinite height at the pole locations, the curved surface of the rubber sheet approximates the magnitude of $H(z)$ for any value of z, as illustrated in Figure 18.7. The poles tend to poke up the surface and the zeros try to pull it down. The slice around the unit circle ($|z| = 1$) approximates the frequency response $H(\Omega)$.

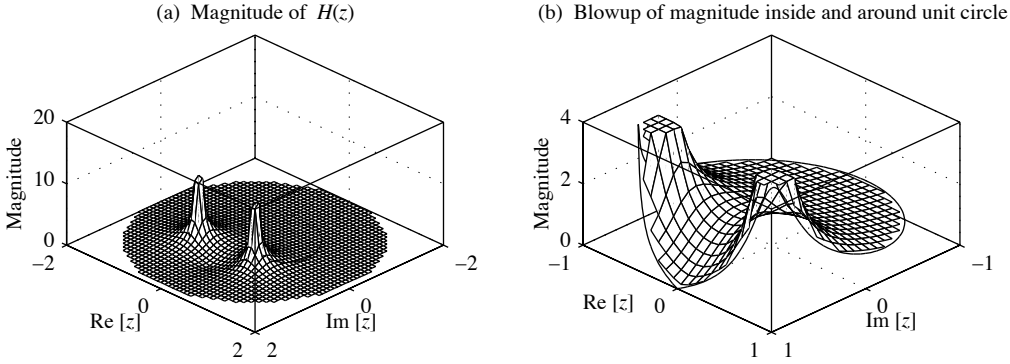


Figure 18.7 A plot of the magnitude of $H(z) = 8(z-1)/(z^2 - 1.2z + 0.72)$ in the z-plane

PGR2026-00033
Krisp Technologies EX1024 Page 660

18.4.2 Implications of Pole-Zero Placement for Filter Design

The basic strategy for pole and zero placement is based on mapping the passband and stopband frequencies on the unit circle and then positioning the poles and zeros based on the following reasoning:

1. **Conjugate symmetry:** All complex poles and zeros must be paired with their complex conjugates.
2. **Causality:** To ensure a causal system, the total number of zeros must be less than or equal to the total number of poles.
3. **Origin:** Poles or zeros at the origin do not affect the magnitude response.
4. **Stability:** For a stable system, the poles must be placed *inside* (not just on) the unit circle. The pole radius is proportional to the gain and inversely proportional to the bandwidth. Poles closer to the unit circle produce a large gain over a narrower bandwidth. Clearly, the passband should contain poles near the unit circle for large passband gains.
   A rule of thumb: For narrowband filters with bandwidth $\Delta\Omega \leq 0.2$ centered about Ω_0, we place conjugate poles at $z = R\exp(\pm j\Omega_0)$, where $R \approx 1 - 0.5\Delta\Omega$ is the pole radius.
5. **Minimum phase:** Zeros can be placed anywhere in the z-plane. To ensure minimum phase, the zeros must lie within the unit circle. Zeros on the unit circle result in a null in the response. Thus, the stopband should contain zeros on or near the unit circle. A good starting choice is to place a zero (on the unit circle) in the middle of the stopband or two zeros at the edges of the stopband.
6. **Transition band:** To achieve a steep transition from the stopband to the passband, a good choice is to pair each stopband zero with a pole along (or near) the same radial line and close to the unit circle.
7. **Pole-zero interaction:** Poles and zeros interact to produce a composite response that may not match qualitative predictions. Their placement may have to be changed, or other poles and zeros may have to be added, to tweak the response. Poles closer to the unit circle, or farther away from each other, produce small interaction. Poles closer to each other, or farther from the unit circle, produce more interaction. The closer we wish to approximate a given response, the more poles and zeros we require, and the higher is the filter order.

Bandstop and bandpass filters with real coefficients must have an order of at least two. Filter design using pole-zero placement involves trial and error. Formal design methods are presented in the next chapter. It is also possible to identify traditional filter types from their pole-zero patterns. The idea is to use the graphical approach to qualitatively observe how the magnitude changes from the lowest frequency $F = 0$ (or $\Omega = 0$) to the highest frequency $F = 0.5$ (or $\Omega = \pi$) and use this to establish the filter type.

REVIEW PANEL 18.6

Identifying Traditional Filter Types from Their Transfer Function $H(z)$

Evaluate the gain $|H(z)|$ at $z = 1$ (dc) and at $z = -1$ (corresponding to $F = 0.5$ or $\Omega = \pi$).

EXAMPLE 18.7 (Filters and Pole-Zero Plots)

(a) Identify the filter types corresponding to the pole-zero plots of Figure E18.7A.

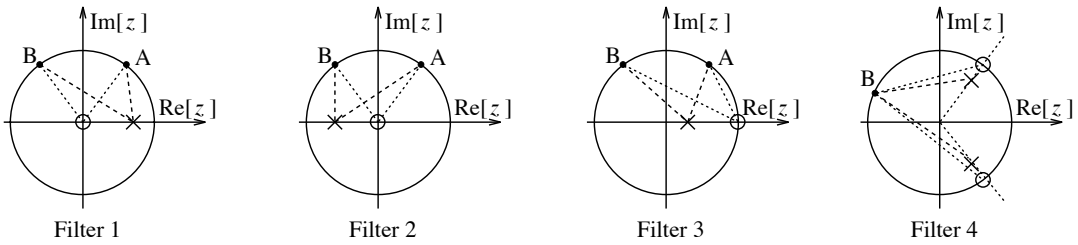


Figure E18.7A Pole-zero plots for Example 18.7(a)

PGR2026-00033
Krisp Technologies EX1024 Page 661

1. For filter 1, the vector length of the numerator is always unity, but the vector length of the denominator keeps increasing as we increase frequency (the points A and B, for example). The magnitude (ratio of the numerator and denominator lengths) thus decreases with frequency and corresponds to a lowpass filter.
2. For filter 2, the vector length of the numerator is always unity, but the vector length of the denominator keeps decreasing as we increase the frequency (the points A and B, for example). The magnitude increases with frequency and corresponds to a highpass filter.
3. For filter 3, the magnitude is zero at $\Omega = 0$. As we increase the frequency (the points A and B, for example), the ratio of the vector lengths of the numerator and denominator increases, and this also corresponds to a highpass filter.
4. For filter 4, the magnitude is zero at the zero location $\Omega = \Omega_0$. At any other frequency (the point B, for example), the ratio of the vector lengths of the numerator and denominator are almost equal and result in almost constant gain. This describes a bandstop filter.

(b) Design a bandpass filter with center frequency = 100 Hz, passband 10 Hz, stopband edges at 50 Hz and 150 Hz, and sampling frequency 400 Hz.

We find $\Omega_0 = \frac{\pi}{2}$, $\Delta\Omega = \frac{\pi}{20}$, $\Omega_s = [\frac{\pi}{4},\ \frac{3\pi}{4}]$, and $R = 1 - 0.5\Delta\Omega = 0.9215$.

Passband: Place poles at $p_{1,2} = Re^{\pm j\Omega_0} = 0.9215e^{\pm j\pi/2} = \pm j0.9215$.

Stopband: Place conjugate zeros at $z_{1,2} = e^{\pm j\pi/4}$ and $z_{3,4} = e^{\pm j3\pi/4}$.

We then obtain the transfer function as

$$H(z) = \frac{(z - e^{j\pi/4})(z - e^{-j\pi/4})(z - e^{j3\pi/4})(z - e^{-j3\pi/4})}{(z - j0.9215)(z + j0.9215)} = \frac{z^4 + 1}{z^2 + 0.8941}$$

Note that this filter is noncausal. To obtain a causal filter $H_1(z)$, we could, for example, use double-poles at each pole location to get

$$H_1(z) = \frac{z^4 + 1}{(z^2 + 0.8941)^2} = \frac{z^4 + 1}{z^4 + 1.6982z^2 + 0.7210}$$

The pole-zero pattern and gain of the modified filter $H_1(z)$ are shown in Figure E18.7B.

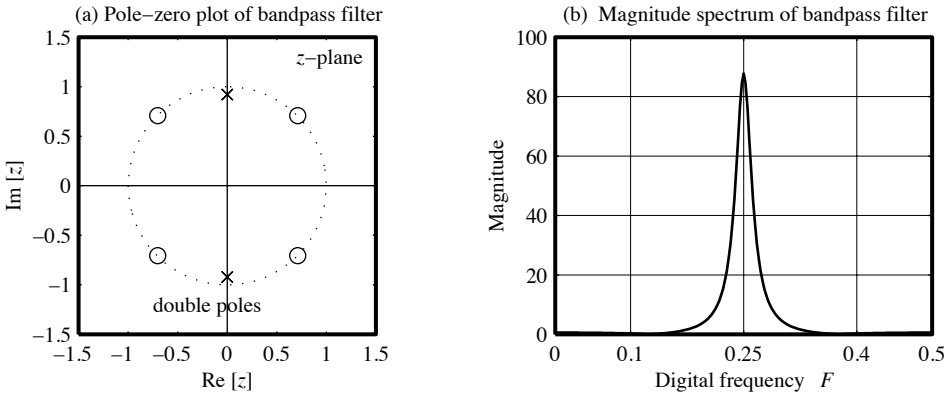


Figure E18.7B Frequency response of the bandpass filter for Example 18.7(b)

PGR2026-00033
Krisp Technologies EX1024 Page 662

(c) Design a notch filter with notch frequency 60 Hz, stopband 5 Hz, and sampling frequency 300 Hz.

We compute $\Omega_0 = \frac{2\pi}{5}$, $\Delta\Omega = \frac{\pi}{30}$, and $R = 1 - 0.5\Delta\Omega = 0.9476$.

Stopband: We place zeros at the notch frequency to get $z_{1,2} = e^{\pm j\Omega_0} = e^{\pm j2\pi/5}$.

Passband: We place poles along the orientation of the zeros at $p_{1,2} = Re^{\pm j\Omega_0} = 0.9476e^{\pm j2\pi/5}$.

We then obtain $H(z)$ as

$$H(z) = \frac{(z - e^{j2\pi/5})(z - e^{-j2\pi/5})}{(z - 0.9476e^{j2\pi/5})(z - 0.9476e^{-j2\pi/5})} = \frac{z^2 - 0.618z + 1}{z^2 - 0.5857 + 0.898}$$

The pole-zero pattern and magnitude spectrum of this filter are shown in Figure E18.7C.

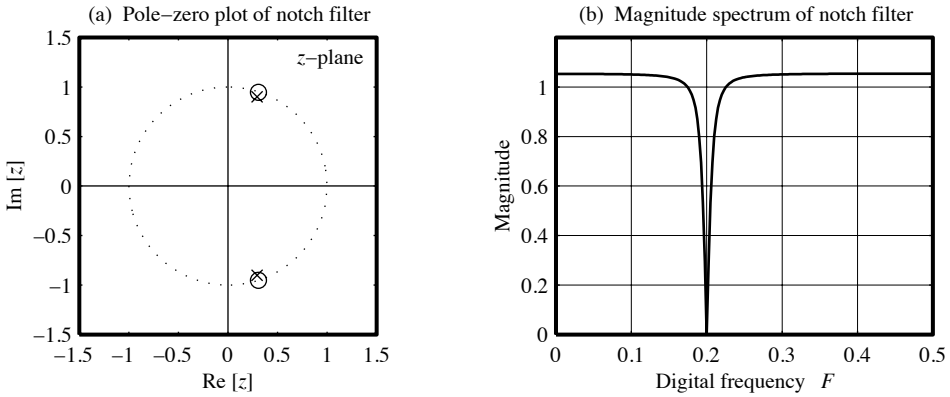


Figure E18.7C Frequency response of the bandstop filter for Example 18.7(c)

18.5 Application-Oriented Examples

In this section we describe several applications of the z-transform and the frequency response to both the analysis of digital filters and to digital filter design.

18.5.1 Equalizers

Audio equalizers are typically used to tailor the sound to suit the taste of the listener. The most common form of equalization is the tone controls (for bass and treble, for example) found on most low-cost audio systems. **Tone controls** employ *shelving filters* that boost or cut the response over a selected frequency band while leaving the rest of the spectrum unaffected (with unity gain). As a result, the filters for the various controls are typically connected *in cascade*. **Graphic equalizers** offer the next level in sophistication and employ a bank of (typically, second-order) bandpass filters covering a fixed number of frequency bands, and with a fixed bandwidth and center frequency for each range. Only the gain of each filter can be adjusted by the user. Each filter isolates a selected frequency range and provides almost zero gain elsewhere. As a result, the individual sections are connected *in parallel*. **Parametric equalizers** offer the ultimate in versatility and comprise filters that allow the user to vary not only the gain but also the filter parameters (such as the cutoff frequency, center frequency, and bandwidth). Each filter in a parametric equalizer affects only a selected portion of the spectrum (providing unity gain elsewhere), and as a result, the individual sections are connected *in cascade*.

PGR2026-00033
Krisp Technologies EX1024 Page 663

18.5.2 Shelving Filters

To vary the bass and treble content, it is common to use first-order *shelving filters* with adjustable gain and cutoff frequency. The transfer function of a first-order lowpass filter whose dc gain is unity and whose response goes to zero at $F = 0.5$ (or $z = -1$) is described by

$$H_{\mathrm{LP}}(z) = \left(\frac{1-\alpha}{2}\right)\frac{z+1}{z-\alpha} \tag{18.14}$$

The half-power or 3-dB cutoff frequency Ω_C of this filter is given by

$$\Omega_C = \cos^{-1}\left(\frac{2\alpha}{1+\alpha^2}\right) \tag{18.15}$$

The transfer function of a first-order highpass filter with the same cutoff frequency Ω_C is simply $H_{\mathrm{HP}}(z) = 1 - H_{\mathrm{LP}}(z)$, and gives

$$H_{\mathrm{HP}}(z) = \left(\frac{1+\alpha}{2}\right)\frac{z-1}{z-\alpha} \tag{18.16}$$

Its gain equals zero at $F = 0$ and unity at $F = 0.5$.

A *lowpass shelving filter* consists of a first-order lowpass filter with adjustable gain G in parallel with a highpass filter. With $H_{\mathrm{LP}}(z) + H_{\mathrm{HP}}(z) = 1$, its transfer function may be written as

$$H_{\mathrm{SL}} = GH_{\mathrm{LP}}(z) + H_{\mathrm{HP}}(z) = 1 + (G-1)H_{\mathrm{LP}}(z) \tag{18.17}$$

A *highpass shelving filter* consists of a first-order highpass filter with adjustable gain G in parallel with a lowpass filter. With $H_{\mathrm{LP}}(z) + H_{\mathrm{HP}}(z) = 1$, its transfer function may be written as

$$H_{\mathrm{SH}} = GH_{\mathrm{HP}}(z) + H_{\mathrm{LP}}(z) = 1 + (G-1)H_{\mathrm{HP}}(z) \tag{18.18}$$

The realizations of these lowpass and highpass shelving filters are shown in Figure 18.8.

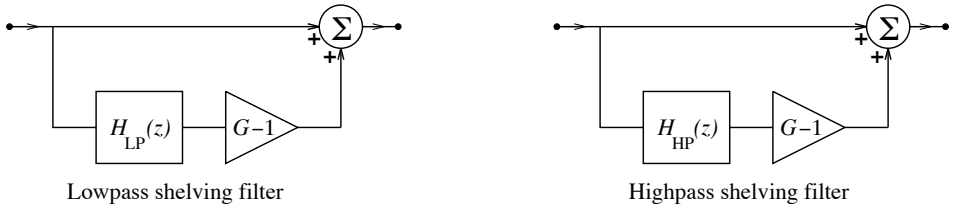


Figure 18.8 Realizations of lowpass and highpass shelving filters

The response of a lowpass and highpass shelving filter is shown for various values of gain (and a fixed α) in Figure 18.9. For $G > 1$, the lowpass shelving filter provides a low-frequency boost, and for $0 < G < 1$ it provides a low-frequency cut. For $G = 1$, we have $H_{\mathrm{SL}} = 1$, and the gain is unity for all frequencies. Similarly, for $G > 1$, the highpass shelving filter provides a high-frequency boost, and for $0 < G < 1$ it provides a high-frequency cut. In either case, the parameter α allows us to adjust the cutoff frequency. Practical realizations of shelving filters and parametric equalizers typically employ allpass structures.

PGR2026-00033
Krisp Technologies EX1024 Page 664

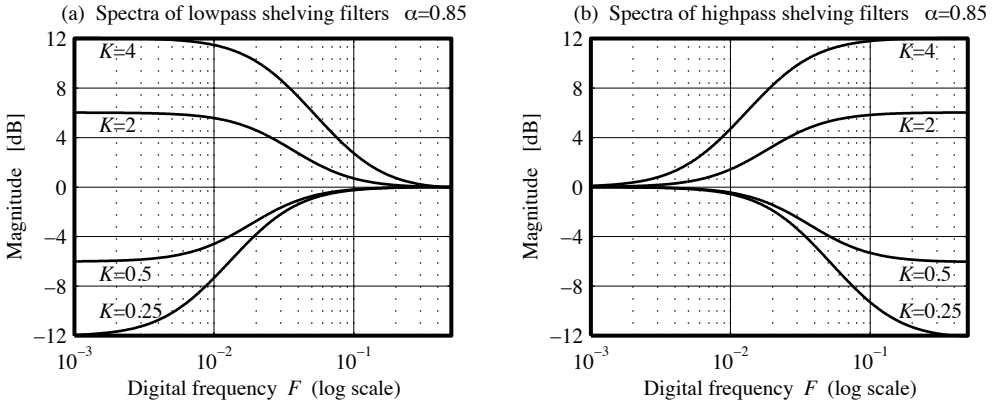


Figure 18.9 Frequency response of lowpass and highpass shelving filters

18.5.3 Digital Oscillators

The impulse response and transfer function of a causal system that generates a pure cosine are

$$h[n] = \cos(n\Omega)u[n] \qquad H(z) = \frac{z^2 - z\cos\Omega}{z^2 - 2z\cos\Omega + 1} \tag{18.19}$$

Similarly, a system whose impulse response is a pure sine is given by

$$h[n] = \sin(n\Omega)u[n] \qquad H(z) = \frac{z\sin\Omega}{z^2 - 2z\cos\Omega + 1} \tag{18.20}$$

The realizations of these two systems, called *digital oscillators*, are shown in Figure 18.10.

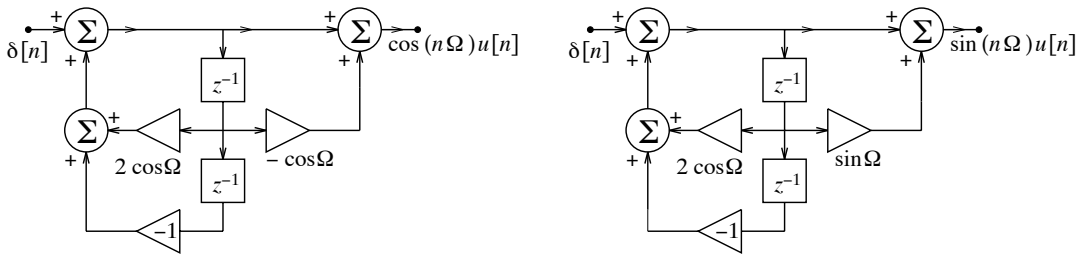


Figure 18.10 Realization of cosine and sine digital oscillators

18.5.4 DTMF Receivers

A touch-tone phone or *dual-tone multi-frequency* (DTMF) transmitter/receiver makes use of digital oscillators to generate audible tones by pressing buttons on a keypad, as shown in Figure 18.11. Pressing a button produces a two-tone signal containing a high- and a low-frequency tone. Each button is associated with a unique pair of low-frequency and high-frequency tones. For example, pressing the button marked 5

PGR2026-00033
Krisp Technologies EX1024 Page 665

would generate a combination of 770-Hz and 1336-Hz tones. There are four low frequencies and four high frequencies. The low- and high-frequency groups have been chosen to ensure that the paired combinations do not interfere with speech. The highest frequency (1633 Hz) is not currently in commercial use. The tones can be generated by using a parallel combination of two programmable digital oscillators, as shown in Figure 18.12.

The sampling rate typically used is $S = 8$ kHz. The digital frequency corresponding to a typical high-frequency tone f_H is $\Omega_H = 2\pi f_H/S$. The code for each button selects the appropriate filter coefficients.

The keys pressed are identified at the receiver by first separating the low- and high-frequency groups using a lowpass filter (with a cutoff frequency of around 1000 Hz) and a highpass filter (with a cutoff frequency of around 1200 Hz) in parallel, and then isolating each tone, using a parallel bank of narrowband bandpass filters tuned to the (eight) individual frequencies. The (eight) outputs are fed to a level detector and decision logic that establishes the presence or absence of a tone. The keys may also be identified by computing the FFT of the tone signal, followed by threshold detection.

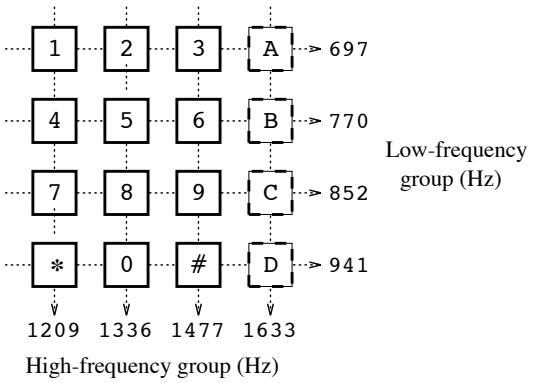


Figure 18.11 Layout of a DTMF touch-tone keypad

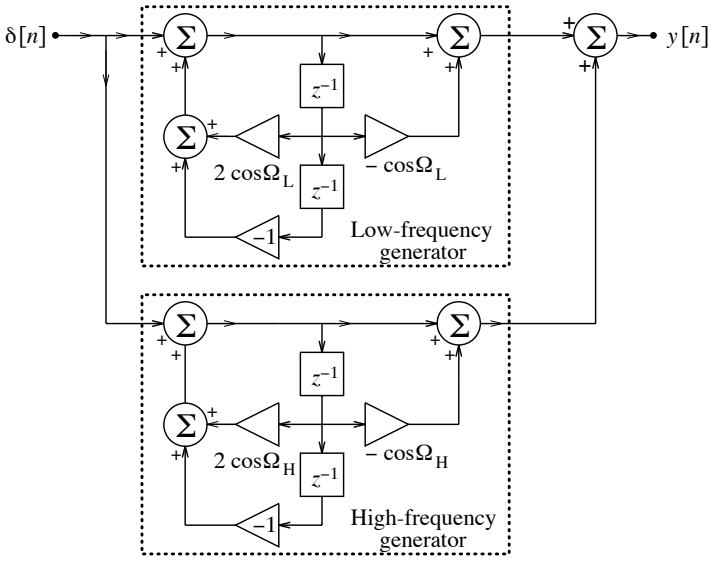


Figure 18.12 Digital oscillators for DTMF tone generation

PGR2026-00033
Krisp Technologies EX1024 Page 666

18.5.5 Digital Resonators

A digital resonator is essentially a narrowband bandpass filter. One way to realize a second-order resonator with a peak at Ω_0 is to place a pair of poles, with angular orientation $\pm\Omega_0$ (at $z = Re^{j\Omega_0}$ and $z = Re^{-j\Omega_0}$), and a pair of zeros at $z = 0$, as shown in Figure 18.13. To ensure stability, we choose the pole radius R to be less than (but close to) unity.

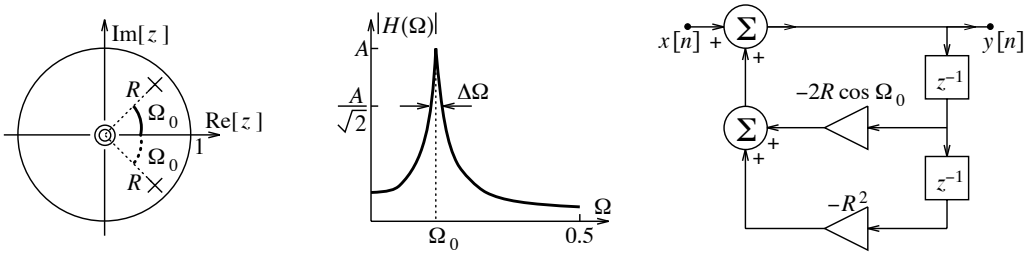


Figure 18.13 A second-order digital resonator

The transfer function of such a digital resonator is

$$H(z) = \frac{z^2}{(z - Re^{j\Omega_0})(z - Re^{-j\Omega_0})} = \frac{z^2}{z^2 - 2zR\cos\Omega_0 + R^2} \tag{18.21}$$

Its magnitude spectrum $|H(\Omega)|$ and realization are also shown in Figure 18.13. The magnitude squared function $|H(\Omega)|^2$ is given by

$$|H(\Omega)|^2 = \left| \frac{e^{j2\Omega}}{(e^{j\Omega} - Re^{j\Omega_0})(e^{j\Omega} - Re^{-j\Omega_0})} \right|^2 \tag{18.22}$$

This simplifies to

$$|H(\Omega)|^2 = \frac{1}{[1 - 2R\cos(\Omega - \Omega_0) + R^2][1 - 2R\cos(\Omega + \Omega_0) + R^2]} \tag{18.23}$$

Its peak value A occurs at (or very close to) the resonant frequency, and is given by

$$A^2 = |H(\Omega_0)|^2 = \frac{1}{(1 - R)^2(1 - 2R\cos 2\Omega_0 + R^2)} \tag{18.24}$$

The half-power bandwidth $\Delta\Omega$ is found by locating the frequencies at which $|H(\Omega)|^2 = 0.5|H(\Omega_0)|^2 = 0.5A^2$, to give

$$\frac{1}{[1 - 2R\cos(\Omega - \Omega_0) + R^2][1 - 2R\cos(\Omega + \Omega_0) + R^2]} = \frac{0.5}{(1 - R)^2(1 - 2R\cos 2\Omega_0 + R^2)} \tag{18.25}$$

Since the half-power frequencies are very close to Ω_0, we have $\Omega + \Omega_0 \approx 2\Omega_0$ and $\Omega - \Omega_0 \approx 0.5\Delta\Omega$, and the above relation simplifies to

$$1 - 2R\cos(0.5\Delta\Omega) + R^2 = 2(1 - R)^2 \tag{18.26}$$

This relation between the pole radius R and bandwidth $\Delta\Omega$ can be simplified when the poles are very close to the unit circle (with $R > 0.9$ or so). The values of R and the peak magnitude A are then reasonably well approximated by

$$R \approx 1 - 0.5\Delta\Omega \qquad\qquad A \approx \frac{1}{(1 - R^2)\sin\Omega_0} \tag{18.27}$$

PGR2026-00033
Krisp Technologies EX1024 Page 667

The peak gain can be normalized to unity by dividing the transfer function by A. The impulse response of this filter can be found by partial fraction expansion of $H(z)/z$, which gives

$$\frac{H(z)}{z} = \frac{z}{(z - Re^{j\Omega_0})(z - Re^{-j\Omega_0})} = \frac{K}{z - Re^{j\Omega_0}} + \frac{K^*}{z - Re^{-j\Omega_0}} \tag{18.28}$$

where

$$K = \left.\frac{z}{z - Re^{-j\Omega_0}}\right|_{z=Re^{j\Omega_0}} = \frac{Re^{j\Omega_0}}{Re^{j\Omega_0} - Re^{-j\Omega_0}} = \frac{Re^{-j\Omega_0}}{2jR\sin\Omega_0} = \frac{e^{j(\Omega_0 - \pi/2)}}{2\sin\Omega_0} \tag{18.29}$$

Then, from lookup tables, we obtain

$$h[n] = \frac{1}{\sin\Omega_0} R^n \cos(n\Omega_0 + \Omega_0 - \tfrac{\pi}{2})u[n] = \frac{1}{\sin\Omega_0} R^n \sin[(n+1)\Omega_0]u[n] \tag{18.30}$$

To null the response at low and high frequencies ($\Omega = 0$ and $\Omega = \pi$), the two zeros in $H(z)$ may be relocated from the origin to $z = 1$ and $z = -1$, and this leads to the modified transfer function $H_1(z)$:

$$H_1(z) = \frac{z^2 - 1}{z^2 - 2zR\cos\Omega_0 + R^2} \tag{18.31}$$

EXAMPLE 18.8 (Digital Resonator Design)

We design a digital resonator with a peak gain of unity at 50 Hz, and a 3-dB bandwidth of 6 Hz, assuming a sampling frequency of 300 Hz.

The digital resonant frequency is $\Omega_0 = \frac{2\pi(50)}{300} = \frac{\pi}{3}$. The 3-dB bandwidth is $\Delta\Omega = \frac{2\pi(6)}{300} = 0.04\pi$. We compute the pole radius as $R = 1 - 0.5\Delta\Omega = 1 - 0.02\pi = 0.9372$. The transfer function of the digital resonator is thus

$$H(z) = \frac{Gz^2}{(z - Re^{j\Omega_0})(z - Re^{-j\Omega_0})} = \frac{Gz^2}{z^2 - 2R\cos\Omega_0 + R^2} = \frac{Gz^2}{z^2 - 0.9372z + 0.8783}$$

For a peak gain of unity, we choose $G = \frac{1}{A} = (1 - R^2)\sin\Omega_0 = 0.1054$. Thus,

$$H(z) = \frac{0.1054z^2}{z^2 - 0.9372z + 0.8783}$$

The magnitude spectrum and passband detail of this filter are shown in Figure E18.8.

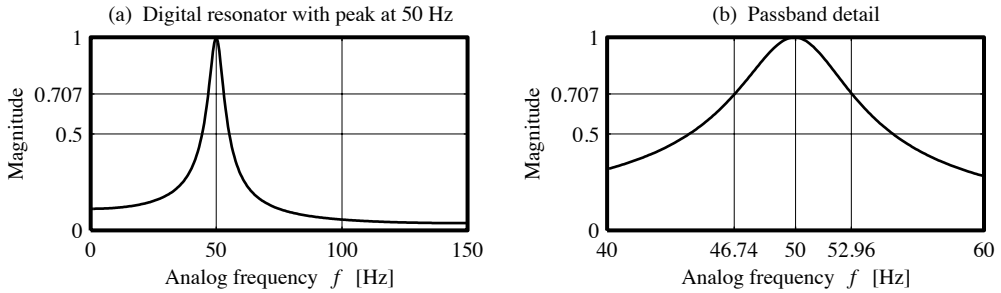


Figure E18.8 Frequency response of the digital resonator for Example 18.8

The passband detail reveals that the half-power frequencies are located at 46.74 Hz and 52.96 Hz, a close match to the bandwidth requirement (of 6 Hz).

PGR2026-00033
Krisp Technologies EX1024 Page 668

18.5.6 Comb Filters

A **comb filter** has a magnitude response that looks much like the rounded teeth of a comb. One such comb filter is described by the transfer function

$$H(z) = 1 - z^{-N} \qquad h[n] = \{1, \underbrace{0,\ 0,\ \ldots,\ 0,\ 0}_{N-1 \text{ zeros}},\ -1\} \tag{18.32}$$

This corresponds to the system difference equation $y[n] = x[n] - x[n-N]$, and represents a linear-phase FIR filter whose impulse response $h[n]$ (which is *odd symmetric* about its midpoint) has $N+1$ samples with $h[0] = 1$, $h[N] = -1$, and all other coefficients zero. There is a pole of multiplicity N at the origin, and the zeros lie on a unit circle with locations specified by

$$z^{-N} = e^{-jN\Omega} = 1 = e^{j2\pi k} \qquad \Omega_k = \frac{2k\pi}{N}, \quad k = 0, 1, \ldots, N-1 \tag{18.33}$$

The pole-zero pattern and magnitude spectrum of this filter are shown for two values of N in Figure 18.14.

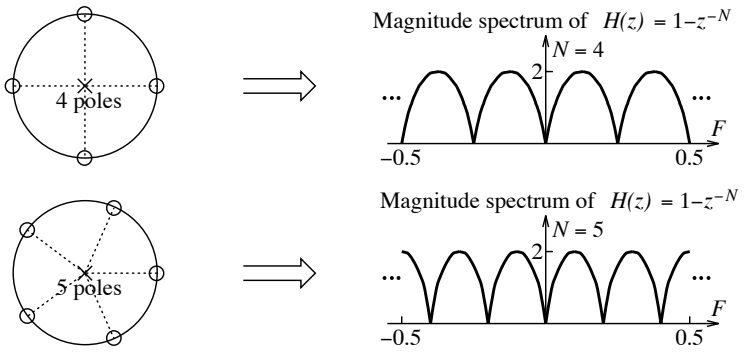


Figure 18.14 Pole-zero plot and frequency response of the comb filter $H(z) = 1 - z^{-N}$

The zeros of $H(z)$ are uniformly spaced $2\pi/N$ radians apart around the unit circle, starting at $\Omega = 0$. For even N, there is also a zero at $\Omega = \pi$. Being an FIR filter, it is always stable for any N. Its frequency response is given by

$$H(F) = 1 - e^{-j2\pi FN} \tag{18.34}$$

Note that $H(0)$ always equals 0, but $H(0.5) = 0$ for even N and $H(0.5) = 2$ for odd N. The frequency response $H(F)$ looks like a comb with N rounded *teeth* over its principal period $-0.5 \le F \le 0.5$.

A more general form of this comb filter is described by

$$H(z) = 1 - \alpha z^{-N} \qquad h[n] = \{1, \underbrace{0,\ 0,\ \ldots,\ 0,\ 0}_{N-1 \text{ zeros}},\ -\alpha\} \tag{18.35}$$

It has N poles at the origin, and its zeros are uniformly spaced $2\pi/N$ radians apart around a circle of radius $R = \alpha^{1/N}$, starting at $\Omega = 0$. Note that this filter is no longer a linear-phase filter because its impulse response is not symmetric about its midpoint. Its frequency response $H(F) = 1 - \alpha e^{-j2\pi FN}$ suggests that $H(0) = 1 - \alpha$ for any N, and $H(0.5) = 1 - \alpha$ for even N and $H(0.5) = 1 + \alpha$ for odd N. Thus, its magnitude varies between $1 - \alpha$ and $1 + \alpha$, as illustrated in Figure 18.15.

PGR2026-00033
Krisp Technologies EX1024 Page 669

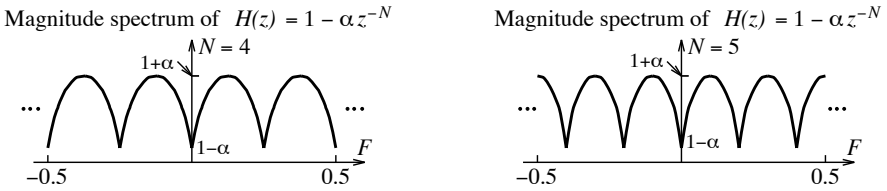


Figure 18.15 Frequency response of the comb filter $H(z) = 1 - \alpha z^{-N}$

Another FIR comb filter is described by the transfer function

$$H(z) = 1 + z^{-N} \qquad h[n] = \{1, \underbrace{0,\ 0,\ \ldots,\ 0,\ 0}_{N-1 \text{ zeros}},\ 1\} \tag{18.36}$$

This corresponds to the system difference equation $y[n] = x[n] + x[n-N]$, and represents a linear-phase FIR filter whose impulse response $h[n]$ (which is *even symmetric* about its midpoint) has $N+1$ samples with $h[0] = h[N] = 1$, and all other coefficients zero. There is a pole of multiplicity N at the origin, and the zero locations are specified by

$$z^{-N} = e^{-jN\Omega} = -1 = e^{j(2k+1)\pi} \qquad \Omega_k = \frac{(2k+1)\pi}{N}, \quad k = 0, 1, \ldots, N-1 \tag{18.37}$$

The pole-zero pattern and magnitude spectrum of this filter are shown for two values of N in Figure 18.16.

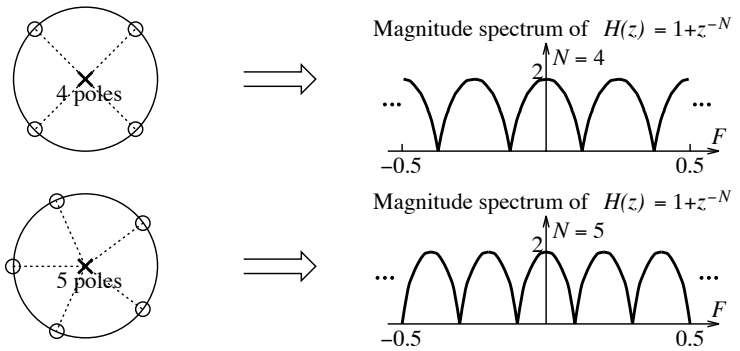


Figure 18.16 Pole-zero plot and frequency response of the comb filter $H(z) = 1 + z^{-N}$

The zeros of $H(z)$ are uniformly spaced $2\pi/N$ radians apart around the unit circle, starting at $\Omega = \pi/N$. For odd N, there is also a zero at $\Omega = \pi$. Its frequency response is given by

$$H(F) = 1 + e^{-j2\pi FN} \tag{18.38}$$

Note that $H(0) = 2$ for any N, but $H(0.5) = 2$ for even N and $H(0.5) = 0$ for odd N.

A more general form of this comb filter is described by

$$H(z) = 1 + \alpha z^{-N} \qquad h[n] = \{1, \underbrace{0,\ 0,\ \ldots,\ 0,\ 0}_{N-1 \text{ zeros}},\ \alpha\} \tag{18.39}$$

PGR2026-00033
Krisp Technologies EX1024 Page 670

It has N poles at the origin and its zeros are uniformly spaced $2\pi/N$ radians apart around a circle of radius $R = \alpha^{1/N}$, starting at $\Omega = \pi/N$. Note that this filter is no longer a linear-phase filter because its impulse response is not symmetric about its midpoint. Its frequency response $H(F) = 1 + \alpha e^{-j2\pi FN}$ suggests that $H(0) = 1+\alpha$ for any N, and $H(0.5) = 1+\alpha$ for even N and $H(0.5) = 1-\alpha$ for odd N. Thus, its magnitude varies between $1-\alpha$ and $1+\alpha$, as illustrated in Figure 18.17.

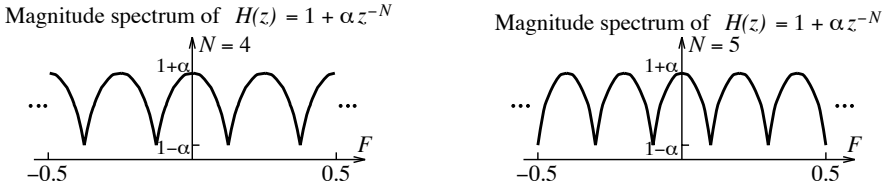


Figure 18.17 Frequency response of the comb filter $H(z) = 1 + \alpha z^{-N}$

18.5.7 Periodic Notch Filters

A **notch filter** with notches at periodic intervals may be constructed by placing zeros on the unit circle at the notch frequencies, and poles in close proximity along the same angular orientations. One such form is

$$H(z) = \frac{N(z)}{N(z/\alpha)} = \frac{1 - z^{-N}}{1 - (z/\alpha)^{-N}} \tag{18.40}$$

Its zeros are uniformly spaced $2\pi/N$ radians apart around a unit circle, starting at $\Omega = 0$, and its poles are along the same angular orientations but on a circle of radius $R = \alpha$. For stability, we require $\alpha < 1$. Of course, the closer α is to unity, the sharper are the notches and the more constant is the response at the other frequencies. Periodic notch filters are often used to remove unwanted components at the power line frequency and its harmonics.

EXAMPLE 18.9 (Periodic Notch Filter Design)

We design a notch filter to filter out 60 Hz and its harmonics from a signal sampled at $S = 300$ Hz.

The digital frequency corresponding to 60 Hz is $F = 60/300 = 0.2$ or $\Omega = 0.4\pi$. There are thus $N = 2\pi/0.4\pi = 5$ notches in the principal range $\pi \le \Omega \le \pi$ (around the unit circle), and the transfer function of the notch filter is

$$H(z) = \frac{N(z)}{N(z/\alpha)} = \frac{1 - z^{-5}}{1 - (z/\alpha)^{-5}} = \frac{z^5 - 1}{z^5 - \alpha^5}$$

Figure E18.9(a) shows the response of this filter for $\alpha = 0.9$ and $\alpha = 0.99$. The choice of α, the pole radius, is arbitrary (as long as $\alpha < 1$), but should be close to unity for sharp notches.

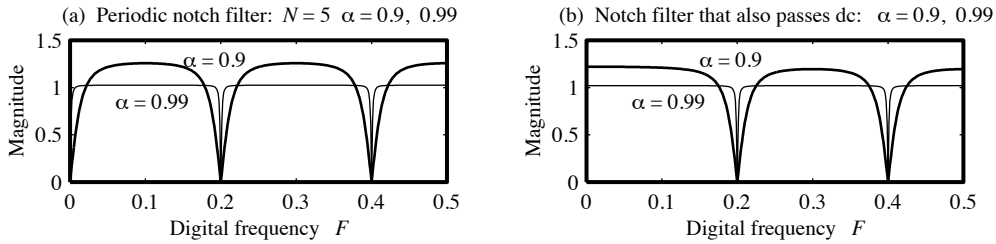


Figure E18.9 Frequency response of the notch filters for Example 18.9

PGR2026-00033
Krisp Technologies EX1024 Page 671

Comment: This notch filter also removes the dc component. If we want to preserve the dc component, we must extract the zero at $z = 1$ (corresponding to $F = 0$) from $N(z)$ (by long division), to give

$$N_1(z) = \frac{1 - z^{-5}}{1 - z^{-1}} = 1 + z^{-1} + z^{-2} + z^{-3} + z^{-4}$$

and use $N_1(z)$ to compute the new transfer function $H_1(z) = N_1(z)/D(z)$ as

$$H_1(z) = \frac{N_1(z)}{N_1(z/\alpha)} = \frac{1 + z^{-1} + z^{-2} + z^{-3} + z^{-4}}{1 + (z/\alpha)^{-1} + (z/\alpha)^{-2} + (z/\alpha)^{-3} + (z/\alpha)^{-4}} = \frac{z^4 + z^3 + z^2 + z + 1}{z^4 + \alpha z^3 + \alpha^2 z^2 + \alpha^3 z + \alpha^4}$$

Figure E18.9(b) compares the response of this filter for $\alpha = 0.9$ and $\alpha = 0.99$, and reveals that the dc component is indeed preserved by this filter.

18.6 Allpass Filters

An **allpass filter** is characterized by a magnitude response that is unity for all frequencies, with $|H(F)| = 1$. Its transfer function $H(z)$ also satisfies the relationship

$$H(z)H(1/z) = 1 \tag{18.41}$$

This implies that each pole of an allpass filter is paired by a conjugate reciprocal zero. As a result, allpass filters cannot be minimum-phase. The cascade of allpass filters is also an allpass filter. An allpass filter of order N has a numerator and denominator of equal order N, with coefficients in reversed order:

$$H_{\text{AP}}(z) = \frac{N(z)}{D(z)} = \frac{C_N + C_{N-1}z^{-1} + \cdots + C_1 z^{N-1} + z^{-N}}{1 + C_1 z^{-1} + C_2 z^{-2} + \cdots + C_N z^{-N}} \tag{18.42}$$

Note that $D(z) = z^{-N}N(1/z)$, and as a result, if the roots of $N(z)$ (the zeros) are at r_k, the roots of $D(z)$ (the poles) are at $1/r_k$, the reciprocal locations.

REVIEW PANEL 18.7

How to Identify an Allpass Filter

From $H(z) = N(z)/D(z)$: The coefficients of $N(z)$ and $D(z)$ appear in reversed order.
From pole-zero plot: Each pole is paired with a conjugate reciprocal zero.

Consider a stable, first-order allpass filter whose transfer function $H(z)$ and frequency response $H(F)$ are described by

$$H(z) = \frac{1 + \alpha z}{z + \alpha} \qquad H_A(F) = \frac{1 + \alpha e^{j2\pi F}}{\alpha + e^{j2\pi F}}, \quad |\alpha| < 1 \tag{18.43}$$

If we factor out $e^{j\pi F}$ from the numerator and denominator of $H(F)$, we obtain the form

$$H(F) = \frac{1 + \alpha e^{j2\pi F}}{\alpha + e^{j2\pi F}} = \frac{e^{-j\pi F} + \alpha e^{j\pi F}}{\alpha e^{-j\pi F} + e^{j\pi F}}, \quad |\alpha| < 1 \tag{18.44}$$

The numerator and denominator are complex conjugates. This implies that their magnitudes are equal (an allpass characteristic) and that the phase of $H(F)$ equals twice the numerator phase. Now, the numerator may be simplified to

$$e^{-j\pi F} + \alpha e^{j\pi F} = \cos(\pi F) - j\sin(\pi F) + \alpha\cos(\pi F) + j\alpha\sin(\pi F) = (1 + \alpha)\cos(\pi F) - j(1 - \alpha)\sin(\pi F)$$

PGR2026-00033
Krisp Technologies EX1024 Page 672

The phase $\phi(F)$ of the allpass filter $H(F)$ equals twice the numerator phase. The phase $\phi(F)$ and phase delay $t_p(F)$ may then be written as

$$\phi(F) = -2\tan^{-1}\left[\frac{1-\alpha}{1+\alpha}\tan(\pi F)\right] \qquad t_p(F) = -\frac{\phi(F)}{2\pi F} = \frac{1}{\pi F}\tan^{-1}\left[\frac{1-\alpha}{1+\alpha}\tan(\pi F)\right] \tag{18.45}$$

A low-frequency approximation for the phase delay is given by

$$t_p(F) = \frac{1-\alpha}{1+\alpha} \qquad \text{(low-frequency approximation)} \tag{18.46}$$

The group delay $t_g(F)$ of this allpass filter equals

$$t_g(F) = -\frac{1}{2\pi}\frac{d\phi(F)}{dF} = \frac{1-\alpha^2}{1+2\alpha\cos(2\pi F)+\alpha^2} \tag{18.47}$$

At $F = 0$ and $F = 0.5$, the group delay is given by

$$t_g(0) = \frac{1-\alpha^2}{1+2\alpha+\alpha^2} = \frac{1-\alpha}{1+\alpha} \qquad t_g(0.5) = \frac{1-\alpha^2}{1-2\alpha+\alpha^2} = \frac{1+\alpha}{1-\alpha} \tag{18.48}$$

18.6.1 Stabilization of Unstable Filters

Allpass filters are often used to stabilize unstable digital filters, while preserving their magnitude response. Consider the unstable filter

$$H_u(z) = \frac{z}{1+\alpha z}, \qquad |\alpha| < 1 \tag{18.49}$$

It has a pole at $z = -1/\alpha$ and is thus unstable. If we cascade $H_u(z)$ with a first-order allpass filter whose transfer function is $H_1(z) = (1+\alpha z)/(z+\alpha^*)$, we obtain

$$H(z) = H_u(z)H_1(z) = \frac{z}{z+\alpha^*} \tag{18.50}$$

The filter $H(z)$ has a pole at $z = -\alpha^*$, and since $|\alpha| < 1$, it is a stable filter.

This idea can easily be extended to unstable filters of arbitrary order described as a cascade of a stable portion $H_s(z)$ and P unstable first-order sections:

$$H_u(z) = H_s(z)\prod_{m=1}^{P}\frac{z}{1+\alpha_m z}, \qquad |\alpha_m| < 1 \tag{18.51}$$

To stabilize $H_u(z)$, we use an allpass filter $H_{\text{AP}}(z)$ that is a cascade of P allpass sections. Its form is

$$H_{\text{AP}}(z) = \prod_{m=1}^{P}\frac{1+\alpha_m z}{z+\alpha_m^*}, \qquad |\alpha_m| < 1 \tag{18.52}$$

The stabilized filter $H(z)$ is then described by the cascade $H_u(z)H_{\text{AP}}(z)$ as

$$H(z) = H_u(z)H_{\text{AP}}(z) = H_s(z)\prod_{m=1}^{P}\frac{z}{z+\alpha_m^*}, \qquad |\alpha_m| < 1 \tag{18.53}$$

There are two advantages to this method. First, the magnitude response of the original filter is unchanged. And second, the order of the new filter is the same as the original. The reason for the *inequality* $|\alpha_m| < 1$ (rather than $|\alpha_m| \le 1$) is that if $H_u(z)$ has a pole on the unit circle, its conjugate reciprocal will also lie on the unit circle and no stabilization is possible.

PGR2026-00033
Krisp Technologies EX1024 Page 673

18.6.2 Minimum-Phase Filters Using Allpass Filters

Even though allpass filters cannot be minimum-phase, they can be used to convert stable nonminimum-phase filters to stable minimum-phase filters. We describe a nonminimum-phase transfer function by a cascade of a minimum-phase part $H_M(z)$ (with all poles and zeros inside the unit circle) and a portion with zeros outside the unit circle such that

$$H_{\text{NM}}(z) = H_M(z) \prod_{m=1}^{P} (z + \alpha_M), \qquad |\alpha_m| > 1 \tag{18.54}$$

We now seek an *unstable* allpass filter with

$$H_{\text{AP}}(z) = \prod_{m=1}^{P} \frac{1 + z\alpha_m^*}{z + \alpha_m}, \qquad |\alpha_m| > 1 \tag{18.55}$$

The cascade of $H_{\text{NM}}(z)$ and $H_{\text{AP}}(z)$ yields a minimum-phase filter with

$$H(z) = H_{\text{NM}}(z)H_{\text{AP}}(z) = H_M(z) \prod_{m=1}^{P} (1 + z\alpha_m^*) \tag{18.56}$$

Once again, $H(z)$ has the same order as the original filter.

18.7 Application-Oriented Examples: Digital Audio Effects

Signal processing techniques allow us not only to compensate for the acoustics of a listening environment, but also to add special musical effects such as delay, echo, and reverb, if desired. The most common example is the audio equalizer that allows us to enhance the bass or treble content of an audio signal. What a listener hears in a room depends not only on the acoustic characteristics of that room, but also on the location of the listener. A piece of music recorded in a concert hall does not sound quite the same elsewhere. The ambience of a concert hall can, however, be duplicated in the living room environment by first compensating for the room acoustics, and then adding the reverberation characteristics of the concert hall. The compensation is achieved by passing the audio signal through an inverse filter corresponding to the impulse response of the room (which can be measured). The concert hall ambiance is then added by a filter that simulates the concert hall acoustics. High-end audio systems allow us not only to measure the room acoustics, but also to simulate the acoustics of a variety of well-known concert hall environments.

In any listening space, what we hear from an audio source consists not only of the direct sound, but also early echoes reflected directly from the walls and other structures, and a late sound or reverberation that describes multiple reflections (that get added to other echoes). This is illustrated in Figure 18.18.

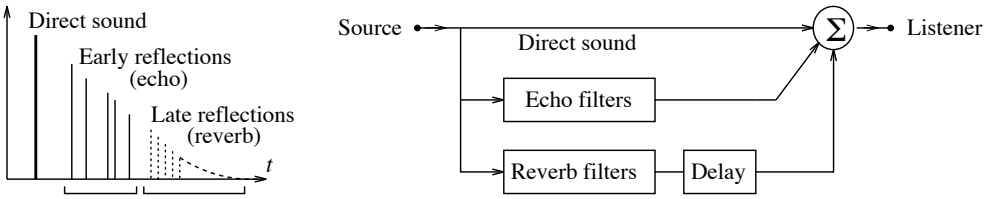


Figure 18.18 Illustrating echo and reverb

The direct sound provides clues to the location of the source, the early echoes provide an indication of the physical size of the listening space, and the reverberation characterizes the warmth and liveliness that

PGR2026-00033
Krisp Technologies EX1024 Page 674

we usually associate with sounds. The amplitude of the echoes and reverberation decays exponentially with time. Together, these characteristics determine the psycho-acoustic qualities we associate with any perceived sound. Typical 60-dB reverberation times (for the impulse response to decay to 0.001 of its peak value) for concert halls are fairly long, up to two seconds. A conceptual model of a listening environment, also shown in Figure 18.18, consists of echo filters and reverb filters.

A single echo can be modeled by a feed-forward system of the form

$$y[n] = x[n] + \alpha x[n-D] \qquad H(z) = 1 + \alpha z^{D} \qquad h[n] = \delta[n] - \alpha\delta[n-D] \tag{18.57}$$

This is just a comb filter in disguise. The zeros of this filter lie on a circle of radius $R = \alpha^{1/D}$, with angular orientations of $\Omega = (2k+1)\pi/D$. Its comb-like magnitude spectrum $H(F)$ shows minima of $1-\alpha$ at the frequencies $F = (2k+1)/D$, and peaks of $1+\alpha$ midway between the dips. To perceive an echo, the index D must correspond to a delay of at least about 50 ms.

A reverb filter that describes multiple reflections has a feedback structure of the form

$$y[n] = \alpha y[n-D] + x[n] \qquad H(z) = \frac{1}{1-\alpha z^{-D}} \tag{18.58}$$

This filter has an *inverse-comb* structure, and its poles lie on a circle of radius $R = \alpha^{1/D}$, with an angular separation of $\Omega = 2\pi/D$. The magnitude spectrum shows peaks of $1/(1-\alpha)$ at the pole frequencies $F = k/D$, and minima of $1/(1+\alpha)$ midway between the peaks.

Conceptually, the two systems just described can form the building blocks for simulating the acoustics of a listening space. Many reverb filters actually use a combination of reverb filters and allpass filters. A typical structure is shown in Figure 18.19. In practice, however, it is more of an art than a science to create realistic effects, and many of the commercial designs are propriety information.

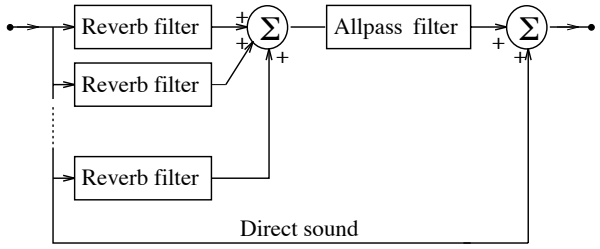


Figure 18.19 Echo and reverb filters for simulating acoustic effects

The reverb filters in Figure 18.19 typically incorporate irregularly spaced delays to allow the blending of echoes, and the allpass filter serves to create the effect of early echoes. Some structures for the reverb filter and allpass filter are shown in Figure 18.20. The first structure is the *plain reverb*. In the second structure, the feedback path incorporates a first-order lowpass filter that accounts for the dependence (increase) of sound absorption with frequency. The allpass filter has the form

$$H(z) = \frac{-\alpha + z^{-L}}{1-\alpha z^{L}} = -\frac{1}{\alpha} + \frac{\frac{1}{\alpha} - \alpha}{1-\alpha z^{-L}} \tag{18.59}$$

The second form of this expression (obtained by long division) shows that, except for the constant term, the allpass filter has the same form as a reverb filter.

PGR2026-00033
Krisp Technologies EX1024 Page 675

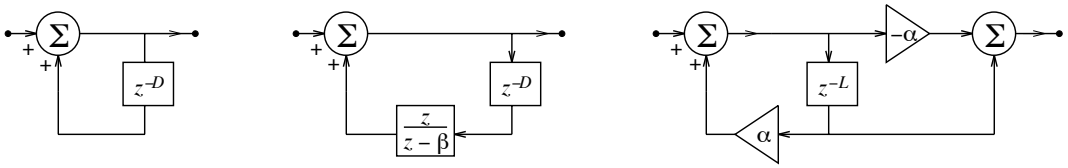


Figure 18.20 Realization of typical reverb filters

18.7.1 Gated Reverb and Reverse Reverb

Two other types of audio effects can also be generated by reverb filters. A **gated reverb** results from the truncation (abrupt or gradual) of the impulse response of a reverb filter, resulting in an FIR filter. A **reverse reverb** is essentially a folded version of the impulse response of a gated reverb filter. The impulse response increases with time, leading to a sound that first gets louder, and then dies out abruptly.

18.7.2 Chorusing, Flanging, and Phasing

The echo filter also serves as the basis for several other audio special effects. Two of these effects, chorusing and phasing, are illustrated in Figure 18.21.

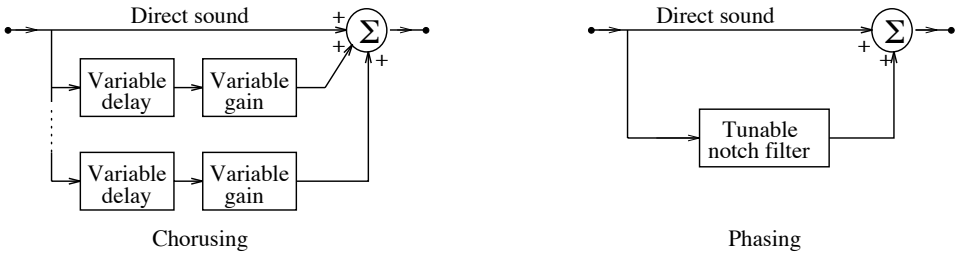


Figure 18.21 Illustrating special audio effects

Chorusing mimics a chorus (or group) singing (or playing) in unison. In practice, of course, the voices (or instruments) are not in perfect synchronization, nor identical in pitch. The chorusing effect can be implemented by a weighted combination of echo filters, each with a time-varying delay d_n of the form

$$y[n] = x[n] + \alpha x[n - d_n] \tag{18.60}$$

Typical delay times used in chorusing are between 20 ms and 30 ms. If the delays are less than 10 ms (but still variable), the resulting "whooshing" sound is known as **flanging**.

Phase shifting or **phasing** also creates many interesting effects, and may be achieved by passing the signal through a notch filter whose frequency can be tuned by the user. It is the sudden phase jumps at the notch frequency that are responsible for the phasing effect. The effects may also be enhanced by the addition of feedback.

18.7.3 Plucked-String Filters

Comb filters, reverb filters, and allpass filters have also been used to synthesize the sounds of plucked instruments, such as the guitar. What we require is a filter that has a comb-like response, with resonances

PGR2026-00033
Krisp Technologies EX1024 Page 676

at multiples of the fundamental frequency of the note. The harmonics should decay exponentially in time, with higher frequencies decaying at a faster rate. A typical structure, first described by Karplus and Strong, is illustrated in Figure 18.22.

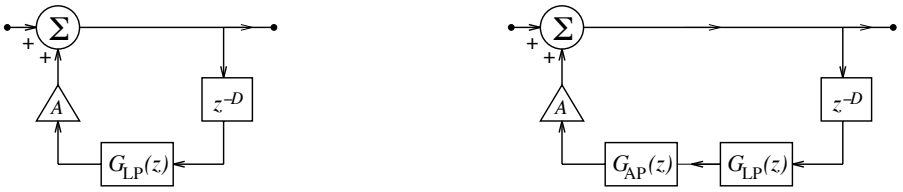


Figure 18.22 The Karplus-Strong plucked-string filter

In the Karplus-Strong structure, the lowpass filter has the transfer function $G_{\text{LP}}(z) = 0.5(1 + z^{-1})$, and contributes a 0.5-sample phase delay. The delay line contributes an additional D-sample delay, and the *loop delay* is thus $D + 0.5$ samples. The overall transfer function of the first structure is

$$H(z) = \frac{1}{1 - Az^{-D}G_{\text{LP}}(z)} \qquad G_{\text{LP}}(z) = 0.5(1 + z^{-1}) \tag{18.61}$$

The frequency response of this Karplus-Strong filter is shown in Figure 18.23, for $D = 8$ and $D = 16$, and clearly reveals a resonant structure with sharp peaks. The lowpass filter $G_{\text{LP}}(z)$ is responsible for the decrease in the amplitude of the peaks, and for making them broader as the frequency increases.

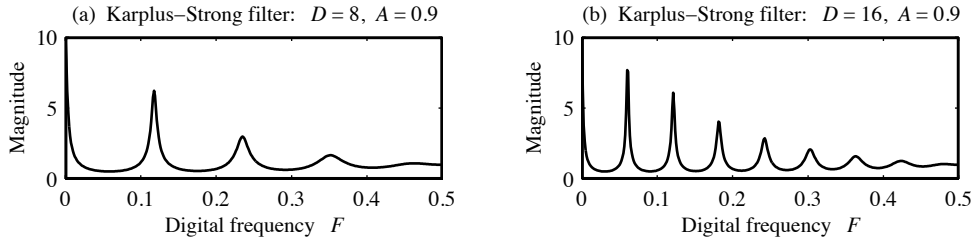


Figure 18.23 Frequency response of the Karplus-Strong plucked-string filter

If A is close to unity, the peaks occur at (or very close to) multiples of the fundamental digital frequency $F_0 = 1/(D + 0.5)$. For a sampling rate S, the note frequency thus corresponds to $f_0 = S/(D + 0.5)$ Hz. However, since D is an integer, we cannot achieve precise tuning (control of the frequency). For example, to generate a 1600-Hz note at a sampling frequency of 14 kHz requires that $D + 0.5 = 8.75$. The closest we can come is by picking $D = 8$ (to give a frequency of $f_0 = \frac{14}{8.5} \approx 1647$ Hz). To generate the exact frequency requires an additional 0.25-sample phase delay. The second structure of Figure 18.22 includes an allpass filter of the form $G_{\text{AP}}(z) = (1+\alpha z)/(z+\alpha)$ in the feedback path that allows us to implement such delays by appropriate choice of the allpass parameter α. For example, to implement the 0.25-sample delay, we would require (using the low-frequency approximation for the phase delay)

$$0.25 = t_p \approx \frac{1 - \alpha}{1 + \alpha} \qquad \alpha = \frac{1 - 0.25}{1 + 0.25} = 0.6 \tag{18.62}$$

PGR2026-00033
Krisp Technologies EX1024 Page 677

CHAPTER 18 PROBLEMS

DRILL AND REINFORCEMENT

18.1 (Realization) Sketch the direct form I, direct form II, and transposed realization for each filter.

(a) $y[n] - \frac{1}{6}y[n-1] - \frac{1}{2}y[n-2] = 3x[n]$ **(b)** $H(z) = \dfrac{z-2}{z^2 - 0.25}$

(c) $y[n] - 3y[n-1] + 2y[n-2] = 2x[n-2]$ **(d)** $H(z) = \dfrac{2z^2 + z - 2}{z^2 - 1}$

18.2 (Realization) Find the transfer function and difference equation for each system realization shown in Figure P18.2.

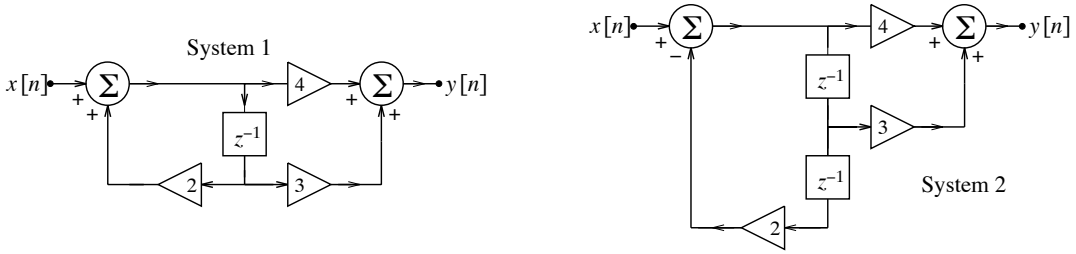


Figure P18.2 Filter realizations for Problem 18.2

18.3 (Inverse Systems) Find the difference equation of the inverse systems for each of the following. Which inverse systems are causal? Which are stable?

(a) $H(z) = (z^2 + \frac{1}{9})/(z^2 - \frac{1}{4})$ **(b)** $H(z) = \dfrac{z+2}{z^2 + 0.25}$

(c) $y[n] - 0.5y[n-1] = x[n] + 2x[n-1]$ **(d)** $h[n] = n(2)^n u[n]$

18.4 (Minimum-Phase Systems) Classify each system as minimum phase, mixed phase, or maximum phase. Which of the systems are stable?

(a) $H(z) = \dfrac{z^2 + \frac{1}{9}}{z^2 - \frac{1}{4}}$ **(b)** $H(z) = \dfrac{z^2 - 4}{z^2 + 9}$

(c) $h[n] = n(2)^n u[n]$ **(d)** $y[n] + y[n-1] + 0.25y[n-2] = x[n] - 2x[n-1]$

18.5 (Minimum-Phase Systems) Find the minimum-phase transfer function corresponding to the systems described by the following:

(a) $H(z)H(1/z) = \dfrac{1}{3z^2 - 10z + 3}$

(b) $H(z)H(1/z) = \dfrac{3z^2 + 10z + 3}{5z^2 + 26z + 5}$

(c) $|H(F)|^2 = \dfrac{1.25 + \cos(2\pi F)}{8.5 + 4\cos(2\pi F)}$

18.6 (System Characteristics) Consider the system $y[n] - \alpha y[n-1] = x[n] - \beta x[n-1]$.

(a) For what values of α and β will the system be stable?

PGR2026-00033
Krisp Technologies EX1024 Page 678

(b) For what values of α and β will the system be minimum-phase?
(c) For what values of α and β will the system be allpass?
(d) For what values of α and β will the system be linear phase?

18.7 (Filter Design by Pole-Zero Placement) Design the following filters by pole-zero placement.

(a) A bandpass filter with a center frequency of $f_0 = 200$ Hz, a 3-dB bandwidth of $\Delta f = 20$ Hz, zero gain at $f = 0$ and $f = 400$ Hz, and a sampling frequency of 800 Hz.
(b) A notch filter with a notch frequency of 1 kHz, a 3-dB stopband of 10 Hz, and sampling frequency 8 kHz.

18.8 (Stabilization by Allpass Filters) The transfer function of a filter is $H(z) = \dfrac{z+3}{z-2}$.

(a) Is this filter stable?
(b) What is the transfer function $A_1(z)$ of a first-order allpass filter that can stabilize this filter? What is the transfer function $H_S(z)$ of the stabilized filter?
(c) If $H_S(z)$ is not minimum phase, pick an allpass filter $A_2(z)$ that converts $H_S(z)$ to a minimum-phase filter $H_M(z)$.
(d) Verify that $|H(F)| = |H_S(F)| = |H_M(F)|$.

REVIEW AND EXPLORATION

18.9 (Realization) Find the transfer function and difference equation for each digital filter realization shown in Figure P18.9.

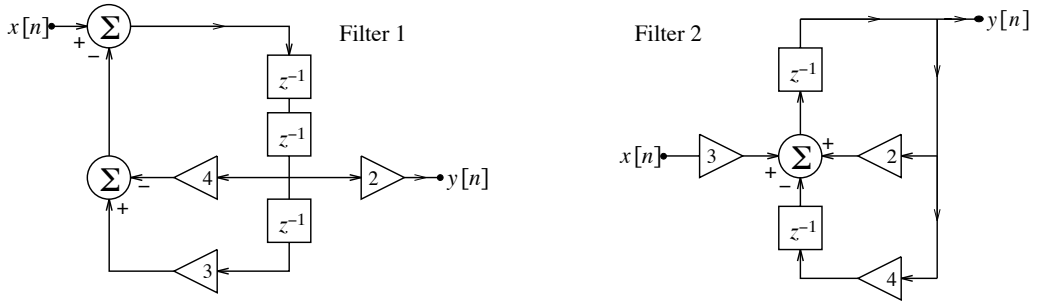


Figure P18.9 Filter realizations for Problem 18.9

18.10 (Feedback Compensation) Feedback compensation is often used to stabilize unstable filters. It is required to stabilize the unstable filter $G(z) = \dfrac{6}{z-1.2}$ by putting it in the forward path of a negative feedback system. The feedback block has the form $H(z) = \dfrac{\alpha}{z-\beta}$.

(a) What values of α and β are required for the overall system to have two poles at $z = 0.4$ and $z = 0.6$? What is the overall transfer function and impulse response?
(b) What values of α and β are required for the overall system to have both poles at $z = 0.6$? What is the overall transfer function and impulse response? How does the double pole affect the impulse response?

PGR2026-00033
Krisp Technologies EX1024 Page 679

18.11 (Recursive and IIR Filters) The terms *recursive* and *IIR* are not always synonymous. A recursive filter could in fact have a finite impulse response and even linear phase. For each of the following recursive filters, find the transfer function $H(z)$ and the impulse response $h[n]$. Which filters (if any) describe IIR filters? Which filters (if any) are linear phase?

(a) $y[n] - y[n-1] = x[n] - x[n-2]$
(b) $y[n] - y[n-1] = x[n] - x[n-1] - 2x[n-2] + 2x[n-3]$

18.12 (Recursive Forms of FIR Filters) An FIR filter may always be recast in recursive form by the simple expedient of including poles and zeros at identical locations. This is equivalent to multiplying the transfer function numerator and denominator by identical factors. For example, the filter $H(z) = 1 - z^{-1}$ is FIR but if we multiply the numerator and denominator by the identical term $1 + z^{-1}$, the new filter and its difference equation become

$$H_1(z) = \frac{(1 - z^{-1})(1 + z^{-1})}{1 + z^{-1}} = \frac{1 - z^{-2}}{1 + z^{-1}} \qquad y[n] + y[n-1] = x[n] - x[n-2]$$

The difference equation can be implemented recursively. Find two different recursive difference equations (with different orders) for each of the following filters.

(a) $h[n] = \{1, \underset{\Downarrow}{2}, 1\}$
(b) $H(z) = \dfrac{z^2 - 2z + 1}{z^2}$
(c) $y[n] = x[n] - x[n-2]$

18.13 (Systems in Cascade and Parallel) Consider the filter realization of Figure P18.13.

(a) Find its transfer function and impulse response if $\alpha \neq \beta$. Is the overall system FIR or IIR?
(b) Find its transfer function and impulse response if $\alpha = \beta$. Is the overall system FIR or IIR?
(c) Find its transfer function and impulse response if $\alpha = \beta = 1$. What does the overall system represent?

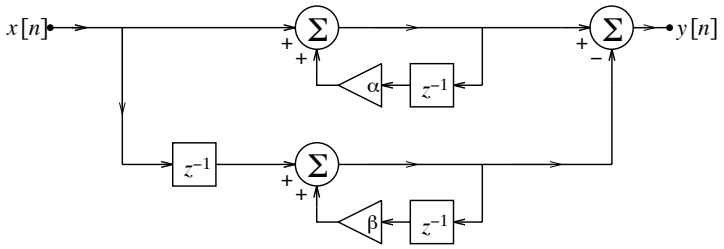


Figure P18.13 Filter realization for Problem 18.13

18.14 (Filter Concepts) Argue for or against the following. Use examples to justify your arguments.

(a) All the finite poles of an FIR filter must lie at $z = 0$.
(b) An FIR filter is always linear phase.
(c) An FIR filter is always stable.
(d) A causal IIR filter can never display linear phase.
(e) A linear-phase sequence is always symmetric about is midpoint.
(f) A minimum-phase filter can never display linear phase.
(g) An allpass filter can never display linear phase.

PGR2026-00033
Krisp Technologies EX1024 Page 680

(h) An allpass filter can never be minimum phase.
(i) The inverse of a minimum-phase filter is also minimum phase.
(j) The inverse of a causal filter is also causal.
(k) For a causal minimum-phase filter to have a a causal inverse, the filter must have as many finite poles as finite zeros.

18.15 (Filter Concepts) Let $H(z) = \dfrac{0.2(z+1)}{z-0.6}$. What type of filter does $H(z)$ describe? Sketch its pole-zero plot. For the following problems, assume that the cutoff frequency of this filter is $F_C = 0.15$.

(a) What filter does $H_1(z) = 1 - H(z)$ describe? Sketch its pole-zero plot. How is the cutoff frequency of this filter related to that of $H(z)$?
(b) What filter does $H_2(z) = H(-z)$ describe? Sketch its pole-zero plot. How is the cutoff frequency of this filter related to that of $H(z)$?
(c) What type of filter does $H_3(z) = 1 - H(z) - H(-z)$ describe? Sketch its pole-zero plot.
(d) Use a combination of the above filters to implement a bandstop filter.

18.16 (Inverse Systems) Consider a system described by $h[n] = 0.5\delta[n] + 0.5\delta[n-1]$.
(a) Sketch the frequency response $H(F)$ of this filter.
(b) In an effort to recover the input $x[n]$, it is proposed to cascade this filter with another filter whose impulse response is $h_1[n] = 0.5\delta[n] - 0.5\delta[n-1]$, as shown:

$$x[n] \longrightarrow \boxed{h[n]} \longrightarrow \boxed{h_1[n]} \longrightarrow y[n]$$

What is the output of the cascaded filter to the input $x[n]$? Sketch the frequency response $H_1(F)$ and the frequency response of the cascaded filter.
(c) What must be the impulse response $h_2[n]$ of a filter connected in cascade with the original filter such that the output of the cascaded filter equals the input $x[n]$, as shown?

$$x[n] \longrightarrow \boxed{h[n]} \longrightarrow \boxed{h_2[n]} \longrightarrow x[n]$$

(d) Are $H_2(F)$ and $H_1(F)$ related in any way?

18.17 (Linear Phase and Symmetry) Assume a sequence $x[n]$ with real coefficients with all its poles at $z = 0$. Argue for or against the following statements. You may want to exploit two useful facts. First, each pair of terms with reciprocal roots such as $(z-\alpha)$ and $(z-1/\alpha)$ yields an even symmetric impulse response sequence. Second, the convolution of symmetric sequences is also endowed with symmetry.
(a) If all the zeros lie on the unit circle, $x[n]$ must be linear phase.
(b) If $x[n]$ is linear phase, its zeros must *always* lie on the unit circle.
(c) If there are no zeros at $z = 1$ and $x[n]$ is linear phase, it is also even symmetric.
(d) If there is one zero at $z = 1$ and $x[n]$ is linear phase, it is also odd symmetric.
(e) If $x[n]$ is even symmetric, there can be no zeros at $z = 1$.
(f) If $x[n]$ is odd symmetric, there must be an odd number of zeros at $z = 1$.

18.18 (Comb Filters) For each comb filter, identify the pole and zero locations and determine whether it is a notch filter or a peaking filter.

(a) $H(z) = \dfrac{z^4 - 0.4096}{z^4 - 0.6561}$ **(b)** $H(z) = \dfrac{z^4 - 1}{z^4 - 0.6561}$

PGR2026-00033
Krisp Technologies EX1024 Page 681

18.19 (Linear-Phase Filters) Argue for or against the following:

(a) The cascade connection of two linear-phase filters is also a linear-phase filter.
(b) The parallel connection of two linear-phase filters is also a linear-phase filter.

18.20 (Minimum-Phase Filters) Argue for or against the following:

(a) The cascade connection of two minimum-phase filters is also a minimum-phase filter.
(b) The parallel connection of two minimum-phase filters is also a minimum-phase filter.

18.21 (Allpass Filters) Argue for or against the following:

(a) The cascade connection of two allpass filters is also an allpass filter.
(b) The parallel connection of two allpass filters is also an allpass filter.

18.22 (Minimum-Phase Systems) Consider the filter $y[n] = x[n] - 0.65x[n-1] + 0.1x[n-2]$.

(a) Find its transfer function $H(z)$ and verify that it is minimum phase.
(b) Find an allpass filter $A(z)$ with the same denominator as $H(z)$.
(c) Is the cascade $H(z)A(z)$ minimum phase? Is it causal? Is it stable?

18.23 (Causality, Stability, and Minimum Phase) Consider two causal, stable, minimum-phase digital filters described by

$$F(z) = \frac{z}{z-0.5} \qquad G(z) = \frac{z-0.5}{z+0.5}$$

Argue that the following filters are also causal, stable, and minimum phase.

(a) The inverse filter $M(z) = 1/F(z)$
(b) The inverse filters $P(z) = 1/G(z)$
(c) The cascade $H(z) = F(z)G(z)$
(d) The inverse of the cascade $R(z) = 1/H(z)$
(e) The parallel connection $N(z) = F(z) + G(z)$

18.24 (Allpass Filters) Consider the filter $H(z) = \dfrac{z+2}{z+0.5}$. The input to this filter is $x[n] = \cos(2n\pi F_0)$.

(a) Is $H(z)$ an allpass filter? If so, what is its gain?
(b) What is the response $y[n]$ and the phase delay if $F_0 = 0$?
(c) What is the response $y[n]$ and the phase delay if $F_0 = 0.25$?
(d) What is the response $y[n]$ and the phase delay if $F_0 = 0.5$?

18.25 (Allpass Filters) Consider two causal, stable, allpass digital filters described by

$$F(z) = \frac{0.5z-1}{0.5-z} \qquad G(z) = \frac{0.5z+1}{0.5+z}$$

(a) Is the filter $L(z) = F^{-1}(z)$ causal? Stable? Allpass?
(b) Is the filter $H(z) = F(z)G(z)$ causal? Stable? Allpass?
(c) Is the filter $M(z) = H^{-1}(z)$ causal? Stable? Allpass?
(d) Is the filter $N(z) = F(z) + G(z)$ causal? Stable? Allpass?

PGR2026-00033
Krisp Technologies EX1024 Page 682

18.26 (Signal Delay) The delay D of a discrete-time energy signal $x[n]$ is defined by

$$D = \frac{\sum_{k=-\infty}^{\infty} kx^2[k]}{\sum_{k=-\infty}^{\infty} x^2[k]}$$

(a) Verify that the delay of the linear-phase sequence $x[n] = \{4, 3, 2, 1, \underset{\Downarrow}{0}, 1, 2, 3, 4\}$ is zero.
(b) Compute the delay of the signals $g[n] = x[n-1]$ and $h[n] = x[n-2]$.
(c) What is the delay of the signal $y[n] = 1.5(0.5)^n u[n] - 2\delta[n]$?
(d) Consider the first-order allpass filter $H(z) = (1+\alpha z)/(z+\alpha)$. Compute the signal delay for its impulse response $h[n]$.

18.27 (First-Order Filters) For each filter, sketch the pole-zero plot, sketch the frequency response to establish the filter type, and evaluate the phase delay at low frequencies. Assume that $\alpha = 0.5$.

(a) $H(z) = \dfrac{z-\alpha}{z+\alpha}$ **(b)** $H(z) = \dfrac{z-1/\alpha}{z+\alpha}$ **(c)** $H(z) = \dfrac{z+1/\alpha}{z+\alpha}$

18.28 (Allpass Filters) Consider a lowpass filter with impulse response $h[n] = (0.5)^n u[n]$. If its input is $x[n] = \cos(0.5n\pi)$, the output will have the form $y[n] = A\cos(0.5n\pi + \theta)$.

(a) Find the values of A and θ.
(b) What should be the transfer function $H_1(z)$ of a first-order allpass filter that can be cascaded with the lowpass filter to correct for the phase distortion and produce the signal $z[n] = B\cos(0.5n\pi)$ at its output?
(c) What should be the gain of the allpass filter in order that $z[n] = x[n]$?

18.29 (Allpass Filters) An unstable digital filter whose transfer function is $H(z) = \dfrac{(z+0.5)(2z+0.5)}{(z+5)(2z+5)}$ is to be stabilized in a way that does not affect its magnitude spectrum.

(a) What must be the transfer function $H_1(z)$ of a filter such that the cascaded filter described by $H_S(z) = H(z)H_1(z)$ is stable?
(b) What is the transfer function $H_S(z)$ of the stabilized filter?
(c) Is $H_S(z)$ causal? Minimum phase? Allpass?

COMPUTATION AND DESIGN

pzdesgui A GUI for Filter Design by Pole-Zero Placement

The graphical user interface `pzdesgui` allows you to visualize the pole-zero plot, impulse response, and frequency response of a digital filter designed by pole-zero placement.
You can place poles and zeros by editing their multiplicity, radius, and angle.
To explore this routine, type `pzdesgui` at the MATLAB prompt.

18.30 (FIR Filter Design) A 22.5-Hz signal is corrupted by 60-Hz hum. It is required to sample this signal at 180 Hz and filter out the interference from the the sampled signal.

PGR2026-00033
Krisp Technologies EX1024 Page 683

(a) Design a minimum-length, linear-phase filter that passes the desired signal with unit gain and completely rejects the interference signal.
(b) Test your design by applying a sampled version of the desired signal, adding 60-Hz interference, filtering the noisy signal, and comparing the desired signal and the filtered signal.

18.31 (Comb Filters) Plot the frequency response of the following filters over $0 \le F \le 0.5$ and describe the action of each filter.

(a) $y[n] = x[n] + \alpha x[n-4], \quad \alpha = 0.5$
(b) $y[n] = x[n] + \alpha x[n-4] + \alpha^2 x[n-8], \quad \alpha = 0.5$
(c) $y[n] = x[n] + \alpha x[n-4] + \alpha^2 x[n-8] + \alpha^3 x[n-12], \quad \alpha = 0.5$
(d) $y[n] = \alpha y[n-4] + x[n], \quad \alpha = 0.5$

18.32 (Filter Design) An ECG signal sampled at 300 Hz is contaminated by interference due to 60-Hz hum. It is required to design a digital filter to remove the interference and provide a dc gain of unity.

(a) Design a 3-point FIR filter (using zero placement) that completely blocks the interfering signal. Plot its frequency response. Does the filter provide adequate gain at other frequencies in the passband? Is this a good design?
(b) Design an IIR filter (using pole-zero placement) that completely blocks the interfering signal. Plot its frequency response. Does the filter provide adequate gain at other frequencies in the passband? Is this a good design?
(c) Generate one period (300 samples) of the ECG signal using the command `yecg=ecgsim(3,9);`. Generate a noisy ECG signal by adding 300 samples of a 60-Hz sinusoid to `yecg`. Obtain filtered signals, using each filter, and compare plots of the filtered signal with the original signal `yecg`. Do the results support your conclusions of parts (a) and (b)? Explain.

18.33 (Nonrecursive Forms of IIR Filters) If we truncate the impulse response of an IIR filter to N terms, we obtain an FIR filter. The larger the truncation index N, the better the FIR filter approximates the underlying IIR filter. Consider the IIR filter described by $y[n] - 0.8y[n-1] = x[n]$.

(a) Find its impulse response $h[n]$ and truncate it to N terms to obtain $h_N[n]$, the impulse response of the approximate FIR equivalent. Would you expect the greatest mismatch in the response of the two filters to identical inputs to occur for lower or higher values of n?
(b) Plot the frequency response $H(F)$ and $H_N(F)$ for $N = 3$. Plot the poles and zeros of the two filters. What differences do you observe?
(c) Plot the frequency response $H(F)$ and $H_N(F)$ for $N = 10$. Plot the poles and zeros of the two filters. Does the response of $H_N(F)$ show a better match to $H(F)$? How do the pole-zero plots compare? What would you expect to see in the pole-zero plot if you increase N to 50? What would you expect to see in the pole-zero plot as $N \to \infty$?

18.34 (LORAN) A LORAN (long-range radio and navigation) system for establishing positions of marine craft uses three transmitters that send out short bursts (10 cycles) of 100-kHz signals in a precise phase relationship. Using phase comparison, a receiver (on the craft) can establish the position (latitude and longitude) of the craft to within a few hundred meters. Suppose the LORAN signal is to be digitally processed by first sampling it at 500 kHz and filtering the sampled signal using a second-order peaking filter with a half-power bandwidth of 1 kHz. Design the peaking filter and use MATLAB to plot its frequency response.

PGR2026-00033
Krisp Technologies EX1024 Page 684

18.35 (Decoding a Mystery Message) During transmission, a message signal gets contaminated by a low-frequency signal and high-frequency noise. The message can be decoded only by displaying it in the time domain. The contaminated signal is provided on disk as `mystery1.mat`. Load this signal into MATLAB (using the command `load mystery1`). In an effort to decode the message, try the following steps and determine what the decoded message says.

(a) Display the contaminated signal. Can you "read" the message?
(b) Display the DFT of the signal to identify the range of the message spectrum.
(c) Design a peaking filter (with unit gain) centered about the message spectrum.
(d) Filter the contaminated signal and display the filtered signal to decode the message.

18.36 (Plucked-String Filters) Figure P18.36 shows three variants of the Karplus-Strong filter for synthesizing plucked-string instruments. Assume that $G_{\text{LP}}(z) = 0.5(1 + z^{-1})$.

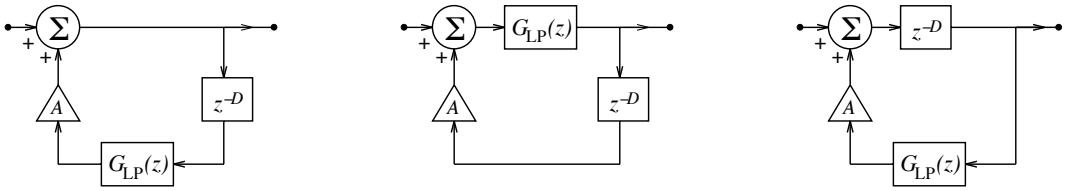


Figure P18.36 Plucked-string filters for Problem 18.36

(a) Find the transfer function of each circuit.
(b) Plot the frequency response of each filter for $D = 12$ (choose an appropriate value for A).
(c) Describe the differences in the frequency response and impulse response of each filter.

18.37 (More Plucked-String Filters) We wish to use the Karplus-Strong filter to synthesize a guitar note played at exactly 880 Hz, using a sampling frequency of 10 kHz, by including a first-order allpass filter $G_{\text{AP}}(z) = (1 + \alpha z)/(z + \alpha)$ in the feedback loop. Choose the second variant from Figure P18.36 and assume that $G_{\text{LP}}(z) = 0.5(1 + z^{-1})$.

(a) What is the value of D and the value of the allpass parameter α?
(b) Plot the frequency response of the designed filter, using an appropriate value for A.
(c) How far off is the fundamental frequency of the designed filter from 880 Hz?
(d) Show that the exact relation for finding the parameter α from the phase delay t_p at any frequency is given by

$$\alpha = \frac{\sin[(1 - t_p)\pi F]}{\sin[(1 + t_p)\pi F]}$$

(e) Use the result of part (d) to compute the exact value of α at the digital frequency corresponding to 880 Hz and plot the frequency response of the designed filter. Is the fundamental frequency of the designed filter any closer to 880 Hz? Will the exact value of α be more useful for lower or higher sampling rates?

18.38 (Generating DTMF Tones) In dual-tone multi-frequency (DTMF) or touch-tone telephone dialing, each number is represented by a dual-frequency tone (as described in the text). It is required to generate the signal at each frequency as a pure cosine, using a digital oscillator operating at a sampling rate of $S = 8192$ Hz. By varying the parameters of the digital filter, it should be possible to generate a signal at any of the required frequencies. A DTMF tone can then be generated by adding the appropriate low-frequency and high-frequency signals.

PGR2026-00033
Krisp Technologies EX1024 Page 685

(a) Design the digital oscillator and use it to generate tones corresponding to all the digits. Each tone should last for 0.1 s.

(b) Use the FFT to obtain the spectrum of each tone and confirm that its frequencies correspond to the appropriate digit.

18.39 (Decoding DTMF Tones) To decode a DTMF signal, we must be able to isolate the tones and then identify the digits corresponding to each tone. This may be accomplished in two ways: using the FFT or by direct filtering.

(a) Generate DTMF tones (by any method) at a sampling rate of 8192 Hz.

(b) For each tone, use the FFT to obtain the spectrum and devise a test that would allow you to identify the digit from the frequencies present in its FFT. How would you automate the decoding process for an arbitrary DTMF signal?

(c) Apply each tone to a parallel bank of bandpass (peaking) filters, compute the output signal energy, compare with an appropriate threshold to establish the presence or absence of a frequency, and relate this information to the corresponding digit. You will need to design eight peaking filters (centered at the appropriate frequencies) to identify the frequencies present and the corresponding digit. How would you automate the decoding process for an arbitrary DTMF signal?

18.40 (Phase Delay and Group Delay of Allpass Filters) Consider the filter $H(z) = \dfrac{z + 1/\alpha}{z + \alpha}$.

(a) Verify that this is an allpass filter.

(b) Pick values of α that correspond to a phase delay of $t_p = 0.1,\ 0.5,\ 0.9$. For each value of α, plot the unwrapped phase, phase delay, and group delay of the filter.

(c) Over what range of digital frequencies is the phase delay a good match to the value of t_p computed in part (b)?

(d) How does the group delay vary with frequency as α is changed?

(e) For each value of α, compute the minimum and maximum values of the phase delay and the group delay and the frequencies at which they occur.

PGR2026-00033
Krisp Technologies EX1024 Page 686

Chapter 19

IIR DIGITAL FILTERS

19.0 Scope and Objectives

Digital filters are widely used in almost all areas of digital signal processing. If linear phase is not critical, infinite impulse response (IIR) filters yield a much smaller filter order for a given application. This chapter begins with an introduction to IIR filters, and the various mappings that are used to convert analog filters to digital filters. It then describes the design of IIR digital filters based on an analog *lowpass prototype* that meets the given specifications, followed by an appropriate *mapping* and *spectral transformation*. The bilinear transformation and its applications are discussed in detail.

19.1 Introduction

Digital filters process discrete-time signals. They are essentially mathematical implementations of a filter equation in software or hardware. They suffer from few limitations. Among their many advantages are high noise immunity, high accuracy (limited only by the roundoff error in the computer arithmetic), easy modification of filter characteristics, freedom from component variations, and, of course, low and constantly decreasing cost. Digital filters are therefore rapidly replacing analog filters in many applications where they can be used effectively. The term *digital filtering* is to be understood in its broadest sense, not only as a smoothing or averaging operation, but also as any processing of the input signal.

19.1.1 The Design Process

Typical magnitude and phase specifications for digital filters are identical to those for analog filters. Additionally, we must also specify a sampling frequency. This frequency is often used to normalize all the other design frequencies. The design process is essentially a three-step process that requires specification of the design parameters, design of the transfer function to meet the design specifications, and realization of the transfer function in software or hardware. The design based on any set of performance specifications is, at best, a compromise at all three levels. At the first level, the actual filter may never meet performance specifications if they are too stringent; at the second, the same set of specifications may lead to several possible realizations; and at the third, quantization and roundoff errors may render the design useless if based on too critical a set of design values. The fewer or less stringent the specifications, the better is the possibility of achieving both design objectives and implementation.

19.1.2 Techniques of Digital Filter Design

Digital filter design revolves around two distinctly different approaches. If linear phase is not critical, IIR filters yield a much smaller filter order for a given application. The design starts with an analog *lowpass prototype* based on the given specifications. It is then converted to the required digital filter, using an

PGR2026-00033
Krisp Technologies EX1024 Page 687

appropriate *mapping* and an appropriate *spectral transformation*. A causal, stable IIR filter can never display linear phase for several reasons. The transfer function of a linear-phase filter must correspond to a symmetric sequence and ensure that $H(z) = \pm H(-1/z)$. For every pole inside the unit circle, there is a reciprocal pole outside the unit circle. This makes the system unstable (if causal) or noncausal (if stable). To make the infinitely long symmetric impulse response sequence of an IIR filter causal, we need an infinite delay, which is not practical, and symmetric truncation (to preserve linear phase) simply transforms the IIR filter into an FIR filter.

Only FIR filters can be designed with linear phase (no phase distortion). Their design is typically based on selecting a symmetric impulse response sequence whose length is chosen to meet design specifications. This choice is often based on iterative techniques or trial and error. For given specifications, FIR filters require many more elements in their realization than do IIR filters.

19.2 IIR Filter Design

There are two related approaches for the design of IIR digital filters. A popular method is based on using methods of well-established analog filter design, followed by a mapping that converts the analog filter to the digital filter. An alternative method is based on designing the digital filter directly, using digital equivalents of analog (or other) approximations. Any transformation of an analog filter to a digital filter should ideally preserve both the response and stability of the analog filter. In practice, this is seldom possible because of the effects of sampling.

19.2.1 Equivalence of Analog and Digital Systems

The impulse response $h(t)$ of an analog system may be approximated by

$$h(t) \approx \tilde{h}_a(t) = t_s \sum_{n=-\infty}^{\infty} h(t)\delta(t - nt_s) = t_s \sum_{n=-\infty}^{\infty} h(nt_s)\delta(t - nt_s) \tag{19.1}$$

Here, t_s is the sampling interval corresponding to the sampling rate $S = 1/t_s$. The discrete-time impulse response $h_s[n]$ describes the samples $h(nt_s)$ of $h(t)$ and may be written as

$$h_s[n] = h(nt_s) = \sum_{k=-\infty}^{\infty} h_s[k]\delta[n - k] \tag{19.2}$$

The Laplace transform $H_a(s)$ of $\tilde{h}_a(t)$ and the z-transform $H_d(z)$ of $h_s[n]$ are

$$H(s) \approx H_a(s) = t_s \sum_{k=-\infty}^{\infty} h(kt_s)e^{-skt_s} \qquad H_d(z) = \sum_{k=-\infty}^{\infty} h_s[k]z^{-k} \tag{19.3}$$

Comparison suggests the equivalence $H_a(s) = t_s H_d(z)$ provided $z^{-k} = e^{-skt_s}$, or

$$z \Rightarrow e^{st_s} \qquad s \Rightarrow \ln(z)/t_s \tag{19.4}$$

These relations describe a **mapping** between the variables z and s. Since $s = \sigma + j\omega$, where ω is the continuous frequency, we can express the complex variable z as

$$z = e^{(\sigma+j\omega)t_s} = e^{\sigma t_s}e^{j\omega t_s} = e^{\sigma t_s}e^{j\Omega} \tag{19.5}$$

Here, $\Omega = \omega t_s = 2\pi f/S = 2\pi F$ is the digital frequency in radians/sample.

PGR2026-00033
Krisp Technologies EX1024 Page 688

The sampled signal $h_s[n]$ has a periodic spectrum given by its DTFT:

$$H_p(f) = S \sum_{k=-\infty}^{\infty} H(f - kS) \tag{19.6}$$

If the analog signal $h(t)$ is band-limited to B and sampled above the Nyquist rate ($S > 2B$), the principal period ($-0.5 \le F \le 0.5$) of $H_p(f)$ equals $SH(f)$, a scaled version of the true spectrum $H(f)$. We may thus relate the analog and digital systems by

$$H(f) = t_s H_p(f) \quad \text{or} \qquad H_a(s)|_{s=j2\pi f} \approx t_s H_d(z)|_{z=e^{j2\pi f/S}}, \qquad |f| < 0.5S \tag{19.7}$$

If $S < 2B$, we have aliasing, and this relationship no longer holds.

19.2.2 The Effects of Aliasing

The relations $z \Rightarrow e^{st_s}$ and $s \Rightarrow \ln(z)/t_s$ do not describe a one-to-one mapping between the s-plane and the z-plane. Since $e^{j\Omega}$ is periodic with period 2π, all frequencies $\Omega_0 \pm 2k\pi$ (corresponding to $\omega_0 \pm k\omega_s$, where $\omega_s = 2\pi S$) are mapped to the same point in the z-plane. A one-to-one mapping is thus possible only if Ω lies in the principal range $-\pi \le \Omega \le \pi$ (or $0 \le \Omega \le 2\pi$), corresponding to the analog frequency range $-0.5\omega_s \le \omega \le 0.5\omega_s$ (or $0 \le \omega \le \omega_s$). Figure 19.1 illustrates how the mapping $z \Rightarrow e^{st_s}$ translates points in the s-domain to corresponding points in the z-domain.

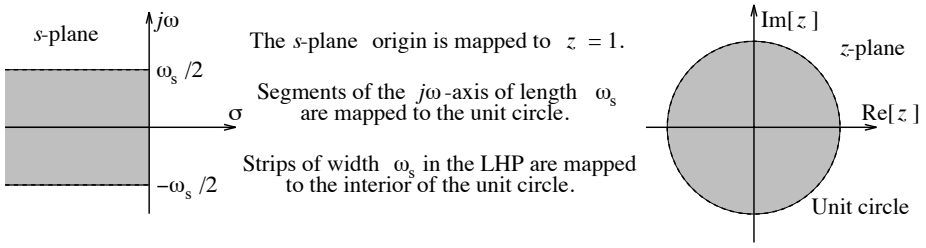


Figure 19.1 Characteristics of the mapping $z \Rightarrow \exp(st_s)$

The origin: The origin $s = 0$ is mapped to $z = 1$, as are all other points corresponding to $s = 0 \pm jk\omega_s$ for which $z = e^{jk\omega_s t_s} = e^{jk2\pi} = 1$.

The $j\omega$-axis: For points on the $j\omega$-axis, $\sigma = 0$, $z = e^{j\Omega}$, and $|z| = 1$. As ω increases from ω_0 to $\omega_0 + \omega_s$, the frequency Ω increases from Ω_0 to $\Omega_0 + 2\pi$, and segments of the $j\omega$-axis of length $\omega_s = 2\pi S$ thus map to the unit circle, *over and over*.

The left half-plane: In the left half-plane, $\sigma < 0$. Thus, $z = e^{\sigma t_s} e^{j\Omega}$ or $|z| = e^{\sigma t_s} < 1$. This describes the interior of the unit circle in the z-plane. In other words, strips of width ω_s in the left half of the s-plane are mapped to the *interior* of the unit circle, *over and over*.

The right half-plane: In the right half-plane, $\sigma > 0$, and we see that $|z| = e^{\sigma t_s} > 1$. Thus, strips of width ω_s in the right half of the s-plane are repeatedly mapped to the *exterior* of the unit circle.

REVIEW PANEL 19.1

The Mapping $z \Rightarrow e^{st_s}$ Is Not a Unique One-to-One Mapping

Strips of width $\omega_s = 2\pi S$ (along the $j\omega$-axis) in the left half of the s-plane map to the interior of the unit circle, over and over.

PGR2026-00033
Krisp Technologies EX1024 Page 689

19.2.3 Practical Mappings

The transcendental nature of the transformation $s \Rightarrow \ln(z)/t_s$ does not permit direct conversion of a rational transfer function $H(s)$ to a rational transfer function $H(z)$. Nor does it permit a one-to-one correspondence for frequencies higher than $0.5S$ Hz. A unique representation in the z-plane is possible only for band-limited signals sampled above the Nyquist rate. Practical mappings are based on one of the following methods:

1. Matching the time response (the response-invariant transformation)
2. Matching terms in a factored $H(s)$ (the matched z-transform)
3. Conversion of system differential equations to difference equations
4. Rational approximations for $z \Rightarrow e^{st_s}$ or $s \Rightarrow \ln(z)/t_s$

In general, each method results in different mapping rules, and leads to different forms for the digital filter $H(z)$ from a given analog filter $H(s)$, and not all methods preserve stability. When comparing the frequency response, it is helpful to remember that the analog frequency range $0 \le f \le 0.5S$ for the frequency response of $H(s)$ corresponds to the digital frequency range $0 \le F \le 0.5$ for the frequency response of $H(z)$. The time-domain response can be compared only at the sampling instants $t = nt_s$.

19.3 Response Matching

The idea behind response matching is to *match* the time-domain analog and digital response for a given input, typically the impulse response or step response. Given the analog filter $H(s)$, and the input $x(t)$ whose invariance we seek, we first find the analog response $y(t)$ as the inverse transform of $H(s)X(s)$. We then sample $x(t)$ and $y(t)$ at intervals t_s to obtain their sampled versions $x[n]$ and $y[n]$. Finally, we compute $H(z) = Y(z)/X(z)$ to obtain the digital filter. The process is illustrated in Figure 19.2.

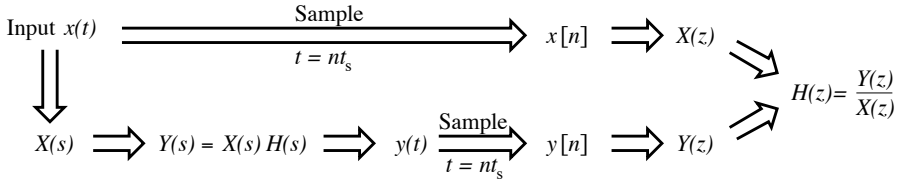


Figure 19.2 The concept of response invariance

Response-invariant matching yields a transfer function that is a good match only for the time-domain response to the input for which it was designed. It may not provide a good match for the response to other inputs. The quality of the approximation depends on the choice of the sampling interval t_s, and a unique correspondence is possible only if the sampling rate $S = 1/t_s$ is above the Nyquist rate (to avoid aliasing). This mapping is thus useful only for analog systems such as lowpass and bandpass filters, whose frequency response is essentially band-limited. This also implies that the analog transfer function $H(s)$ must be *strictly proper* (with numerator degree *less than* the denominator degree).

REVIEW PANEL 19.2

Response-Invariant Mappings Match the Time Response of the Analog and Digital Filter

The response $y(t)$ of $H(s)$ matches the response $y[n]$ of $H(z)$ at the sampling instants $t = nt_s$.

PGR2026-00033
Krisp Technologies EX1024 Page 690

EXAMPLE 19.1 (Response-Invariant Mappings)

(a) Convert $H(s) = \dfrac{1}{s+1}$ to a digital filter $H(z)$, using impulse invariance, with $t_s = 1$ s.

For impulse invariance, we select the input as $x(t) = \delta(t)$. We then find

$$X(s) = 1 \qquad Y(s) = H(s)X(s) = \frac{1}{s+1} \qquad y(t) = e^{-t}u(t)$$

The sampled versions of the input and output are

$$x[n] = \delta[n] \qquad y[n] = e^{-nt_s}u[n]$$

Taking the ratio of their z-transforms and using $t_s = 1$, we obtain

$$H(z) = \frac{Y(z)}{X(z)} = Y(z) = \frac{z}{z - e^{-t_s}} = \frac{z}{z - e^{-1}} = \frac{z}{z - 0.3679}$$

The frequency response of $H(s)$ and $H(z)$ is compared in Figure E19.1(a).

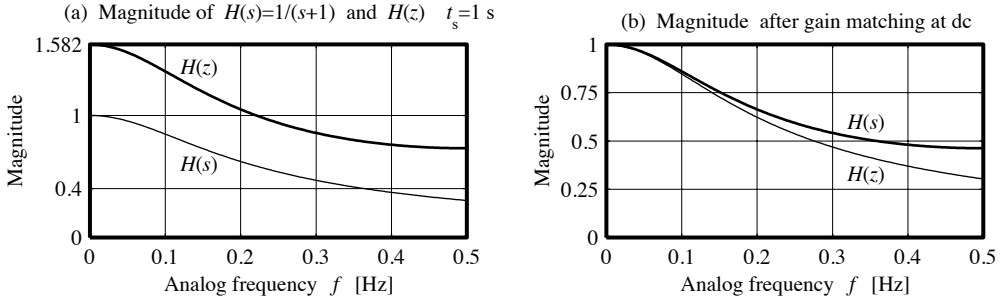


Figure E19.1 Frequency response of $H(s)$ and $H(z)$ for Example 19.1(a)

The dc gain of $H(s)$ (at $s = 0$) is unity, but that of $H(z)$ (at $z = 1$) is 1.582. Even if we normalize the dc gain of $H(z)$ to unity, as in Figure E19.1(b), we see that the frequency response of the analog and digital filters is different. However, the analog impulse response $h(t) = e^{-t}$ matches $h[n] = e^{-n}u[n]$ at the sampling instants $t = nt_s = n$. A perfect match for the time-domain response, for which the filter was designed, lies at the heart of response invariant mappings. The time-domain response to any other inputs will be different. For example, the step response of the analog filter is

$$S(s) = \frac{1}{s(s+1)} = \frac{1}{s} - \frac{1}{s+1} \qquad s(t) = (1 - e^{-t})u(t)$$

To find the step response $S(z)$ of the digital filter whose input is $u[n] \Leftrightarrow z/(z-1)$, we use partial fractions on $S(z)/z$ to obtain

$$S(z) = \frac{z^2}{(z-1)(z-e^{-1})} = \frac{ze/(e-1)}{z-1} + \frac{z/(1-e)}{z-e^{-1}} \qquad s[n] = \frac{e}{e-1}u[n] + \frac{1}{1-e}e^{-n}u[n]$$

The sampled version of $s(t)$ is quite different from $s[n]$. Figure E19.1A reveals that, at the sampling instants $t = nt_s$, the impulse response of the two filters shows a perfect match, but the step response does not, and neither will the time-domain response to any other input.

PGR2026-00033
Krisp Technologies EX1024 Page 691

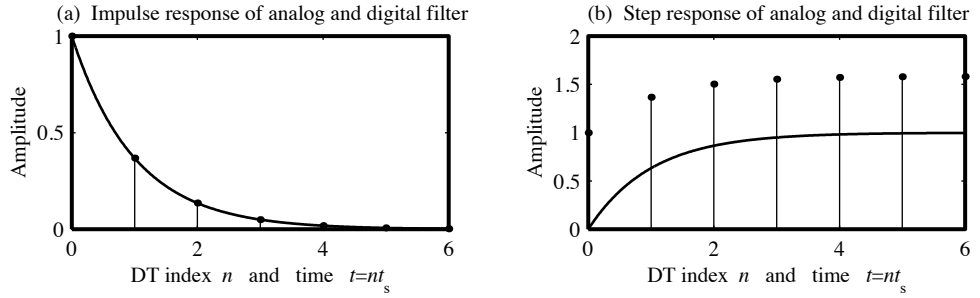


Figure E19.1A Impulse response and step response of $H(s)$ and $H(z)$ for Example 19.1(a)

(b) Convert $H(s) = \dfrac{4}{(s+1)(s+2)}$ to $H(z)$, using various response-invariant transformations.

1. **Impulse invariance:** We choose $x(t) = \delta(t)$. Then, $X(s) = 1$, and

$$Y(s) = H(s)X(s) = \frac{4}{(s+1)(s+2)} = \frac{4}{s+1} - \frac{4}{s+2} \qquad y(t) = 4e^{-t}u(t) - 4e^{-2t}u(t)$$

The sampled input and output are then

$$x[n] = \delta[n] \qquad y[n] = 4e^{-nt_s}u[n] - 4e^{-2nt_s}u[n]$$

The ratio of their z-transforms yields the transfer function of the digital filter as

$$H_I(z) = \frac{Y(z)}{X(z)} = Y(z) = \frac{4z}{z - e^{-t_s}} - \frac{4z}{z - e^{-2t_s}}$$

2. **Step invariance:** We choose $x(t) = u(t)$. Then, $X(s) = 1/s$, and

$$Y(s) = H(s)X(s) = \frac{4}{s(s+1)(s+2)} = \frac{2}{s} - \frac{4}{s+1} + \frac{2}{s+2} \qquad y(t) = (2 - 4e^{-t} + 2e^{-2t})u(t)$$

The sampled input and output are then

$$x[n] = u[n] \qquad y[n] = (2 - 4e^{-nt_s} + 2e^{-2nt_s})u[n]$$

Their z-transforms give

$$X(z) = \frac{z}{z-1} \qquad Y(z) = \frac{2z}{z-1} - \frac{4z}{z - e^{-t_s}} + \frac{2z}{z - e^{-2t_s}}$$

The ratio of their z-transforms yields the transfer function of the digital filter as

$$H_S(z) = \frac{Y(z)}{X(z)} = 2 - \frac{4(z-1)}{z - e^{-t_s}} + \frac{2(z-1)}{z - e^{-2t_s}}$$

3. **Ramp invariance:** We choose $x(t) = r(t) = tu(t)$. Then, $X(s) = 1/s^2$, and

$$Y(s) = \frac{4}{s^2(s+1)(s+2)} = \frac{-3}{s} + \frac{2}{s^2} + \frac{4}{s+1} - \frac{1}{s+2} \qquad y(t) = (-3 + 2t + 4e^{-t} - e^{-2t})u(t)$$

PGR2026-00033
Krisp Technologies EX1024 Page 692

The sampled input and output are then

$$x[n] = nt_s u[n] \qquad y[n] = (-3 + 2nt_s + 4e^{-nt_s} - e^{-2nt_s})u[n]$$

Their z-transforms give

$$X(z) = \frac{zt_s}{(z-1)^2} \qquad Y(z) = \frac{-3z}{z-1} + \frac{2zt_s}{(z-1)^2} + \frac{4z}{z-e^{-t_s}} - \frac{z}{z-e^{-2t_s}}$$

The ratio of their z-transforms yields the transfer function of the digital filter as

$$H_R(z) = \frac{-3(z-1)}{t_s} + 2 + \frac{4(z-1)^2}{t_s(z-e^{-t_s})} - \frac{(z-1)^2}{t_s(z-e^{-2t_s})}$$

19.3.1 The Impulse-Invariant Transformation

The impulse-invariant mapping yields some useful design relations. We start with a first-order analog filter described by $H(s) = 1/(s+p)$. The impulse response $h(t)$, and its sampled version $h[n]$, are

$$h(t) = e^{-pt}u(t) \qquad h[n] = e^{-pnt_s}u[n] = \left(e^{-pt_s}\right)^n u[n] \tag{19.8}$$

The z-transform of $h[n]$ (which has the form $\alpha^n u[n]$, where $\alpha = e^{-pt_s}$), yields the transfer function $H(z)$ of the digital filter as

$$H(z) = \frac{z}{z-e^{-pt_s}}, \qquad |z| > e^{-pt_s} \tag{19.9}$$

This relation suggests that we can go directly from $H(s)$ to $H(z)$, using the mapping

$$\frac{1}{s+p} \Longrightarrow \frac{z}{z-e^{-pt_s}} \tag{19.10}$$

We can now extend this result to filters of higher order. If $H(s)$ is in partial fraction form, we can obtain simple expressions for impulse-invariant mapping. If $H(s)$ has no repeated roots, it can be described as a sum of first-order terms, using partial fraction expansion, and each term can be converted by the impulse-invariant mapping to give

$$H(s) = \sum_{k=1}^{N} \frac{A_k}{s+p_k} \qquad H(z) = \sum_{k=1}^{N} \frac{zA_k}{z-e^{-p_k t_s}}, \qquad \text{ROC: } |z| > e^{-|p|_{\max} t_s} \tag{19.11}$$

Here, the region of convergence of $H(z)$ is in terms of the largest pole magnitude $|p|_{\max}$ of $H(s)$.

REVIEW PANEL 19.3

Impulse-Invariant Design Requires $H(s)$ in Partial Fraction Form

$$\frac{1}{s+p_k} \Longrightarrow \frac{z}{z-e^{-p_k t_s}} \quad \text{(for each term in the partial fraction expansion)}$$

If the denominator of $H(s)$ also contains repeated roots, we start with a typical kth term $H_k(s)$ with a root of multiplicity M, and find

$$H_k(s) = \frac{A_k}{(s+p_k)^M} \qquad h_k(t) = \frac{A_k}{(M-1)!} t^{M-1} e^{-p_k t} u(t) \tag{19.12}$$

PGR2026-00033
Krisp Technologies EX1024 Page 693

The sampled version $h_k[n]$, and its z-transform, can then be found by the times-n property of the z-transform. Similarly, quadratic terms corresponding to complex conjugate poles in $H(s)$ may also be simplified to obtain a real form. These results are listed in Table 19.1. Note that *impulse-invariant design requires H(s) in partial fraction form and yields a digital filter H(z) in the same form.* It must be reassembled if we need a composite rational function form. The left half-plane poles of $H(s)$ (corresponding to $p_k > 0$) map into poles of $H(z)$ that lie inside the unit circle (corresponding to $z = e^{-p_k t_s} < 1$). Thus, a stable analog filter $H(s)$ is transformed into a stable digital filter $H(z)$.

REVIEW PANEL 19.4

Impulse-Invariant Mappings Are Prone to Aliasing but Preserve Stability

The analog impulse response $h(t)$ matches $h[n]$ at the sampling instants.
Impulse-invariant mappings are not suitable for highpass or bandstop filter design.

Table 19.1 Impulse-Invariant Transformations

Term	**Form of $H(s)$**	$H(z)$ **(with $\alpha = e^{-pt_s}$)**
Distinct	$\dfrac{A}{(s+p)}$	$\dfrac{Az}{(z-\alpha)}$
Complex conjugate	$\dfrac{Ae^{j\Omega}}{s+p+jq} + \dfrac{Ae^{-j\Omega}}{s+p-jq}$	$\dfrac{2z^2 A\cos(\Omega) - 2A\alpha z\cos(\Omega + qt_s)}{z^2 - 2\alpha z\cos(qt_s) + \alpha^2}$
Repeated twice	$\dfrac{A}{(s+p)^2}$	$At_s\dfrac{\alpha z}{(z-\alpha)^2}$
Repeated thrice	$\dfrac{A}{(s+p)^3}$	$0.5At_s^2\dfrac{\alpha z(z+\alpha)}{(z-\alpha)^3}$

EXAMPLE 19.2 (Impulse-Invariant Mappings)

(a) Convert $H(s) = \dfrac{2s-1}{s^2+5s+4}$ to $H(z)$, using impulse invariance at $S = 2$ Hz.

First, by partial fractions, we obtain

$$H(s) = \frac{2s-1}{s^2+5s+4} = \frac{2s-1}{(s+1)(s+4)} = \frac{3}{s+4} - \frac{1}{s+1}$$

The impulse-invariant transformation, with $t_s = 1/S = 0.5$ s, gives

$$H(z) = \frac{3z}{z-e^{-4t_s}} - \frac{z}{z-e^{-t_s}} = \frac{3z}{z-e^{-2}} - \frac{z}{z-e^{-0.5}} = \frac{2z^2 - 1.6843z}{z^2 - 0.7419z + 0.0831}$$

(b) Convert $H(s) = \dfrac{4}{(s+1)(s^2+4s+5)}$ to $H(z)$, using impulse invariance, with $t_s = 0.5$ s.

The partial fraction form for $H(s)$ is

$$H(s) = \frac{2}{s+1} + \frac{-1-j}{s+2+j} + \frac{-1+j}{s+2-j}$$

PGR2026-00033
Krisp Technologies EX1024 Page 694

For the second term, we write $K = (-1 - j) = \sqrt{2}e^{-j3\pi/4} = Ae^{j\Omega}$. Thus, $A = \sqrt{2}$ and $\Omega = -3\pi/4$. We also have $p = 2$, $q = 1$, and $\alpha = e^{-pt_s} = 1/e$. With these values, Table 19.1 gives

$$H(z) = \frac{2z}{z - 1/\sqrt{e}} + \frac{2\sqrt{2}z^2\cos(-\frac{3\pi}{4}) - 2\sqrt{2}(z/e)\cos(0.5 - \frac{3\pi}{4})}{z^2 - 2(z/e)\cos(0.5) + e^{-2}}$$

This result simplifies to

$$H(z) = \frac{0.2146z^2 + 0.0930z}{z^3 - 1.2522z^2 + 0.5270z - 0.0821}$$

Comment: The first step involved partial fractions. Note that we cannot compute $H(z)$ as the cascade of the impulse-invariant digital filters for $H_1(s) = \dfrac{4}{s+1}$ and $H_2(s) = \dfrac{1}{s^2 + 4s + 5}$.

19.3.2 Modifications to Impulse-Invariant Design

Gain Matching

The mapping $H(s) = 1/(s + p_k) \Rightarrow H(z) = z/(z - e^{-p_k t_s})$ reveals that the dc gain of the analog term $H(s)$ (with $s = 0$) equals $1/p_k$ but the dc gain of the digital term $H(z)$ (with $z = 1$) is $1/(1 - e^{-p_k t_s})$. If the sampling interval t_s is small enough such that $p_k t_s \ll 1$, we can use the approximation $e^{-p_k t_s} \approx 1 - p_k t_s$ to give the dc gain of $H(z)$ as $1/p_k t_s$. This suggests that in order to match the dc gain of the two filters, $H(z)$ must be scaled (multiplied) by t_s. This scaling is automatically accounted for if we normalize the frequency specifications by the sampling frequency S before designing the filter, because it is equivalent to choosing $t_s = 1$ during the mapping. In any case, it is customary to compare $H(s)$ with $t_s H(z)$, or an appropriately scaled version $KH(z)$, where the constant K is chosen to match the gain of $H(s)$ and $KH(z)$ at a convenient frequency. Keep in mind, however, that any scale factor K also scales the impulse response $h[n]$ by K, and the design is then no longer strictly impulse invariant.

REVIEW PANEL 19.5

How to Account for Gain Matching in Impulse-Invariant Design

For high sampling rates S, the gain of $H(z)$ is S times larger than the gain of $H(s)$.

Accounting for Sampling Errors

The impulse-invariant method suffers from errors in sampling $h(t)$ if it shows a jump at $t = 0$. If $h(0)$ is not zero, the sampled value at the origin should be chosen as $0.5h(0)$. As a result, the impulse response of the digital filter must be modified to $h_M[n] = h[n] - 0.5h(0)\delta[n]$. This leads to the modified transfer function $H_M(z) = H(z) - 0.5h(0)$. The simplest way to find $h(0)$ is to use the initial value theorem $h(0) = \lim_{s\to\infty} sH(s)$. Since $h(0)$ is nonzero only if the degree N of the denominator of $H(s)$ exceeds the degree M of its numerator by 1, we need this modification only if $N - M = 1$.

EXAMPLE 19.3 (Modified Impulse-Invariant Design)

(a) (Impulse-Invariant Design)

Convert the analog filter $H(s) = \dfrac{1}{s+1}$, with a cutoff frequency of 1 rad/s, to a digital filter with a cutoff frequency of $f_c = 10$ Hz and $S = 60$ Hz, using impulse invariance and gain matching.

There are actually two ways to do this:

PGR2026-00033
Krisp Technologies EX1024 Page 695

1. We normalize by the sampling frequency S, which allows us to use $t_s = 1$ in all subsequent computations. Normalization gives $\Omega_C = 2\pi f_C/S = \frac{\pi}{3}$. We denormalize $H(s)$ to Ω_C to get

$$H_1(s) = H\Big(\frac{s}{\Omega_C}\Big) = \frac{\frac{\pi}{3}}{s + \frac{\pi}{3}}$$

   Finally, with $t_s = 1$, impulse invariance gives

$$H_1(z) = \frac{\frac{\pi}{3}z}{z - e^{-\pi/3}} = \frac{1.0472z}{z - 0.3509}$$

2. We first denormalize $H(s)$ to f_C to get

$$H_2(s) = H\Big(\frac{s}{2\pi f_C}\Big) = \frac{20\pi}{s + 20\pi}$$

   We use impulse invariance and multiply the resulting digital filter transfer function by t_s to get

$$H_2(z) = t_s\frac{20\pi z}{z - e^{-20\pi t_s}} = \frac{z\frac{\pi}{3}}{z - e^{-\pi/3}} == \frac{1.0472z}{z - 0.3509}$$

Comment: Both approaches yield identical results. The first method automatically accounts for the gain matching. For a perfect gain match at dc, we should multiply $H(z)$ by the gain factor

$$K = \frac{1}{H(1)} = \frac{1 - 0.3509}{1.0472} = 0.6198$$

(b) (Modified Impulse-Invariant Design)

Convert $H(s) = \dfrac{1}{s+1}$ to a digital filter, with $t_s = 1$ s, using modified impulse invariance to account for sampling errors and gain matching at dc.

Using impulse invariance, the transfer function of the digital filter is

$$H(z) = \frac{z}{z - e^{-1}} = \frac{z}{z - 0.3679}$$

Since $h(t) = e^{-t}u(t)$, $h(t)$ has a jump of 1 unit at $t = 0$, and thus $h(0) = 1$. The modified impulse-invariant mapping thus gives

$$H_M(z) = H(z) - 0.5h(0) = \frac{z}{z - e^{-1}} - 0.5 = \frac{0.5(z + e^{-1})}{z - e^{-1}} = \frac{0.5(z + 0.3679)}{z - 0.3679}$$

The dc gain of $H(s)$ is unity. We compute the dc gain of $H(z)$ and $H_M(z)$ (with $z = 1$) as

$$H(z)|_{z=1} = \frac{1}{1 - e^{-1}} = 1.582 \qquad H_M(z)|_{z=1} = 0.5\frac{1 + e^{-1}}{1 - e^{-1}} = 1.082$$

For unit dc gain, the transfer functions of the original and modified digital filter become

$$H_1(z) = \frac{z}{1.582(z - e^{-1})} = \frac{0.6321z}{z - 0.3679} \qquad H_{M1}(z) = \frac{0.5(z + e^{-1})}{1.082(z - e^{-1})} = \frac{0.4621(z + 0.3679)}{z - 0.3679}$$

PGR2026-00033
Krisp Technologies EX1024 Page 696

Figure E19.3B compares the response of $H(s)$, $H(z)$, $H_1(z)$, $H_M(z)$, and $H_{M1}(z)$. It clearly reveals the improvement due to each modification.

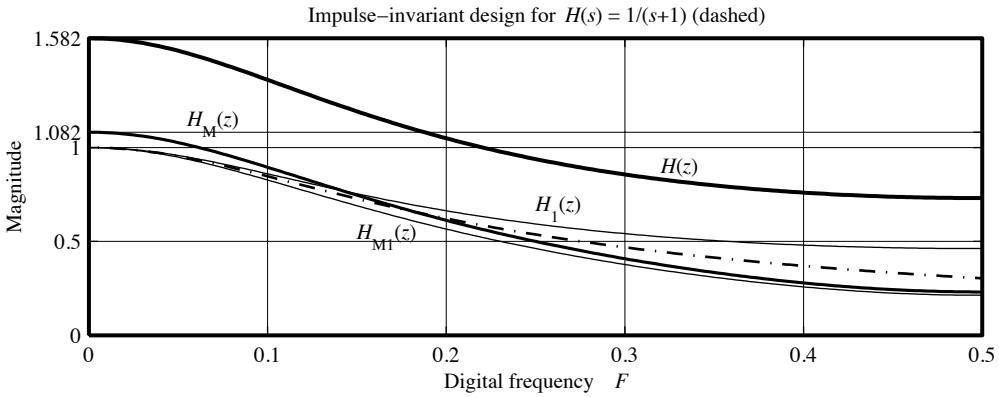


Figure E19.3B Response of the various filters for Example 19.3(b)

(c) (Modified Impulse-Invariant Design)

Convert the analog filter $H(s) = \dfrac{2s-1}{s^2+5s+4}$ to a digital filter, with $t_s = 0.5$ s, using modified impulse invariance to account for sampling errors.

Since the numerator degree is $M = 1$ and the denominator degree is $N = 2$, we have $N - M = 1$, and a modification is needed. The initial value theorem gives

$$h(0) = \lim_{s\to\infty} sH(s) = \lim_{s\to\infty} \frac{2s^2 - s}{s^2+5s+4} = \frac{2 - 1/s}{1 + 5/s + 4/s^2} = 2$$

The transfer function of the digital filter using impulse-invariant mapping was found in part (a) as

$$H(z) = \frac{3z}{z-e^{-2}} - \frac{z}{z-e^{-1}} = \frac{2z^2 - 1.6843z}{z^2 - 0.7419z + 0.0831}$$

The modified transfer function is thus

$$H_M(z) = H(z) - 0.5h(0) = \frac{3z}{z-e^{-2}} - \frac{z}{z-e^{-1}} - 1 = \frac{z^2 - 0.9424z - 0.0831}{z^2 - 0.7419z + 0.0831}$$

(d) (Modified Impulse-Invariant Design)

Convert $H(s) = \dfrac{4}{(s+1)(s^2+4s+5)}$ to a digital filter, with $t_s = 0.5$ s, using modified impulse invariance (if required) to account for sampling errors.

Since the numerator degree is $M = 0$ and the denominator degree is $N = 3$, we have $N - M = 3$. Since this does not equal 1, the initial value is zero, and no modification is required. The transfer function of the digital filter using impulse-invariant mapping is thus identical to the one found in Example 19.2(b):

$$H(z) = \frac{0.2146z^2 + 0.0930z}{z^3 - 1.2522z^2 + 0.5270z - 0.0821}$$

PGR2026-00033
Krisp Technologies EX1024 Page 697

19.4 The Matched z-Transform for Factored Forms

The impulse-invariant mapping may be expressed as

$$\frac{1}{s+\alpha} \Longrightarrow \frac{z}{z-e^{-\alpha t_s}} \qquad s+\alpha \Longrightarrow \frac{z-e^{-\alpha t_s}}{z} \tag{19.13}$$

The **matched z-transform** uses this mapping to convert each numerator and denominator term of a *factored* $H(s)$ to yield the digital filter $H(z)$, also in factored form, as

$$H(s) = K_0 \frac{\prod_{i=1}^{M}(s-z_i)}{\prod_{k=1}^{N}(s-p_k)} \qquad H(z) = (AK_0 z^P)\frac{\prod_{i=1}^{M}(z-e^{z_i t_s})}{\prod_{k=1}^{N}(z-e^{p_k t_s})} \tag{19.14}$$

The power P of z^P in $H(z)$ is $P = N - M$, the difference in the degree of the denominator and numerator polynomials of $H(s)$. The constant A is chosen to match the gains of $H(s)$ and $H(z)$ at some convenient frequency (typically, dc).

For complex roots, we can replace each conjugate pair using the mapping

$$(s+p-jq)(s+p+jq) \Longrightarrow \frac{z^2 - 2ze^{-pt_s}\cos(qt_s) + e^{-2pt_s}}{z^2} \tag{19.15}$$

Since poles in the left half of the s-plane are mapped inside the unit circle in the z-plane, the matched z-transform preserves stability. It converts an all-pole analog system to an all-pole digital system but may not preserve the frequency response of the analog system. As with the impulse-invariant mapping, the matched z-transform also suffers from aliasing errors.

REVIEW PANEL 19.6

The Matched z-Transform Requires $H(s)$ in Factored Form

$(s+p_k) \Longrightarrow \dfrac{z-e^{-p_k t_s}}{z}$ (for each factor of $H(s)$ in factored form)

The matched z-transform is not suitable for highpass or bandpass filter design.

19.4.1 Modifications to Matched z-Transform Design

The matched z-transform maps the zeros of $H(s)$ at $s = \infty$ (corresponding to the highest analog frequency $f = \infty$) to $z = 0$. This yields the term z^P in the design relation. In the modified matched z-transform, some or all of these zeros are mapped to $z = -1$ (corresponding to the highest digital frequency $F = 0.5$ or the analog frequency $f = 0.5S$). Two modifications to $H(z)$ are in common use:

1. Move *all* zeros from $z = 0$ to $z = -1$ (or replace z^P by $(z+1)^P$) in $H(z)$.
2. Move *all but one* of the zeros at $z = 0$ to $z = -1$ (or replace z^P by $z(z+1)^{P-1}$ in $H(z)$).

These modifications allow us to use the matched z-transform even for highpass and bandstop filters.

REVIEW PANEL 19.7

Two Modifications to Matched z-Transform Design

1. Move *all* zeros from $z = 0$ to $z = -1$. **2.** Move *all but one* of the zeros at $z = 0$ to $z = -1$.

PGR2026-00033
Krisp Technologies EX1024 Page 698

EXAMPLE 19.4 (The Matched z-Transform)

Convert $H(s) = \dfrac{4}{(s+1)(s+2)}$ to $H(z)$ by the matched z-transform and its modifications, with $t_s = 0.5$ s.

(a) No modification: The matched z-transform yields

$$H(z) = \frac{Az^2}{(z - e^{-t_s})(z - e^{-2t_s})} = \frac{Az^2}{(z - e^{-0.5})(z - e^{-1})} = \frac{Az^2}{z^2 - 0.9744z + 0.2231}$$

(b) First modification: Replace both zeros in $H(z)$ (the term z^2) by $(z+1)^2$:

$$H_1(z) = \frac{A_1(z+1)^2}{(z - e^{-t_s})(z - e^{-2t_s})} = \frac{A_1(z+1)^2}{(z - e^{-0.5})(z - e^{-1})} = \frac{A_1(z+1)^2}{z^2 - 0.9744z + 0.2231}$$

(c) Second modification: Replace only one zero in $H(z)$ (the term z^2) by $z+1$:

$$H_2(z) = \frac{A_2z(z+1)}{(z - e^{-t_s})(z - e^{-2t_s})} = \frac{A_2z(z+1)}{(z - e^{-0.5})(z - e^{-1})} = \frac{A_2z(z+1)}{z^2 - 0.9744z + 0.2231}$$

Comment: In each case, the constants (A, A_1, and A_2) may be chosen for a desired gain.

19.5 Mappings from Discrete Algorithms

Discrete-time algorithms are often used to develop mappings to convert analog filters to digital filters. In the discrete domain, the operations of integration and differentiation are replaced by numerical integration and numerical differences, respectively. The mappings are derived by equating the transfer function $H(s)$ of the ideal analog operation with the transfer function $H(z)$ of the corresponding discrete operation.

19.5.1 Mappings from Difference Algorithms

Numerical differences are often used to convert differential equations to difference equations. Three such difference algorithms are listed in Table 19.2.

Table 19.2 Numerical Difference Algorithms

Difference	**Numerical Algorithm**	**Mapping for s**
Backward	$y[n] = \dfrac{x[n] - x[n-1]}{t_s}$	$s = \dfrac{z-1}{zt_s}$
Forward	$y[n] = \dfrac{x[n+1] - x[n]}{t_s}$	$s = \dfrac{z-1}{t_s}$
Central	$y[n] = \dfrac{x[n+1] - x[n-1]}{2t_s}$	$s = \dfrac{z^2-1}{2zt_s}$

Numerical differences approximate the slope (derivative) at a point, as illustrated for the backward-difference and forward-difference algorithms in Figure 19.3.

PGR2026-00033
Krisp Technologies EX1024 Page 699

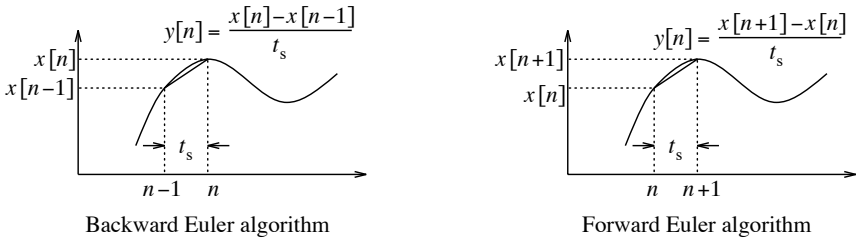


Figure 19.3 Backward-difference and forward-difference algorithms

The mappings that result from the operations (also listed in Table 19.2) are based on comparing the ideal derivative operator $H(s) = s$ with the transfer function $H(z)$ of each difference algorithm, as follows:

Backward difference: $y[n] = \dfrac{x[n] - x[n-1]}{t_s}$

$$Y(z) = \frac{X(z) - z^{-1}X(z)}{t_s} \qquad H(z) = \frac{Y(z)}{X(z)} = \frac{z-1}{zt_s} \qquad s = \frac{z-1}{zt_s} \tag{19.16}$$

Forward difference: $y[n] = \dfrac{x[n+1] - x[n]}{t_s}$

$$Y(z) = \frac{zX(z) - X(z)}{t_s} \qquad H(z) = \frac{Y(z)}{X(z)} = \frac{z-1}{t_s} \qquad s = \frac{z-1}{t_s} \tag{19.17}$$

Central difference: $y[n] = \dfrac{x[n+1] - x[n-1]}{2t_s}$

$$Y(z) = \frac{zX(z) - z^{-1}X(z)}{2t_s} \qquad H(z) = \frac{Y(z)}{X(z)} = \frac{z^2-1}{2zt_s} \qquad s = \frac{z^2-1}{2zt_s} \tag{19.18}$$

19.5.2 Stability Properties of the Backward-Difference Algorithm

The backward-difference mapping $s \to (z-1)/zt_s$ results in

$$z = 0.5 + 0.5\frac{1+st_s}{1-st_s} \qquad z - 0.5 = 0.5\frac{1+st_s}{1-st_s} \tag{19.19}$$

To find where the $j\omega$-axis maps, we set $\sigma = 0$ to obtain

$$z - 0.5 = 0.5\frac{1+j\omega t_s}{1-j\omega t_s} \qquad |z - 0.5| = 0.5 \tag{19.20}$$

Thus, the $j\omega$-axis is mapped into a circle of radius 0.5, centered at $z = 0.5$, as shown in Figure 19.4.

Since this region is within the unit circle, the mapping preserves stability. It does, however, restrict the pole locations of the digital filter. Since the frequencies are mapped into a smaller circle, this mapping is a good approximation only in the vicinity of $z = 1$ or $\Omega \approx 0$ (where it approximates the unit circle), which implies high sampling rates.

PGR2026-00033
Krisp Technologies EX1024 Page 700

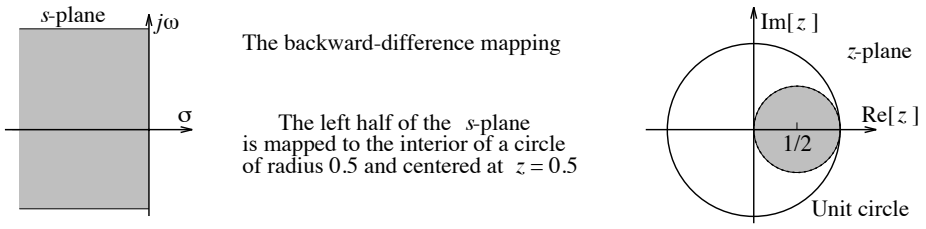


Figure 19.4 Mapping region for the mapping based on the backward difference

19.5.3 The Forward-Difference Algorithm

The mapping for the forward-difference is $s \rightarrow (z-1)/t_s$. With $z = u + jv$, we obtain

$$z = u + jv = 1 + t_s(\sigma + j\omega) \tag{19.21}$$

If $\sigma = 0$, then $u = 1$, and the $j\omega$-axis maps to $z = 1$. If $\sigma > 0$, then $u > 1$, and the right half of the s-plane maps to the right of $z = 1$. If $\sigma < 0$, then $u < 1$, and the left half of the s-plane maps to the left of $z = 1$, as shown in Figure 19.5. This region (in the z-plane) includes not only the unit circle but also a vast region outside it. Thus, a stable analog filter with poles in the left half of the s-plane may result in an unstable digital filter $H(z)$ with poles anywhere to the left of $z = 1$ (but not inside the unit circle) in the z-plane !

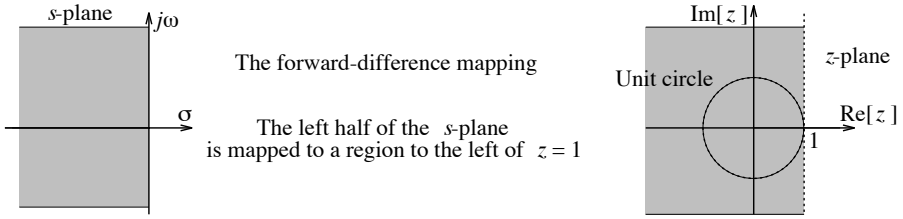


Figure 19.5 Mapping region for the mapping based on the forward difference

REVIEW PANEL 19.8

Mappings Based on the Backward Difference and Forward Difference

Backward difference: $s = \dfrac{z-1}{zt_s}$ (stable) **Forward difference:** $s = \dfrac{z-1}{t_s}$ (not always stable)

EXAMPLE 19.5 (Mappings from Difference Algorithms)

(a) We convert the stable analog filter $H(s) = \dfrac{1}{s+\alpha}$, $\alpha > 0$, to a digital filter, using the backward-difference mapping $s = (z-1)/zt_s$ to obtain

$$H(z) = \frac{1}{\alpha + (z-1)/zt_s} = \frac{zt_s}{(1+\alpha t_s)z - 1}$$

The digital filter has a pole at $z = 1/(1+\alpha t_s)$. Since this is always less than unity if $\alpha > 0$ (for a stable $H(s)$) and $t_s > 0$, we have a stable $H(z)$.

PGR2026-00033
Krisp Technologies EX1024 Page 701

(b) We convert the stable analog filter $H(s) = \dfrac{1}{s+\alpha},\ \alpha > 0$, to a digital filter, using the forward-difference mapping $s = (z-1)/t_s$ to obtain

$$H(z) = \frac{1}{\alpha + (z-1)/t_s} = \frac{t_s}{z - (1 - \alpha t_s)}$$

The digital filter has a pole at $z = 1 - \alpha t_s$ and is thus stable only if $0 < \alpha t_s < 2$ (to ensure $|z| < 1$). Since $\alpha > 0$ and $t_s > 0$, we are assured a stable system only if $\alpha < 2/t_s$. This implies that the sampling rate S must be chosen to ensure that $S > 0.5\alpha$.

(c) We convert the stable analog filter $H(s) = \dfrac{1}{s+\alpha},\ \alpha > 0$, to a digital filter, using the central-difference mapping $s = (z^2-1)/2zt_s$ to obtain

$$H(z) = \frac{1}{\alpha + (z^2-1)/2zt_s} = \frac{2zt_s}{z^2 + 2\alpha t_s z - 1}$$

The digital filter has a pair of poles at $z = -\alpha t_s \pm \sqrt{(\alpha t_s)^2 + 1}$. The magnitude of one of the poles is always greater than unity, and the digital filter is thus unstable for any $\alpha > 0$.

Comment: Clearly, from a stability viewpoint, only the mapping based on the backward difference is useful for the filter $H(s) = 1/(s+\alpha)$. In fact, this mapping preserves stability for any stable analog filter.

19.5.4 Mappings from Integration Algorithms

Two commonly used algorithms for numerical integration are based on the rectangular rule and trapezoidal rule. These integration algorithms, listed in Table 19.3, estimate the "area" $y[n]$ from $y[n-1]$ by using step interpolation (for the rectangular rule) or linear interpolation (for the trapezoidal rule) between the samples of $x[n]$, as illustrated in Figure 19.6.

Table 19.3 Numerical Integration Algorithms

Rule	**Numerical Algorithm**	**Mapping for s**
Rectangular	$y[n] = y[n-1] + t_s x[n]$	$s = \frac{1}{t_s}\left(\frac{z-1}{z}\right)$
Trapezoidal	$y[n] = y[n-1] + \frac{2}{t_s}\left(x[n] + x[n-1]\right)$	$s = \frac{2}{t_s}\left(\frac{z-1}{z+1}\right)$

The mappings resulting from these operators are also listed in Table 19.3 and are based on comparing the transfer function $H(s) = 1/s$ of the ideal integrator with the transfer function $H(z)$ of each integration algorithm, as follows:

Rectangular rule: $y[n] = y[n-1] + t_s x[n]$

$$Y(z) = z^{-1}Y(z) + t_s X(z) \qquad H(z) = \frac{Y(z)}{X(z)} = \frac{zt_s}{z-1} \qquad s = \frac{z-1}{zt_s} \tag{19.22}$$

We remark that the rectangular algorithm for integration and the backward difference for the derivative generate identical mappings.

PGR2026-00033
Krisp Technologies EX1024 Page 702

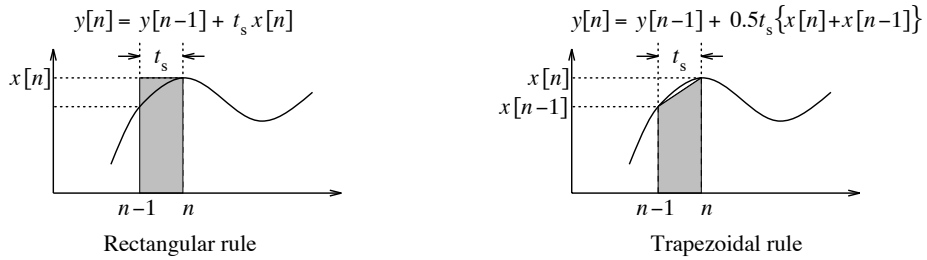


Figure 19.6 Illustrating numerical integration algorithms

Trapezoidal rule: $y[n] = y[n-1] + 0.5t_s(x[n] + x[n-1])$

$$Y(z) = z^{-1}Y(z) + 0.5t_s[X(z) + z^{-1}X(z) \qquad H(z) = \frac{Y(z)}{X(z)} = \frac{0.5t_s(z+1)}{z-1} \qquad s = \frac{2}{t_s}\left[\frac{z-1}{z+1}\right] \tag{19.23}$$

The mapping based on the trapezoidal rule is also called **Tustin's rule**.

19.5.5 Stability Properties of Integration-Algorithm Mappings

Mappings based on the rectangular and trapezoidal algorithms always yield a stable $H(z)$ for any stable $H(s)$, and any choice of t_s. The rectangular rule is equivalent to the backward-difference algorithm. It thus maps the left half of the s-plane to the interior of a circle of radius 0.5 in the z-plane, centered at $z = 0.5$, and is always stable. For the trapezoidal rule, we express z in terms of $s = \sigma + j\omega$ to get

$$z = \frac{2 + st_s}{2 - st_s} = \frac{2 + \sigma t_s + j\omega t_s}{2 - \sigma t_s - j\omega t_s} \tag{19.24}$$

If $\sigma = 0$, we get $|z| = 1$, and for $\sigma < 0$, $|z| < 1$. Thus, the $j\omega$-axis is mapped to the unit circle, and the left half of the s-plane is mapped into the interior of the unit circle, as shown in Figure 19.7. This means that a stable analog system will always yield a stable digital system using this transformation. If $\omega = 0$, we have $z = 1$, and the dc gain of both the analog and digital filters is identical.

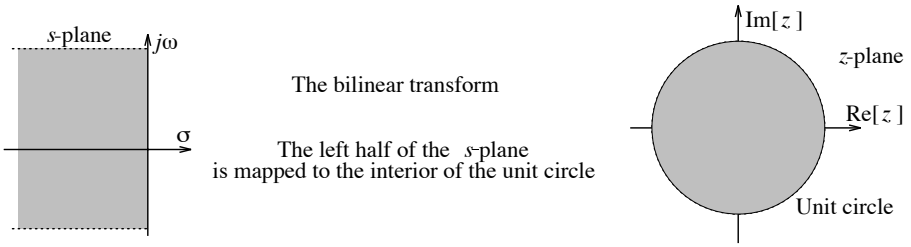


Figure 19.7 Mapping region for the trapezoidal integration algorithm

Discrete difference and integration algorithms are good approximations only for small digital frequencies ($F < 0.1$, say) or high sampling rates S (small t_s) that may be well in excess of the Nyquist rate. This is why the sampling rate is a critical factor in the frequency-domain performance of these algorithms. Another factor is stability. For example, the mapping based on the central-difference algorithm is not very useful because it always produces an unstable digital filter. Algorithms based on trapezoidal integration and the backward difference are popular because they always produce stable digital filters.

PGR2026-00033
Krisp Technologies EX1024 Page 703

REVIEW PANEL 19.9

Mappings Based on Numerical Integration Algorithms

Rectangular rule: $s = \dfrac{z-1}{zt_s}$ (always stable)

Trapezoidal rule: $s = \dfrac{2}{t_s}\left[\dfrac{z-1}{z+1}\right]$ (always stable)

EXAMPLE 19.6 (Mappings from Integration Algorithms)

(a) Convert $H(s) = \dfrac{1}{s+\alpha}$, $\alpha > 0$, to a digital filter $H(z)$, using the trapezoidal numerical integration algorithm, and comment on the stability of $H(z)$.

Using the the mapping based on the trapezoidal rule, we obtain

$$H(z) = H(s)\Big|_{s=\frac{2(z-1)}{t_s(z+1)}} = \frac{t_s(z+1)}{(2+\alpha t_s)z - (2-\alpha t_s)}$$

The pole location of $H(z)$ is $z = (2-\alpha t_s)/(2+\alpha t_s)$. Since this is always less than unity (if $\alpha > 0$ and $t_s > 0$), we have a stable $H(z)$.

(b) Simpson's algorithm for numerical integration finds $y[n]$ over two time steps from $y[n-2]$ and is given by

$$y[n] = y[n-2] + \tfrac{t_s}{3}(x[n] + 4x[n-1] + x[n-2])$$

Derive a mapping based on Simpson's rule, use it to convert $H(s) = \dfrac{1}{s+\alpha}$, $\alpha > 0$, to a digital filter $H(z)$, and comment on the stability of $H(z)$.

The transfer function $H_S(z)$ of this algorithm is found as follows:

$$Y(z) = z^{-2}Y(z) + \tfrac{t_s}{3}(1 + 4z^{-1} + z^{-2})X(z) \qquad H_S(z) = t_s\frac{z^2+4z+1}{3(z^2-1)}$$

Comparison with the transfer function of the ideal integrator $H(s) = 1/s$ gives

$$s = \frac{3}{t_s}\left[\frac{z^2-1}{z^2+4z+1}\right]$$

If we convert $H(s) = 1/(s+\alpha)$, using this mapping, we obtain

$$H(z) = \frac{t_s(z^2+4z+1)}{(3+\alpha t_s)z^2 + 4\alpha t_s z - (3-\alpha t_s)}$$

The poles of $H(z)$ are the roots of $(3+\alpha t_s)z^2 + 4\alpha t_s z - (3-\alpha t_s) = 0$. The magnitude of one of these roots is always greater than unity (if $\alpha > 0$ and $t_s > 0$), and $H(z)$ is thus an unstable filter.

PGR2026-00033
Krisp Technologies EX1024 Page 704

19.5.6 Mappings from Rational Approximations

Some of the mappings that we have derived from numerical algorithms may also be viewed as rational approximations of the transformations $z \to e^{st_s}$ and $s \to \ln(z)/t_s$. The forward-difference mapping is based on a first-order approximation for $z = e^{st_s}$ and yields

$$z = e^{st_s} \approx 1 + st_s, \quad st_s \ll 1 \qquad\qquad s \approx \frac{z-1}{t_s} \tag{19.25}$$

The backward-difference mapping is based on a first-order approximation for $z^{-1} = e^{-st_s}$ and yields

$$\frac{1}{z} = e^{-st_s} \approx 1 - st_s, \quad st_s \ll 1 \qquad\qquad s \approx \frac{z-1}{zt_s} \tag{19.26}$$

The trapezoidal mapping is based on a first-order rational function approximation of $s = \ln(z)/t_s$, with $\ln(z)$ described by a power series, and yields

$$s = \frac{\ln(z)}{t_s} = \frac{2}{t_s}\left[\frac{z-1}{z+1}\right] + \frac{2}{3t_s}\left[\frac{z-1}{z+1}\right]^2 + \frac{2}{5t_s}\left[\frac{z-1}{z+1}\right]^2 + \cdots \approx \frac{2}{t_s}\left[\frac{z-1}{z+1}\right] \tag{19.27}$$

19.6 The Bilinear Transformation

If we generalize the mapping based on the trapezoidal rule by letting $C = 2/t_s$, we obtain the **bilinear transformation**, defined by

$$s = C\frac{z-1}{z+1} \qquad\qquad z = \frac{C+s}{C-s} \tag{19.28}$$

If we let $\sigma = 0$, we obtain the complex variable z in the form

$$z = \frac{C+j\omega}{C-j\omega} = e^{j2\tan^{-1}(\omega/C)} \tag{19.29}$$

Since $z = e^{j\Omega}$, where $\Omega = 2\pi F$ is the digital frequency, we find

$$\Omega = 2\tan^{-1}(\omega/C) \qquad\qquad \omega = C\tan(0.5\Omega) \tag{19.30}$$

This is a *nonlinear* relation between the analog frequency ω and the digital frequency Ω. When $\omega = 0, \Omega = 0$, and as $\omega \to \infty$, $\Omega \to \pi$. It is thus a one-to-one mapping that *nonlinearly compresses* the analog frequency range $-\infty < f < \infty$ to the digital frequency range $-\pi < \Omega < \pi$. It avoids the effects of aliasing at the expense of distorting, compressing, or *warping* the analog frequencies, as shown in Figure 19.8.

The higher the frequency, the more severe is the warping. We can compensate for this warping (but not eliminate it) if we **prewarp** the frequency specifications before designing the analog system $H(s)$ or applying the bilinear transformation. Prewarping of the frequencies prior to analog design is just a scaling (stretching) operation based on the inverse of the warping relation, and is given by

$$\omega = C\tan(0.5\Omega) \tag{19.31}$$

Figure 19.9 shows a plot of ω versus Ω for various values of C, compared with the linear relation $\omega = \Omega$. The analog and digital frequencies always show a match at the origin ($\omega = \Omega = 0$), and at *one other value* dictated by the choice of C.

We point out that the nonlinear stretching effect of the prewarping often results in a filter of lower order, especially if the sampling frequency is not high enough. For high sampling rates, it turns out that the prewarping has little effect and may even be redundant.

PGR2026-00033
Krisp Technologies EX1024 Page 705

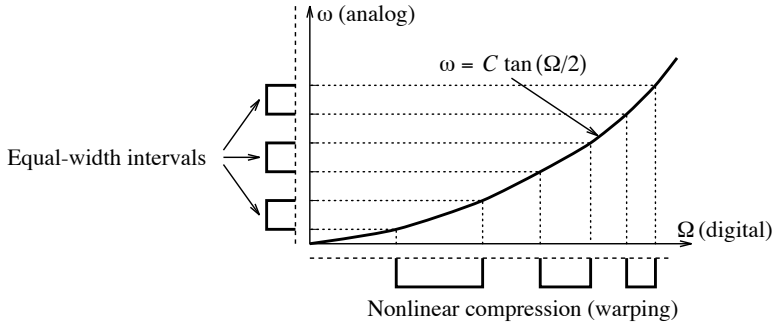


Figure 19.8 The warping effect of the bilinear transformation

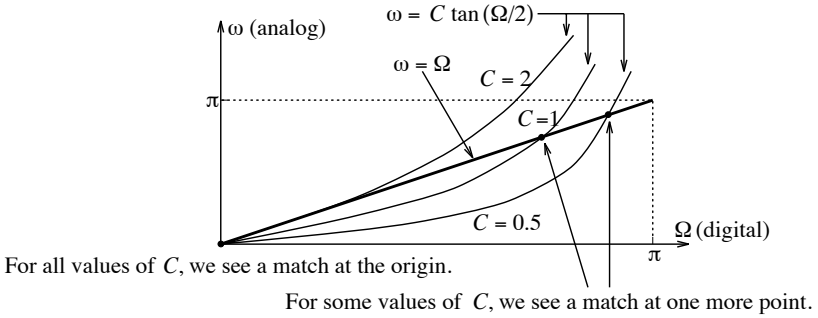


Figure 19.9 The warping relation $\omega = C\tan(0.5\Omega)$ for various choices of C

The popularity of the bilinear transformation stems from its simple, stable, one-to-one mapping. It avoids problems caused by aliasing and can thus be used even for highpass and bandstop filters. Though it does suffer from warping effects, it can also compensate for these effects, using a simple relation.

REVIEW PANEL 19.10

The Bilinear Transformation Avoids Aliasing at the Expense of Nonlinear Warping

Bilinear transformation: $s = C\dfrac{z-1}{z+1}$ **Warping relation:** $\omega = C\tan(0.5\Omega)$

19.6.1 Using the Bilinear Transformation

Given an analog transfer function $H(s)$ whose response at the analog frequency ω_A is to be matched to $H(z)$ at the digital frequency Ω_D, we may design $H(z)$ in one of two ways:

1. We pick C by matching ω_A and the prewarped frequency Ω_D, and obtain $H(z)$ from $H(s)$, using the transformation $s = C\dfrac{z-1}{z+1}$. This process may be summarized as follows:

$$\omega_A = C\tan(0.5\Omega_D) \qquad C = \frac{\omega_A}{\tan(0.5\Omega_D)} \qquad H(z) = H(s)\big|_{s=C(z-1)/(z+1)} \tag{19.32}$$

PGR2026-00033
Krisp Technologies EX1024 Page 706

2. We pick a convenient value for C (say, $C = 1$). This actually matches the response at an arbitrary prewarped frequency ω_x given by

$$\omega_x = \tan(0.5\Omega_D) \tag{19.33}$$

Next, we frequency scale $H(s)$ to $H_1(s) = H(s\omega_A/\omega_x)$, and obtain $H(z)$ from $H_1(s)$, using the transformation $s = \dfrac{z-1}{z+1}$ (with $C = 1$). This process may be summarized as follows (for $C = 1$):

$$\omega_x = \tan(0.5\Omega_D) \qquad H_1(s) = H(s)\big|_{s=s\omega_A/\omega_x} \qquad H(z) = H_1(s)\big|_{s=(z-1)/(z+1)} \tag{19.34}$$

The two methods yield an identical digital filter $H(z)$. The first method does away with the scaling of $H(s)$, and the second method allows a convenient choice for C.

REVIEW PANEL 19.11

The Bilinear Transformation Allows Gain Matching by Appropriate Choice of C

Choose $C = \dfrac{\omega_A}{\tan(0.5\Omega_D)}$ to match the gain of $H(s)$ at ω_A to the gain of $H(z)$ at Ω_D.

EXAMPLE 19.7 (Using the Bilinear Transformation)

(a) Consider a Bessel filter described by $H(s) = \dfrac{3}{s^2 + 3s + 3}$. Design a digital filter whose magnitude at $f_0 = 3$ kHz equals the magnitude of $H(s)$ at $\omega_A = 4$ rad/s if the sampling rate $S = 12$ kHz.

The digital frequency is $\Omega = 2\pi f_0/S = 0.5\pi$. We can now proceed in one of two ways:

1. **Method 1:** We select C by choosing the prewarped frequency to equal $\omega_A = 4$:

$$\omega_A = 4 = C\tan(0.5\Omega) \quad \text{or} \quad C = \frac{4}{\tan(0.5\Omega)} = 4$$

We transform $H(s)$ to $H(z)$, using $s = C\dfrac{z-1}{z+1} = \dfrac{4(z-1)}{z+1}$, to obtain

$$H(z) = H(s)\big|_{s=4(z-1)/(z+1)} = \frac{3(z+1)^2}{31z^2 - 26z + 7}$$

2. **Method 2:** We choose $C = 1$, say, and evaluate $\omega_x = \tan(0.5\Omega) = \tan(0.25\pi) = 1$.
Next, we frequency scale $H(s)$ to

$$H_1(s) = H(s\omega_A/\omega_x) = H(4s) = \frac{3}{16s^2 + 12s + 3}$$

Finally, we transform $H_1(s)$ to $H(z)$, using $s = \dfrac{z-1}{z+1}$, to obtain

$$H(z) = H(s)\big|_{s=(z-1)/(z+1)} = \frac{3(z+1)^2}{31z^2 - 26z + 7}$$

The magnitude $|H(s)|$ at $s = j\omega = j4$ matches the magnitude of $|H(z)|$ at $z = e^{j\Omega} = e^{j\pi/2} = j$. We find that

$$|H(s)|_{s=j4} = \left|\frac{3}{-13 + j12}\right| = 0.1696 \qquad |H(z)|_{z=j} = \left|\frac{3(j+1)^2}{-24 - j26}\right| = 0.1696$$

PGR2026-00033
Krisp Technologies EX1024 Page 707

Figure E19.7(a) compares the magnitude of $H(s)$ and $H(z)$. The linear phase of the Bessel filter is not preserved during the transformation (unless the sampling frequency is very high).

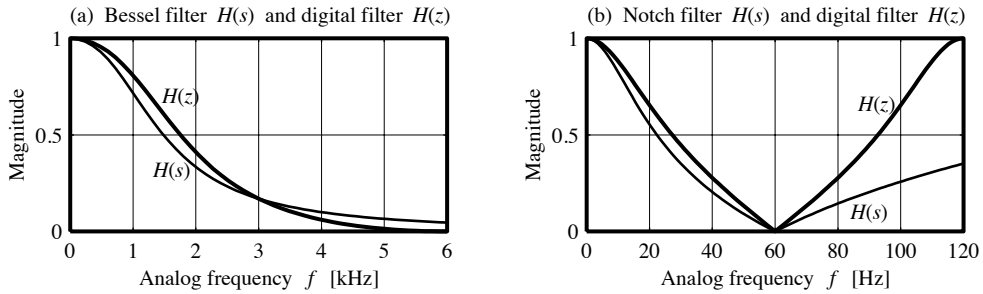


Figure E19.7 Magnitude of the analog and digital filters for Example 19.7(a and b)

(b) The twin-T notch filter $H(s) = \dfrac{s^2+1}{s^2+4s+1}$ has a notch frequency $\omega_0 = 1$ rad/s. Design a digital notch filter with $S = 240$ Hz and a notch frequency $f = 60$ Hz.

The digital notch frequency is $\Omega = 2\pi f/S = 0.5\pi$. We pick C by matching the analog notch frequency ω_0 and the prewarped digital notch frequency Ω to get

$$\omega_0 = C\tan(0.5\Omega) \qquad 1 = C\tan(0.25\pi) \qquad C = 1$$

Finally, we convert $H(s)$ to $H(z)$, using $s = C\dfrac{z-1}{z+1} = \dfrac{z-1}{z+1}$, to get

$$H(z) = H(s)\big|_{s=(z-1)/(z+1)} = \frac{(z-1)^2+(z+1)^2}{(z-1)^2+4(z^2-1)+(z+1)^2} = \frac{z^2+1}{3z^2-1}$$

We confirm that $H(s) = 0$ at $s = j\omega_0 = j$ and $H(z) = 0$ at $z = e^{j\Omega} = e^{j\pi/2} = j$. Figure E19.7(b) shows the magnitude of the two filters and the perfect match at $f = 60$ Hz (or $F = 0.25$).

19.7 Spectral Transformations for IIR Filters

As with analog filters, the design of IIR filters usually starts with an analog lowpass prototype, which is converted to a digital lowpass prototype by an appropriate mapping and transformed to the required filter type by an appropriate spectral transformation. For the bilinear mapping, we may even perform the mapping and the spectral transformation (in a single step) on the analog prototype itself.

19.7.1 Digital-to-Digital Transformations

If a digital lowpass prototype has been designed, the digital-to-digital (D2D) transformations of Table 19.4 can be used to convert it to the required filter type. These transformations preserve stability by mapping the unit circle (and all points within it) into itself.

As with analog transformations, the lowpass-to-bandpass (LP2BP) and lowpass-to-bandstop (LP2BS) transformations yield a digital filter with twice the order of the lowpass prototype. The lowpass-to-lowpass (LP2LP) transformation is actually a special case of the more general allpass transformation:

$$z \Rightarrow \pm\frac{z-\alpha}{1-\alpha z}, \qquad |\alpha| < 1 \text{ (and } \alpha \text{ real)} \tag{19.35}$$

PGR2026-00033
Krisp Technologies EX1024 Page 708

Table 19.4 Digital-to-Digital (D2D) Frequency Transformations

Note: The digital lowpass prototype cutoff frequency is Ω_D. All digital frequencies are normalized to $\Omega = 2\pi f/S$.			
Form	**Band Edge(s)**	**Mapping $z \rightarrow$**	**Mapping Parameters**
LP2LP	Ω_C	$\dfrac{z-\alpha}{1-\alpha z}$	$\alpha = \sin[0.5(\Omega_D - \Omega_C)]/\sin[0.5(\Omega_D + \Omega_C)]$
LP2HP	Ω_C	$\dfrac{-(z+\alpha)}{1+\alpha z}$	$\alpha = -\cos[0.5(\Omega_D + \Omega_C)]/\cos[0.5(\Omega_D - \Omega_C)]$
LP2BP	$[\Omega_1,\ \Omega_2]$	$\dfrac{-(z^2 + A_1 z + A_2)}{A_2 z^2 + A_1 z + 1}$	$K = \tan(0.5\Omega_D)/\tan[0.5(\Omega_2 - \Omega_1)]$ $\alpha = -\cos[0.5(\Omega_2 + \Omega_1)]/\cos[0.5(\Omega_2 - \Omega_1)]$ $A_1 = 2\alpha K/(K+1) \qquad A_2 = (K-1)/(K+1)$
LP2BS	$[\Omega_1,\ \Omega_2]$	$\dfrac{(z^2 + A_1 z + A_2)}{A_2 z^2 + A_1 z + 1}$	$K = \tan(0.5\Omega_D)\tan[0.5(\Omega_2 - \Omega_1)]$ $\alpha = -\cos[0.5(\Omega_2 + \Omega_1)]/\cos[0.5(\Omega_2 - \Omega_1)]$ $A_1 = 2\alpha/(K+1) \qquad A_2 = -(K-1)/(K+1)$

EXAMPLE 19.8 (Using D2D Transformations)

A lowpass filter $H(z) = \dfrac{3(z+1)^2}{31z^2 - 26z + 7}$ operates at $S = 8$ kHz, and its cutoff frequency is $f_C = 2$ kHz.

(a) Use $H(z)$ to design a highpass filter with a cutoff frequency of 1 kHz.

We find $\Omega_D = 2\pi f_C/S = 0.5\pi$ and $\Omega_C = 0.25\pi$. The LP2HP transformation (Table 19.4) requires

$$\alpha = -\frac{\cos[0.5(\Omega_D + \Omega_C)]}{\cos[0.5(\Omega_D - \Omega_C)]} = -\frac{\cos(3\pi/8)}{\cos(\pi/8)} = -0.4142$$

The LP2HP spectral transformation is thus $z \rightarrow \dfrac{-(z+\alpha)}{1+\alpha z} = \dfrac{-(z-0.4142)}{1-0.4142z}$ and yields

$$H_{\text{HP}}(z) = \frac{0.28(z-1)^2}{z^2 - 0.0476z + 0.0723}$$

(b) Use $H(z)$ to design a bandpass filter with band edges of 1 kHz and 3 kHz.

The various digital frequencies are $\Omega_1 = 0.25\pi,\ \Omega_2 = 0.75\pi,\ \Omega_2 - \Omega_1 = 0.5\pi$, and $\Omega_2 + \Omega_1 = \pi$.

From Table 19.4, the parameters needed for the LP2BP transformation are

$$K = \frac{\tan(\pi/4)}{\tan(\pi/4)} = 1 \qquad \alpha = -\frac{\cos(\pi/2)}{\cos(\pi/4)} = 0 \qquad A_1 = 0 \qquad A_2 = 0$$

The LP2BP transformation is thus $z \rightarrow -z^2$ and yields

$$H_{\text{BP}}(z) = \frac{3(z^2-1)^2}{31z^4 + 26z^2 + 7}$$

PGR2026-00033
Krisp Technologies EX1024 Page 709

(c) Use $H(z)$ to design a bandstop filter with band edges of 1.5 kHz and 2.5 kHz.

Once again, we need $\Omega_1 = 3\pi/8,\ \Omega_2 = 5\pi/8,\ \Omega_2 - \Omega_1 = \pi/4$, and $\Omega_2 + \Omega_1 = \pi$.

From Table 19.4, the LP2BS transformation requires the parameters

$$K = \frac{\tan(\pi/8)}{\tan(\pi/4)} = 0.4142 \qquad \alpha = -\frac{\cos(\pi/2)}{\cos(\pi/8)} = 0 \qquad A_1 = 0 \qquad A_2 = 0.4142$$

The LP2BS transformation is thus $z \rightarrow \dfrac{z^2 + 0.4142}{0.4142z^2 + 1}$ and yields

$$H_{\text{BS}}(z) = \frac{0.28(z^2+1)^2}{z^4 + 0.0476z^2 + 0.0723}$$

Figure E19.8 compares the magnitudes of each filter designed in this example.

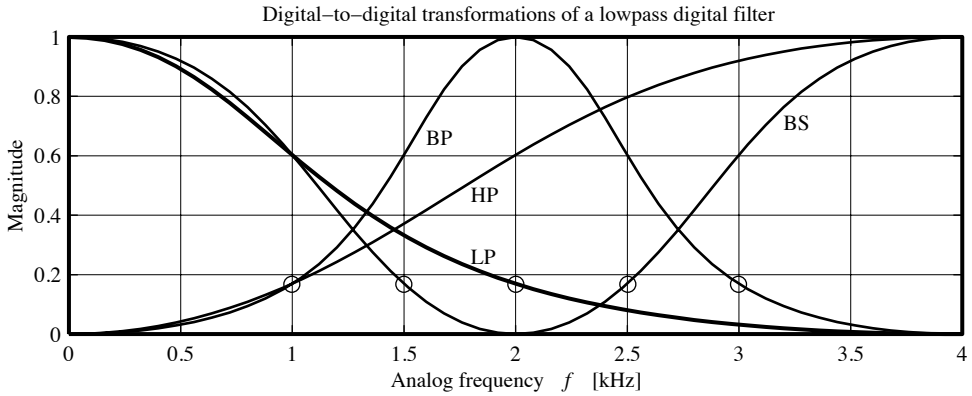


Figure E19.8 The digital filters for Example 19.8

19.7.2 Direct (A2D) Transformations for Bilinear Design

All stable transformations that are also free of aliasing introduce warping effects. Only the bilinear mapping offers a simple relation to compensate for the warping. Combining the bilinear mapping with the D2D transformations yields the analog-to-digital (A2D) transformations of Table 19.5 for bilinear design. These can be used to convert a prewarped analog lowpass prototype (*with a cutoff frequency of* 1 rad/s) directly to the required digital filter.

19.7.3 Bilinear Transformation for Peaking and Notch Filters

If we wish to use the bilinear transformation to design a second-order digital peaking (bandpass) filter with a 3-dB bandwidth of $\Delta\Omega$ and a center frequency of Ω_0, we start with the lowpass analog prototype $H(s) = 1/(s+1)$ (whose cutoff frequency is 1 rad/s), and apply the A2D LP2BP transformation, to obtain

$$H_{\text{BP}}(z) = \frac{C}{1+C}\left(\frac{z^2-1}{z^2 - \frac{2\beta}{1+C}z + \frac{1-C}{1+C}}\right) \qquad \beta = \cos\Omega_0 \quad C = \tan(0.5\Delta\Omega) \tag{19.36}$$

Similarly, if we wish to use the bilinear transformation to design a second-order digital notch (bandstop) filter with a 3-dB notch bandwidth of $\Delta\Omega$ and a notch frequency of Ω_0, we once again start with the lowpass

PGR2026-00033
Krisp Technologies EX1024 Page 710

Table 19.5 Direct Analog-to-Digital (A2D) Transformations for Bilinear Design

Note: The analog lowpass prototype *prewarped* cutoff frequency is 1 rad/s. The digital frequencies are normalized ($\Omega = 2\pi f/S$) but are *not prewarped.*			
Form	**Band Edge(s)**	**Mapping $s \rightarrow$**	**Mapping Parameters**
LP2LP	Ω_C	$\dfrac{z-1}{C(z+1)}$	$C = \tan(0.5\Omega_C)$
LP2HP	Ω_C	$\dfrac{C(z+1)}{z-1}$	$C = \tan(0.5\Omega_C)$
LP2BP	$\Omega_1 < \Omega_0 < \Omega_2$	$\dfrac{z^2 - 2\beta z + 1}{C(z^2-1)}$	$C = \tan[0.5(\Omega_2 - \Omega_1)],\ \beta = \cos\Omega_0$ or $\beta = \cos[0.5(\Omega_2 + \Omega_1)]/\cos[0.5(\Omega_2 - \Omega_1)]$
LP2BS	$\Omega_1 < \Omega_0 < \Omega_2$	$\dfrac{C(z^2-1)}{z^2 - 2\beta z + 1}$	$C = \tan[0.5(\Omega_2 - \Omega_1)],\ \beta = \cos\Omega_0$ or $\beta = \cos[0.5(\Omega_2 + \Omega_1)]/\cos[0.5(\Omega_2 - \Omega_1)]$

analog prototype $H(s) = 1/(s+1)$, and apply the A2D LP2BS transformation, to obtain

$$H_{\text{BS}}(z) = \frac{1}{1+C}\left(\frac{z^2 - 2\beta z + 1}{z^2 - \frac{2\beta}{1+C}z + \frac{1-C}{1+C}}\right) \qquad \beta = \cos\Omega_0 \quad C = \tan(0.5\Delta\Omega) \tag{19.37}$$

If either design calls for an A-dB bandwidth of $\Delta\Omega$, the constant C is replaced by KC, where

$$K = \left(\frac{1}{10^{0.1A} - 1}\right)^{1/2} \qquad (A \text{ in dB}) \tag{19.38}$$

This is equivalent to denormailzing the lowpass prototype such that its gain corresponds to an attenuation of A decibels at unit radian frequency. For a 3-dB bandwidth, we obtain $K = 1$ as expected. These design relations prove quite helpful in the quick design of notch and peaking filters.

The center frequency Ω_0 is used to determine the parameter $\beta = \cos\Omega_0$. If only the band edges Ω_1 and Ω_2 are specified, but the center frequency Ω_0 is not, β may also be found from the alternative relation of Table 19.5 in terms of Ω_1 and Ω_2. The center frequency of the designed filter will then be based on the geometric symmetry of the prewarped frequencies and can be computed from

$$\tan(0.5\Omega_0) = \sqrt{\tan(0.5\Omega_1)\tan(0.5\Omega_2)} \tag{19.39}$$

In fact, the digital band edges Ω_1 and Ω_2 do not show geometric symmetry with respect to the center frequency Ω_0. We can find Ω_1 and Ω_2 in terms of $\Delta\Omega$ and Ω_0 by equating the two expressions for finding β (in Table 19.5) to obtain

$$\beta = \cos\Omega_0 = \frac{\cos[0.5(\Omega_2 + \Omega_1)]}{\cos[0.5(\Omega_2 - \Omega_1)]} \tag{19.40}$$

With $\Delta\Omega = \Omega_2 - \Omega_1$, we get

$$\Omega_2 = 0.5\Delta\Omega + \cos^{-1}[\cos\Omega_0 \cos(0.5\Delta\Omega)] \qquad \Omega_1 = \Omega_2 - \Delta\Omega \tag{19.41}$$

PGR2026-00033
Krisp Technologies EX1024 Page 711

EXAMPLE 19.9 (Bilinear Design of Second-Order Filters)

(a) Let us design a peaking filter with a 3-dB bandwidth of 5 kHz and a center frequency of 6 kHz. The sampling frequency is 25 kHz.

The digital frequencies are $\Delta\Omega = 2\pi(5/25) = 0.4\pi$ and $\Omega_0 = 2\pi(6/25) = 0.48\pi$.

We compute $C = \tan(0.5\Delta\Omega) = 0.7265$ and $\beta = \cos\Omega_0 = 0.0628$. Substituting these into the form for the required filter, we obtain

$$H(z) = \frac{0.4208(z^2 - 1)}{z^2 - 0.0727z + 0.1584}$$

Figure E19.9(a) shows the magnitude spectrum. The center frequency is 6 kHz, as expected. The band edges Ω_1 and Ω_2 may be computed from

$$\Omega_2 = 0.5\Delta\Omega + \cos^{-1}[\cos\Omega_0 \cos(0.5\Delta\Omega)] = 2.1483 \qquad \Omega_1 = \Omega_2 - \Delta\Omega = 0.8917$$

These correspond to the frequencies $f_1 = \frac{S\Omega_1}{2\pi} = 3.55$ kHz and $f_2 = \frac{S\Omega_2}{2\pi} = 8.55$ kHz.

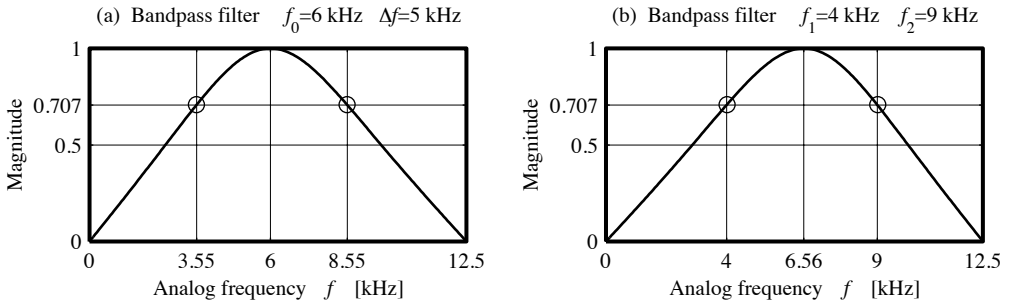


Figure E19.9 Response of bandpass filters for Example 19.9

(a and b)

(b) Let us design a peaking (bandpass) filter with a 3-dB band edges of 4 kHz and 9 kHz. The sampling frequency is 25 kHz.

The digital frequencies are $\Omega_1 = 2\pi(4/25) = 0.32\pi$, $\Omega_2 = 2\pi(6/25) = 0.72\pi$, and $\Delta\Omega = 0.4\pi$.

We find $C = \tan(0.5\Delta\Omega) = 0.7265$ and $\beta = \cos[0.5(\Omega_2+\Omega_1)]/\cos[0.5(\Omega_2-\Omega_1)] = -0.0776$. Substituting these into the form for the required filter, we obtain

$$H(z) = \frac{0.4208(z^2 - 1)}{z^2 + 0.0899z + 0.1584}$$

Figure E19.9(b) shows the magnitude spectrum. The band edges are at 4 kHz and 9 kHz as expected. The center frequency, however, is at 6.56 kHz. This is because the digital center frequency Ω_0 must be computed from $\beta = \cos\Omega_0 = -0.0776$. This gives $\Omega_0 = \cos^{-1}(-0.0776) = 1.6485$. This corresponds to $f_0 = S\Omega_0/(2\pi) = 6.5591$ kHz.

Comment: We could also have computed Ω_0 from

$$\tan(0.5\Omega_0) = \sqrt{\tan(0.5\Omega_1)\tan(0.5\Omega_2)} = 1.0809$$

Then, $\Omega_0 = 2\tan^{-1}(1.0809) = 1.6485$, as before.

PGR2026-00033
Krisp Technologies EX1024 Page 712

(c) We design a peaking filter with a center frequency of 40 Hz and a 6-dB bandwidth of 2 Hz, operating at a sampling rate of 200 Hz.

We compute $\Delta\Omega = 2\pi(2/200) = 0.02\pi$ and $\Omega_0 = 2\pi(40/200) = 0.4\pi$ and $\beta = \cos\Omega_0 = 0.3090$. Since we are given the 6-dB bandwidth, we compute K and C as follows:

$$K = \left(\frac{1}{10^{0.1A} - 1}\right)^{1/2} = \left(\frac{1}{10^{0.6} - 1}\right)^{1/2} = 0.577 \qquad C = K\tan(0.5\Delta\Omega) = 0.0182$$

Substituting these into the form for the required filter, we obtain

$$H(z) = \frac{0.0179(z^2 - 1)}{z^2 - 0.6070z + 0.9642}$$

Figure E19.9C shows the magnitude spectrum. The blowup reveals that the 6-dB bandwidth (where the gain is 0.5) equals 2 Hz, as required.

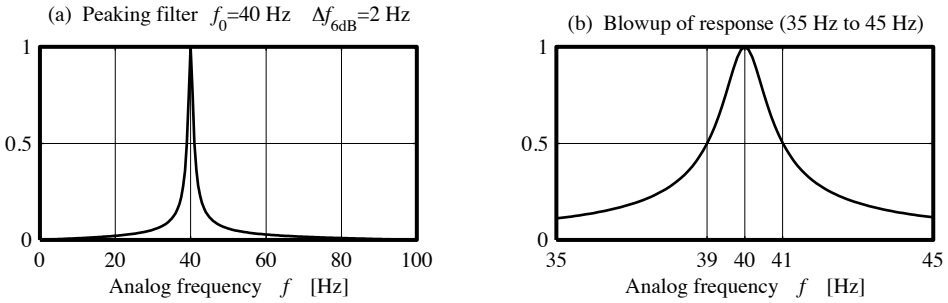


Figure E19.9C Response of peaking filter for Example 19.9(c)

EXAMPLE 19.10 (Interference Rejection)

We wish to design a filter to remove 60-Hz interference in an ECG signal sampled at 300 Hz. A 2-s recording of the noisy signal is shown in Figure E19.10A.

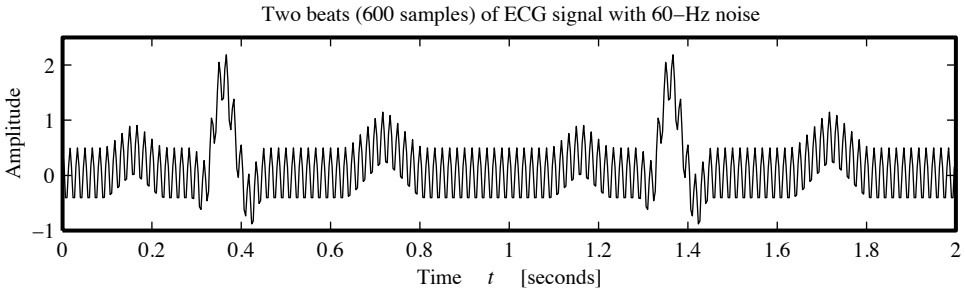


Figure E19.10A Simulated ECG signal with 60-Hz interference for Example 19.10

If we design a high-Q notch filter with $Q = 50$ and a notch at $f_0 = 60$ Hz, we have a notch bandwidth of $\Delta f = f_0/Q = 1.2$ Hz. The digital notch frequency is $\Omega_0 = 2\pi f_0/S = 2\pi(60/300) = 0.4\pi$, and the digital bandwidth is $\Delta\Omega = 2\pi\Delta f/S = 2\pi(1.2/300) = 0.008\pi$.

We find $C = \tan(0.5\Delta\Omega) = 0.0126$ and $\beta = \cos\Omega_0 = 0.3090$. Substituting these into the form for the notch filter, we obtain

$$H_1(z) = \frac{0.9876(z^2 - 0.6180z + 1)}{z^2 - 0.6104z + 0.9752}$$

PGR2026-00033
Krisp Technologies EX1024 Page 713

A low-Q design with $Q = 5$ gives $\Delta f = f_0/Q = 12$ Hz, and $\Delta\Omega = 2\pi(12/300) = 0.08\pi$.

We find $C = \tan(0.5\Delta\Omega) = 0.1263$ and, with $\beta = \cos\Omega_0 = 0.3090$ as before, we obtain

$$H_2(z) = \frac{0.8878(z^2 - 0.6180z + 1)}{z^2 - 0.5487z + 0.7757}$$

Figure E19.10B shows the magnitude spectrum of the two filters. Naturally, the filter $H_1(z)$ (with the higher Q) exhibits the sharper notch.

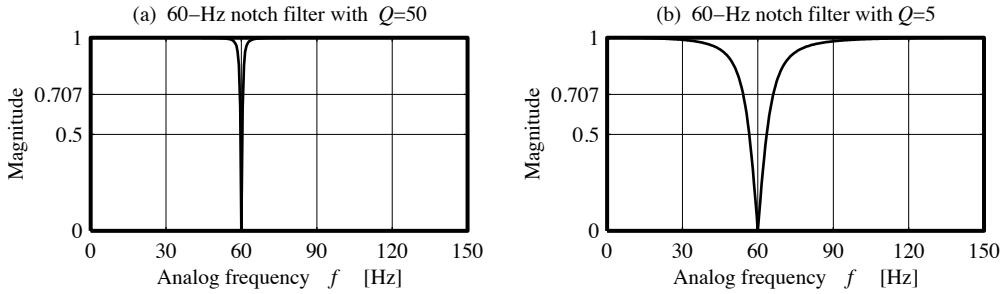


Figure E19.10B Response of the notch filters for Example 19.10

The filtered ECG signal corresponding to these two notch filters is shown in Figure E19.10C. Although both filters are effective in removing the 60-Hz noise, the filter $H_2(z)$ (with the lower Q) shows a much shorter start-up transient (because the highly oscillatory transient response of the high-Q filter $H_1(z)$ takes much longer to reach steady state).

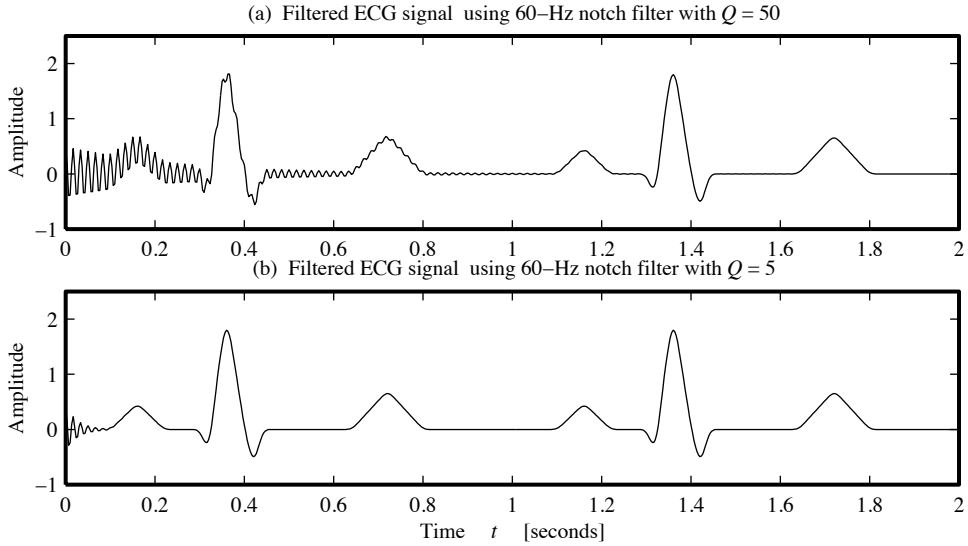


Figure E19.10C Output of the notch filters for Example 19.10

19.7.4 Application-Oriented Example: Audio Equalizers

Graphic equalizers permit adjustment of the tonal quality of the sound (in each channel of a stereo system) to suit the personal preference of a listener and represent the next step in sophistication, after tone controls.

PGR2026-00033
Krisp Technologies EX1024 Page 714

They employ a bank of (typically, second-order) bandpass filters covering the audio frequency spectrum, and with a fixed bandwidth and center frequency for each range. Only the gain of each filter can be adjusted by the user. Each filter isolates a selected frequency range and provides almost zero gain elsewhere. As a result, the individual sections are connected in *parallel*, as shown in Figure 19.10.

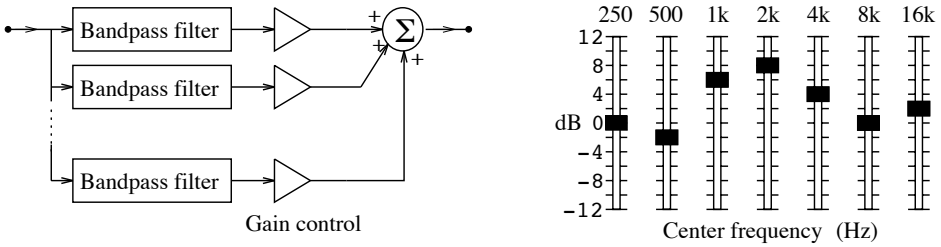


Figure 19.10 A graphic equalizer (left) and the display panel (right)

The input signal is split into as many channels as there are frequency bands, and the weighted sum of the outputs of each filter yields the equalized signal. A control panel, usually calibrated in decibels, allows for gain adjustment by sliders, as illustrated in Figure 19.10. The slider settings provide a rough visual indication of the equalized response, hence the name *graphic equalizer*. The design of each second-order bandpass section is based on its center frequency and its bandwidth or quality factor Q. A typical set of center frequencies for a ten-band equalizer is [31.5, 63, 125, 250, 500, 1000, 2000, 4000, 8000, 16000] Hz. A typical range for the gain of each filter is ± 12 dB (or 0.25 times to 4 times the nominal gain).

EXAMPLE 19.11 (A Digital Graphic Equalizer)

We design a five-band audio equalizer operating at a sampling frequency of 8192 Hz, with center frequencies of [120, 240, 480, 960, 1920] Hz. We select the 3-dB bandwidth of each section as 0.75 times its center frequency to cover the whole frequency range without overlap. This is equivalent to choosing each section with $Q = 4/3$. Each section is designed as a second-order bandpass IIR filter, using the bilinear transformation. Figure E19.11 shows the spectrum of each section and the overall response. The constant Q design becomes much more apparent when the gain is plotted in dB against $\log(f)$, as shown in Figure E19.11(b). A change in the gain of any section results in a change in the equalized response.

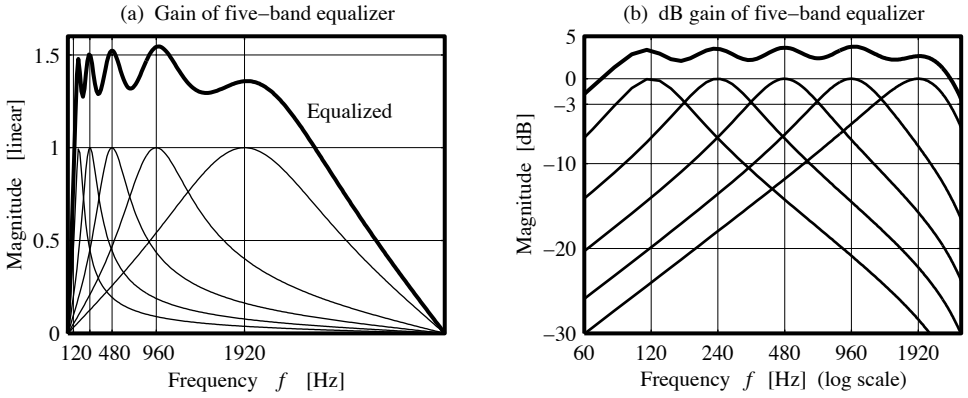


Figure E19.11 Frequency response of the audio equalizer for Example 19.11

PGR2026-00033
Krisp Technologies EX1024 Page 715

19.7.5 Parametric Equalizers

Parametric equalizers offer the ultimate in versatility and allow the user to vary not only the gain but also the filter parameters (such as the cutoff frequency, center frequency, and bandwidth). Each filter in a parametric equalizer affects only a selected portion of the spectrum (while providing unity gain elsewhere), and as a result, the individual sections are connected in *cascade*. Most parametric equalizers are based on second-order IIR peaking or notch filters whose transfer functions we reproduce from Section 19.7.3:

$$H_{\text{BP}}(z) = \frac{C}{1+C}\left(\frac{z^2-1}{z^2-\frac{2\beta}{1+C}z+\frac{1-C}{1+C}}\right) \qquad H_{\text{BS}}(z) = \frac{1}{1+C}\left(\frac{z^2-2\beta z+1}{z^2-\frac{2\beta}{1+C}z+\frac{1-C}{1+C}}\right) \tag{19.42}$$

The center frequency and bandwidth for the filters are given by

$$\Omega_0 = \cos^{-1}\beta \qquad \Delta\Omega = \cos^{-1}\left(\frac{2\alpha}{1+\alpha^2}\right), \quad \text{where} \quad \alpha = \frac{1-C}{1+C} \tag{19.43}$$

It is interesting to note that $H_{\text{BS}}(z) = 1 - H_{\text{BP}}(z)$. A tunable second-order equalizer stage $H_{\text{PAR}}(z)$ consists of the bandpass filter $H_{\text{BP}}(z)$, with a peak gain of G, in parallel with the bandstop filter $H_{\text{BS}}(z)$:

$$H_{\text{PAR}}(z) = GH_{\text{BP}}(z) + H_{\text{BS}}(z) = 1 + (G-1)H_{\text{BP}}(z) \tag{19.44}$$

A realization of this filter is shown in Figure 19.11.

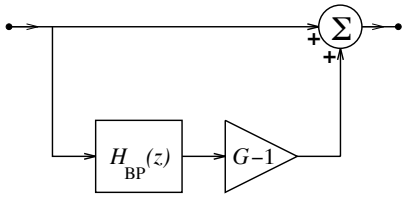


Figure 19.11 A second-order parametric equalizer

For $G = 1$, we have $H_{\text{PAR}} = 1$, and the gain is unity for all frequencies. The parameters α and β allow us to adjust the bandwidth and the center frequency, respectively. The response of this filter is shown for various values of these parameters in Figure 19.12.

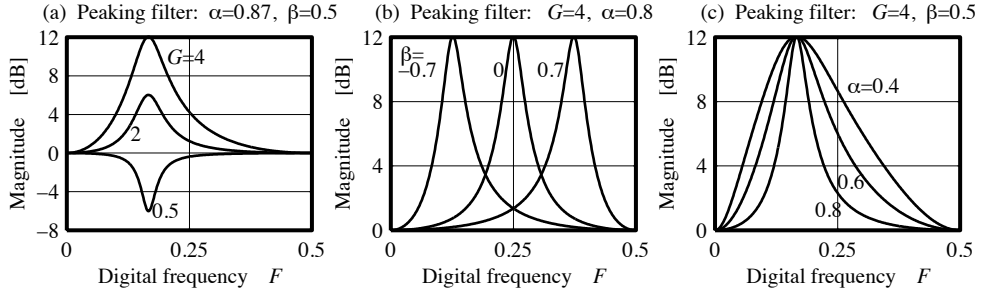


Figure 19.12 Typical response of a parametric equalizer

PGR2026-00033
Krisp Technologies EX1024 Page 716

19.8 Design Recipe for IIR Filters

There are several approaches to the design of IIR filters, using the mappings and spectral transformations described in this chapter. The first approach, illustrated in Figure 19.13, is based on developing the analog filter $H(s)$ followed by the required mapping to convert $H(s)$ to $H(z)$.

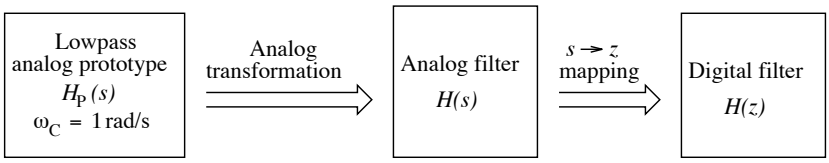


Figure 19.13 Converting an analog filter to a digital filter

A major disadvantage of this approach is that it cannot be used for mappings that suffer from aliasing problems (such as the impulse-invariant mapping) to design highpass or bandpass filters.

The second, indirect approach, illustrated in Figure 19.14, tries to overcome this problem by designing only the lowpass prototype $H_P(s)$ in the analog domain. This is followed by the required mapping to obtain a digital lowpass prototype $H_P(z)$. The final step is the spectral (D2D) transformation of $H_P(z)$ to the required digital filter $H(z)$.

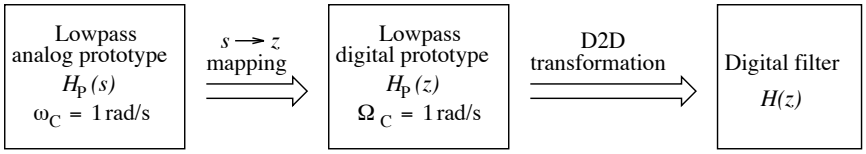


Figure 19.14 Indirect conversion of an analog filter to a digital filter

This approach allows us to use any mappings, including those (such as response invariance) that may otherwise lead to excessive aliasing for highpass and bandstop filters. Designing $H_P(z)$ also allows us to match its dc magnitude with $H_P(s)$ for subsequent comparison.

A third approach that applies only to the bilinear transformation is illustrated in Figure 19.15. We prewarp the frequencies, design an analog lowpass prototype (from *prewarped* specifications), and apply A2D transformations to obtain the required digital filter $H(z)$.

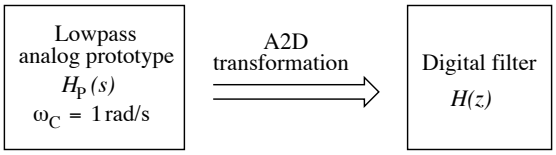


Figure 19.15 Conversion of analog lowpass prototype directly to a digital filter

A Step-by-Step Approach

Given the passband and stopband edges, the passband and stopband attenuation, and the sampling frequency S, a standard recipe for the design of IIR filters is as follows:

PGR2026-00033
Krisp Technologies EX1024 Page 717

1. Normalize (divide) the design band edges by S. This allows us to use a sampling interval $t_s = 1$ in subsequent design. For bilinear design, we also *prewarp* the normalized band edges.
2. Use the normalized band edges and attenuation specifications to design an *analog lowpass prototype* $H_P(s)$ whose cutoff frequency is $\omega_C = 1$ rad/s.
3. Apply the chosen mapping (with $t_s = 1$) to convert $H_P(s)$ to a *digital lowpass prototype* filter $H_P(z)$ with $\Omega_D = 1$.
4. Use D2D transformations (with $\Omega_D = 1$) to convert $H_P(z)$ to $H(z)$.
5. For bilinear design, we can also convert $H_P(s)$ to $H(z)$ directly (using A2D transformations).

REVIEW PANEL 19.12

Design Recipe for IIR Digital Filters

Normalize (divide) band edges by S (and prewarp if using bilinear design).
Use the normalized band edges to design *analog lowpass prototype* $H_P(s)$ with $\omega_C = 1$ rad/s.
Apply the chosen mapping (with $t_s = 1$) to convert $H_P(s)$ to $H_P(z)$ with $\Omega_D = 1$.
Use D2D transformations to convert $H_P(z)$ to $H(z)$.
For bilinear design, convert $H_P(s)$ to $H(z)$ (using A2D transformations).

EXAMPLE 19.12 (IIR Filter Design)

Design a Chebyshev IIR filter to meet the following specifications:
Passband edges at [1.8, 3.2] kHz, stopband edges at [1.6, 4.8] kHz, $A_p = 2$ dB, $A_s = 20$ dB, and sampling frequency $S = 12$ kHz.

(a) (Indirect Bilinear Design)

The normalized band edges $[\Omega_1,\ \Omega_2,\ \Omega_3,\ \Omega_4]$, in increasing order, are

$$[\Omega_1,\ \Omega_2,\ \Omega_3,\ \Omega_4] = 2\pi[1.6,\ 1.8,\ 3.2,\ 4.8]/12 = [0.84,\ 0.94,\ 1.68,\ 2.51]$$

The passband edges are $[\Omega_1, \Omega_2] = [0.94, 1.68]$. We choose $C = 2$ and prewarp each band-edge frequency using $\omega = 2\tan(0.5\Omega)$ to give the prewarped values [0.89, 1.019, 2.221, 6.155].
The prewarped passband edges are $[\Omega_{p1},\ \Omega_{p2}] = [1.019,\ 2.221]$.

We design an analog filter meeting the prewarped specifications. Such a filter was actually designed in Example 13.5(c), yielding the lowpass prototype and actual transfer function as

$$H_P(s) = \frac{0.1634}{s^4 + 0.7162s^3 + 1.2565s^2 + 0.5168s + 0.2058}$$

$$H_{\text{BP}}(s) = \frac{0.34s^4}{s^8 + 0.86s^7 + 10.87s^6 + 6.75s^5 + 39.39s^4 + 15.27s^3 + 55.69s^2 + 9.99s + 26.25}$$

Finally, we transform the bandpass filter $H_{\text{BP}}(s)$ to $H(z)$, using $s \to 2(z-1)/(z+1)$, to obtain

$$H(z) = \frac{0.0026z^8 - 0.0095z^6 + 0.0142z^4 - 0.0095z^2 + 0.0026}{z^8 - 1.94z^7 + 4.44z^6 - 5.08z^5 + 6.24z^4 - 4.47z^3 + 3.44z^2 - 1.305z + 0.59}$$

Figure E19.12A compares the response of the digital filter $H(z)$ with the analog filter $H_{\text{BP}}(s)$, and with a digital filter designed from the *unwarped* frequencies.

PGR2026-00033
Krisp Technologies EX1024 Page 718

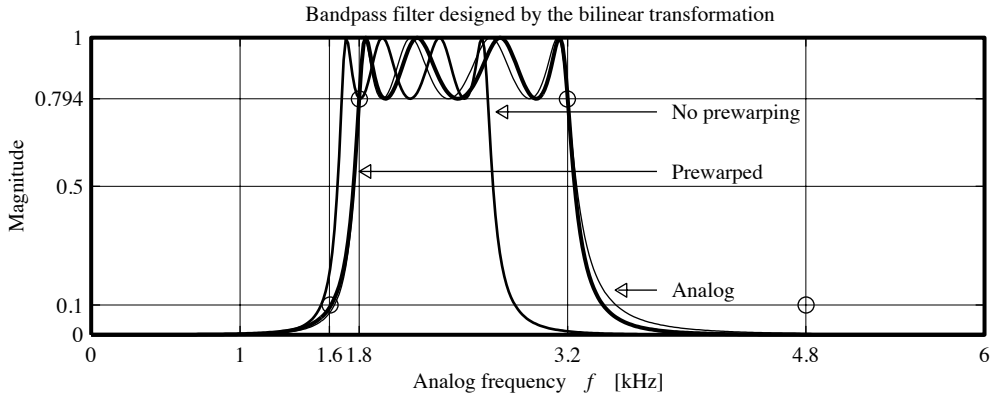


Figure E19.12A Bandpass filter for Example 19.12 designed by the bilinear transformation

---

(b) (Direct A2D Design)

For bilinear design, we can also convert $H_P(s)$ directly.

We use the A2D LP2BP transformation with the *unwarped* passband edges $[\Omega_1,\ \Omega_2] = [0.94,\ 1.68]$.

The constants C and β (from Table 19.5) are found as

$$C = \tan[0.5(\Omega_2 - \Omega_1)] = 0.3839 \qquad \beta = \frac{\cos[0.5(\Omega_2 + \Omega_1)]}{\cos[0.5(\Omega_2 - \Omega_1)]} = 0.277$$

We transform the prototype analog filter $H_P(s)$ to obtain

$$H(z) = \frac{0.0026z^8 - 0.0095z^6 + 0.0142z^4 - 0.0095z^2 + 0.0026}{z^8 - 1.94z^7 + 4.44z^6 - 5.08z^5 + 6.24z^4 - 4.47z^3 + 3.44z^2 - 1.305z + 0.59}$$

This expression is identical to the transfer function $H(z)$ of part (a).

(c) (Design Using Other Mappings)

We can also design the digital filter based on other mappings by using the following steps:

1. Use the normalized (but *unwarped*) band edges [0.89, 0.94, 1.68, 2.51] to design an analog lowpass prototype $H_P(s)$ with $\omega_C = 1$ rad/s (fortunately, the unwarped and prewarped specifications yield the same $H_P(s)$ for the specifications of this problem).
2. Convert $H_P(s)$ to $H_P(z)$ with $\Omega_D = 1$, using the chosen mapping, with $t_s = 1$. For the backward-difference, for example, we would use $s = (z-1)/zt_s = (z-1)/z$. To use the impulse-invariant mapping, we would have to first convert $H_P(s)$ to partial fraction form.
3. Convert $H_P(z)$ to $H(z)$, using the D2D LP2BP transformation with $\Omega_D = 1$, and the *unwarped* passband edges $[\Omega_1,\ \Omega_2] = [0.94,\ 1.68]$.

Figure E19.12C compares the response of two such designs, using the impulse-invariant mapping and the backward-difference mapping (both with gain matching at dc). The design based on the backward-difference mapping shows a poor match to the analog filter.

PGR2026-00033
Krisp Technologies EX1024 Page 719

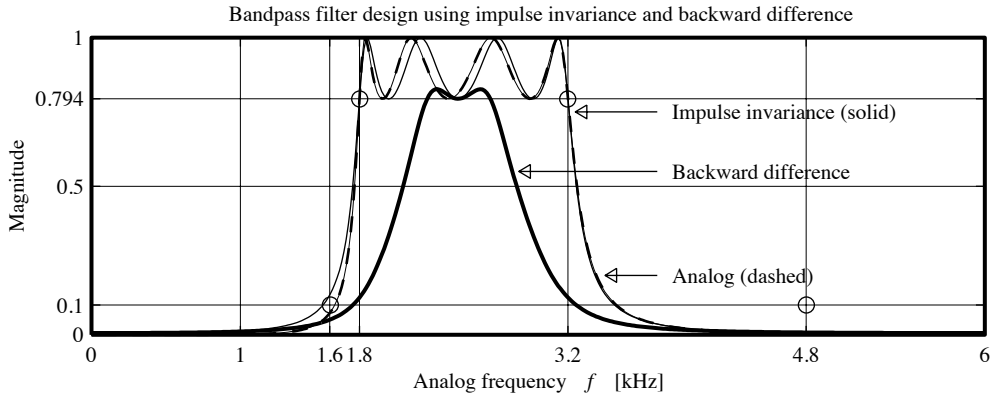


Figure E19.12C Bandpass filter for Example 19.12 designed by impulse invariance and backward difference

19.8.1 Effects of Coefficient Quantization

For IIR filters, the effects of rounding or truncating the filter coefficients can range from minor changes in the frequency response to serious problems, including instability. Consider the stable analog filter:

$$H(s) = \frac{(s+0.5)(s+1.5)}{(s+1)(s+2)(s+4.5)(s+8)(s+12)} \tag{19.45}$$

Bilinear transformation of $H(s)$ with $t_s = 0.01$ s yields the digital transfer function $H(z) = B(z)/A(z)$ whose denominator coefficients to double precision (A_k) and truncated to seven significant digits (A_k^t) are given by

Filter Coefficients A_k	**Truncated** A_k^t
1.144168420199997e+0	1.144168e+0
−5.418904483999996e+0	−5.418904e+0
1.026166736200000e+1	1.026166e+1
−9.712186808000000e+0	−9.712186e+0
4.594164261000004e+0	4.594164e+0
−8.689086648000011e−1	−8.689086e−1

The poles of $H(z)$ all lie within the unit circle, and the designed filter is thus stable. However, if we use the truncated coefficients A_k^t to compute the roots, the filter becomes unstable because one pole moves out of the unit circle! The bottom line is that stability is an important issue in the design of IIR digital filters.

19.8.2 Concluding Remarks

IIR filters are well suited to applications requiring frequency-selective filters with sharp cutoffs or where linear phase is relatively unimportant. Examples include graphic equalizers for digital audio, tone generators for digital touch-tone receivers, and filters for digital telephones. The main advantages are standardized easy design, and low filter order. On the other hand, IIR filters cannot exhibit linear phase and are quite susceptible to the effects of coefficient quantization. If linear phase is important as in biomedical signal processing, or stability is paramount as in many adaptive filtering schemes, it is best to use FIR filters.

PGR2026-00033
Krisp Technologies EX1024 Page 720

CHAPTER 19 PROBLEMS

DRILL AND REINFORCEMENT

19.1 (Response Invariance) Consider the analog filter $H(s) = \dfrac{1}{s+2}$.

(a) Convert $H(s)$ to a digital filter $H(z)$, using impulse invariance. Assume that the sampling frequency is $S = 2$ Hz.

(b) Will the impulse response $h[n]$ match the impulse response $h(t)$ of the analog filter at the sampling instants? Should it? Explain.

(c) Will the step response $s[n]$ match the step response $s(t)$ of the analog filter at the sampling instants? Should it? Explain.

19.2 (Response Invariance) Consider the analog filter $H(s) = \dfrac{1}{s+2}$.

(a) Convert $H(s)$ to a digital filter $H(z)$, using step invariance at a sampling frequency of $S = 2$ Hz.

(b) Will the impulse response $h[n]$ match the impulse response $h(t)$ of the analog filter at the sampling instants? Should it? Explain.

(c) Will the step response $s[n]$ match the step response $s(t)$ of the analog filter at the sampling instants? Should it? Explain.

19.3 (Response Invariance) Consider the analog filter $H(s) = \dfrac{1}{s+2}$.

(a) Convert $H(s)$ to a digital filter $H(z)$, using ramp invariance at a sampling frequency of $S = 2$ Hz.

(b) Will the impulse response $h[n]$ match the impulse response $h(t)$ of the analog filter at the sampling instants? Should it? Explain.

(c) Will the step response $s[n]$ match the step response $s(t)$ of the analog filter at the sampling instants? Should it? Explain.

(d) Will the response $v[n]$ to a unit ramp match the unit-ramp response $v(t)$ of the analog filter at the sampling instants? Should it? Explain.

19.4 (Response Invariance) Consider the analog filter $H(s) = \dfrac{s+1}{(s+1)^2 + \pi^2}$.

(a) Convert $H(s)$ to a digital filter $H(z)$, using impulse invariance. Assume that the sampling frequency is $S = 2$ Hz.

(b) Convert $H(s)$ to a digital filter $H(z)$, using invariance to the input $x(t) = e^{-t}u(t)$ at a sampling frequency of $S = 2$ Hz.

19.5 (Impulse Invariance) Use the impulse-invariant transformation with $t_s = 1$ s to transform the following analog filters to digital filters.

(a) $H(s) = \dfrac{1}{s+2}$ **(b)** $H(s) = \dfrac{2}{s+1} + \dfrac{2}{s+2}$ **(c)** $H(s) = \dfrac{1}{(s+1)(s+2)}$

19.6 (Impulse-Invariant Design) We are given the analog lowpass filter $H(s) = \dfrac{1}{s+1}$ whose cutoff frequency is known to be 1 rad/s. It is required to use this filter as the basis for designing a digital filter by the impulse-invariant transformation. The digital filter is to have a cutoff frequency of 50 Hz and operate at a sampling frequency of 200 Hz.

PGR2026-00033
Krisp Technologies EX1024 Page 721

(a) What is the transfer function $H(z)$ of the digital filter if no gain matching is used?
(b) What is the transfer function $H(z)$ of the digital filter if the gain of the analog filter and digital filter are matched at dc? Does the gain of the two filters match at their respective cutoff frequencies?
(c) What is the transfer function $H(z)$ of the digital filter if the gain of the analog filter at its cutoff frequency (1 rad/s) is matched to the gain of the digital filter at its cutoff frequency (50 Hz)? Does the gain of the two filters match at dc?

19.7 (Matched z-Transform) Use the matched z-transform $s + \alpha \Longrightarrow \dfrac{z - e^{-\alpha t_s}}{z}$, with $t_s = 0.5$ s and gain matching at dc, to transform each analog filter $H(z)$ to a digital filter $H(z)$.

(a) $H(s) = \dfrac{1}{s+2}$ **(b)** $H(s) = \dfrac{1}{(s+1)(s+2)}$
(c) $H(s) = \dfrac{1}{s+1} + \dfrac{1}{s+2}$ **(d)** $H(s) = \dfrac{s+1}{(s+1)^2 + \pi^2}$

19.8 (Backward Euler Algorithm) The backward Euler algorithm for numerical integration is given by $y[n] = y[n-1] + t_s x[n]$.
(a) Derive a mapping rule for converting an analog filter to a digital filter, based on this algorithm.
(b) Apply the mapping to convert the analog filter $H(s) = \dfrac{4}{s+4}$ to a digital filter $H(z)$, using a sampling interval of $t_s = 0.5$ s.

19.9 (Mapping from Difference Algorithms) Consider the analog filter $H(s) = \dfrac{1}{s+\alpha}$.
(a) For what values of α is this filter stable?
(b) Convert $H(s)$ to a digital filter $H(z)$, using the mapping based on the forward difference at a sampling rate S. Is $H(z)$ always stable if $H(s)$ is stable?
(c) Convert $H(s)$ to a digital filter $H(z)$, using the mapping based on the backward difference at a sampling rate S. Is $H(z)$ always stable if $H(s)$ is stable?

19.10 (Bilinear Transformation) Consider the lowpass analog Bessel filter $H(s) = \dfrac{3}{s^2 + 3s + 3}$.
(a) Use the bilinear transformation to convert this analog filter $H(s)$ to a digital filter $H(z)$ at a sampling rate of $S = 2$ Hz.
(b) Use $H(s)$ and the bilinear transformation to design a digital lowpass filter $H(z)$ whose gain at $f = 20$ kHz matches the gain of $H(s)$ at $\omega = 3$ rad/s. The sampling frequency is $S = 80$ kHz.

19.11 (Bilinear Transformation) Consider the analog filter $H(s) = \dfrac{s}{s^2 + s + 1}$.
(a) What type of filter does $H(s)$ describe?
(b) Use $H(s)$ and the bilinear transformation to design a digital filter $H(z)$ operating at $S = 1$ kHz such that its gain at 250 Hz matches the gain of $H(s)$ at $\omega = 1$ rad/s. What type of filter does $H(z)$ describe?
(c) Use $H(s)$ and the bilinear transformation to design a digital filter $H(z)$ operating at $S = 10$ Hz such that gains of $H(z)$ and $H(s)$ match at 1 Hz. What type of filter does $H(z)$ describe?

19.12 (Spectral Transformation of Digital Filters) The digital lowpass filter $H(z) = \dfrac{z+1}{z^2 - z + 0.2}$ has a cutoff frequency $f = 0.5$ kHz and operates at a sampling frequency $S = 10$ kHz. Use this filter to design the following:

PGR2026-00033
Krisp Technologies EX1024 Page 722

(a) A lowpass digital filter with a cutoff frequency of 2 kHz
(b) A highpass digital filter with a cutoff frequency of 1 kHz
(c) A bandpass digital filter with band edges of 1 kHz and 3 kHz
(d) A bandstop digital filter with band edges of 1.5 kHz and 3.5 kHz

19.13 (Spectral Transformation of Analog Prototypes) The analog lowpass filter $H(s) = \dfrac{2}{s^2 + 2s + 2}$ has a cutoff frequency of 1 rad/s. Use this prototype to design the following digital filters.
(a) A lowpass filter with a passband edge of 100 Hz and $S = 1$ kHz
(b) A highpass filter with a cutoff frequency of 500 Hz and $S = 2$ kHz
(c) A bandpass filter with band edges at 400 Hz and 800 Hz, and $S = 3$ kHz
(d) A bandstop filter with band edges at 1 kHz and 1200 Hz, and $S = 4$ kHz

19.14 (IIR Filter Design) Design IIR filters that meet each of the following sets of specifications. Assume a passband attenuation of $A_p = 2$ dB and a stopband attenuation of $A_s = 30$ dB.
(a) A Butterworth lowpass filter with passband edge at 1 kHz, stopband edge at 3 kHz, and $S = 10$ kHz, using the backward Euler transformation.
(b) A Butterworth highpass filter with passband edge at 400 Hz, stopband edge at 100 Hz, and $S = 2$ kHz, using the impulse-invariant transformation.
(c) A Chebyshev bandpass filter with passband edges at 800 Hz and 1600 Hz, stopband edges at 400 Hz and 2 kHz, and $S = 5$ kHz, using the bilinear transformation.
(d) An inverse Chebyshev bandstop filter with passband edges at 200 Hz and 1.2 kHz, stopband edges at 500 Hz and 700 Hz, and $S = 4$ kHz, using the bilinear transformation.

19.15 (Group Delay) A digital filter $H(z)$ is designed from an analog filter $H(s)$, using the bilinear transformation $\omega_a = C\tan(\Omega_d/2)$.
(a) Show that the group delays T_a and T_d of $H(s)$ and $H(z)$ are related by

$$T_d = 0.5C(1 + \omega_a^2)T_a$$

(b) Design a digital filter $H(z)$ from the analog filter $H(s) = \dfrac{5}{s+5}$ at a sampling frequency of $S = 4$ Hz such that the gain of $H(s)$ at $\omega = 2$ rad/s matches the gain of $H(z)$ at 1 Hz.
(c) What is the group delay T_a of the analog filter $H(s)$?
(d) Use the above results to find the group delay T_d of the digital filter $H(z)$ designed in part (b).

19.16 (Digital-to-Analog Mappings) The bilinear transformation allows us to use a linear mapping to transform a digital filter $H(z)$ to an analog equivalent $H(s)$.
(a) Develop such a mapping based on the bilinear transformation.
(b) Use this mapping to convert a digital filter $H(z) = \dfrac{z}{z - 0.5}$ operating at $S = 2$ Hz to its analog equivalent $H(s)$.

REVIEW AND EXPLORATION

19.17 (Impulse Invariance) The impulse-invariant method allows us to take a digital filter described by $H_1(z) = \dfrac{z}{z - \alpha t_1}$ at a sampling interval of t_1 and convert this to a new digital filter $H_2(z) = \dfrac{z}{z - \alpha t_2}$ at a different sampling interval t_2.

PGR2026-00033
Krisp Technologies EX1024 Page 723

(a) Using the fact that $s = \dfrac{\ln(\alpha t_1)}{t_1} = \dfrac{\ln(\alpha t_2)}{t_2}$, show that $\alpha t_2 = (\alpha t_1)^M$, where $M = t_2/t_1$.

(b) Use the result of part (a) to convert the digital filter $H_1(z) = \dfrac{z}{z-0.5} + \dfrac{z}{z-0.25}$, with $t_s = 1\,\text{s}$, to a digital filter $H_2(z)$, with $t_s = 0.5\,\text{s}$.

19.18 (Matched *z*-Transform) The analog filter $H(s) = \dfrac{4s(s+1)}{(s+2)(s+3)}$ is to be converted to a digital filter $H(z)$ at a sampling rate of $S = 4$ Hz.

(a) Convert $H(s)$ to a digital filter, using the matched z-transform $s + \alpha \Longrightarrow \dfrac{z - e^{-\alpha t_s}}{z}$.

(b) Convert $H(s)$ to a digital filter, using the *modified* matched z-transform by moving all zeros at the origin ($z = 0$) to $z = -1$.

(c) Convert $H(s)$ to a digital filter, using the *modified* matched z-transform, by moving *all but one* zero at the origin ($z = 0$) to $z = -1$.

19.19 (Simpson's Algorithm) Simpson's numerical integration algorithm is described by

$$y[n] = y[n-2] + \tfrac{t_s}{3}(x[n] + 4x[n-1] + x[n-2])$$

(a) Derive a mapping rule to convert an analog filter $H(s)$ to a digital filter $H(z)$, based on this algorithm.

(b) Let $H(s) = \dfrac{1}{s+1}$. Convert $H(s)$ to $H(z)$, using the mapping derived in part (a).

(c) Is the filter $H(z)$ designed in part (b) stable for any choice of $t_s > 0$?

19.20 (Bilinear Transformation) A second-order Butterworth lowpass analog filter with a half-power frequency of 1 rad/s is converted to a digital filter $H(z)$, using the bilinear transformation at a sampling rate of $S = 1$ Hz.

(a) What is the transfer function $H(s)$ of the analog filter?

(b) What is the transfer function $H(z)$ of the digital filter?

(c) Are the dc gains of $H(z)$ and $H(s)$ identical? Should they be? Explain.

(d) Are the dc gains $H(z)$ and $H(s)$ at their respective half-power frequencies identical? Explain.

19.21 (Digital-to-Analog Mappings) In addition to the bilinear transformation, the backward Euler method also allows a linear mapping to transform a digital filter $H(z)$ to an analog equivalent $H(s)$.

(a) Develop such a mapping based on the backward Euler algorithm.

(b) Use this mapping to convert a digital filter $H(z) = \dfrac{z}{z-0.5}$ operating at $S = 2$ Hz to its analog equivalent $H(s)$.

19.22 (Digital-to-Analog Mappings) The forward Euler method also allows a linear mapping to transform a digital filter $H(z)$ to an analog equivalent $H(s)$.

(a) Develop such a mapping based on the forward Euler algorithm.

(b) Use this mapping to convert a digital filter $H(z) = \dfrac{z}{z-0.5}$ operating at $S = 2$ Hz to its analog equivalent $H(s)$.

PGR2026-00033
Krisp Technologies EX1024 Page 724

19.23 (Digital-to-Analog Mappings) Two other methods that allow us to convert a digital filter $H(z)$ to an analog equivalent $H(z)$ are impulse invariance and the matched z-transform $s+\alpha \Longrightarrow \dfrac{z-e^{-\alpha t_s}}{z}$. Let $H(z)=\dfrac{z(z+1)}{(z-0.25)(z-0.5)}$. Find the analog filter $H(s)$ from which $H(z)$ was developed, assuming a sampling frequency of $S=2$ Hz and

(a) Impulse invariance. **(b)** Matched z-transform. **(c)** Bilinear transformation.

19.24 (Filter Specifications) A hi-fi audio signal band-limited to 20 kHz is contaminated by high-frequency noise between 70 kHz and 110 kHz. We wish to design a digital filter to reduce the noise by a factor of 100, with no appreciable signal loss. Pick the minimum sampling frequency that avoids aliasing of the noise spectrum into the signal spectrum and develop the specifications for the digital filter.

19.25 (IIR Filter Design) A digital filter is required to have a monotonic response in the passband and stopband. The half-power frequency is to be 4 kHz, and the attenuation past 5 kHz is to exceed 20 dB. Design the digital filter, using impulse invariance and a sampling frequency of 15 kHz.

19.26 (Notch Filters) A notch filter is required to remove 50-Hz interference. Design such a filter, using the bilinear transformation and assuming a bandwidth of 4 Hz and a sampling rate of 300 Hz. Compute the gain of this filter at 40 Hz, 50 Hz, and 60 Hz.

19.27 (Peaking Filters) A peaking filter is required to isolate a 100-Hz signal with unit gain. Design this filter, using the bilinear transformation and assuming a bandwidth of 5 Hz and a sampling rate of 500 Hz. Compute the gain of this filter at 90 Hz, 100 Hz, and 110 Hz.

19.28 (IIR Filter Design) A fourth-order digital filter is required to have a passband between 8 kHz and 12 kHz and a maximum passband ripple that equals 5% of the peak magnitude. Design the digital filter, using the bilinear transformation if the sampling frequency is assumed to be 40 kHz.

19.29 (IIR Filter Design) Lead-lag systems are often used in control systems and have the generic form $H(s)=\dfrac{1+s\tau_1}{1+s\tau_2}$. Use a sampling frequency of $S=10$ Hz and the bilinear transformation to design IIR filters from this lead-lag compensator if

(a) $\tau_1=1$ s, $\tau_2=10$ s. **(b)** $\tau_1=10$ s, $\tau_2=1$ s.

19.30 (Pade Approximations) A delay of t_s may be approximated by $e^{-st_s}\approx 1-st_s+\dfrac{(st_s)^2}{2!}-\cdots$. An nth-order Pade approximation is based on a rational function of order n that minimizes the truncation error of this approximation. The first-order and second-order Pade approximations are

$$P_1(s)=\frac{1-\frac{1}{2}st_s}{1+\frac{1}{2}st_s} \qquad P_2(s)=\frac{1-\frac{1}{2}st_s+\frac{1}{12}(st_s)^2}{1+\frac{1}{2}st_s+\frac{1}{12}(st_s)^2}$$

Since e^{-st_s} describes a delay of one sample (or z^{-1}), Pade approximations can be used to generate inverse mappings for converting a digital filter $H(z)$ to an analog filter $H(s)$.

(a) Generate mappings for converting a digital filter $H(z)$ to an analog filter $H(s)$ based on the first-order and second-order Pade approximations.

(b) Use each mapping to convert $H(z)=\dfrac{z}{z-0.5}$ to $H(s)$, assuming $t_s=0.5$ s.

(c) Show that the first-order mapping is bilinear. Is this mapping related in any way to the bilinear transformation?

PGR2026-00033
Krisp Technologies EX1024 Page 725

COMPUTATION AND DESIGN

diirgui A GUI for IIR Filter Design and Visualization

The graphical user interface `diirgui` allows you to design and visualize IIR filters.
You can select from various forms (highpass, lowpass, etc.), types (Butterworth, elliptic, etc.) and mapping methods (impulse invariant, bilinear transformation, etc.).
The plots are displayed along with the filter order and design attenuation at the band edges.
To explore this routine, type `diirgui` at the MATLAB prompt.

19.31 (IIR Filter Design) It is required to design a lowpass digital filter $H(z)$ from the analog filter $H(s) = \frac{1}{s+1}$. The sampling rate is $S = 1$ kHz. The half-power frequency of $H(z)$ is to be $\Omega_C = \pi/4$.

(a) Use impulse invariance to design $H(z)$ such that gain of the two filters matches at dc. Compare the frequency response of both filters (after appropriate frequency scaling). Which filter would you expect to yield better performance? To confirm your expectations, define the (in-band) signal to (out-of-band) noise ratio (SNR) in dB as

$$\text{SNR } = 20\log\left(\frac{\text{signal level}}{\text{noise level}}\right) \text{ dB}$$

(b) What is the SNR at the input and output of each filter if the input is $x(t) = \cos(0.2\omega_C t) + \cos(1.2\omega_C t)$ for $H(s)$ and $x[n] = \cos(0.2n\Omega_C) + \cos(1.2n\Omega_C)$ for $H(z)$?

(c) What is the SNR at the input and output of each filter if the input is $x(t) = \cos(0.2\omega_C t) + \cos(3\omega_C t)$ for $H(s)$ and $x[n] = \cos(0.2n\Omega_C) + \cos(3n\Omega_C)$ for $H(z)$?

(d) Use the bilinear transformation to design another filter $H_1(z)$ such that gain of the two filters matches at dc. Repeat the computations of parts (a) and (b) for this filter. Of the two digital filters $H(z)$ and $H_1(z)$, which one would you recommend using, and why?

19.32 (The Effect of Group Delay) The nonlinear phase of IIR filters is responsible for signal distortion. Consider a lowpass filter with a 1-dB passband edge at $f = 1$ kHz, a 50-dB stopband edge at $f = 2$ kHz, and a sampling frequency of $S = 10$ kHz.

(a) Design a Butterworth filter $H_B(z)$ and an elliptic filter $H_E(z)$ to meet these specifications. Using the MATLAB routine `grpdelay` (or otherwise), compute and plot the group delay of each filter. Which filter has the lower order? Which filter has a more nearly constant group delay in the passband? Which filter would cause the least phase distortion in the passband? What are the group delays N_B and N_E (expressed as the number of samples) of the two filters?

(b) Generate the signal $x[n] = 3\sin(0.03n\pi) + \sin(0.09n\pi) + 0.6\sin(0.15n\pi)$ over $0 \le n \le 100$. Use the ADSP routine `filter` to compute the response $y_B[n]$ and $y_E[n]$ of each filter. Plot the filter outputs $y_B[n]$ and $y_E[n]$ (delayed by N_B and N_E, respectively) and the input $x[n]$ on the same plot to compare results. Does the filter with the more nearly constant group delay also result in smaller signal distortion?

(c) Are all the frequency components of the input signal in the filter passband? If so, how can you justify that the distortion is caused by the nonconstant group delay and not by the filter attenuation in the passband?

19.33 (LORAN) A LORAN (long-range radio and navigation) system for establishing positions of marine craft uses three transmitters that send out short bursts (10 cycles) of 100-kHz signals in a precise phase

PGR2026-00033
Krisp Technologies EX1024 Page 726

relationship. Using phase comparison, a receiver (on the craft) can establish the position (latitude and longitude) of the craft to within a few hundred meters. Suppose the LORAN signal is to be digitally processed by first sampling it at 500 kHz and filtering the sampled signal using a second order peaking filter with a half-power bandwidth of 100 Hz. Use the bilinear transformation to design the filter from an analog filter with unit half-power frequency. Compare your design with the digital filter designed in Problem 18.34 (to meet the same specifications).

19.34 (Decoding a Mystery Message) During transmission, a message signal gets contaminated by a low-frequency signal and high-frequency noise. The message can be decoded only by displaying it in the time domain. The contaminated signal $x[n]$ is provided on disk as `mystery1.mat`. Load this signal into MATLAB (use the command `load mystery1`). In an effort to decode the message, try the following methods and determine what the decoded message says.

(a) Display the contaminated signal. Can you "read" the message? Display the DFT of the signal to identify the range of the message spectrum.

(b) Use the bilinear transformation to design a second-order IIR bandpass filter capable of extracting the message spectrum. Filter the contaminated signal and display the filtered signal to decode the message. Use both the `filter` (filtering) and `filifilt` (zero-phase filtering) commands.

(c) As an alternative method, first zero out the DFT component corresponding to the low-frequency contamination and obtain the IDFT $y[n]$. Next, design a lowpass IIR filter (using impulse invariance) to reject the high-frequency noise. Filter the signal $y[n]$ and display the filtered signal to decode the message. Use both the `filter` and `filtfilt` commands.

(d) Which of the two methods allows better visual detection of the message? Which of the two filtering routines (in each method) allows better visual detection of the message?

19.35 (Interpolation) The signal $x[n] = \cos(2\pi F_0 n)$ is to be interpolated by 5 using up-sampling followed by lowpass filtering. Let $F_0 = 0.4$.

(a) Generate and plot 20 samples of $x[n]$ and up-sample by 5.

(b) What must be the cutoff frequency F_C and gain A of a lowpass filter that follows the up-sampler to produce the interpolated output?

(c) Design a fifth-order digital Butterworth filter (using the bilinear transformation) whose half-power frequency equals F_C and whose peak gain equals A.

(d) Filter the up-sampled signal through this filter and plot the result. Is the result an interpolated version of the input signal? Do the peak amplitudes of the interpolated signal and original signal match? Should they? Explain.

19.36 (Coefficient Quantization) Consider the analog filter described by

$$H(s) = \frac{(s+0.5)(s+1.5)}{(s+1)(s+2)(s+4.5)(s+8)(s+12)}$$

(a) Is this filter stable? Why?

(b) Use the bilinear transformation with $S = 100$ Hz to convert this to a digital filter $H(z)$.

(c) Truncate the filter coefficients to seven significant digits to generate the filter $H_2(z)$.

(d) Compare the frequency response of $H(z)$ and $H_2(z)$. Are there any significant differences?

(e) Is the filter $H_2(z)$ stable? Should it be? Explain.

(f) Suppose the coefficients are to be quantized to B bits by rounding. What is the smallest number of bits B required in order to preserve the stability of the quantized filter?

PGR2026-00033
Krisp Technologies EX1024 Page 727

19.37 (Numerical Integration Algorithms) It is claimed that mapping rules to convert an analog filter to a digital filter, based on numerical integration algorithms that approximate the area $y[n]$ from $y[n-2]$ or $y[n-3]$ (two or more time steps away), do not usually preserve stability. Consider the following integration algorithms.

(1) $y[n] = y[n-1] + \frac{t_s}{12}(5x[n] + 8x[n-1] - x[n-2])$ (Adams-Moulton rule)

(2) $y[n] = y[n-2] + \frac{t_s}{3}(x[n] + 4x[n-1] + x[n-2])$ (Simpson's rule)

(3) $y[n] = y[n-3] + \frac{3t_s}{8}(x[n] + 3x[n-1] + 3x[n-2] + x[n-3])$ (Simpson's three-eighths rule)

Derive mapping rules for each algorithm, convert the analog filter $H(s) = \dfrac{1}{s+1}$ to a digital filter using each mapping with $S = 5$ Hz, and use MATLAB to compare their frequency response. Which of these mappings (if any) allow us to convert a stable analog filter to a stable digital filter? Is the claim justified?

19.38 (RIAA Equalization) Audio signals usually undergo a high-frequency boost (and low-frequency cut) before being used to make the master for commercial production of phonograph records. During playback, the signal from the phono cartridge is fed to a preamplifier (equalizer) that restores the original signal. The frequency response of the preamplifier is based on the RIAA (Recording Industry Association of America) equalization curve whose Bode plot is shown in Figure P19.38, with break frequencies at 50, 500, and 2122 Hz.

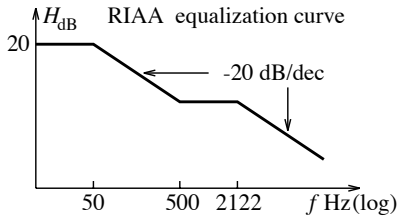


Figure P19.38 Figure for Problem 19.38

(a) What is the transfer function $H(s)$ of the RIAA equalizer?

(b) It is required to implement RIAA equalization using a digital filter. Assume that the signal from the cartridge is band-limited to 15 kHz. Design an IIR filter $H_I(z)$, using impulse invariance, that implements the equalization characteristic.

(c) Use the bilinear transformation to design an IIR filter $H_B(z)$ that implements the equalization characteristic. Assume that the gain of $H(s)$ and $H_B(z)$ are to match at 1 kHz.

(d) Compare the performance of $H(s)$ with $H_B(z)$ and $H_I(z)$ and comment on the results. Which IIR design method results in better implementation?

19.39 (An Audio Equalizer) Many hi-fi systems are equipped with graphic equalizers to tailor the frequency response. Consider the design of a four-band graphic equalizer. The first section is to be a lowpass filter with a passband edge of 300 Hz. The next two sections are bandpass filters with passband edges of [300, 1000] Hz and [1000, 3000] Hz, respectively. The fourth section is a highpass filter with a stopband edge of 3 kHz. Use a sampling rate of 20 kHz and the bilinear transformation to design second-order IIR Butterworth filters to implement these filters. Superimpose plots of the frequency response of each section and their parallel combination. Explain how you might adjust the gain of each section to vary over ± 12 dB.

PGR2026-00033
Krisp Technologies EX1024 Page 728

Chapter 20

FIR DIGITAL FILTERS

20.0 Scope and Objectives

FIR filters are always stable and can be designed with linear phase, and hence no phase distortion. Their design is typically based on selecting a symmetric impulse response sequence whose length is chosen to meet design specifications. For given specifications, FIR filters require many more elements in their realization than do IIR filters. This chapter starts with an introduction to symmetric sequences and their application to FIR filters. It then describes various methods of FIR filter design, including the window method and frequency sampling. It introduces special-purpose FIR filters such as differentiators and Hilbert transformers, and concludes with applications to multistage interpolation/decimation and adaptive signal processing.

20.1 Symmetric Sequences and Linear Phase

Symmetric sequences possess linear phase and result in a *constant delay* with *no amplitude distortion*. This is an important consideration in filter design. The DTFT of a real, even symmetric sequence $x[n]$ is of the form $H(F) = A(F)$ and always real, and the DTFT of a real, odd symmetric sequence is of the form $H(F) = jA(F)$ and purely imaginary. Symmetric sequences also imply noncausal filters. However, such filters can be made causal if the impulse response is suitably delayed. A time shift of $x[n]$ to $x[n - M]$ introduces only a linear phase of $\phi(F) = e^{-j2M\pi F}$. The DTFT of sequences that are symmetric about their midpoint is said to possess **generalized linear phase**. Generalized linear phase is illustrated in Figure 20.1.

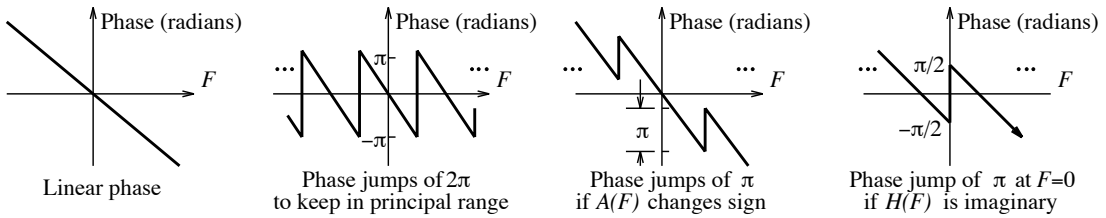


Figure 20.1 Examples of linear phase and generalized linear phase

The term *generalized* means that $\phi(F)$ may include a jump (of π at $F = 0$, if $H(F)$ is imaginary). There may also be phase jumps of 2π (if the phase is restricted to the principal range $\pi \leq \phi(F) \leq \pi$). If we plot the magnitude $|H(F)|$, there will also be phase jumps of π (where the amplitude $A(F)$ changes sign).

REVIEW PANEL 20.1

Linear-Phase Filters Provide a Constant Group Delay

PGR2026-00033
Krisp Technologies EX1024 Page 729

20.1.1 Classification of Linear-Phase Sequences

The length N of finite symmetric sequences can be odd or even, since the center of symmetry may fall on a sample point (for odd N) or midway between samples (for even N). This results in four possible types of symmetric sequences.

Type 1 Sequences

A type 1 sequence $h_1[n]$ and its amplitude spectrum $A_1(F)$ are illustrated in Figure 20.2.

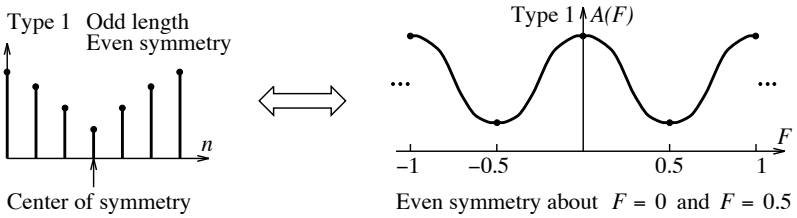


Figure 20.2 Features of a type 1 symmetric sequence

This sequence is even symmetric with odd length N, and a center of symmetry at the integer value $M = (N-1)/2$. Using Euler's relation, its frequency response $H_1(F)$ may be expressed as

$$H_1(F) = \left[h[M] + 2 \sum_{k=0}^{M-1} h[k] \cos[(M-k)2\pi F] \right] e^{-j2\pi MF} = A_1(F) e^{-j2\pi MF} \tag{20.1}$$

Thus, $H_1(F)$ shows a linear phase of $-2\pi MF$, and a constant group delay of M. The amplitude spectrum $A_1(F)$ is even symmetric about both $F = 0$ and $F = 0.5$, and both $|H_1(0)|$ and $|H_1(0.5)|$ can be nonzero.

Type 2 Sequences

A type 2 sequence $h_2[n]$ and its amplitude spectrum $A_2(F)$ are illustrated in Figure 20.3.

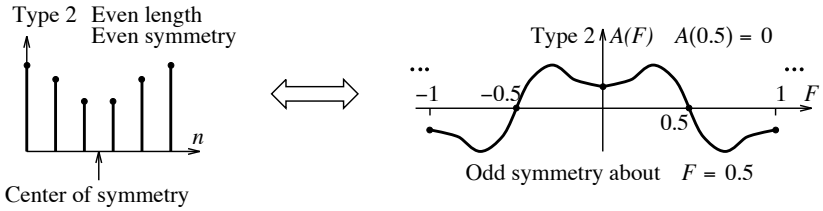


Figure 20.3 Features of a type 2 symmetric sequence

This sequence is also even symmetric but of even length N, and a center of symmetry at the half-integer value $M = (N-1)/2$. Using Euler's relation, its frequency response $H_2(F)$ may be expressed as

$$H_2(F) = \left[2 \sum_{k=0}^{M-1/2} h[k] \cos[(M-k)2\pi F] \right] e^{-j2\pi MF} = A_2(F) e^{-j2\pi MF} \tag{20.2}$$

PGR2026-00033
Krisp Technologies EX1024 Page 730

Thus, $H_2(F)$ also shows a linear phase of $-2\pi MF$, and a constant group delay of M. The amplitude spectrum $A_2(F)$ is even symmetric about $F = 0$, and odd symmetric about $F = 0.5$, and as a result, $|H_2(0.5)|$ is always zero.

Type 3 Sequences

A type 3 sequence $h_3[n]$ and its amplitude spectrum $A_3(F)$ are illustrated in Figure 20.4.

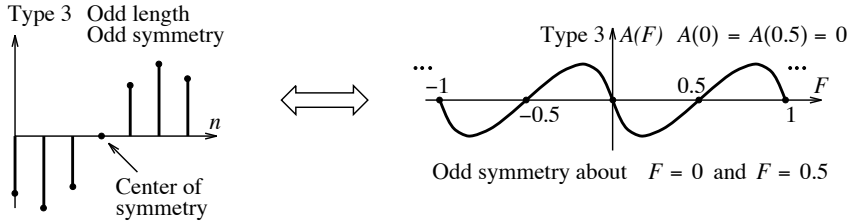


Figure 20.4 Features of a type 3 symmetric sequence

This sequence is odd symmetric with odd length N, and a center of symmetry at the integer value $M = (N-1)/2$. Using Euler's relation, its frequency response $H_3(F)$ may be expressed as

$$H_3(F) = j\left[2\sum_{k=0}^{M-1} h[k]\sin[(M-k)2\pi F]\right]e^{-j2\pi MF} = A_3(F)e^{j(0.5\pi - 2\pi MF)} \tag{20.3}$$

Thus, $H_3(F)$ shows a *generalized* linear phase of $\frac{\pi}{2} - 2\pi MF$, and a constant group delay of M. The amplitude spectrum $A_3(F)$ is odd symmetric about both $F = 0$ and $F = 0.5$, and as a result, $|H_3(0)|$ and $|H_3(0.5)|$ are always zero.

Type 4 Sequences

A type 4 sequence $h_4[n]$ and its amplitude spectrum $A_4(F)$ are illustrated in Figure 20.5.

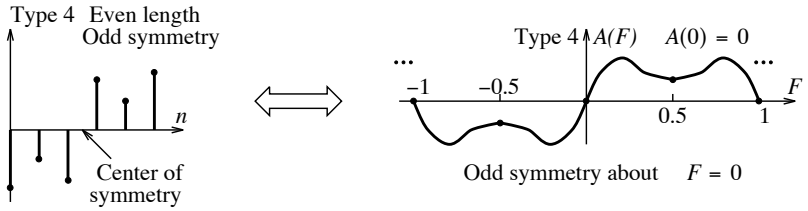


Figure 20.5 Features of a type 4 symmetric sequence

This sequence is odd symmetric with even length N, and a center of symmetry at the half-integer value $M = (N-1)/2$. Using Euler's relation, its frequency response $H_4(F)$ may be expressed as

$$H_4(F) = j\left[2\sum_{k=0}^{M-1/2} h[k]\sin[2(M-k)\pi F]\right]e^{-j2\pi MF} = A_4(F)e^{j(0.5\pi - 2\pi MF)} \tag{20.4}$$

PGR2026-00033
Krisp Technologies EX1024 Page 731

Thus, $H_4(F)$ also shows a generalized linear phase of $\frac{\pi}{2} - 2\pi MF$, and a constant group delay of M. The amplitude spectrum $A_4(F)$ is odd symmetric about $F = 0$, and even symmetric about $F = 0.5$, and as a result, $|H_4(0)|$ is always zero.

20.1.2 Poles and Zeros of Linear-Phase Sequences

The poles of any finite-length sequence must lie at $z = 0$. The zeros of linear-phase sequences must occur in conjugate reciprocal pairs. Real zeros at $z = 1$ or $z = -1$ need not be paired (they form their own reciprocals), but all other real zeros must be paired with their reciprocals. Complex zeros on the unit circle must be paired with their conjugates (that also form their reciprocals), and complex zeros anywhere else must occur in conjugate reciprocal quadruples. To identify the type of sequence from its pole-zero plot, all we need to do is check for the presence of zeros at $z = \pm 1$ and count their *number*. A type 2 sequence must have an odd number of zeros at $z = -1$, a type 3 sequence must have an odd number of zeros at $z = -1$ and $z = 1$, and a type 4 sequence must have *an odd number of zeros* at $z = 1$. The number of other zeros, if present (at $z = 1$ for type 1 and type 2, or $z = -1$ for type 1 or type 4), must be even. The reason is simple. If we exclude $z = \pm 1$, the remaining zeros occur in pairs or quadruples and always yield a sequence with *odd length and even symmetry*. Including a zero at $z = 1$ or $z = -1$ increases the length by 1. An odd number of zeros at $z = 1$ describes an odd symmetric sequence, and its inclusion (think convolution) also yields an odd symmetric sequence. These ideas are illustrated in Figure 20.6.

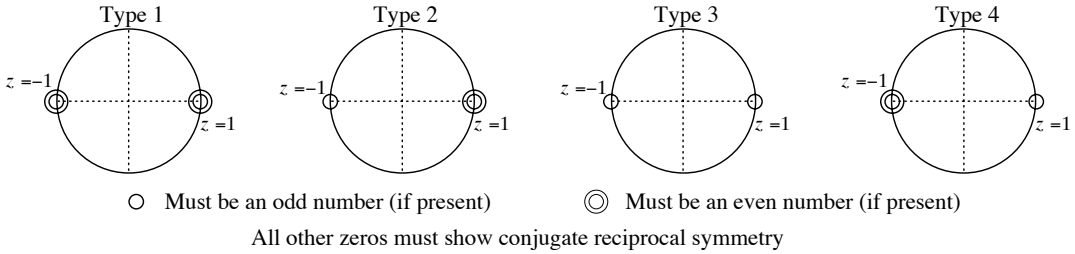


Figure 20.6 Identifying the sequence type from its zeros at $z = \pm 1$

REVIEW PANEL 20.2

How to Identify a Linear-Phase Sequence

From the sequence $x[n]$: The signal $x[n]$ must show symmetry about is midpoint.
From the pole-zero plot: All poles are at $z = 0$. Zeros show conjugate reciprocal symmetry.

Type 1: An even number of zeros (or no zeros) at $z = -1$ and/or $z = 1$
Type 2: Odd number of zeros at $z = -1$ (and even number at $z = 1$, if present)
Type 3: Odd number of zeros at $z = 1$ and odd number of zeros at $z = -1$
Type 4: Odd number of zeros at $z = 1$ (and even number at $z = -1$, if present)

EXAMPLE 20.1 (Identifying Linear-Phase Sequences)

(a) Find all the zero locations of a type 1 sequence, assuming the smallest length, if it is known that there is a zero at $z = 0.5e^{j\pi/3}$ and a zero at $z = 1$.

Due to conjugate reciprocal symmetry, the zero at $z = 0.5e^{j\pi/3}$ implies the quadruple zeros

$$z = 0.5e^{j\pi/3} \qquad z = 0.5e^{-j\pi/3} \qquad z = 2e^{j\pi/3} \qquad z = 2e^{-j\pi/3}$$

PGR2026-00033
Krisp Technologies EX1024 Page 732

For a type 1 sequence, the number of zeros at $z = 1$ must be even. So, there must be another zero at $z = 1$. Thus, we have a total of six zeros.

(b) Find the transfer function and impulse response of a causal type 3 linear-phase filter, assuming the smallest length and smallest delay, if it is known that there is a zero at $z = j$ and two zeros at $z = 1$.

The zero at $z = j$ must be paired with its conjugate (and reciprocal) $z = -j$.

A type 3 sequence requires an odd number of zeros at $z = 1$ and $z = -1$. So, the minimum number of zeros required is one zero at $z = -1$ and three zeros at $z = 1$ (with two already present).
The transfer function has the form

$$H(z) = (z + j)(z - j)(z + 1)(z - 1)^3 = z^6 - 2z^5 + z^4 - z^2 + 2z - 1$$

The transfer function of the causal filter with the minimum delay is

$$H_C(z) = 1 - 2z^{-1} + z^{-2} - z^{-4} + 2z^{-5} - z^{-6} \qquad h_C[n] = \{\overset{\Downarrow}{1},\ -2,\ 1,\ 0,\ -1,\ 2,\ -1\}$$

20.1.3 Applications of Linear-Phase Sequences

Table 20.1 summarizes the amplitude response characteristics of the four types of linear-phase sequences and their use in FIR digital filter design. Type 1 sequences are by far the most widely used because they allow us to design any filter type by appropriate choice of filter coefficients. Type 2 sequences can be used for lowpass and bandpass filters, but not for bandstop or highpass filters (whose response is not zero at $F = 0.5$). Type 3 sequences are useful primarily for bandpass filters, differentiators, and Hilbert transformers. Type 4 sequences are suitable for highpass or bandpass filters, and for differentiators and Hilbert transformers.

Bandstop filters (whose response is nonzero at $F = 0$ and $F = 0.5$) can be designed only with type 1 sequences. Only antisymmetric sequences (whose transfer function is imaginary) or their causal versions (which correspond to type 3 and type 4 sequences) can be used to design digital differentiators and Hilbert transformers.

Table 20.1 Applications of Symmetric Sequences

Type	$H(F) = 0$ **(or** $H(\Omega) = 0$**) at**	**Application**
1		All filter types. Only sequence for BSF
2	$F = 0.5$ $(\Omega = \pi)$	Only LPF and BPF
3	$F = 0$ $(\Omega = 0)$, $F = 0.5$ $(\Omega = \pi)$	BPF, differentiators, Hilbert transformers
4	$F = 0$ $(\Omega = 0)$	HPF, BPF, differentiators, Hilbert transformers

20.1.4 FIR Filter Design

The design of FIR filters involves the selection of a finite sequence that best represents the impulse response of an ideal filter. FIR filters are always stable. Even more important, FIR filters are capable of perfectly linear phase (a pure time delay), meaning total freedom from phase distortion. For given specifications, however, FIR filters typically require a much higher filter order or length than do IIR filters. And sometimes we must go to great lengths to ensure linear phase! The three most commonly used methods for FIR filter design are window-based design using the impulse response of ideal filters, frequency sampling, and iterative design based on optimal constraints.

PGR2026-00033
Krisp Technologies EX1024 Page 733

20.2 Window-Based Design

The window method starts by selecting the impulse response $h_N[n]$ as a symmetrically truncated version of the impulse response $h[n]$ of an ideal filter with frequency response $H(F)$. The impulse response of an ideal lowpass filter with a cutoff frequency F_C is $h[n] = 2F_C \operatorname{sinc}(2nF_C)$. Its symmetric truncation yields

$$h_N[n] = 2F_C \operatorname{sinc}(2nF_C), \qquad |n| \le 0.5(N-1) \tag{20.5}$$

Note that for an even length N, the index n is not an integer. Even though the designed filter is an approximation to the ideal filter, it is the best approximation (in the mean square sense) compared to any other filter of the same length. The problem is that it shows certain undesirable characteristics. Truncation of the ideal impulse response $h[n]$ is equivalent to multiplication of $h[n]$ by a rectangular window $w[n]$ of length N. The spectrum of the windowed impulse response $h_W[n] = h[n]w[n]$ is the (periodic) convolution of $H(F)$ and $W(F)$. Since $W(F)$ has the form of a Dirichlet kernel, this spectrum shows overshoot and ripples (the Gibbs effect). It is the abrupt truncation of the rectangular window that leads to overshoot and ripples in the magnitude spectrum. To reduce or eliminate the Gibbs effect, we use tapered windows.

REVIEW PANEL 20.3

The Spectrum of the Truncated Ideal Impulse Response Shows the Gibbs Effect

$h_N[n] = 2F_C \operatorname{sinc}(2nF_C), \qquad |n| \le 0.5(N-1)$

The Gibbs effect (spectral overshoot and oscillation) is due to abrupt truncation of $h[n]$.

Tapered windows reduce (or eliminate) the spectral overshoot and oscillation.

20.2.1 Characteristics of Window Functions

The amplitude response of symmetric, finite-duration windows invariably shows a mainlobe and decaying sidelobes that may be entirely positive or that may alternate in sign. The spectral measures for a typical window are illustrated in Figure 20.7.

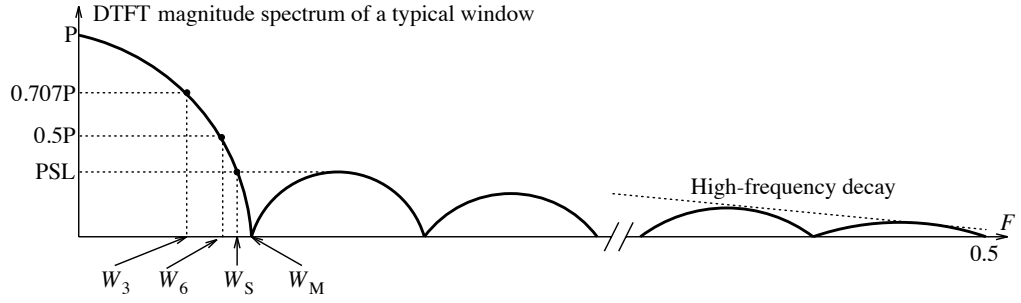


Figure 20.7 Spectrum of a typical window

Amplitude-based measures for a window include the peak sidelobe level (PSL), usually in decibels, and the decay rate D_S in dB/dec. Frequency-based measures include the mainlobe width W_M, the 3-dB and 6-dB widths (W_3 and W_6), and the width W_S to reach the peak sidelobe level. The windows commonly used in FIR filter design and their spectral features are listed in Table 20.2 and illustrated in Figure 20.8. As the window length N increases, the width parameters decrease, but the peak sidelobe level remains more or less constant. Ideally, the spectrum of a window should approximate an impulse and be confined to as narrow a mainlobe as possible, with as little energy in the sidelobes as possible.

PGR2026-00033
Krisp Technologies EX1024 Page 734

Table 20.2 Some Windows for FIR Filter Design

Note: $I_0(x)$ is the modified Bessel function of order zero.	
Window	**Expression** $w[n], \quad -0.5(N-1) \le n \le 0.5(N-1)$
Boxcar	1
Cosine	$\cos\Big(\dfrac{n\pi}{N-1}\Big)$
Riemann	$\text{sinc}^L\Big(\dfrac{2n}{N-1}\Big), \quad L > 0$
Bartlett	$1 - \dfrac{2\lvert n\rvert}{N-1}$
von Hann (Hanning)	$0.5 + 0.5\cos\Big(\dfrac{2n\pi}{N-1}\Big)$
Hamming	$0.54 + 0.46\cos\Big(\dfrac{2n\pi}{N-1}\Big)$
Blackman	$0.42 + 0.5\cos\Big(\dfrac{2n\pi}{N-1}\Big) + 0.08\cos\Big(\dfrac{4n\pi}{N-1}\Big)$
Kaiser	$\dfrac{I_0(\pi\beta\sqrt{1-4[n/(N-1)]^2})}{I_0(\pi\beta)}$

Spectral Characteristics of Window Functions								
Window	G_P	G_S/G_P	A_{SL} **(dB)**	W_M	W_S	W_6	W_3	D_S
Boxcar	1	0.2172	13.3	1	0.81	0.6	0.44	20
Cosine	0.6366	0.0708	23	1.5	1.35	0.81	0.59	40
Riemann	0.5895	0.0478	26.4	1.64	1.5	0.86	0.62	40
Bartlett	0.5	0.0472	26.5	2	1.62	0.88	0.63	40
von Hann (Hanning)	0.5	0.0267	31.5	2	1.87	1.0	0.72	60
Hamming	0.54	0.0073	42.7	2	1.91	0.9	0.65	20
Blackman	0.42	0.0012	58.1	3	2.82	1.14	0.82	60
Kaiser ($\beta = 2.6$)	0.4314	0.0010	60	2.98	2.72	1.11	0.80	20

NOTATION:
G_P: Peak gain of mainlobe
A_{SL}: Sidelobe attenuation ($\frac{G_P}{G_S}$) in dB
W_3 3-dB half-width
G_S: Peak sidelobe gain
W_6: 6-dB half-width
W_M: Half-width of mainlobe
D_S: High-frequency attenuation (dB/decade)
W_S: Half-width of mainlobe to reach P_S

Notes:
1. All widths (W_M, W_S, W_6, W_3) must be normalized (divided) by the window length N.
2. Values for the Kaiser window depend on the parameter β. Empirically determined relations are

$$G_P = \frac{|\text{sinc}(j\beta)|}{I_0(\pi\beta)}, \qquad \frac{G_S}{G_P} = \frac{0.22\pi\beta}{\sinh(\pi\beta)}, \qquad W_M = (1+\beta^2)^{1/2}, \qquad W_S = (0.661+\beta^2)^{1/2}$$

PGR2026-00033
Krisp Technologies EX1024 Page 735

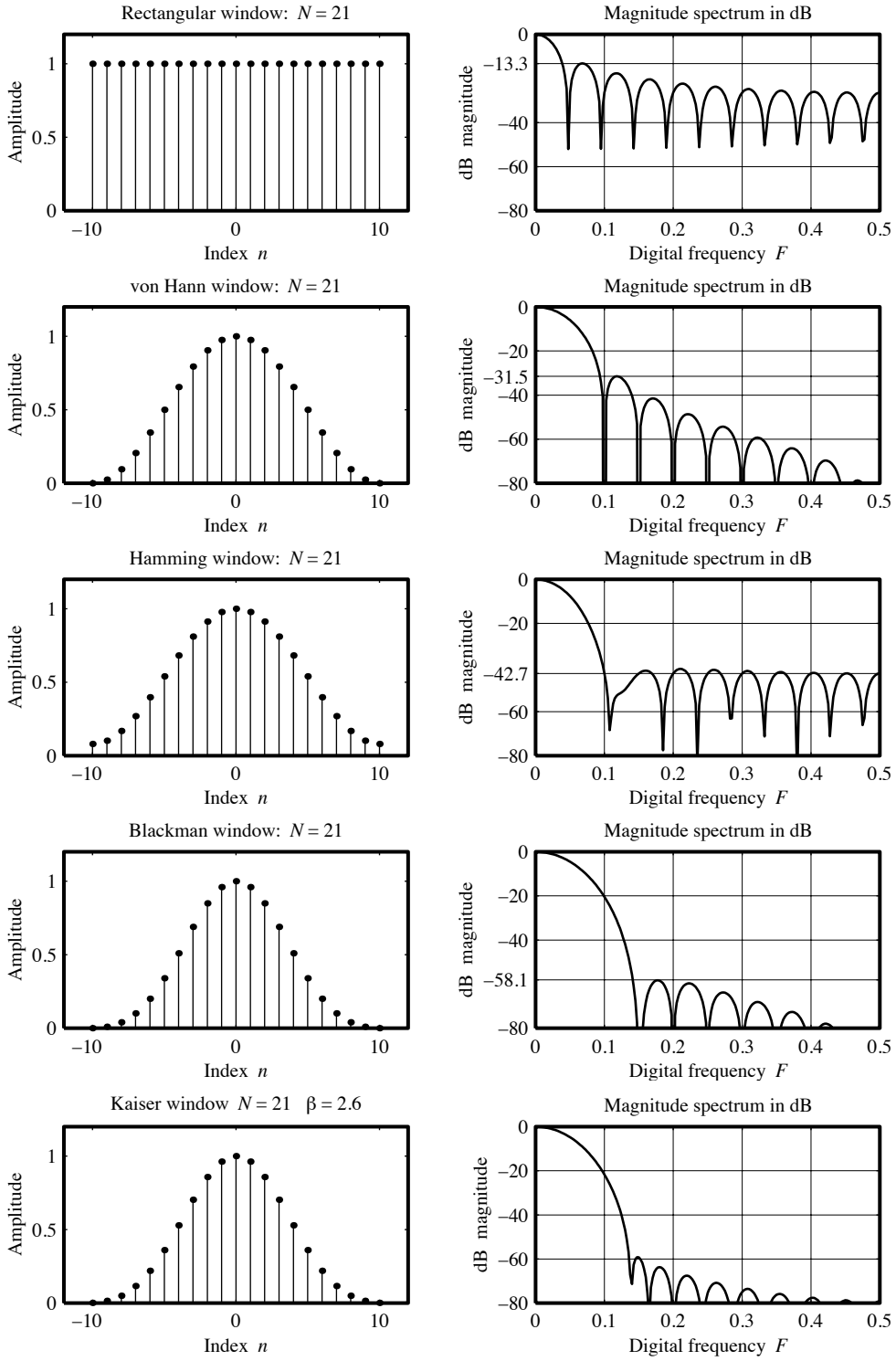


Figure 20.8 Commonly used DTFT windows and their spectra

PGR2026-00033
Krisp Technologies EX1024 Page 736

Most windows have been developed with some optimality criterion in mind. Ultimately, the trade-off is a compromise between the conflicting requirements of a narrow mainlobe (or a small transition width), and small sidelobe levels. Some windows are based on combinations of simpler windows. For example, the von Hann (or Hanning) window is the sum of a rectangular and a cosine window, the Bartlett window is the convolution of two rectangular windows, and the $\cos^a$ window is the product of a von Hann and a cosine window. Other windows are designed to emphasize certain desirable features. The von Hann window improves the high-frequency decay (at the expense of a larger peak sidelobe level). The Hamming window minimizes the sidelobe level (at the expense of a slower high-frequency decay). The Kaiser window has a variable parameter β that controls the peak sidelobe level. Still other windows are based on simple mathematical forms or easy application. For example, the $\cos^a$ windows have easily recognizable transforms, and the von Hann window is easy to apply as a convolution in the frequency domain. An optimal time-limited window should maximize the energy in its spectrum over a given frequency band. In the continuous-time domain, this constraint leads to a window based on *prolate spheroidal wave functions* of the first order. The Kaiser window best approximates such an optimal window in the discrete domain.

20.2.2 Some Other Windows

The Dolph (also called Chebyshev) window corresponds to the inverse DTFT of a spectrum whose sidelobes remain constant (without decay) at a specified level and whose mainlobe width is the smallest for a given length. The time-domain expression for this window is cumbersome. This window can be computed from

$$w[n] = \text{IDFT}\left\{T_{N-1}\left[\alpha\cos\left(\frac{n\pi}{N-1}\right)\right]\right\} \qquad \alpha = \cosh\left[\frac{\cosh^{-1}(10^{A/20})}{N-1}\right] \quad \text{(odd } n \text{ only)} \tag{20.6}$$

Here, $T_n(x)$ is the Chebyshev polynomial of order n, and A is the sidelobe attenuation in decibels.

Table 20.3 Characteristics of Harris Windows

Notes: A_{SL} is the normalized peak sidelobe attenuation in dB (decibels).
W_S is the half-width of the main lobe to reach P_S.
D_S is the high frequency decay rate in dB/octave (6 dB/oct = 20 dB/dec).

Type	b_0	b_1	b_2	b_3	A_{SL} **(dB)**	$W_S = K/N$	D_S
0	0.375	0.5	0.125	0.0	46.74	$K = 2.89$	30
1	10/32	15/32	6/32	1/32	60.95	$K = 3.90$	42
2	0.40897	0.5	0.09103	0.0	64.19	$K = 2.94$	18
3	0.42323	0.49755	0.07922	0.0	70.82	$K = 2.95$	6
4	0.4243801	0.4973406	0.0782793	0.0	71.48	$K = 2.95$	6
5	0.338946	0.481973	0.161054	0.018027	82.6	$K = 3.95$	30
6	0.355768	0.487396	0.144232	0.012604	92.32	$K = 3.96$	18
7	0.3635819	0.4891775	0.1365995	0.0106411	98.17	$K = 3.96$	6

Harris windows offer a variety of peak sidelobe levels and high-frequency decay rates. Their characteristics are listed in Table 20.3. Harris windows have the general form

$$w[n] = b_0 - b_1\cos(m) + b_2\cos(2m) - b_3\cos(3m), \quad \text{where} \quad m = \frac{2n\pi}{N-1}, \quad |n| \le \frac{N-1}{2} \tag{20.7}$$

PGR2026-00033
Krisp Technologies EX1024 Page 737

20.2.3 What Windowing Means

Symmetric windowing of the impulse response of an ideal filter is accomplished by windows that are themselves symmetric, as illustrated in Figure 20.9.

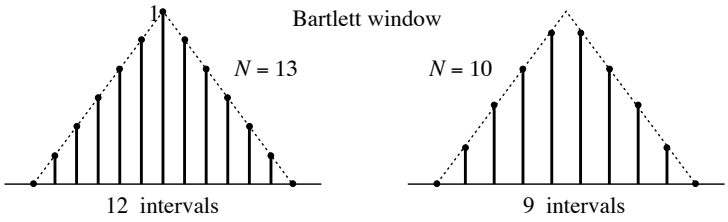


Figure 20.9 The mechanics of windowing

An N-point FIR window uses $N-1$ intervals to generate N samples (including both end samples). Once selected, the window sequence must be symmetrically positioned with respect to the symmetric impulse response sequence. Windowing is then simply a *pointwise multiplication* of the two sequences.

The symmetrically windowed impulse response of an ideal lowpass filter may be written as

$$h_W[n] = 2F_C \,\mathrm{sinc}(2nF_C)w[n], \qquad -0.5(N-1) \le n \le 0.5(N-1) \tag{20.8}$$

For even N, the index n is not an integer, even though it is incremented every unit, and we require a *non-integer* delay to produce a causal sequence. If the end samples of $w[n]$ equal zero, so will those of $h_W[n]$, and the true filter length will be less by 2.

EXAMPLE 20.2 (Truncation and Windowing)
Consider a lowpass filter with a cutoff frequency of 5 kHz and a sampling frequency of 20 kHz. Then, $F_C = 0.25$ and $h[n] = 2F_C \,\mathrm{sinc}(2nF_C) = 0.5\,\mathrm{sinc}(0.5n)$.

(a) With $N = 9$ and a Bartlett window $w = 1 - \frac{2|n|}{N-1},\ -4 \le n \le 4$, we find

$$\begin{aligned} h[n] &= \{0,\ -0.1061,\ 0,\ 0.3183,\ \overset{\Downarrow}{0.5},\ 0.3183,\ 0,\ -0.1061,\ 0\} \\ w[n] &= \{0,\ 0.25,\ 0.5,\ 0.75,\ \overset{\Downarrow}{1},\ 0.75,\ 0.5,\ 0.25,\ 0\} \end{aligned}$$

Pointwise multiplication of h_N and w_N gives

$$h_W[n] = \{0,\ -0.0265,\ 0,\ 0.2387,\ \overset{\Downarrow}{0.5},\ 0.2387,\ 0,\ -0.0265,\ 0\}$$

If we include the end samples, the causal filter requires a delay of four samples to give

$$H_C(z) = 0 - 0.0265z^{-1} + 0.2387z^{-3} + 0.5z^{-4} + 0.2387z^{-5} - 0.0265z^{-7}$$

Since $S = 20$ kHz, we have $t_s = 1/S = 0.05$ ms, and a four-sample delay corresponds to an actual delay of $4(0.05) = 0.2$ ms. If we ignore the end samples, the causal filter requires a delay of only three (not four) samples to give

$$H_C(z) = -0.0265 + 0.2387z^{-2} + 0.5z^{-3} + 0.2387z^{-4} - 0.0265z^{-6}$$

This corresponds to a delay of 0.15 ms.

PGR2026-00033
Krisp Technologies EX1024 Page 738

(b) With $N = 6$, we have $F_C = 0.25$ and $h_N[n] = 2F_C \operatorname{sinc}(2nF_C) = 0.5 \operatorname{sinc}(0.5n),\ -2.5 \le n \le 2.5$.
For a von Hann (Hanning) window, $w[n] = 0.5 + 0.5\cos(\frac{2n\pi}{N-1}) = 0.5 + 0.5\cos(0.4n\pi),\ \ -2.5 \le n \le 2.5$.

We then compute values of $h[n]$, $w[n]$, and $h_W[n] = h[n]w[n]$ as

$h[n]$	=	-0.09	0.1501	0.4502	0.4502	0.1501	-0.09
$w[n]$	=	0.0	0.3455	0.9045	0.9045	0.3455	0.0
$h_W[n]$	=	0.0	0.0518	0.4072	0.4072	0.0518	0.0

Since the first sample of $h_W[n]$ is zero, the *minimum delay* required for a causal filter is only 1.5 (and not 2.5) samples, and the transfer function of the causal sequence is

$$H_C(z) = 0.0518 + 0.4072z^{-1} + 0.4072z^{-2} + 0.0518z^{-3}$$

20.2.4 Some Design Issues

As with analog and IIR filters, we start with the design of lowpass FIR filters. The design of other types of filters is then based on appropriate spectral transformations. The spectrum of a typical filter using window-based design is illustrated in Figure 20.10.

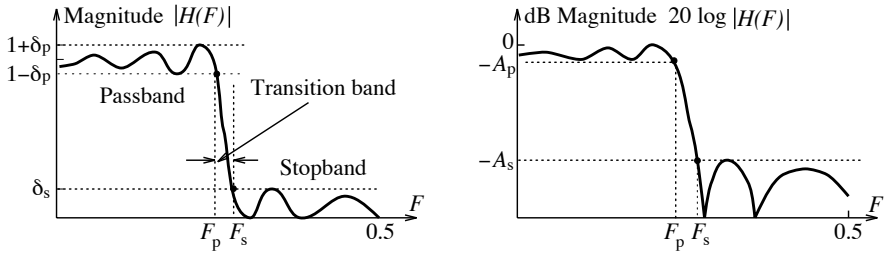


Figure 20.10 The features of a typical filter

The gain specifications of an FIR filter are often based on the passband ripple δ_p (the maximum deviation from unit gain) and the stopband ripple δ_s (the maximum deviation from zero gain). Since the decibel gain is usually normalized with respect to the peak gain, the passband and stopband attenuation A_p and A_s (in decibels) are related to these ripple parameters by

$$A_p(\text{dB}) = -20\log\left(\frac{1-\delta_p}{1+\delta_p}\right) \qquad A_s(\text{dB}) = -20\log\left(\frac{\delta_s}{1+\delta_p}\right) \approx -20\log\delta_s, \quad \delta_p \ll 1 \tag{20.9}$$

To convert attenuation specifications (in decibels) to values for the ripple parameters, we use

$$\delta_p = \frac{10^{A_p/20} - 1}{10^{A_p/20} + 1} \qquad \delta_s = (1+\delta_p)10^{-A_s/20} \approx 10^{-A_s/20}, \quad \delta_p \ll 1 \tag{20.10}$$

Window-based design calls for normalization of the design frequencies by the sampling frequency, developing the lowpass filter by windowing the impulse response of an ideal filter, and conversion to the required filter type, using spectral transformations. The method may appear deceptively simple (and it is), but some issues can only be addressed qualitatively. For example, the choice of cutoff frequency is affected by the window length N. The smallest length N that meets specifications depends on the choice of window, and the choice of the window, in turn, depends on the (stopband) attenuation specifications.

PGR2026-00033
Krisp Technologies EX1024 Page 739

20.2.5 Characteristics of the Windowed Spectrum

When we multiply the ideal impulse response $h[n] = 2F_C \operatorname{sinc}(2nF_C)$ by a window $w[n]$ of length N in the time domain, the spectrum of the windowed impulse response $h_W[n] = h[n]w[n]$ is the (periodic) convolution of $H(F)$ and $W(F)$, as illustrated in Figure 20.11.

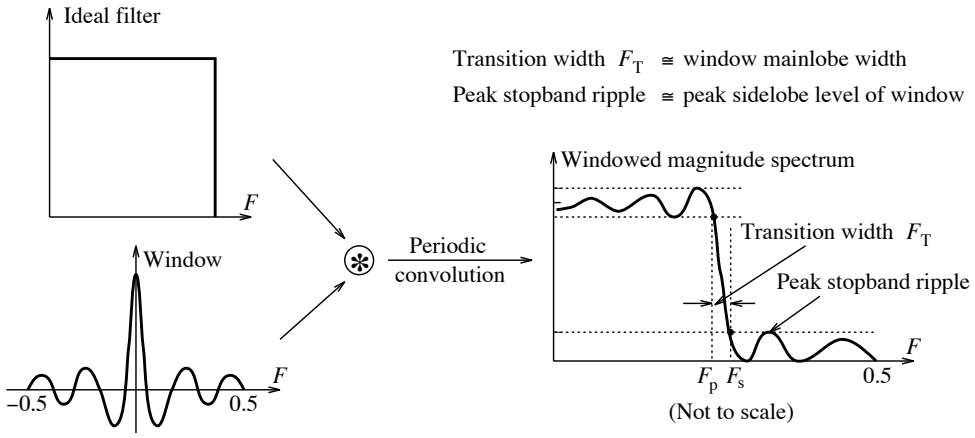


Figure 20.11 The spectrum of a windowed ideal filter

The ideal spectrum has a jump discontinuity at $F = F_C$. The windowed spectrum shows overshoot and ripples, and a finite transition width but no abrupt jump. Its magnitude at $F = F_C$ equals 0.5 (corresponding to an attenuation of 6 dB).

Table 20.4 lists the characteristics of the *windowed spectrum* for various windows. It excludes windows (such as the Bartlett window) whose amplitude spectrum is entirely positive (because they result in a complete elimination of overshoot and the Gibbs effect). Since both the window function and the impulse response are symmetric sequences, the spectrum of the windowed filter is also endowed with symmetry. Here are some general observations about the windowed spectrum:

1. Even though the peak passband ripple equals the peak stopband ripple ($\delta_p = \delta_s$), the passband (or stopband) ripples are not of equal magnitude.

2. The peak stopband level of the windowed spectrum is typically slightly less than the peak sidelobe level of the window itself. In other words, the filter stopband attenuation (listed as A_{WS} in Table 20.4) is typically greater (by a few decibels) than the peak sidelobe attenuation of the window (listed as A_{SL} in Tables 20.2 and 20.3). The peak sidelobe level, the peak passband ripple, and the passband attenuation (listed as A_{WP} in Table 20.4) remain more or less constant with N.

3. The peak-to-peak width across the transition band is roughly equal to the mainlobe width of the window (listed as W_M in Tables 20.2 and 20.3). The actual transition width (listed as F_W in Table 20.4) of the windowed spectrum (when the response first reaches $1 - \delta_p$ and δ_s) is less than this width. The transition width F_{WS} is inversely related to the window length N (with $F_{\mathrm{WS}} \approx C/N$, where C is more or less a constant for each window).

The numbers vary in the literature, and the values in Table 20.4 were found here by using an ideal impulse response $h[n] = 0.5 \operatorname{sinc}(0.5n)$, with $F_C = 0.25$, windowed by a 51-point window. The magnitude specifications are normalized with respect to the peak magnitude. The passband and stopband attenuation are computed from the passband and stopband ripple (with $\delta_p = \delta_s$), using the relations already given.

PGR2026-00033
Krisp Technologies EX1024 Page 740

Table 20.4 Characteristics of the Windowed Spectrum

Window	**Peak Ripple** $\delta_p = \delta_s$	**Passband Attenuation** A_{WP} (dB)	**Peak Sidelobe Attenuation** A_{WS} (dB)	**Transition Width** $F_{\text{WS}} \approx C/N$
Boxcar	0.0897	1.5618	21.7	$C = 0.92$
Cosine	0.0207	0.36	33.8	$C = 2.1$
Riemann	0.0120	0.2087	38.5	$C = 2.5$
von Hann (Hanning)	0.0063	0.1103	44	$C = 3.21$
Hamming	0.0022	0.0384	53	$C = 3.47$
Blackman	$(1.71)10^{-4}$	$(2.97)10^{-3}$	75.3	$C = 5.71$
Dolph ($R = 40$ dB)	0.0036	0.0620	49	$C = 3.16$
Dolph ($R = 50$ dB)	$(9.54)10^{-4}$	0.0166	60.4	$C = 3.88$
Dolph ($R = 60$ dB)	$(2.50)10^{-4}$	0.0043	72	$C = 4.6$
Harris (0)	$(8.55)10^{-4}$	0.0148	61.4	$C = 5.36$
Harris (1)	$(1.41)10^{-4}$	$(2.44)10^{-3}$	77	$C = 7.45$
Harris (2)	$(1.18)10^{-4}$	$(2.06)10^{-3}$	78.5	$C = 5.6$
Harris (3)	$(8.97)10^{-5}$	$(1.56)10^{-3}$	81	$C = 5.6$
Harris (4)	$(9.24)10^{-5}$	$(1.61)10^{-3}$	81	$C = 5.6$
Harris (5)	$(9.96)10^{-6}$	$(1.73)10^{-4}$	100	$C = 7.75$
Harris (6)	$(1.94)10^{-6}$	$(3.38)10^{-5}$	114	$C = 7.96$
Harris (7)	$(5.26)10^{-6}$	$(9.15)10^{-5}$	106	$C = 7.85$

20.2.6 Selection of Window and Design Parameters

The choice of a window is based primarily on the design stopband specification A_s. The peak sidelobe attenuation A_{WS} of the windowed spectrum (listed in Table 20.4) should match (or exceed) the specified stopband attenuation A_s. Similarly, the peak passband attenuation A_{WP} of the windowed spectrum should not exceed the specified passband attenuation A_p, a condition that is often ignored because it is usually satisfied for most practical specifications.

The windowed spectrum is the convolution of the spectra of the impulse response and the window function, and this spectrum changes as we change the filter length. An optimal (in the mean square sense) impulse response and an optimal (in any sense) window may not together yield a windowed response with optimal features. Window selection is at best an art, and at worst a matter of trial and error. Just so you know, the three most commonly used windows are the von Hann (Hanning), Hamming, and Kaiser windows.

Choosing the Filter Length

The transition width of the windowed spectrum decreases with the length N. There is no accurate way to establish the minimum filter length N that meets design specifications. However, empirical estimates are

PGR2026-00033
Krisp Technologies EX1024 Page 741

based on matching the given transition width specification F_T to the transition width $F_{\text{WS}} = C/N$ of the windowed spectrum (as listed in Table 20.4)

$$F_T = F_s - F_p = F_{\text{WS}} \approx \frac{C}{N} \qquad N = \frac{C}{F_s - F_p} \tag{20.11}$$

Here, F_p and F_s are the *digital* passband and stopband frequencies. The window length depends on the choice of window (which dictates the choice of C). The closer the match between the stopband attenuation A_s and the stopband attenuation A_{WS} of the windowed spectrum, the smaller is the window length N. In any case, for a given window, this relation typically *overestimates* the smallest filter length, and we can often decrease this length and still meet design specifications.

The Kaiser Window

Empirical relations have been developed to estimate the filter length N of FIR filters based on the Kaiser window. We first compute the peak passband ripple δ_p, and the peak stopband ripple δ_s, and choose the smallest of these as the *ripple parameter* δ:

$$\delta_p = \frac{10^{A_p/20} - 1}{10^{A_p/20} + 1} \qquad \delta_s = 10^{-A_s/20} \qquad \delta = \min(\delta_p, \delta_s) \tag{20.12}$$

The ripple parameter δ is used to recompute the actual stopband attenuation A_{s0} in decibels:

$$A_{s0} = -20 \log \delta \ \ \text{dB} \tag{20.13}$$

Finally, the length N is well approximated by

$$N \geq \begin{cases} \dfrac{A_{s0} - 7.95}{14.36(F_s - F_p)} + 1, & A_{s0} \geq 21 \text{ dB} \\ \dfrac{0.9222}{F_s - F_p} + 1, & A_{s0} < 21 \text{ dB} \end{cases} \tag{20.14}$$

The Kaiser window parameter β is estimated from the actual stopband attenuation A_{s0}, as follows:

$$\beta = \begin{cases} 0.0351(A_{s0} - 8.7), & A_{s0} > 50 \text{ dB} \\ 0.186(A_{s0} - 21)^{0.4} + 0.0251(A_{s0} - 21), & 21 \text{ dB} \leq A_{s0} \leq 50 \text{ dB} \\ 0, & A_{s0} < 21 \text{ dB} \end{cases} \tag{20.15}$$

Choosing the Cutoff Frequency

A common choice for the cutoff frequency (used in the expression for $h[n]$) is $F_C = 0.5(F_p + F_s)$. The actual frequency that meets specifications for the smallest length is often less than this value. The cutoff frequency is affected by the filter length N. A design that ensures the minimum length N is based on starting with the above value of F_C, and then changing (typically reducing) the length and/or tweaking (typically decreasing) F_C until we just meet specifications (typically at the passband edge).

20.2.7 Spectral Transformations

The design of FIR filters other than lowpass starts with a lowpass prototype, followed by appropriate spectral transformations to convert it to the required filter type. These spectral transformations are developed from the shifting and modulation properties of the DTFT. Unlike analog and IIR filters, these transformations

PGR2026-00033
Krisp Technologies EX1024 Page 742

do not change the filter order (or length). The starting point is an ideal lowpass filter, with *unit passband gain* and a cutoff frequency of F_C, whose windowed impulse response $h_{\text{LP}}[n]$ is given by

$$h_{\text{LP}}[n] = 2F_C \,\text{sinc}(2nF_C)w[n], \qquad -0.5(N-1) \le n \le 0.5(N-1) \tag{20.16}$$

Here, $w[n]$ is a window function of length N. If N is even, the index n takes on non-integer values. The lowpass-to-highpass (LP2HP) transformation may be achieved in two ways, as illustrated in Figure 20.12.

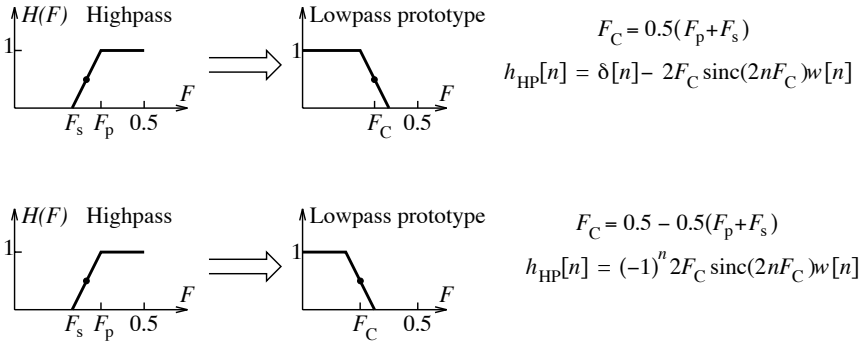


Figure 20.12 The lowpass-to-highpass transformation

The first transformation of Figure 20.12 is valid only if the filter length N is odd, and reads

$$h_{\text{HP}}[n] = \delta[n] - h_{\text{LP}}[n] = \delta[n] - 2F_C \,\text{sinc}(2nF_C)w[n], \qquad -\frac{N-1}{2} \le n \le \frac{N-1}{2} \tag{20.17}$$

Its cutoff frequency $F_H = 0.5(F_p + F_s)$ equals the cutoff frequency F_C of the lowpass filter. The second transformation (assuming a *causal* lowpass prototype $h_{\text{LP}}[n]$) is valid for any length N, and reads

$$h_{\text{HP}}[n] = (-1)^n h_{\text{LP}}[n] = 2(-1)^n F_C \,\text{sinc}(2nF_C)w[n], \qquad 0 \le n \le N-1 \tag{20.18}$$

However, its cutoff frequency is $0.5 - 0.5(F_p + F_s)$. In order to design a highpass filter with a cutoff frequency of $F_H = 0.5(F_p + F_s)$, we must start with a lowpass filter whose cutoff frequency is $F_C = 0.5 - 0.5(F_p + F_s)$.

The LP2BP and LP2BS transformations preserve arithmetic (not geometric) symmetry about the center frequency F_0. If the band edges are $[F_1,\ F_2,\ F_3,\ F_4]$ in increasing order, arithmetic symmetry means that $F_1 + F_4 = F_2 + F_3 = 2F_0$ and implies equal transition widths. If the transition widths are not equal, we must relocate a band edge to make both transition widths equal to the smaller width, as shown in Figure 20.13.

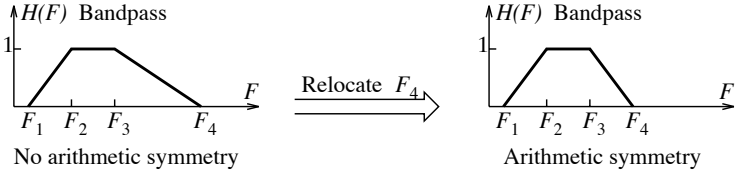


Figure 20.13 How to ensure arithmetic symmetry of the band edges

The LP2BP and LP2BS transformations are illustrated in Figure 20.14. In each transformation, the center frequency F_0, and the cutoff frequency F_C of the lowpass filter, are given by

$$F_0 = 0.5(F_2 + F_3) \qquad\qquad F_C = 0.5(F_3 + F_4) - F_0 \tag{20.19}$$

PGR2026-00033
Krisp Technologies EX1024 Page 743

The lowpass-to-bandpass (LP2BP) transformation is given by

$$h_{\mathrm{BP}}[n] = 2\cos(2n\pi F_0)h_{\mathrm{LP}}[n] = 4F_C\,\mathrm{sinc}(2nF_C)\cos(2\pi nF_0)w[n], \quad -\frac{N-1}{2} \le n \le \frac{N-1}{2} \tag{20.20}$$

The lowpass-to-bandstop (LP2BS) transformation is given by

$$h_{\mathrm{BS}}[n] = \delta[n] - h_{\mathrm{BS}}[n] = \delta[n] - 4F_C\,\mathrm{sinc}(2nF_C)\cos(2\pi nF_0)w[n], \quad -\frac{N-1}{2} \le n \le \frac{N-1}{2} \tag{20.21}$$

Note that the LP2BS transformation requires a type 4 sequence (with even symmetry and odd length N).

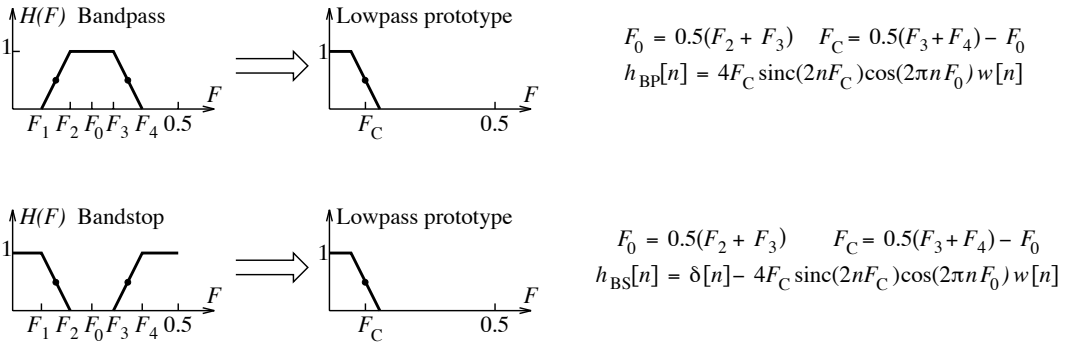


Figure 20.14 The lowpass-to-bandpass and lowpass-to-bandstop transformations

REVIEW PANEL 20.4

Impulse Response of Windowed Filters Over $|n| \le 0.5(N-1)$

$h_P[n] = 2F_C\,\mathrm{sinc}(2nF_C)$ $\quad h_{\mathrm{HP}}[n] = \delta[n] - h_P[n]$ $\quad F_C$ = cutoff frequency of lowpass prototype

$h_{\mathrm{BP}}[n] = 2h_P[n]\cos(2\pi nF_0)$ $\quad h_{\mathrm{BS}}[n] = \delta[n] - h_{\mathrm{BP}}[n]$ $\quad F_0$ = center frequency (for BPF and BSF)

REVIEW PANEL 20.5

Recipe for Window-Based FIR Filter Design

Normalize the analog design frequencies by the sampling frequency S.
Obtain the band edges F_p and F_s of the lowpass prototype.
Choose the lowpass prototype cutoff as $F_C = 0.5(F_p + F_s)$.
Choose a window (from Table 20.4) that satisfies $A_{\mathrm{WS}} \ge A_s$ and $A_{\mathrm{WP}} \le A_s$.
Compute the window length N from $F_T = F_s - F_p = F_{\mathrm{WS}} = \frac{C}{N}$ (with C as in Table 20.4).
Compute the prototype impulse response $h[n] = 2F_C\,\mathrm{sinc}[2nF_C]$, $|n| \le 0.5(N-1)$.
Window $h[n]$ and apply spectral transformations (if needed) to convert to required filter type.

Minimum-length design: Adjust N and/or F_C until the design specifications are just met.

EXAMPLE 20.3 (FIR Filter Design Using Windows)

(a) Design an FIR filter to meet the following specifications:

$f_p = 2$ kHz, $f_s = 4$ kHz, $A_p = 2$ dB, $A_s = 40$ dB, and sampling frequency $S = 20$ kHz.

PGR2026-00033
Krisp Technologies EX1024 Page 744

This describes a lowpass filter. The digital frequencies are $F_p = f_p/S = 0.1$ and $F_s = f_s/S = 0.2$.

With $A_s = 40$ dB, possible choices for a window are (from Table 20.4) von Hann (with $A_{\text{WS}} = 44$ dB) and Blackman (with $A_{\text{WS}} = 75.3$ dB). Using $F_{\text{WS}} \approx (F_s - F_p) = C/N$, the approximate filter lengths for these windows (using the values of C from Table 20.4) are:

$$\text{von Hann:} N \approx \tfrac{3.21}{0.1} \approx 33 \qquad \text{Blackman:} N \approx \tfrac{5.71}{0.1} \approx 58$$

We choose the cutoff frequency as $F_C = 0.5(F_p + F_s) = 0.15$. The impulse then response equals

$$h_N[n] = 2F_C \,\text{sinc}(2nF_C) = 0.3 \,\text{sinc}(0.3n)$$

Windowing gives the impulse response of the required lowpass filter

$$h_{\text{LP}}[n] = w[n]h_N[n] = 0.3w[n]\text{sinc}(0.3n)$$

As Figure E20.3A(a) shows, the design specifications are indeed met by each filter (but the lengths are actually overestimated). The Blackman window requires a larger length because of the larger difference between A_s and A_{WS}. It also has the larger transition width.

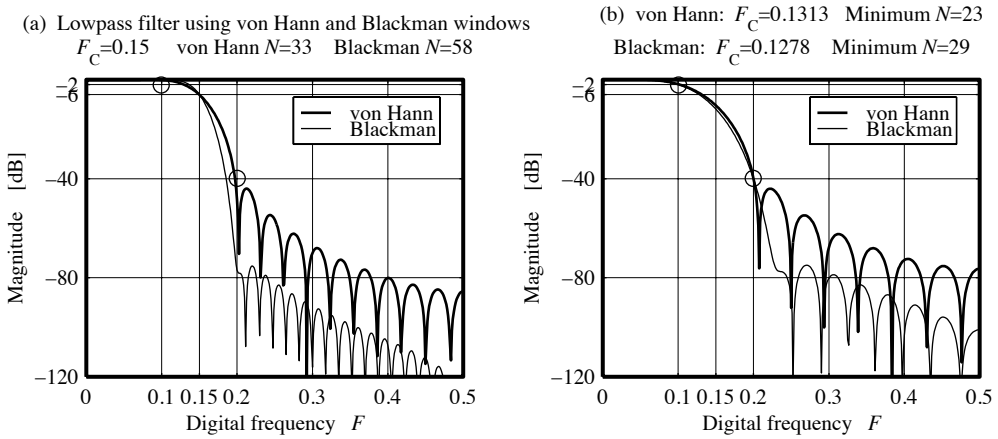


Figure E20.3A Lowpass FIR filters for Example 20.3(a and b)

(b) (Minimum-Length Design) By trial and error, the cutoff frequency and the smallest length that just meet specifications turn out to be $F_C = 0.1313$, $N = 23$, for the von Hann window, and $F_C = 0.1278$, $N = 29$, for the Blackman window. Figure E20.3A(b) shows the response of these minimum-length filters. The passband and stopband attenuation are [1.9, 40.5] dB for the von Hann window, and [1.98, 40.1] dB for the Blackman window. Even though the filter lengths are much smaller, each filter does meet the design specifications.

(c) Design an FIR filter to meet the following specifications:

$f_p = 4$ kHz, $f_s = 2$ kHz, $A_p = 2$ dB, $A_s = 40$ dB, and sampling frequency $S = 20$ kHz.

The specifications describe a highpass filter. The digital frequencies are $F_p = f_p/S = 0.2$ and $F_s = f_s/S = 0.1$. The transition width is $F_T = 0.2 - 0.1 = 0.1$.

With $A_s = 40$ dB, possible choices for a window (see Table 20.4) are Hamming and Blackman. Using $F_T \approx F_{\text{WS}} = C/N$, the approximate filter lengths for these windows are

$$\text{Hamming:} N \approx \tfrac{3.47}{0.1} \approx 35 \qquad \text{Blackman:} N \approx \tfrac{5.71}{0.1} \approx 58$$

PGR2026-00033
Krisp Technologies EX1024 Page 745

We can now design the highpass filter in one of two ways:

1. Choose the cutoff frequency of the lowpass filter as $F_C = 0.5(F_p + F_s) = 0.15$. The impulse response $h_N[n]$ then equals

   $$h_N[n] = 2F_C \,\text{sinc}(2nF_C) = 0.3 \,\text{sinc}(0.3n)$$

   The windowed response is thus $h_W[n] = 0.3w[n]\text{sinc}(0.3n)$.

   The impulse of the required highpass filter is then

   $$h_{\text{HP}}[n] = \delta[n] - h_W[n] = \delta[n] - 0.3w[n]\text{sinc}(0.3n)$$

2. Choose the cutoff frequency of the lowpass filter as $F_C = 0.5 - 0.5(F_p + F_s) = 0.35$.

   Then, the impulse response equals $h_N[n] = 2F_C \,\text{sinc}(2nF_C) = 0.7 \,\text{sinc}(0.7n)$.

   The windowed response is thus $h_W[n] = 0.7w[n]\text{sinc}(0.7n)$.

   The impulse of the required highpass filter is then

   $$h_{\text{HP}}[n] = (-1)^n h_W[n] = 0.7(-1)^n w[n]\text{sinc}(0.7n)$$

The two methods yield identical results. As Figure E20.3B(a) shows, the design specifications are indeed met by each window, but the lengths are actually overestimated.

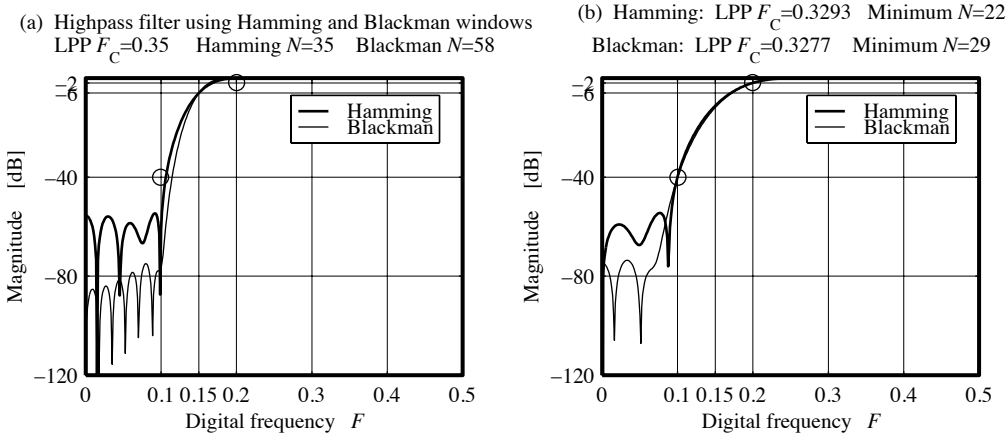


Figure E20.3B Highpass FIR filters for Example 20.3(c and d)

(d) (Minimum-Length Design) By trial and error, the cutoff frequency, and the smallest length that just meet specifications, turn out to be $F_C = 0.3293$, $N = 22$, for the Hamming window, and $F_C = 0.3277$, $N = 29$, for the Blackman window. Figure E20.3B(b) shows the response of the minimum-length filters. The passband and stopband attenuation are [1.94, 40.01] dB for the Hamming window, and [1.99, 40.18] dB for the Blackman window. Each filter meets the design specifications, even though the filter lengths are much smaller than the values computed from the design relations.

(e) Design an FIR filter to meet the following specifications:

$A_p = 3$ dB, $A_s = 45$ dB, passband: [4, 8] kHz, stopband: [2, 12] kHz, and $S = 25$ kHz.

The specifications describe a bandpass filter. If we assume a fixed passband, the center frequency lies at the center of the passband and is given by $f_0 = 0.5(4 + 8) = 6$ kHz.

PGR2026-00033
Krisp Technologies EX1024 Page 746

The specifications do not show arithmetic symmetry. The smaller transition width is 2 kHz. For arithmetic symmetry, we therefore choose the band edges as [2, 4, 8, 10] kHz.

The digital frequencies are: passband [0.16, 0.32], stopband [0.08, 0.4], and $F_0 = 0.24$.

The lowpass band edges become $F_p = 0.5(F_{p2} - F_{p1}) = 0.08$ and $F_s = 0.5(F_{s2} - F_{s1}) = 0.16$.

With $A_s = 45$ dB, one of the windows we can use (from Table 20.4) is the Hamming window.

For this window, we estimate $F_{\text{WS}} \approx (F_s - F_p) = C/N$ to obtain $N = 3.47/0.08 = 44$.

We choose the cutoff frequency as $F_C = 0.5(F_p + F_s) = 0.12$.

The lowpass impulse response is $h_N[n] = 2F_C \operatorname{sinc}(2nF_C) = 0.24 \operatorname{sinc}(0.24n),\ -21.5 \le n \le 21.5$.

Windowing this gives $h_W[n] = h_N[n]w[n]$, and the LP2BP transformation gives

$$h_{\text{BP}}[n] = 2\cos(2\pi nF_0)h_W[n] = 2\cos(0.48\pi n)h_W[n]$$

Its frequency response is shown in Figure E20.3C(a) and confirms that the specifications are met.

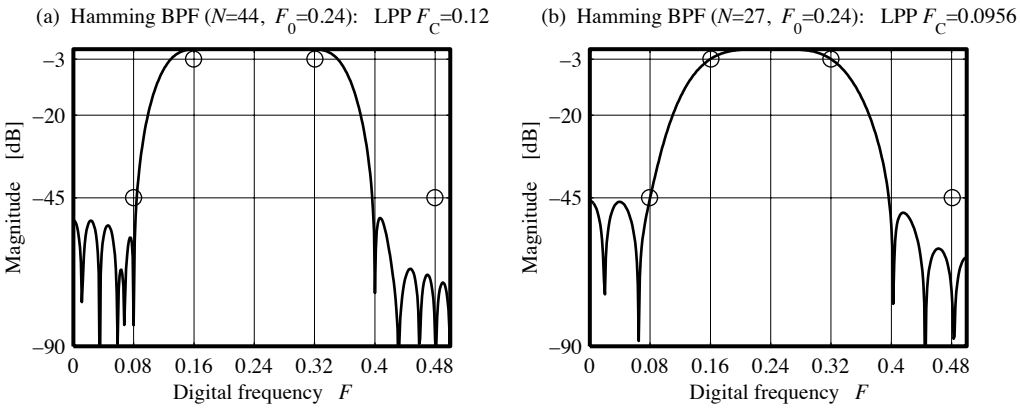


Figure E20.3C Bandpass FIR filters for Example 20.3(e)

The cutoff frequency and smallest filter length that meets specifications turn out to be smaller. By decreasing N and F_C, we find that the specifications are just met with $N = 27$ and $F_C = 0.0956$. For these values, the lowpass filter is $h_N[n] = 2F_C \operatorname{sinc}(2nF_C) = 0.1912 \operatorname{sinc}(0.1912n),\ -13 \le n \le 13$.

Windowing and bandpass transformation yields the filter whose spectrum is shown in Figure E20.3C(b). The attenuation is 3.01 dB at 4 kHz and 8 kHz, 45.01 dB at 2 kHz, and 73.47 dB at 12 kHz.

20.3 Half-Band FIR Filters

A **half-band** FIR filter has an odd-length impulse response $h[n]$ whose alternate samples are zero. The main advantage of half-band filters is that their realization requires only about half the number of multipliers. The impulse response of an ideal lowpass filter is $h[n] = 2F_C \operatorname{sinc}(2nF_C)$. If we choose $F_C = 0.25$, we obtain

$$h[n] = 2F_C \operatorname{sinc}(2nF_C) = 0.5 \operatorname{sinc}(0.5n), \quad |n| \le 0.5(N-1) \tag{20.22}$$

Thus, $h[n] = 0$ for even n, and the filter length N is always odd. Being a type 1 sequence, its transfer function $H(F)$ displays even symmetry about $F = 0$. It is also antisymmetric about $F = 0.25$, with

$$H(F) = 1 - H(0.5 - F) \tag{20.23}$$

PGR2026-00033
Krisp Technologies EX1024 Page 747

A highpass half-band filter also requires $F_C = 0.25$. If we choose $F_C = 0.5(F_p + F_s)$, the sampling frequency S must equal $2(f_p + f_s)$ to ensure $F_C = 0.25$, and cannot be selected arbitrarily. Examples of lowpass and highpass half-band filters are shown in Figure 20.15. Note that the peak passband ripple and the peak stopband ripple are of equal magnitude, with $\delta_p = \delta_s = \delta$ (as they are for any symmetric window).

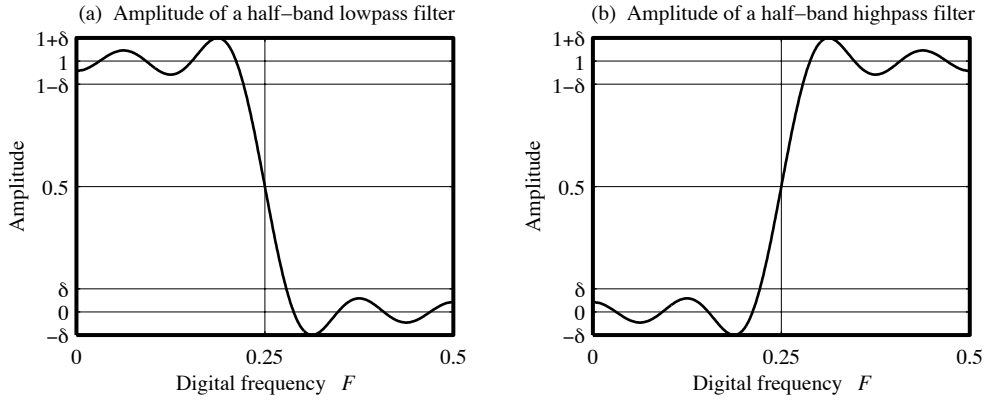


Figure 20.15 Characteristics of lowpass and highpass half-band filters

Since the impulse response of bandstop and bandpass filters contains the term $2\cos(2n\pi F_0)h_{\text{LP}}[n]$, a choice of $F_0 = 0.25$ (for the center frequency) ensures that the odd-indexed terms vanish. Once again, the sampling frequency S cannot be arbitrarily chosen, and must equal $4f_0$ to ensure $F_0 = 0.25$. Even though the choice of sampling rate may cause aliasing, the aliasing will be restricted primarily to the transition band between f_p and f_s, where its effects are not critical.

Except for the restrictions in the choice of sampling rate S (which dictates the choice of F_C for lowpass and highpass filters, or F_0 for bandpass and bandstop filters) and an odd length sequence, the design of half-band filters follows the same steps as window-based design.

REVIEW PANEL 20.6

Recipe for Design of Half-Band FIR Filters

Fix $S = 2(f_p + f_s)$ (for LPF and HPF) or $S = 4f_0$ (for BPF and BSF).
Find F_p and F_s for the lowpass prototype and pick $F_C = 0.5(F_p + F_s)$.
Choose an *odd* filter length N.
Compute the prototype impulse response $h[n] = 2F_C\,\text{sinc}[2nF_C],\ |n| \le 0.5(N-1)$.
Window the impulse response and apply spectral transformations to convert to required filter.
If the specifications are exceeded, decrease N (in steps of 2) until specifications are just met.

EXAMPLE 20.4 (Half-Band FIR Filter Design)

(a) Design a lowpass half-band filter to meet the following specifications:

Passband edge: 8 kHz, stopband edge: 16 kHz, $A_p = 1$ dB, and $A_s = 50$ dB.

We choose $S = 2(f_p + f_s) = 48$ kHz. The digital band edges are $F_p = \frac{1}{6}$, $F_s = \frac{1}{3}$, and $F_C = 0.25$. The impulse response of the filter is $h[n] = 2F_C\,\text{sinc}(2nF_C) = 0.5\,\text{sinc}(0.5n)$.

PGR2026-00033
Krisp Technologies EX1024 Page 748

1. If we use the Kaiser window, we compute the filter length N and the Kaiser parameter β as follows:

$$\delta_p = \frac{10^{A_p/20}-1}{10^{A_p/20}+1} = 0.0575 \quad \delta_s = 10^{-A_s/20} = 0.00316 \quad \delta = 0.00316 \quad A_{s0} = -20\log\delta = 50$$

$$N = \frac{A_{s0}-7.95}{14.36(F_s-F_p)} + 1 = 18.57 \approx 19 \qquad \beta = 0.0351(A_{s0}-8.7) = 1.4431$$

The impulse response is therefore $h_N[n] = 0.5\,\text{sinc}(0.5n), \quad -9 \le n \le 9$.
Windowing $h_N[n]$ gives the required impulse response $h_W[n]$.
Figure E20.4A(a) shows that this filter does meet specifications with an attenuation of 0.045 dB at 8 kHz and 52.06 dB at 16 kHz.

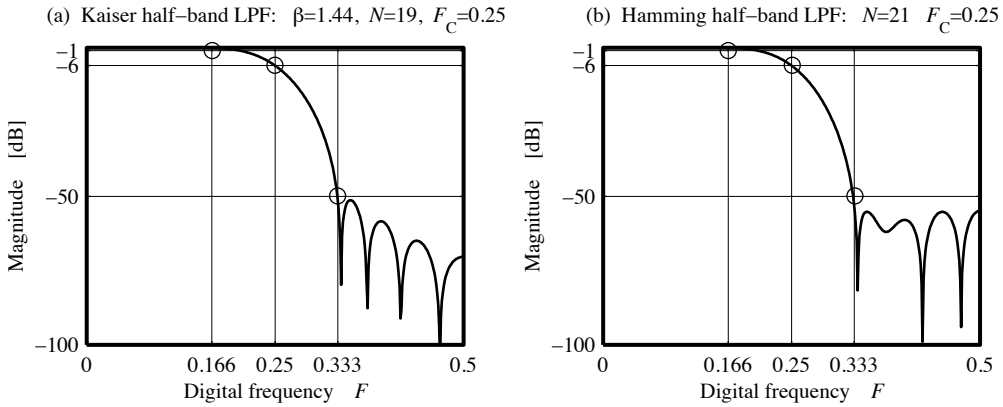


Figure E20.4A Lowpass half-band filters for Example 20.4(a)

2. If we choose a Hamming window, we use Table 20.4 to approximate the *odd* filter length as

$$F_{\text{WS}} = F_s - F_p = \frac{C}{N} \qquad N = \frac{C}{F_s - F_p} = \frac{3.47}{1/6} \approx 21$$

This value of N meets specifications and also turns out to be the smallest length that does. Its response is plotted in Figure E20.4A(b). This filter shows an attenuation of 0.033 dB at 8 kHz and an attenuation of 53.9 dB at 16 kHz.

(b) Design a bandstop half-band filter to meet the following specifications:

stopband edges: [2, 3] kHz, passband edges: [1, 4] kHz, $A_p = 1$ dB, and $A_s = 50$ dB.

Since both the passband and the stopband are symmetric, we have $f_0 = 0.5(2+3) = 2.5$ kHz.
We then choose the sampling frequency as $S = 4f_0 = 10$ kHz. The digital frequencies are

stopband edges = [0.2, 0.3], passband edges = [0.1, 0.4], $F_0 = 0.25$.

The specifications for the lowpass prototype are

$$F_p = 0.5(F_{s2} - F_{s1}) = 0.05, \qquad F_s = 0.5(F_{p2} - F_{p1}) = 0.15, \qquad F_C = 0.5(F_p + F_s) = 0.1$$

The impulse response of the prototype is $h[n] = 2F_C\,\text{sinc}(2nF_C) = 0.2\,\text{sinc}(0.2n)$.

PGR2026-00033
Krisp Technologies EX1024 Page 749

1. If we use the Kaiser window, we must compute the filter length N and the Kaiser parameter β as follows:

$$\delta_p = \frac{10^{A_p/20} - 1}{10^{A_p/20} + 1} = 0.0575 \quad \delta_s = 10^{-A_s/20} = 0.00316 \quad \delta = 0.00316 \quad A_{s0} = -20\log\delta = 50$$

$$N = \frac{A_{s0} - 7.95}{14.36(F_s - F_p)} + 1 = 30.28 \approx 31 \qquad \beta = 0.0351(A_{s0} - 8.7) = 1.4431$$

The prototype impulse response is $h_N[n] = 0.2\,\mathrm{sinc}(0.2n), \quad -15 \le n \le 15$. Windowing $h_N[n]$ gives $h_W[n]$. We transform $h_W[n]$ to the bandstop form $h_{\mathrm{BP}}[n]$, using $F_0 = 0.25$, to give

$$h_{\mathrm{BP}}[n] = \delta[n] - 2\cos(2\pi n F_0)h_W[n] = \delta[n] - 2\cos(0.5\pi n)h_W[n]$$

Figure E20.4B(a) shows that this filter does meet specifications, with an attenuation of 0.046 dB at 2 kHz and 3 kHz, and 53.02 dB at 1 kHz and 4 kHz.

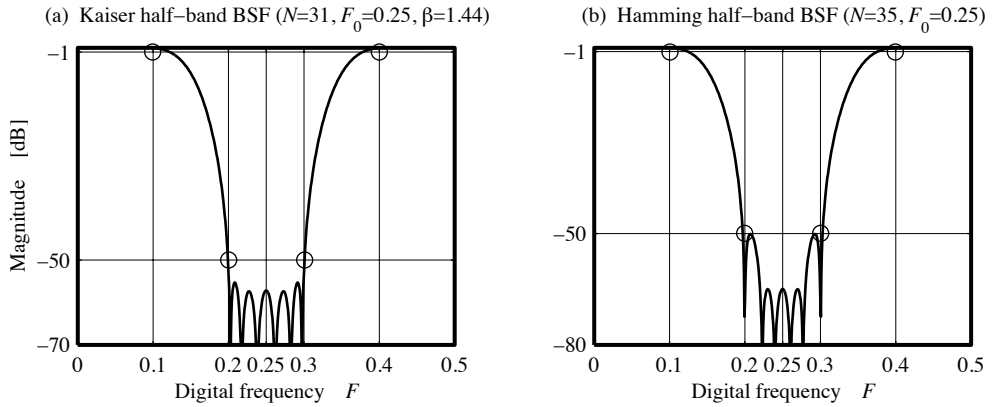


Figure E20.4B Bandstop half-band filters for Example 20.4(b)

2. For a Hamming window, we use Table 20.4 to approximate the *odd* filter length as

$$F_{\mathrm{WS}} = F_s - F_p = \frac{C}{N} \qquad N = \frac{C}{F_s - F_p} = \frac{3.47}{0.1} \approx 35$$

This is also the smallest filter length that meets specifications. The magnitude response is shown in Figure E20.4B(b). We see an attenuation of 0.033 dB at [2, 3] kHz and 69.22 dB at [1, 4] kHz.

20.4 FIR Filter Design by Frequency Sampling

In window-based design, we start with the impulse response $h[n]$ of an ideal filter, and use truncation and windowing to obtain an acceptable frequency response $H(F)$. The design of FIR filters by frequency sampling starts with the required form for $H(F)$, and uses interpolation and the DFT to obtain $h[n]$. In this sense, it is more versatile, since arbitrary frequency response forms can be handled with ease.

Recall that a continuous (but band-limited) signal $h(t)$ can be perfectly reconstructed from its samples (taken above the Nyquist rate), using a sinc interpolating function (that equals zero at the sampling instants). If $h(t)$ is not band-limited, we get a perfect match only at the sampling instants.

PGR2026-00033
Krisp Technologies EX1024 Page 750

By analogy, the continuous (but periodic) spectrum $H(F)$ can also be recovered from its frequency samples, using a periodic extension of the sinc interpolating function (that equals zero at the sampling points). The reconstructed spectrum $H_N(F)$ will show an exact match to a desired $H(F)$ at the sampling instants, even though $H_N(F)$ could vary wildly at other frequencies. This is the basis for FIR filter design by frequency sampling. Given the desired form for $H(F)$, we sample it at N frequencies, and find the IDFT of the N-point sequence $H[k],\ k = 0, 1, \ldots, N-1$. The following design guidelines stem both from design aspects as well as computational aspects of the IDFT itself.

1. The N samples of $H(F)$ must correspond to the digital frequency range $0 \le F < 1$, with

$$H[k] = H(F)\big|_{F=k/N}, \qquad k = 0,\ 1,\ 2, \ldots, N-1 \tag{20.24}$$

   The reason is that most DFT and IDFT algorithms require samples in the range $0 \le k \le N-1$.

2. Since $h[n]$ must be real, its DFT $H[k]$ must possess conjugate symmetry about $k = 0.5N$ (this is a DFT requirement). Note that conjugate symmetry will always leave $H[0]$ unpaired. It can be set to any real value, in keeping with the required filter type (this is a design requirement). For example, we must choose $H[0] = 0$ for bandpass or highpass filters.

3. For even length N, the computed end-samples of $h[n]$ may not turn out to be symmetric. To ensure symmetry, we must force $h[0]$ to equal $h[N]$ (setting both to $0.5h[0]$, for example).

4. For $h[n]$ to be causal, we must delay it (this is a design requirement). This translates to a linear phase shift to produce the sequence $|H[k]|e^{j\phi[k]}$. In keeping with conjugate symmetry about the index $k = 0.5N$, the phase for the first $N/2$ samples of $H[k]$ will be given by

$$\phi[k] = \frac{-\pi k(N-1)}{N}, \qquad k = 0,\ 1,\ 2, \ldots, 0.5(N-1) \tag{20.25}$$

   Note that for type 3 and type 4 (antisymmetric) sequences, we must also add a constant phase of 0.5π to $\phi[k]$ (up to $k = 0.5N$). The remaining samples of $H[k]$ are found by conjugate symmetry.

5. To minimize the Gibbs effect near discontinuities in $H(F)$, we may allow the sample values to vary gradually between jumps (this is a design guideline). This is equivalent to introducing a finite transition width. The choice of the sample values in the transition band can affect the response dramatically.

EXAMPLE 20.5 (FIR Filter Design by Frequency Sampling)

(a) Consider the design of a lowpass filter shown in Figure E20.5(a). Let us sample the ideal $H(F)$ (shown dark) over $0 \le F < 1$, with $N = 10$ samples. The magnitude of the sampled sequence $H[k]$ is

$$|H[k]| = \{1,\ 1,\ 1,\ 0,\ 0,\ \underbrace{0,}_{k=5}\ 0,\ 0,\ 1,\ 1\}$$

The actual (phase-shifted) sequence is $H[k] = |H[k]|e^{j\phi[k]}$, where

$$\phi[k] = \frac{-\pi k(N-1)}{N} = -0.9\pi k, \quad k \le 5$$

Note that $H[k]$ must be conjugate symmetric about $k = 0.5N = 5$, with $H[k] = H^*[N-k]$.
Now, $H[k] = 0,\ k = 4, 5, 6, 7$, and the remaining samples are $H[0] = 1e^{j0} = 1$ and

$$H[1] = 1e^{j\phi[1]} = e^{-j0.9\pi} \quad H[9] = H^*[1] = e^{j0.9\pi} \qquad H[2] = 1e^{j\phi[3]} = e^{-j1.8\pi} \quad H[8] = H^*[2] = e^{j1.8\pi}$$

PGR2026-00033
Krisp Technologies EX1024 Page 751

The inverse DFT of $H[k]$ yields the symmetric real impulse response sequence $h_1[n]$, with

$$h_1[n] = \{0.0716,\ -0.0794,\ 0.1,\ 0.1558,\ 0.452,\ 0.452,\ 0.1558,\ 0.1,\ -0.0794,\ 0.0716\}$$

Its DTFT magnitude $H_1(F)$, shown light in Figure E20.5(a), reveals a perfect match at the sampling points but has a large overshoot near the cutoff frequency. To reduce the overshoot, let us pick

$$H[2] = 0.5e^{j\phi[2]} = 0.5e^{-1.8\pi} \qquad H[8] = H^*[2] = 0.5e^{1.8\pi}$$

The inverse DFT of this new set of samples yields the new impulse response sequence $h_2[n]$:

$$h_2[n] = \{-0.0093,\ -0.0485,\ 0,\ 0.1867,\ 0.3711,\ 0.3711,\ 0.1867,\ 0,\ -0.0485,\ -0.0093\}$$

Its frequency response $H_2(F)$, in Figure E20.5(a), not only shows a perfect match at the sampling points but also a reduced overshoot, which we obtain at the expense of a broader transition width.

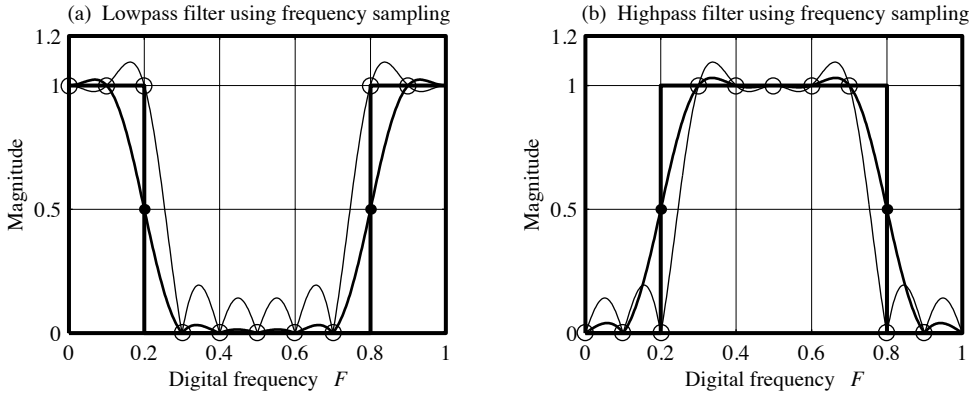


Figure E20.5 Lowpass and highpass filters for Example 20.5

(a and b)

(b) Consider the design of a highpass filter shown in Figure E20.5(b). Let us sample the ideal $H(F)$ (shown dark) over $0 \le F < 1$, with $N = 10$ samples. The magnitude of the sampled sequence $H[k]$ is

$$|H[k]| = \{0,\ 0,\ 0,\ 1,\ 1,\ \underbrace{1,}_{k=5}\ 1,\ 1,\ 0,\ 0\}$$

The actual (phase-shifted) sequence is $H[k] = |H[k]|e^{j\phi[k]}$. Since the impulse response $h[n]$ must be antisymmetric (for a highpass filter), $\phi[k]$ includes an additional phase of 0.5π and is given by

$$\phi[k] = \frac{-\pi k(N-1)}{N} + 0.5\pi = 0.5\pi - 0.9\pi k, \quad k \le 5$$

Note that $H[k]$ is conjugate symmetric about $k = 0.5N = 5$, with $H[k] = H^*[N-k]$.
Now, $H[k] = 0,\ k = 0, 1, 2, 8, 9$, and $H[5] = 1e^{j\phi[5]} = 1$. The remaining samples are

$$H[3] = 1e^{j\phi[3]} = e^{-j2.2\pi} \quad H[7] = H^*[3] = e^{j2.2\pi} \qquad H[4] = 1e^{j\phi[4]} = e^{-j3.1\pi} \quad H[6] = H^*[4] = e^{j3.1\pi}$$

The inverse DFT of $H[k]$ yields the antisymmetric real impulse response sequence $h_1[n]$, with

$$h_1[n] = \{0.0716,\ 0.0794,\ -0.1,\ -0.1558,\ 0.452,\ -0.452,\ 0.1558,\ 0.1,\ -0.0794,\ -0.0716\}$$

PGR2026-00033
Krisp Technologies EX1024 Page 752

Its DTFT magnitude $H_1(F)$, shown light in Figure E20.5(b), reveals a perfect match at the sampling points but a large overshoot near the cutoff frequency. To reduce the overshoot, let us choose

$$H[2] = 0.5e^{j\phi[2]} = 0.5e^{-1.3\pi} \qquad H[8] = 0.5e^{1.3\pi}$$

The inverse DFT of this new set of samples yields the new impulse response sequence $h_2[n]$:

$$h_2[n] = \{0.0128,\ -0.0157,\ -0.1,\ -0.0606,\ 0.5108,\ -0.5108,\ 0.0606,\ 0.1,\ 0.0157,\ -0.0128\}$$

Its frequency response $H_2(F)$, in Figure E20.5(b), not only shows a perfect match at the sampling points but also a reduced overshoot, which we obtain at the expense of a broader transition width.

20.4.1 Frequency Sampling and Windowing

The frequency-sampling method can be used to design filters with arbitrary frequency-response shapes. We can even combine this versatility with the advantages of window-based design. Given the response specification $H(F)$, we sample it at a *large* number of points M and find the IDFT to obtain the M-point impulse response $h[n]$. The choice $M = 512$ is not unusual. Since $h[n]$ is unacceptably long, we truncate it to a smaller length N by windowing $h[n]$. The choice of window is based on the same considerations that apply to window-based design. If the design does not meet specifications, we can change N and/or adjust the sample values in the transition band and repeat the process. Naturally, this design is best carried out on a computer. The sample values around the transitions are adjusted to minimize the approximation error. This idea forms a special case of the more general optimization method called *linear programming*. But when it comes right down to choosing between the various computer-aided optimization methods, by far the most widely used is the *equiripple optimal approximation* method, which we describe in the next section.

20.4.2 Implementing Frequency-Sampling FIR Filters

We can readily implement frequency-sampling FIR filters by a nonrecursive structure, once we know the impulse response $h[n]$. We can even implement a recursive realization without the need for finding $h[n]$, as follows. Suppose the frequency samples $H_N[k]$ of $H(F)$ are approximated by an N-point DFT of the filter impulse response $h_N[n]$. We may then write

$$H_N[k] = \sum_{k=0}^{N-1} h_N[n]e^{-j2\pi nk/N} \tag{20.26}$$

Its impulse response $h_N[n]$ may be found using the inverse DFT as

$$h_N[n] = \frac{1}{N}\sum_{k=0}^{N-1} H_N[k]e^{j2\pi nk/N} \tag{20.27}$$

The filter transfer function $H(z)$ is the z-transform of $h_N[n]$:

$$H(z) = \sum_{n=0}^{N-1} h_N[n]z^{-n} = \sum_{n=0}^{N-1} z^{-n}\left[\frac{1}{N}\sum_{k=0}^{N-1} H_N[k]e^{j2\pi nk/N}\right] \tag{20.28}$$

Interchanging summations, setting $z^{-n}e^{j2\pi nk/N} = [z^{-1}e^{j2\pi k/N}]^n$, and using the closed form for the finite geometric sum, we obtain

$$H(z) = \frac{1}{N}\sum_{k=0}^{N-1} H_N[k]\left[\frac{1-z^{-N}}{1-z^{-1}e^{j2\pi k/N}}\right] \tag{20.29}$$

PGR2026-00033
Krisp Technologies EX1024 Page 753

The frequency response corresponding to $H(z)$ is

$$H(F) = \frac{1}{N}\sum_{k=0}^{N-1} H_N[k]\frac{1-e^{-j2\pi FN}}{1-e^{-j2\pi(F-k/N)}} \tag{20.30}$$

If we factor out $\exp(-j\pi FN)$ from the numerator, $\exp[-j\pi(F-k/N)]$ from the denominator, and use Euler's relation, we can simplify this result to

$$H(F) = \sum_{k=0}^{N-1} H_N[k]\frac{\text{sinc}[N(F-\frac{k}{N})]}{\text{sinc}[(F-\frac{k}{N})]}e^{-j\pi(N-1)(F-k/N)} = \sum_{k=0}^{N-1} H_N[k]W[F-\tfrac{k}{N}] \tag{20.31}$$

Here, $W[F-\frac{k}{N}]$ describes a sinc interpolating function, defined by

$$W[F-\tfrac{k}{N}] = \frac{\text{sinc}[N(F-\frac{k}{N})]}{\text{sinc}[(F-\frac{k}{N})]}e^{-j\pi(N-1)(F-k/N)} \tag{20.32}$$

and reconstructs $H(F)$ from its samples $H_N[k]$ taken at intervals $1/N$. It equals 1 when $F = k/N$, and zero otherwise. As a result, $H_N(F)$ equals the desired $H(F)$ at the sampling instants, even though $H_N(F)$ could vary wildly at other frequencies. This is the concept behind frequency sampling.

The transfer function $H(z)$ may also be written as the product of two transfer functions:

$$H(z) = H_1(z)H_2(z) = \left[\frac{1-z^{-N}}{N}\right]\sum_{k=0}^{N-1}\left[\frac{H_N[k]}{1-z^{-1}e^{j2\pi k/N}}\right] \tag{20.33}$$

This form of $H(z)$ suggests a method of recursive implementation of FIR filters. We cascade a comb filter, described by $H_1(z)$, with a parallel combination of N first-order resonators, described by $H_2(z)$. Note that each resonator has a complex pole on the unit circle and the resonator poles actually lie at the same locations as the zeros of the comb filter. Each pair of terms corresponding to complex conjugate poles may be combined to form a second-order system with real coefficients for easier implementation.

Why implement FIR filters recursively? There are several reasons. In some cases, this may reduce the number of arithmetic operations. In other cases, it may reduce the number of delay elements required. Since the pole and zero locations depend only on N, such filters can be used for all FIR filters of length L by changing only the multiplicative coefficients.

Even with these advantages, things can go wrong. In theory, the poles and zeros balance each other. In practice, quantization errors may move some poles outside the unit circle and lead to system instability. One remedy is to multiply the poles and zeros by a real number ρ slightly smaller than unity to relocate the poles and zeros. The transfer function then becomes

$$H(z) = \frac{1-(\rho z)^{-N}}{N}\sum_{k=0}^{N-1}\left[\frac{H_N[k]}{1-(\rho z)^{-1}e^{j2\pi k/N}}\right] \tag{20.34}$$

With $\rho = 1-\epsilon$, typically used values for ϵ range from 2^{-12} to 2^{-27} (roughly 10^{-4} to 10^{-9}) and have been shown to improve stability with little change in the frequency response.

20.5 Design of Optimal Linear-Phase FIR Filters

Quite like analog filters, the design of optimal linear-phase FIR filters requires that we minimize the maximum error in the approximation. Optimal design of FIR filters is also based on a Chebyshev approximation. We

PGR2026-00033
Krisp Technologies EX1024 Page 754

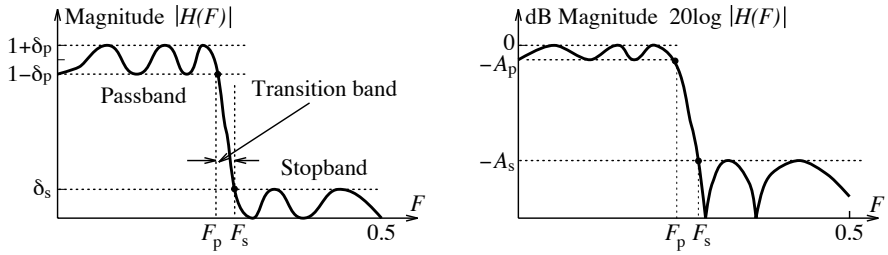


Figure 20.16 An optimal filter has ripples of equal magnitude in the passband and stopband

should therefore expect such a design to yield the smallest filter length, and a response that is equiripple in both the passband and the stopband. A typical spectrum is shown in Figure 20.16.

There are three important concepts relevant to optimal design:

1. The error between the approximation $H(F)$ and the desired response $D(F)$ must be equiripple. The error curve must show equal maxima and minima with alternating zero crossings. The more the number of points where the error goes to zero (the zero crossings), the higher is the order of the approximating polynomial and the longer is the filter length.

2. The frequency response $H(F)$ of a filter whose impulse response $h[n]$ is a symmetric sequence can always be put in the form

$$H(F) = Q(F) \sum_{n=0}^{M} \alpha_n \cos(2\pi nF) = Q(F)P(F) \tag{20.35}$$

   Here, $Q(F)$ equals 1 (type 1), $\cos(\pi F)$ (type 2), $\sin(2\pi F)$ (type 3), or $\sin(\pi F)$ (type 4); M is related to the filter length N with $M = \text{int}(\frac{N-1}{2})$ (types 1, 2, 4) or $M = \text{int}(\frac{N-3}{2})$ (type 3); and the α_n are related to the impulse response coefficients $h[n]$. The quantity $P(F)$ may also be expressed as as a power series in $\cos(2\pi F)$ (or as a sum of Chebyshev polynomials). If we can select the α_n to best meet optimal constraints, we can design $H(F)$ as an optimal approximation to $D(F)$.

3. The *alternation theorem* offers the clue to selecting the α_n.

20.5.1 The Alternation Theorem

We start by approximating $D(F)$ by the Chebyshev polynomial form for $H(F)$ and define the *weighted* approximation error $\epsilon(F)$ as

$$\epsilon(F) = W(F)[D(F) - H(F)] \tag{20.36}$$

Here, $W(F)$ represents a set of weight factors that can be used to select different error bounds in the passband and stopband. The nature of $D(F)$ and $W(F)$ depends on the type of the filter required. The idea is to select the α_k (in the expression for $H(F)$) so as to minimize the maximum absolute error $|\epsilon|_{\max}$. The alternation theorem points the way (though it does not tells us how). In essence, it says that we must be able to find at least $M + 2$ frequencies $F_k, k = 1, 2, \ldots, M + 2$, called the **extremal frequencies** or **alternations** where

1. The error alternates between two equal maxima and minima (extrema):

$$\epsilon(F_k) = -\epsilon(F_{k+1}), \qquad k = 1, 2, \ldots, M + 1 \tag{20.37}$$

PGR2026-00033
Krisp Technologies EX1024 Page 755

2. The error at the frequencies F_k equals the *maximum absolute error*:

$$|\epsilon(F_k)| = |\epsilon(F)|_{\max}, \qquad k = 1, 2, \ldots, M+2 \tag{20.38}$$

In other words, we require $M+2$ extrema (including the band edges) where the error attains its maximum absolute value. These frequencies yield the smallest filter length (number of coefficients α_k) for optimal design. In some instances, we may get $M+3$ extremal frequencies leading to so called *extra ripple* filters.

The design strategy to find the extremal frequencies invariably requires iterative methods. The most popular is the algorithm of Parks and McClellan, which in turn relies on the so-called **Remez exchange algorithm**.

The Parks-McClellan algorithm requires the band edge frequencies F_p and F_s, the ratio $K = \delta_p/\delta_s$ of the passband and stopband ripple, and the filter length N. It returns the coefficients α_k and the actual design values of δ_p and δ_s for the given filter length N. If these values of δ_p and δ_s are not acceptable (or do not meet requirements), we can increase N (or change the ratio K), and repeat the design.

A good starting estimate for the filter length N is given by a relation similar to the Kaiser relation for half-band filters, and reads

$$N = 1 + \frac{-10\log(\delta_p\delta_s) - 13}{14.6F_T} \qquad \delta_p = \frac{10^{A_p/20} - 1}{10^{A_p/20} + 1} \qquad \delta_s = 10^{-A_s/20} \tag{20.39}$$

Here, F_T is the digital transition width. More accurate (but more involved) design relations are also available.

To explain the algorithm, consider a lowpass filter design. To approximate an ideal lowpass filter, we choose $D(F)$ and $W(F)$ as

$$D(F) = \begin{cases} 1, & 0 \le F \le F_p \\ 0, & F_s \le F \le 0.5 \end{cases} \qquad W(F) = \begin{cases} 1, & 0 \le F \le F_p \\ K = \delta_p/\delta_s, & F_s \le F \le 0.5 \end{cases} \tag{20.40}$$

To find the α_k in $H(F)$, we use the **Remez exchange algorithm**. Here is how it works. We start with a trial set of $M+2$ frequencies $F_k, k = 1, 2, \ldots, M+2$. To force the alternation condition, we must satisfy $\epsilon(F_k) = -\epsilon(F_{k+1}), k = 1, 2, \ldots, M+1$. Since the maximum error is yet unknown, we let $\rho = \epsilon(F_k) = -\epsilon(F_{k+1})$. We now have $M+1$ unknown coefficients α_k and the unknown ρ, a total of $M+2$. We solve for these by using the $M+2$ frequencies, to generate the $M+2$ equations:

$$-(-1)^k\rho = W(F_k)[D(F_k) - H(F_k)], \qquad k = 1, 2, \ldots, M+2 \tag{20.41}$$

Here, the quantity $(-1)^k$ brings out the alternating nature of the error.

Once the α_k are found, the right-hand side of this equation is known in its entirety and is used to compute the extremal frequencies. The problem is that these frequencies may no longer satisfy the alternation condition. So we must go back and evaluate a new set of α_k and ρ, using the computed frequencies. We continue this process until the computed frequencies also turn out to be the actual extremal frequencies (to within a given tolerance, of course).

Do you see why it is called the *exchange* algorithm? First, we exchange an old set of frequencies F_k for a new one. Then we exchange an old set of α_k for a new one. Since the α_k and F_k actually describe the impulse response and frequency response of the filter, we are in essence going back and forth between the two domains until the coefficients α_k yield a spectrum with the desired optimal characteristics.

Many time-saving steps have been suggested to speed the computation of the extremal frequencies, and the α_k, at each iteration. The Parks-McClellan algorithm is arguably one of the most popular methods of filter design in the industry, and many of the better commercial software packages on signal processing include it in their stock list of programs. Having said that, we must also point out two disadvantages of this method. First, the filter length must still be estimated by empirical means. And second, we have no control over the actual ripple that the design yields. The only remedy, if this ripple is unacceptable, is to start afresh with a different set of weight functions or with a different filter length.

PGR2026-00033
Krisp Technologies EX1024 Page 756

20.5.2 Optimal Half-Band Filters

Since about half the coefficients of a half-band filter are zero, the computational burden can be reduced by developing a filter that contains only the nonzero coefficients followed by zero interpolation. To design a half-band lowpass filter $h_{\text{HB}}[n]$ with band edges F_p and F_s, and ripple δ, an estimate of the filter length $N = 4k - 1$ is first found from the given design specifications. Next, we design a lowpass prototype $h_P[n]$ of *even length* $0.5(N + 1)$, with band edges of $2F_p$ and 0.5 (in effect, with no stopband). This filter describes a type 2 sequence whose response is zero at $F = 0.5$ and whose ripple is twice the design ripple. Finally, we insert zeros between adjacent samples of $0.5h_P[n]$ and set the (zero-valued) center coefficient to 0.5 to obtain the required half-band filter $h_{\text{HB}}[n]$ with the required ripple.

EXAMPLE 20.6 (Optimal FIR Filter Design)

(a) We design an optimal bandstop filter with stopband edges of [2, 3] kHz, passband edges of [1, 4] kHz, $A_p = 1$ dB, $A_s = 50$ dB, and a sampling frequency of $S = 10$ kHz

We find the digital passband edges as [0.2, 0.3] and the stopband edges as [0.1, 0.4].
The transition width is $F_T = 0.1$. We find the approximate filter length N, as follows:

$$\delta_p = \frac{10^{A_p/20} - 1}{10^{A_p/20} + 1} = 0.0575 \qquad \delta_s = 10^{-A_s/20} = 0.00316 \qquad N = 1 + \frac{-10\log(\delta_p\delta_s) - 13}{14.6F_T} \approx 17.7$$

Choosing the next odd length gives $N = 19$. The bandstop specifications are actually met by a filter with $N = 21$. The response of the designed filter is shown in Figure E20.6(a).

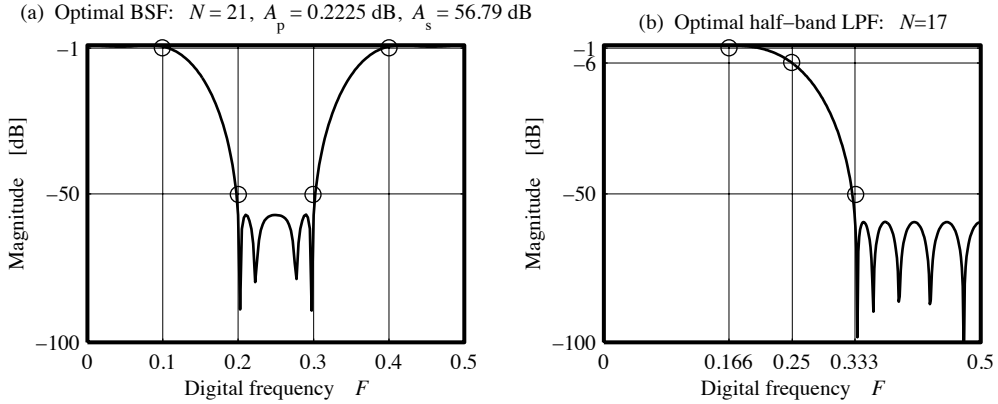


Figure E20.6 Optimal filters for Example 20.6

(a and b)

(b) (Optimal Half-Band Filter Design) We design an optimal half-band filter to meet the following specifications:

Passband edge: 8 kHz, stopband edge: 16 kHz, $A_p = 1$ dB, and $A_s = 50$ dB.

We choose $S = 2(f_p + f_s) = 48$ kHz. The digital band edges are $F_p = \frac{1}{6}$, $F_s = \frac{1}{3}$, and $F_C = 0.25$.
Next, we find the minimum ripple δ and the approximate filter length N as

$$\delta_p = \frac{10^{A_p/20} - 1}{10^{A_p/20} + 1} = 0.0575 \qquad \delta_s = 10^{-A_s/20} = 0.00316 \qquad \delta = 0.00316$$

PGR2026-00033
Krisp Technologies EX1024 Page 757

$$N = 1 + \frac{-10\log\delta^2 - 13}{14.6(F_s - F_p)} = 1 + \frac{-20\log(0.00316) - 13}{14.6(1/6)} \approx 16.2 \quad \Rightarrow N = 19$$

The choice $N = 19$ is based on the form $N = 4k - 1$. Next, we design the lowpass prototype $h_P[n]$ of even length $M = 0.5(N + 1) = 10$ and band edges at $2F_p = 2(1/6) = 1/3$ and $F_s = 0.5$. The result is

$$h_P[n] = \{0.0074,\ -0.0267,\ 0.0708,\ -0.173,\ 0.6226,\ 0.6226,\ -0.173,\ 0.0708,\ -0.0267,\ 0.0074\}$$

Finally, we zero-interpolate $0.5h_P[n]$ and choose the center coefficient as 0.5 to obtain $H_{\rm HB}[n]$:

$$\{0.0037, 0, -0.0133, 0, 0.0354,\ 0, -0.0865, 0,\ 0.3113,\ 0,\ 0.5,\ 0.3113, 0, -0.0865, 0, 0.0354, 0, -0.0133, 0, 0.0037\}$$

Its length is $N = 19$, as required. Figure E20.6(b) shows that this filter does meet specifications, with an attenuation of 0.02 dB at 8 kHz and 59.5 dB at 16 kHz.

20.6 Application: Multistage Interpolation and Decimation

Recall that a sampling rate increase, or interpolation, by N involves up-sampling (inserting $N - 1$ zeros between samples), followed by lowpass filtering with a gain of N. If the interpolation factor N is large, it is much more economical to carry out the interpolation in stages because it results in a smaller overall filter length or order. For example, if N can be factored as $N = I_1 I_2 I_3$, then interpolation by N can be accomplished in three stages with individual interpolation factors of I_1, I_2, and I_3. At a typical Ith stage, the output sampling rate is given by $S_{\rm out} = IS_{\rm in}$, as shown in Figure 20.17.

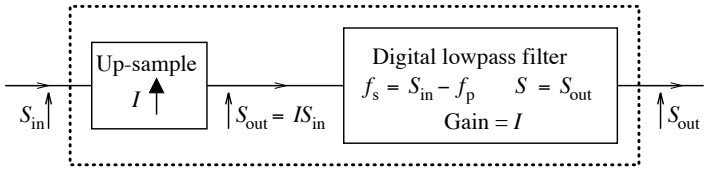


Figure 20.17 One stage of a typical multistage interpolating filter

The filter serves to remove the spectral replications due to the zero interpolation by the up-sampler. These replicas occur at multiples of the input sampling rate. As a result, the filter stopband edge is computed from the input sampling rate $S_{\rm in}$ as $f_s = S_{\rm in} - f_p$, while the filter passband edge remains fixed (by the given specifications). The filter sampling rate corresponds to the output sampling rate $S_{\rm out}$ and the filter gain equals the interpolation factor I_k. At each successive stage (except the first), the spectral images occur at higher and higher frequencies, and their removal requires filters whose transition bands get wider with each stage, leading to less complex filters with smaller filter lengths. Although it is not easy to establish the optimum values for the interpolating factors and their order for the smallest overall filter length, it turns out that interpolating factors in increasing order generally yield smaller overall lengths, and any multistage design results in a substantial reduction in the filter length as compared to a single-stage design.

EXAMPLE 20.7 (The Concept of Multistage Interpolation)

(a) Consider a signal band-limited to 1.8 kHz and sampled at 4 kHz. It is required to raise the sampling rate to 48 kHz. This requires interpolation by 12. The value of the passband edge is $f_p = 1.8$ kHz for either a single-stage design or multistage design.

PGR2026-00033
Krisp Technologies EX1024 Page 758

For a single-stage interpolator, the output sampling rate is $S_{\text{out}} = 48$ kHz, and we thus require a filter with a stopband edge of $f_s = S_{\text{in}} - f_p = 4 - 1.8 = 2.2$ kHz, a sampling rate of $S = S_{\text{out}} = 48$ kHz, and a gain of 12. If we use a crude approximation for the filter length as $L \approx 4/F_T$, where $F_T = (f_s - f_p)/S$ is the digital transition width, we obtain $L = 4(48/0.4) = 480$.

(b) If we choose two-stage interpolation with $I_1 = 3$ and $I_2 = 4$, at each stage we compute the important parameters as follows:

Stage	S_{in} (kHz)	**Interpolation Factor**	$S_{\text{out}} = S$ (kHz)	f_p (kHz)	$f_s = S_{\text{in}} - f_p$ (kHz)	**Filter Length** $L \approx 4S/(f_s - f_p)$
1	4	$I_1 = 3$	12	1.8	2.2	$48/0.4 = 120$
2	12	$I_2 = 4$	48	1.8	10.2	$192/8.4 = 23$

The total filter length is thus 143.

(c) If we choose three-stage interpolation with $I_1 = 2,\ I_2 = 3$, and $I_3 = 2$, at each stage we compute the important parameters, as follows:

Stage	S_{in} (kHz)	**Interpolation Factor**	$S_{\text{out}} = S$ (kHz)	f_p (kHz)	$f_s = S_{\text{in}} - f_p$ (kHz)	**Filter Length** $L \approx 4S/(f_s - f_p)$
1	4	$I_1 = 2$	8	1.8	2.2	$32/0.4 = 80$
2	8	$I_2 = 3$	24	1.8	6.2	$96/4.4 = 22$
3	24	$I_2 = 2$	48	1.8	22.2	$192/20.4 = 10$

The total filter length is thus 112.

(d) If we choose three-stage interpolation but with the different order $I_1 = 3,\ I_2 = 2$, and $I_3 = 2$, at each stage we compute the important parameters, as follows:

Stage	S_{in} (kHz)	**Interpolation Factor**	$S_{\text{out}} = S$ (kHz)	f_p (kHz)	$f_s = S_{\text{in}} - f_p$ (kHz)	**Filter Length** $L \approx 4S/(f_s - f_p)$
1	4	$I_1 = 3$	12	1.8	2.2	$48/0.4 = 120$
2	12	$I_2 = 2$	24	1.8	10.2	$96/8.4 = 12$
3	24	$I_2 = 2$	48	1.8	22.2	$192/20.4 = 10$

The total filter length is thus 142.

Any multistage design results in a substantial reduction in the filter length as compared to a single-stage design, and a smaller interpolation factor in the first stage of a multistage design does seem to yield smaller overall lengths. Also remember that the filter lengths are only a crude approximation to illustrate the relative merits of each design and the actual filter lengths will depend on the attenuation specifications.

PGR2026-00033
Krisp Technologies EX1024 Page 759

For multistage operations, the actual filter lengths depend not only on the order of the interpolating factors, but also on the given attenuation specifications A_p and A_s. Since attenuations in decibels add in a cascaded system, the passband attenuation A_p is usually distributed among the various stages to ensure an overall attenuation that matches specifications. The stopband specification needs no such adjustment. Since the signal is attenuated even further at each stage, the overall stopband attenuation always exceeds specifications. For interpolation, we require a filter whose gain is scaled (multiplied) by the interpolation factor of the stage.

EXAMPLE 20.8 (Design of Interpolating Filters)

(a) Consider a signal band-limited to 1.8 kHz and sampled at 4 kHz. It is required to raise the sampling rate to 48 kHz. The passband attenuation should not exceed 0.6 dB, and the minimum stopband attenuation should be 50 dB. Design an interpolation filter using a single-stage interpolator.

The single stage requires interpolation by 12. The value of the passband edge is $f_p = 1.8$ kHz. The output sampling rate is $S_{\text{out}} = 48$ kHz, and we thus require a filter with a stopband edge of $f_s = S_{\text{in}} - f_p = 4 - 1.8 = 2.2$ kHz and a sampling rate of $S = S_{\text{out}} = 48$ kHz. We ignore the filter gain of 12 in the following computations. To compute the filter length, we first find the ripple parameters

$$\delta_p = \frac{10^{A_p/20} - 1}{10^{A_p/20} + 1} = \frac{10^{0.6/20} - 1}{10^{0.6/20} + 1} = 0.0345 \qquad \delta_s = 10^{-A_s/20} = 10^{-50/20} = 0.00316$$

and then approximate the length N by

$$N \approx \frac{-10\log(\delta_p\delta_s) - 13}{14.6(F_s - F_p)} + 1 = \frac{S[-10\log(\delta_p\delta_s) - 13]}{14.6(f_s - f_p)} + 1 = 230$$

The actual filter length of the optimal filter turns out to be $N = 233$, and the filter shows a passband attenuation of 0.597 dB and a stopband attenuation of 50.05 dB.

(b) Repeat the design using a three-stage interpolator with $I_1 = 2,\ I_2 = 3$, and $I_3 = 2$. How do the results compare with those of the single-stage design?

We distribute the passband attenuation (in decibels) equally among the three stages. Thus, $A_p = 0.2\,\text{dB}$ for each stage, and the ripple parameters for each stage are

$$\delta_p = \frac{10^{0.2/20} - 1}{10^{0.2/20} + 1} = 0.0115 \qquad \delta_s = 10^{-50/20} = 0.00316$$

For each stage, the important parameters and the filter length are listed in the following table:

Stage	S_{in} (kHz)	**Interpolation Factor**	$S_{\text{out}} = S$ (kHz)	f_p (kHz)	$f_s = S_{\text{in}} - f_p$ (kHz)	**Filter Length** L
1	4	$I_1 = 2$	8	1.8	2.2	44
2	8	$I_2 = 3$	24	1.8	6.2	13
3	24	$I_2 = 2$	48	1.8	22.2	7

In this table, for example, we compute the filter length for the first stage as

$$L \approx \frac{-10\log(\delta_p\delta_s) - 13}{14.6(F_s - F_p)} + 1 = \frac{S[-10\log(\delta_p\delta_s) - 13]}{14.6(f_s - f_p)} + 1 = 44$$

PGR2026-00033
Krisp Technologies EX1024 Page 760

The actual filter lengths of the optimal filters turn out to be 47 (with design attenuations of 0.19 dB and 50.31 dB), 13 (with design attenuations of 0.18 dB and 51.09 dB), and 4 (with design attenuations of 0.18 dB and 50.91 dB). The overall filter length is only 64. This is about four times less than the filter length for single-stage design.

20.6.1 Multistage Decimation

Decimation by M involves lowpass filtering (with a gain of unity), followed by down-sampling (discarding $M-1$ samples and retaining every Mth sample). The process is essentially the inverse of interpolation. If the decimation factor M is large, decimation in stages results in a smaller overall filter length or order. If M can be factored as $M = D_1D_2D_3$, then decimation by M can be accomplished in three stages with individual decimation factors of D_1, D_2, and D_3. At a typical stage, the output sampling rate is given by $S_{\text{out}} = S_{\text{in}}/D$, where D is the decimation factor for that stage. This is illustrated in Figure 20.18.

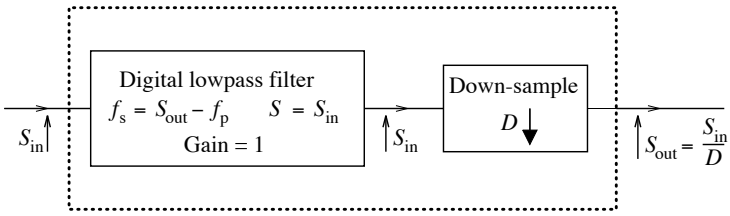


Figure 20.18 One stage of a typical multistage decimating filter

The decimation filter has a gain of unity and operates at the input sampling rate S_{in}, and its stopband edge is computed from the output sampling rate as $f_s = S_{\text{out}} - f_p$. At each successive stage (except the first), the transition bands get narrower with each stage.

The overall filter length does depend on the order in which the decimating factors are used. Although it is not easy to establish the optimum values for the decimation factors and their order for the smallest overall filter length, it turns out that decimation factors in decreasing order generally yield smaller overall lengths, and any multistage design results in a substantial reduction in the filter length as compared to a single-stage design.

The actual filter lengths also depend on the given attenuation specifications. Since attenuations in decibels add in a cascaded system, the passband attenuation A_p is usually distributed among the various stages to ensure an overall value that matches specifications.

EXAMPLE 20.9 (The Concept of Multistage Decimation)

(a) Consider a signal band-limited to 1.8 kHz and sampled at 48 kHz. It is required to reduce the sampling rate to 4 kHz. This requires decimation by 12. The passband edge is $f_p = 1.8$ kHz and remains unchanged for a single stage-design or multistage design.

For a single-stage decimator, the output sampling rate is $S_{\text{out}} = 4$ kHz, and we thus require a filter with a stopband edge of $f_s = S_{\text{out}} - f_p = 4 - 1.8 = 2.2$ kHz, a sampling rate of $S = S_{\text{in}} = 48$ kHz, and a gain of unity. If we use a crude approximation for the filter length as $L \approx 4/F_T$, where $F_T = (f_s - f_p)/S$ is the digital transition width, we obtain $L = 48/0.4 = 120$.

PGR2026-00033
Krisp Technologies EX1024 Page 761

(b) If we choose two-stage decimation with $D_1 = 4$ and $D_2 = 3$, at each stage we compute the important parameters, as follows:

Stage	$S_{\text{in}} = S$ (kHz)	**Decimation Factor**	S_{out} (kHz)	f_p (kHz)	$f_s = S_{\text{out}} - f_p$ (kHz)	**Filter Length** $L \approx 4S/(f_s - f_p)$
1	48	$D_1 = 4$	12	1.8	10.2	$192/8.4 = 23$
2	12	$D_2 = 3$	4	1.8	2.2	$48/0.4 = 120$

The total filter length is thus 143.

(c) If we choose three-stage decimation with $D_1 = 2,\ D_2 = 3$, and $D_3 = 2$, at each stage we compute the important parameters, as follows:

Stage	$S_{\text{in}} = S$ (kHz)	**Decimation Factor**	S_{out} (kHz)	f_p (kHz)	$f_s = S_{\text{out}} - f_p$ (kHz)	**Filter Length** $L \approx 4S/(f_s - f_p)$
1	48	$D_1 = 2$	24	1.8	22.2	$192/20.4 = 10$
2	24	$D_2 = 3$	8	1.8	6.2	$96/4.4 = 22$
3	8	$D_3 = 2$	4	1.8	2.2	$32/0.4 = 80$

The total filter length is thus 112.

(d) If we choose three-stage decimation but with the different order $D_1 = 2,\ D_2 = 2$, and $D_3 = 3$, at each stage we compute the important parameters, as follows:

Stage	$S_{\text{in}} = S$ (kHz)	**Decimation Factor**	S_{out} (kHz)	f_p (kHz)	$f_s = S_{\text{out}} - f_p$ (kHz)	**Filter Length** $L \approx 4S/(f_s - f_p)$
1	48	$D_1 = 2$	24	1.8	22.2	$192/20.4 = 10$
2	24	$D_2 = 2$	12	1.8	10.2	$96/8.4 = 12$
3	12	$D_3 = 3$	4	1.8	2.2	$48/0.4 = 120$

The total filter length is thus 142.

Note how decimation uses the same filter frequency specifications as does interpolation for a given split, except in reversed order. Any multistage design results in a substantial reduction in the filter length as compared to a single-stage design. Also remember that the filter lengths chosen here are only a crude approximation (in order to illustrate the relative merits of each design), and the actual filter lengths will depend on the attenuation specifications.

20.7 Maximally Flat FIR Filters

Linear-phase FIR filters can also be designed with **maximally flat** frequency response. Such filters are usually used in situations where accurate filtering is needed at low frequencies (near dc). The design of lowpass maximally flat filters uses a closed form for the transfer function $H(F)$ given by

$$H(F) = \cos^{2K}(\pi F) \sum_{n=0}^{L-1} d_n \sin^{2n}(\pi F) \qquad d_n = \frac{(K+n-1)!}{(K-1)!\ n!} = C_n^{K+n-1} \tag{20.42}$$

PGR2026-00033
Krisp Technologies EX1024 Page 762

Here, the d_n have the form of binomial coefficients as indicated. Note that $2L-1$ derivatives of $|H(F)|^2$ are zero at $F=0$ and $2K-1$ derivatives are zero at $F=0.5$. This is the basis for the maximally flat response of the filter. The filter length equals $N=2(K+L)-1$, and is thus odd. The integers K and L are determined from the passband and stopband frequencies F_p and F_s that correspond to gains of 0.95 and 0.05 (or attenuations of about 0.5 dB and 26 dB), respectively. Here is an empirical design method:

1. Define the cutoff frequency as $F_C=0.5(F_p+F_s)$ and let $F_T=F_s-F_p$.
2. Obtain a first estimate of the odd filter length as $N_0=1+0.5/F_T^2$.
3. Define the parameter α as $\alpha=\cos^2(\pi F_C)$.
4. Find the best rational approximation $\alpha\approx K/M_{\min}$, with $0.5(N_0-1)\le M_{\min}\le(N_0-1)$.
5. Evaluate L and the true filter length N from $L=M_{\min}-K$ and $N=2M_{\min}-1$.
6. Find $h[n]$ as the N-point inverse DFT of $H(F),\ F=0,1/N,\ldots,(N-1)/N$.

EXAMPLE 20.10 (Maximally Flat FIR Filter Design)
Consider the design of a maximally flat lowpass FIR filter with normalized frequencies $F_p=0.2$ and $F_s=0.4$.

We have $F_C=0.3$ and $F_T=0.2$. We compute $N_0=1+0.5/F_T^2=13.5\approx 15$, and $\alpha=\cos^2(\pi F_C)=0.3455$. The best rational approximation to α works out to be $\alpha\approx 5/14=K/M_{\min}$.
With $K=5$ and $M_{\min}=14$, we get $L=M_{\min}-K=9$, and the filter length $N=2M_{\min}-1=27$.
Figure E20.10 shows the impulse response and $|H(F)|$ of the designed filter. The response is maximally flat, with no ripples in the passband.

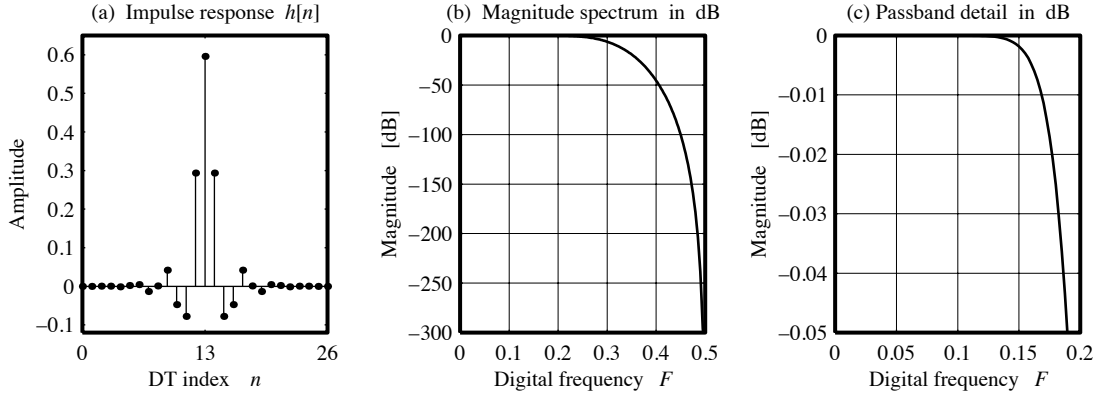


Figure E20.10 Features of the maximally flat lowpass filter for Example 20.10

20.8 FIR Differentiators and Hilbert Transformers

An ideal digital differentiator is described by $H(F)=j2\pi F,\ |F|\le 0.5$. In practical situations, we seldom require filters that differentiate for the full frequency range up to $|F|=0.5$. If we require differentiation only up to a cutoff frequency of F_C, then

$$H(F)=j2\pi F,\qquad |F|\le F_C \tag{20.43}$$

PGR2026-00033
Krisp Technologies EX1024 Page 763

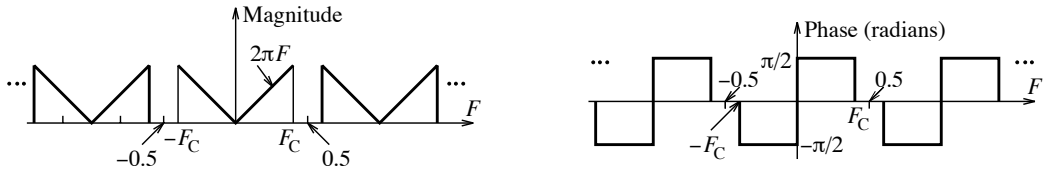


Figure 20.19 Magnitude and phase spectrum of an ideal differentiator

The magnitude and phase spectrum of such a differentiator are shown in Figure 20.19. Since $H(F)$ is odd, $h[0] = 0$. To find $h[n],\ n \neq 0$ we use the inverse DTFT to obtain

$$h[n] = j \int_{-F_C}^{F_C} 2\pi F e^{j2n\pi F} dF \tag{20.44}$$

Invoking Euler's relation and symmetry, this simplifies to

$$h[n] = -4\pi \int_0^{F_C} F \sin(2n\pi F)\, dF = \frac{2n\pi F_C \cos(2n\pi F_C) - \sin(2n\pi F_C)}{\pi n^2} \tag{20.45}$$

For $F_C = 0.5$, this yields $h[0] = 0$ and $h[n] = \dfrac{\cos(n\pi)}{n},\ n \neq 0.$

20.8.1 Hilbert Transformers

A **Hilbert transformer** is used to shift the phase of a signal by $-\frac{\pi}{2}$ (or $-90°$) over the frequency range $|F| \leq 0.5$. In practice, we seldom require filters that shift the phase for the full frequency range up to $|F| = 0.5$. If we require phase shifting only up to a cutoff frequency of F_C, the Hilbert transformer may be described by

$$H(F) = -j\,\mathrm{sgn}(F), \qquad |F| \leq F_C \tag{20.46}$$

Its magnitude and phase spectrum are shown in Figure 20.20.

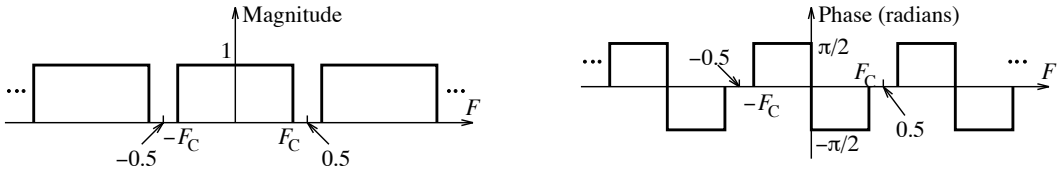


Figure 20.20 Magnitude and phase spectrum of an ideal Hilbert transformer

With $h[0] = 0$, we use the inverse DTFT to find its impulse response $h[n],\ n \neq 0$, as

$$h[n] = \int_{-F_C}^{F_C} -j\,\mathrm{sgn}(F) e^{j2n\pi F} dF = 2\int_0^{F_C} \sin(2n\pi F)\, dF = \frac{1 - \cos(2n\pi F_C)}{n\pi} \tag{20.47}$$

For $F_C = 0.5$, this reduces to $h[n] = \dfrac{1 - \cos(n\pi)}{n\pi},\ n \neq 0.$

PGR2026-00033
Krisp Technologies EX1024 Page 764

20.8.2 Design of FIR Differentiators and Hilbert Transformers

To design an FIR differentiator, we must truncate $h[n]$ to $h_N[n]$ and choose a type 3 or type 4 sequence, since $H(F)$ is purely imaginary. To ensure odd symmetry, the filter coefficients may be computed only for $n > 0$, and the same values (in reversed order) are used for $n < 0$. If N is odd, we must also include the sample $h[0]$. We may window $h_N[n]$ to minimize the overshoot and ripple in the spectrum $H_N(F)$. And, finally, to ensure causality, we must introduce a delay of $(N-1)/2$ samples. Figure 20.21(a) shows the magnitude response of Hamming-windowed FIR differentiators for both even and odd lengths N. Note that $H(0)$ is always zero, but $H(0.5) = 0$ only for type 3 (odd-length) sequences.

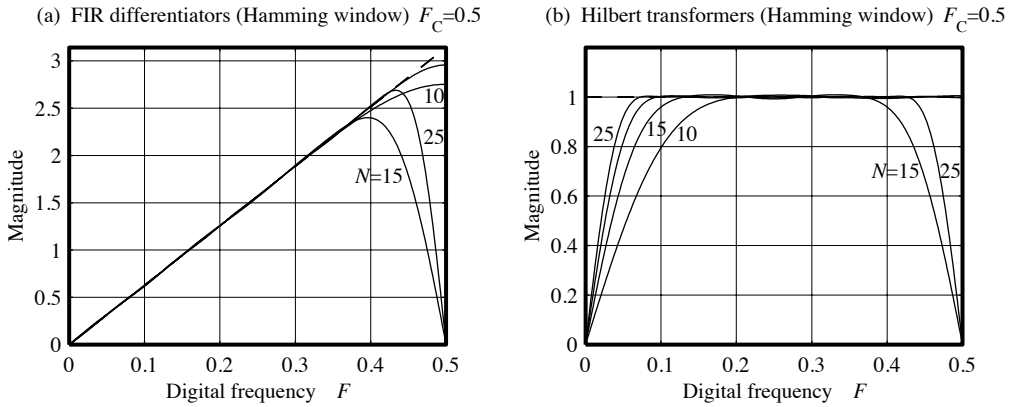


Figure 20.21 Magnitude spectra of ideal differentiators and Hilbert transformers

The design of Hilbert transformers closely parallels the design of FIR differentiators. The sequence $h[n]$ is truncated to $h_N[n]$. The chosen filter must correspond to a type 3 or type 4 sequence, since $H(F)$ is imaginary. To ensure odd symmetry, the filter coefficients may be computed only for $n > 0$, with the same values (in reversed order) used for $n < 0$. If N is odd, we must also include the sample $h[0]$. We may window $h_N[n]$ to minimize the ripples (due to the Gibbs effect) in the spectrum $H_N(F)$. The Hamming window is a common choice, but others may also be used. To make the filter causal, we introduce a delay of $(N-1)/2$ samples. The magnitude spectrum of a Hilbert transformer becomes flatter with increasing filter length N. Figure 20.21(b) shows the magnitude response of Hilbert transformers for both even and odd lengths N. Note that $H(0)$ is always zero, but $H(0.5) = 0$ only for type 3 (odd-length) sequences.

20.9 Least Squares and Adaptive Signal Processing

The concept of least squares can be formulated as the solution to a set of algebraic equations expressed in matrix form as

$$\mathbf{Xb} = \mathbf{Y} \tag{20.48}$$

If the number of unknowns equals the number of equations, and the matrix $\mathbf{X}$ is non-singular such that its inverse $\mathbf{X}^{-1}$ exists (the unique case), the solution is simply $\mathbf{b} = \mathbf{X}^{-1}\mathbf{Y}$. However, if the number of equations exceeds the number of unknowns (the over-determined case), no solution is possible, and only an approximate result may be obtained using, for example, least squares minimization. If the number of unknowns exceeds the number of equations (the under-determined case), many possible solutions exist, including the one where the mean square error is minimized.

PGR2026-00033
Krisp Technologies EX1024 Page 765

In many practical situations, the set of equations is over-determined and thus amenable to a least-squares solution. To solve for $\mathbf{b}$, we simply premultiply both sides by $\mathbf{X}^T$ to obtain

$$\mathbf{X}^T\mathbf{X}\mathbf{b} = \mathbf{X}^T\mathbf{Y} \tag{20.49}$$

If the inverse of $\mathbf{X}^T\mathbf{X}$ exists, the solution is given by

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \tag{20.50}$$

The matrix $\mathbf{X}^T\mathbf{X}$ is called the **covariance matrix**.

20.9.1 Adaptive Filtering

The idea of least squares finds widespread application in **adaptive signal processing**, a field that includes adaptive filtering, deconvolution, and system identification, and encompasses many disciplines. Typically, the goal is to devise a digital filter whose coefficients $b[k]$ can be adjusted to optimize its performance (in the face of changing signal characteristics or to combat the effects of noise, for example). A representative system for adaptive filtering is shown in Figure 20.22. The measured signal $x[n]$ (which may be contaminated by noise) is fed to an FIR filter whose output $\hat{y}[n]$ is compared with the desired signal $y[n]$ to generate the error signal $e[n]$. The error signal is used to update the filter coefficients $b[k]$ through an adaptation algorithm in a way that minimizes the error $e[n]$, and thus provides an optimal estimate of the desired signal $y[n]$.

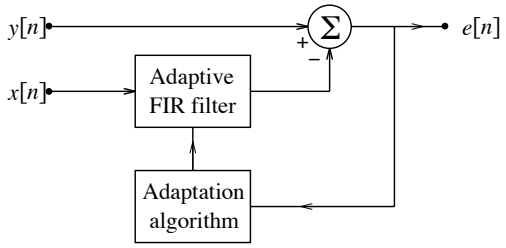


Figure 20.22 The structure of a typical adaptive filtering system

In convolution form, the output $\hat{\mathbf{Y}}$ of the $(M+1)$-point adaptive FIR filter is described by the model

$$\hat{y}[n] = \sum_{k=0}^{M} b[k]x[n-k], \qquad n = 0, 1, \ldots, M+N+1 \tag{20.51}$$

This set of $M+N+1$ equations may be cast in matrix form as

$$\begin{bmatrix} \hat{y}[0] \\ \vdots \\ \hat{y}[N] \end{bmatrix} = \begin{bmatrix} x[0] & 0 & \cdots & \cdots & \cdots \\ x[1] & x[0] & 0 & \cdots & \cdots \\ x[2] & x[1] & x[0] & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x[N] & x[N-1] & x[N-2] & \cdots & x[N-M] \\ 0 & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \cdots \end{bmatrix} \begin{bmatrix} b[0] \\ \vdots \\ b[M] \end{bmatrix} \tag{20.52}$$

PGR2026-00033
Krisp Technologies EX1024 Page 766

In vector notation, it has the form $\hat{\mathbf{Y}} = \mathbf{Xb}$, where $\mathbf{b}$ is an $(M+1)\times 1$ column matrix of the filter coefficients $b[0]$ through $b[M]$, $\hat{\mathbf{Y}}$ is an $(M+N+1)\times 1$ matrix of the output samples, and $\mathbf{X}$ is an $(M+N+1)\times(M+1)$ **Toeplitz** (constant diagonal) matrix whose columns are successively shifted replicas of the input sequence. The $(M+1)$ filter coefficients in $\mathbf{b}$ are chosen to ensure that the output $\hat{\mathbf{Y}}$ of the model provides an optimal estimate of the desired output $\mathbf{Y}$. Clearly, this problem is amenable to a least squares solution.

Since the data is being acquired continuously in many practical situations, the least squares solution is implemented on-line or in real time using iterative or recursive numerical algorithms. Two common approaches are the **recursive least squares** (RLS) algorithm and the **least mean squares** (LMS) algorithm. Both start with an assumed set of filter coefficients and update this set as each new input sample arrives.

The RLS algorithm: In the RLS algorithm, the filter coefficients are updated by weighting the input samples (typically in an exponential manner) so as to emphasize more recent inputs.

The LMS algorithm: The LMS algorithm, though not directly related to least squares, uses the so-called *method of steepest descent* to generate updates that converge about the least squares solution. Although it may not converge as rapidly as the RLS algorithm, it is far more popular due to its ease of implementation. In fact, the updating equation (for the nth update) has the simple form

$$b_n[k] = b_{n-1}[k] + 2\mu(y[n] - \hat{y}[n])x[n-k] = b_{n-1}[k] + 2\mu e[n]x[n-k], \quad 0 \le k \le M \tag{20.53}$$

The parameter μ governs both the rate of convergence and the stability of the algorithm. Larger values result in faster convergence, but the filter coefficients tend to oscillate about the optimum values. Typically, μ is restricted to the range $0 < \mu < \frac{1}{\sigma_x}$, where σ_x, the variance of $x[n]$, provides a measure of the power in the input signal $x[n]$.

20.9.2 Applications of Adaptive Filtering

Adaptive filtering forms the basis for many signal-processing applications including system identification, noise cancellation, and channel equalization. We conclude with a brief introduction.

System Identification

In system identification, the goal is to identify the transfer function (or impulse response) of an unknown system. Both the adaptive filter and the unknown system are excited by the same input, and the signal $y[n]$ represents the output of the unknown system. Minimizing $e[n]$ implies that the output of the unknown system and the adaptive filter are very close, and the adaptive filter coefficients describe an FIR approximation to the unknown system.

Noise Cancellation

In adaptive noise-cancellation systems, the goal is to improve the quality of a desired signal $y[n]$ that may be contaminated by noise. The signal $x[n]$ is a noise signal, and the adaptive filter minimizes the power in $e[n]$. Since the noise power and signal power add (if they are uncorrelated), the signal $e[n]$ (with its power minimized) also represents a cleaner estimate of the desired signal $y[n]$.

Channel Equalization

In adaptive channel equalization, the goal is to allow a modem to adapt to different telephone lines (so as to prevent distortion and intersymbol interference). A known *training* signal $y[n]$ is transmitted at the start of each call, and $x[n]$ is the output of the telephone channel. The error signal $e[n]$ is used to generate an FIR filter (an inverse system) that cancels out the effects of the telephone channel. Once found, the filter coefficients are fixed, and the modem operates with the fixed filter.

PGR2026-00033
Krisp Technologies EX1024 Page 767

CHAPTER 20 PROBLEMS

DRILL AND REINFORCEMENT

20.1 (Symmetric Sequences) Find $H(z)$ and $H(F)$ for each sequence and establish the type of FIR filter it describes by checking values of $H(F)$ at $F = 0$ and $F = 0.5$.

(a) $h[n] = \{\overset{\Downarrow}{1}, 0, 1\}$ **(b)** $h[n] = \{\overset{\Downarrow}{1}, 2, 2, 1\}$

(c) $h[n] = \{\overset{\Downarrow}{1}, 0, -1\}$ **(d)** $h[n] = \{\overset{\Downarrow}{-1}, 2, -2, 1\}$

20.2 (Symmetric Sequences) What types of sequences can we use to design the following filters?

(a) Lowpass **(b)** Highpass **(c)** Bandpass **(d)** Bandstop

20.3 (Truncation and Windowing) Consider a windowed lowpass FIR filter with cutoff frequency 5 kHz and sampling frequency $S = 20$ kHz. Find the truncated, windowed sequence, the minimum delay (in samples and in seconds) to make the filter causal, and the transfer function $H(z)$ of the causal filter if

(a) $N = 7$, and we use a Bartlett window.
(b) $N = 8$, and we use a von Hann (Hanning) window.
(c) $N = 9$, and we use a Hamming window.

20.4 (Spectral Transformations) Assuming a sampling frequency of 40 kHz and a fixed passband, find the specifications for a digital lowpass FIR prototype and the subsequent spectral transformation to convert to the required filter type for the following filters.

(a) Highpass: passband edge at 10 kHz, stopband edge at 4 kHz
(b) Bandpass: passband edges at 6 kHz and 10 kHz, stopband edges at 2 kHz and 12 kHz
(c) Bandstop: passband edges 8 kHz and 16 kHz, stopband edges 12 kHz and 14 kHz

20.5 (Window-Based FIR Filter Design) We wish to design a window-based linear-phase FIR filter. What is the *approximate* filter length N required if the filter to be designed is

(a) Lowpass: $f_p = 1$ kHz, $f_s = 2$ kHz, $S = 10$ kHz, using a von Hann (Hanning) window?
(b) Highpass: $f_p = 2$ kHz, $f_s = 1$ kHz, $S = 8$ kHz, using a Blackman window?
(c) Bandpass: $f_p = [4, 8]$ kHz, $f_s = [2, 12]$ kHz, $S = 25$ kHz, using a Hamming window?
(d) Bandstop: $f_p = [2, 12]$ kHz, $f_s = [4, 8]$ kHz, $S = 25$ kHz, using a Hamming window?

20.6 (Half-Band FIR Filter Design) A lowpass half-band FIR filter is to be designed using a von Hann window. Assume a filter length $N = 11$ and find its windowed, causal impulse response sequence and the transfer function $H(z)$ of the causal filter.

20.7 (Half-Band FIR Filter Design) Design the following half-band FIR filters, using a Kaiser window.

(a) Lowpass filter: 3-dB frequency 4 kHz, stopband edge 8 kHz, and $A_s = 40$ dB
(b) Highpass filter: 3-dB frequency 6 kHz, stopband edge 3 kHz, and $A_s = 50$ dB
(c) Bandpass filter: passband edges at [2, 3] kHz, stopband edges at [1, 4] kHz, $A_p = 1$ dB, and $A_s = 35$ dB
(d) Bandstop filter: stopband edges at [2, 3] kHz, passband edges at [1, 4] kHz, $A_p = 1$ dB, and $A_s = 35$ dB

PGR2026-00033
Krisp Technologies EX1024 Page 768

20.8 (Frequency-Sampling FIR Filter Design) Consider the frequency-sampling design of a lowpass FIR filter with $F_C = 0.25$.

(a) Choose eight samples over the range $0 \leq F < 1$ and set up the frequency response $H[k]$ of the filter.

(b) Compute the impulse response $h[n]$ for the filter.

(c) To reduce the overshoot, modify $H[3]$ and recompute $h[n]$.

20.9 (Maximally Flat FIR Filter Design) Design a maximally flat lowpass FIR filter with normalized frequencies $F_p = 0.1$ and $F_s = 0.4$, and find its frequency response $H(F)$.

20.10 (FIR Differentiators) Find the impulse response of a digital FIR differentiator with

(a) $N = 6$, cutoff frequency $F_C = 0.4$, and no window.

(b) $N = 6$, cutoff frequency $F_C = 0.4$, and a Hamming window.

(c) $N = 5$, cutoff frequency $F_C = 0.5$, and a Hamming window.

20.11 (FIR Hilbert Transformers) Find the impulse response of an FIR Hilbert transformer with

(a) $N = 6$, cutoff frequency $F_C = 0.4$, and no window.

(b) $N = 6$, cutoff frequency $F_C = 0.4$, and a von Hann window.

(c) $N = 7$, cutoff frequency $F_C = 0.5$, and a von Hann window.

20.12 (Multistage Interpolation) It is required to design a three-stage interpolator. The interpolating filters are to be designed with identical passband and stopband attenuation and are required to provide an overall attenuation of no more than 3 dB in the passband and at least 50 dB in the stopband. Specify the passband and stopband attenuation of each filter.

REVIEW AND EXPLORATION

20.13 (Linear-Phase Sequences) The first few values of the impulse response sequence of a linear-phase filter are $h[n] = \{2,\ -3,\ 4,\ 1, \ldots\}$. Determine the complete sequence, assuming the smallest length for $h[n]$, and assuming that the sequence is the following:

(a) Type 1 **(b)** Type 2 **(c)** Type 3 **(d)** Type 4

20.14 (Linear Phase and Symmetry) Assume a sequence $h[n]$ with real coefficients with all its poles at $z = 0$. Argue for or against the following statements. You may want to exploit two useful facts. First, each pair of terms with reciprocal roots such as $(z - \alpha)$ and $(z - 1/\alpha)$ yields an even symmetric impulse response sequence. Second, the convolution of symmetric sequences is also endowed with symmetry.

(a) If all the zeros lie on the unit circle, $h[n]$ must be linear phase.

(b) If $h[n]$ is linear phase its zeros must *always* lie on the unit circle.

(c) If $h[n]$ is odd symmetric, there must be an odd number of zeros at $z = 1$.

20.15 (Linear Phase and Symmetry) Assume a linear-phase sequence $h[n]$ and refute the following statements by providing simple examples using zero locations only at $z = \pm 1$.

(a) If $h[n]$ has zeros at $z = -1$ it must be a type 2 sequence.

(b) If $h[n]$ has zeros at $z = -1$ and $z = 1$, it must be a type 3 sequence.

(c) If $h[n]$ has zeros at $z = 1$ it must be a type 4 sequence.

PGR2026-00033
Krisp Technologies EX1024 Page 769

20.16 (Linear Phase and Symmetry) The locations of the zeros at $z = \pm 1$ and their number provides useful clues about the type of a linear-phase sequence. What is the sequence type for the following zero locations at $z = \pm 1$? Other zero locations are arbitrary but in keeping with linear phase and conjugate reciprocal symmetry.

(a) No zeros at $z = \pm 1$
(b) One zero at $z = -1$, none at $z = 1$
(c) Two zeros at $z = -1$, one zero at $z = 1$
(d) One zero at $z = 1$, none at $z = -1$
(e) Two zeros at $z = 1$, none at $z = -1$
(f) One zero at $z = 1$, one zero at $z = -1$
(g) Two zeros at $z = 1$, one zero at $z = -1$

20.17 (Linear-Phase Sequences) What is the smallest length linear-phase sequence that meets the requirements listed? Identify all the zero locations and the type of linear-phase sequence.

(a) Zero location: $z = e^{j0.25\pi}$; even symmetry; odd length
(b) Zero location: $z = 0.5e^{j0.25\pi}$; even symmetry; even length
(c) Zero location: $z = e^{j0.25\pi}$; odd symmetry; even length
(d) Zero location: $z = 0.5e^{j0.25\pi}$; odd symmetry; odd length

20.18 (Linear-Phase Sequences) The zeros of various finite-length linear-phase filters are given. Assuming the smallest length, identify the sequence type, find the transfer function of each filter, and identify the filter type.

(a) Zero location: $z = 0.5e^{j0.25\pi}$
(b) Zero location: $z = e^{j0.25\pi}$
(c) Zero locations: $z = 0.5,\ z = e^{j0.25\pi}$
(d) Zero locations: $z = 0.5,\ z = 1$; odd symmetry
(e) Zero locations: $z = 0.5,\ z = 1$; oven symmetry

20.19 (IIR Filters and Linear Phase) Even though IIR filters cannot be designed with linear phase, they can actually be used to provide no phase distortion. Consider a sequence $x[n]$. We fold $x[n]$, feed it to a filter $H(F)$, and fold the filter output to obtain $y_1[n]$. We also feed $x[n]$ directly to the filter $H(F)$ to obtain $y_2[n]$. The signals $y_1[n]$ and $y_2[n]$ are summed to give $y[n]$.

(a) Sketch a block diagram of this system.
(b) How is $Y(F)$ related to $X(F)$?
(c) Show that the signal $y[n]$ suffers no phase distortion.

20.20 (FIR Filter Specifications) Figure P20.20 shows the magnitude and phase characteristics of a *causal* FIR filter designed at a sampling frequency of 10 kHz.

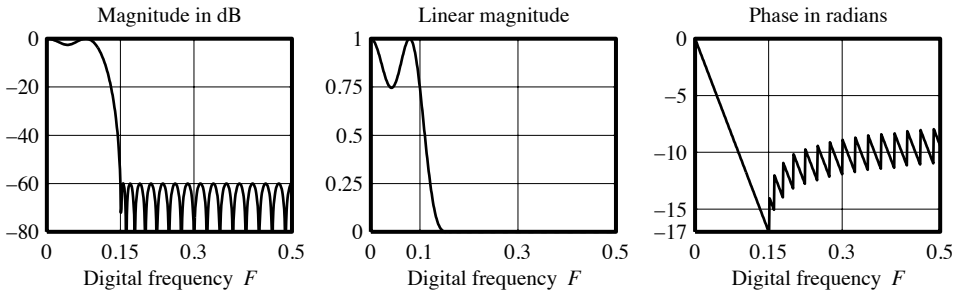


Figure P20.20 Filter characteristics for Problem 20.20

PGR2026-00033
Krisp Technologies EX1024 Page 770

(a) What are the values of the passband ripple δ_p and stopband ripple δ_s?
(b) What are the values of the attenuation A_s and A_p in decibels?
(c) What are the frequencies of the passband edge and stopband edge?
(d) What is the group delay of the filter?
(e) What is the filter length N?
(f) Could this filter have been designed using the window method? Explain.
(g) Could this filter have been designed using the optimal method? Explain.

COMPUTATION AND DESIGN

lpsiggui A GUI for Visualization of Linear-Phase Sequences

The graphical user interface `lpsiggui` allows you to visualize a linear-phase signal and its pole-zero plot, amplitude spectrum, and phase spectrum.
The sequence type is indicated on the plots.
To explore this routine, type `lpsiggui` at the MATLAB prompt.

dfirgui A GUI for FIR Filter Design and Visualization

The graphical user interface `dfirgui` allows you to design FIR filters and visualize their magnitude and phase spectrum.
You can select from various forms (highpass, lowpass, etc.) and methods (window-based, half-band, and optimal).
The plots are displayed along with the filter order and design attenuation at the band edges.
You can change the filter length and other relevant design parameters.
To explore this routine, type `dfirgui` at the MATLAB prompt.

20.21 (FIR Filter Design) It is desired to reduce the frequency content of a hi-fi audio signal band-limited to 20 kHz and sampled at 44.1 kHz for purposes of AM transmission. Only frequencies up to 10 kHz are of interest. Frequencies past 15 kHz are to be attenuated by at least 55 dB, and the passband loss is to be less than 10%. Design a digital filter using the Kaiser window that meets these specifications.

20.22 (FIR Filter Design) It is desired to eliminate 60-Hz interference from an ECG signal whose significant frequencies extend to 35 Hz.
(a) What is the minimum sampling frequency we can use to avoid in-band aliasing?
(b) If the 60-Hz interference is to be suppressed by a factor of at least 100, with no appreciable signal loss, what should be the filter specifications?
(c) Design the filter using a Hamming window and plot its frequency response.
(d) Test your filter on the signal $x(t) = \cos(40\pi t) + \cos(70\pi t) + \cos(120\pi t)$. Plot and compare the frequency response of the sampled test signal and the filtered signal to confirm that your design objectives are met.

20.23 (Digital Filter Design) We wish to design a lowpass filter for processing speech signals. The specifications call for a passband of 4 kHz and a stopband of 5 kHz. The passband attenuation is to be less than 1 dB, and the stopband gain is to be less than 0.01. The sampling frequency is 40 kHz.
(a) Design FIR filters, using the window method (with Hamming and Kaiser windows) and using optimal design. Which of these filters has the minimum length?

PGR2026-00033
Krisp Technologies EX1024 Page 771

(b) Design IIR Butterworth and elliptic filters, using the bilinear transformation, to meet the same set of specifications. Which of these filters has the minimum order? Which has the best delay characteristics?

(c) How does the complexity of the IIR filters compare with that of the FIR filters designed with the same specifications? What are the trade-offs in using an IIR filter over an FIR filter?

20.24 (The Effect of Group Delay) The nonlinear phase of IIR filters is responsible for signal distortion. Consider a lowpass filter with a 1-dB passband edge at $f_p = 1$ kHz, a 50-dB stopband edge at $f_p = 2\,\text{kHz}$, and a sampling frequency of $S = 10$ kHz.

(a) Design a Butterworth filter $H_B(z)$, an elliptic filter $H_E(z)$, and an optimal FIR filter $H_O(z)$ to meet these specifications. Using the MATLAB routine `grpdelay` (or otherwise), compute and plot the group delay of each filter. Which filter has the best (most nearly constant) group delay in the passband? Which filter would cause the least phase distortion in the passband? What are the group delays $N_B,\ N_E$, and N_O (expressed as the number of samples) of the three filters?

(b) Generate the signal $x[n] = 3\sin(0.03n\pi) + \sin(0.09n\pi) + 0.6\sin(0.15n\pi)$ over $0 \le n \le 100$. Use the ADSP routine `filter` to compute the response $y_B[n],\ y_E[n]$, and $y_O[n]$ of each filter. Plot the filter outputs $y_B[n],\ y_E[n]$, and $y_O[n]$ (delayed by $N_B,\ N_E$, and N_O, respectively) and the input $x[n]$ on the same plot to compare results. Which filter results in the smallest signal distortion?

(c) Are all the frequency components of the input signal in the filter passband? If so, how can you justify that what you observe as distortion is actually the result of the nonconstant group delay and not the filter attenuation in the passband?

20.25 (Raised Cosine Filters) The impulse response of a raised cosine filter has the form

$$h_R[n] = h[n]\frac{\cos(2n\pi RF_C)}{1-(4nRF_C)^2}$$

where the *roll-off factor* R satisfies $0 < R < 1$ and $h[n] = 2F_C\,\text{sinc}(2nF_C)$ is the impulse response of an ideal lowpass filter.

(a) Let $F_C = 0.2$. Generate the impulse response of an ideal lowpass filter with length 21 and the impulse response of the corresponding raised cosine filter with $R = 0.2,\ 0.5,\ 0.9$. Plot the magnitude spectra of each filter over $0 \le F \le 1$ on the same plot, using linear and decibel scales. How does the response in the passband and stopband of the raised cosine filter differ from that of the ideal filter? How does the transition width and peak sidelobe attenuation of the raised cosine filter compare with that of the ideal filter for different values of R? What is the effect of increasing R on the frequency response?

(b) Compare the frequency response of $h_R[n]$ with that of an ideal lowpass filter with $F_C = 0.25$. Is the raised cosine filter related to this ideal filter?

20.26 (Interpolation) The signal $x[n] = \cos(2\pi F_0 n)$ is to be interpolated by 5, using up-sampling followed by lowpass filtering. Let $F_0 = 0.4$.

(a) Generate and plot 20 samples of $x[n]$ and up-sample by 5.

(b) What must be the cutoff frequency F_C and gain A of a lowpass filter that follows the up-sampler to produce the interpolated output?

(c) Design an FIR filter (using the window method or optimal design) to meet these specifications.

(d) Filter the up-sampled signal through this filter and plot the result.

(e) Is the filter output an interpolated version of the input signal? Do the peak amplitude of the interpolated signal and original signal match? Should they? Explain.

PGR2026-00033
Krisp Technologies EX1024 Page 772

20.27 (Multistage Interpolation) To relax the design requirements for the analog reconstruction filter, many compact disc systems employ oversampling during the DSP stages. Assume that audio signals are band-limited to 20 kHz and sampled at 44.1 kHz. Assume a maximum passband attenuation of 1 dB and a minimum stopband attenuation of 50 dB.

(a) Design a single-stage optimal interpolating filter that increases the sampling rate to 176.4 kHz.

(b) Design multistage optimal interpolating filters that increase the sampling rate to 176.4 kHz.

(c) Which of the two designs would you recommend?

(d) For each design, explain how you might incorporate compensating filters during the DSP stage to offset the effects of the sinc distortion caused by the zero-order-hold reconstruction device.

20.28 (Multistage Interpolation) The sampling rate of a speech signal band-limited to 3.4 kHz and sampled at 8 kHz is to be increased to 48 kHz. Design three different schemes that will achieve this rate increase and compare their performance. Use optimal FIR filter design where required and assume a maximum passband attenuation of 1 dB and a minimum stopband attenuation of 45 dB.

20.29 (Multistage Decimation) The sampling rate of a speech signal band-limited to 3.4 kHz and sampled at 48 kHz is to be decreased to 8 kHz. Design three different schemes that will achieve this rate decrease and compare their performance. Use optimal FIR filter design where required and assume a maximum passband attenuation of 1 dB and a minimum stopband attenuation of 45 dB. How do these filters compare with the filters designed for multistage interpolation in Problem 20.28.

20.30 (Filtering Concepts) This problem deals with time-frequency plots of a combination of sinusoids and their filtered versions.

(a) Generate 600 samples of the signal $x[n] = \cos(0.1n\pi) + \cos(0.4n\pi) + \cos(0.7n\pi)$ comprising the sum of three pure cosines at $F = 0.05,\ 0.2,\ 0.35$. Use the MATLAB command `fft` to plot its DFT magnitude. Use the ADSP routine `timefreq` to display its time-frequency plot. What do the plots reveal? Now design an optimal lowpass filter with a 1-dB passband edge at $F = 0.1$ and a 50-dB stopband edge at $F = 0.15$ and filter $x[n]$ through this filter to obtain the filtered signal $x_f[n]$. Plot its DFT magnitude and display its time-frequency plot. What do the plots reveal? Does the filter perform its function? Plot $x_f[n]$ over a length that enables you to identify its period. Does the period of $x_f[n]$ match your expectations?

(b) Generate 200 samples each of the three signals $y_1[n] = \cos(0.1n\pi)$, $y_2[n] = \cos(0.4n\pi)$, and $y_3[n] = \cos(0.7n\pi)$. Concatenate them to form the 600-sample signal $y[n] = \{y_1[n],\ y_2[n],\ y_3[n]\}$. Plot its DFT magnitude and display its time-frequency plot. What do the plots reveal? In what way does the DFT magnitude plot differ from part (a)? In what way does the time-frequency plot differ from part (a)? Use the optimal lowpass filter designed in part (a) to filter $y[n]$, obtain the filtered signal $y_f[n]$, plot its DFT magnitude, and display its time-frequency plot. What do the plots reveal? In what way does the DFT magnitude plot differ from part (a)? In what way does the time-frequency plot differ from part (a)? Does the filter perform its function? Plot $y_f[n]$ over a length that enables you to identify its period. Does the period of $y_f[n]$ match your expectations?

20.31 (Decoding a Mystery Message) During transmission, a message signal gets contaminated by a low-frequency signal and high-frequency noise. The message can be decoded only by displaying it in the time domain. The contaminated signal $x[n]$ is provided on disk as `mystery1.mat`. Load this signal into MATLAB (use the command `load mystery1`). In an effort to decode the message, try the following methods and determine what the decoded message says.

PGR2026-00033
Krisp Technologies EX1024 Page 773

(a) Display the contaminated signal. Can you "read" the message? Display the DFT of the signal to identify the range of the message spectrum.

(b) Design an optimal FIR bandpass filter capable of extracting the message spectrum. Filter the contaminated signal and display the filtered signal to decode the message. Use both the `filter` (linear-phase filtering) and `filtfilt` (zero-phase filtering) commands.

(c) As an alternative method, first zero out the DFT component corresponding to the low-frequency contamination and obtain its IDFT $y[n]$. Next, design an optimal lowpass FIR filter to reject the high-frequency noise. Filter the signal $y[n]$ and display the filtered signal to decode the message. Use both the `filter` and `filtfilt` commands.

(d) Which of the two methods allows better visual detection of the message? Which of the two filtering routines (in each method) allows better visual detection of the message?

20.32 (Filtering of a Chirp Signal) This problem deals with time-frequency plots of a chirp signal and its filtered versions.

(a) Use the ADSP routine `chirp` to generate 500 samples of a chirp signal $x[n]$ whose frequency varies from $F = 0$ to $F = 0.12$. Use the MATLAB command `fft` to plot its DFT magnitude. Use the ADSP routine `timefreq` to display its time-frequency plot. What do the plots reveal? Plot $x[n]$ and confirm that its frequency is increasing with time.

(b) Design an optimal lowpass filter with a 1-dB passband edge at $F = 0.04$ and a 40-dB stopband edge at $F = 0.1$ and use the MATLAB command `filtfilt` to obtain the zero-phase filtered signal $y_1[n]$. Plot its DFT magnitude and display its time-frequency plot. What do the plots reveal? Plot $y_1[n]$ and $x[n]$ on the same plot and compare. Does the filter perform its function?

(c) Design an optimal highpass filter with a 1-dB passband edge at $F = 0.06$ and a 40-dB stopband edge at $F = 0.01$ and use the MATLAB command `filtfilt` to obtain the zero-phase filtered signal $y_2[n]$. Plot its DFT magnitude and display its time-frequency plot. What do the plots reveal? Plot $y_2[n]$ and $x[n]$ on the same plot and compare. Does the filter perform its function?

20.33 (Filtering of a Chirp Signal) This problem deals with time-frequency plots of a chirp signal and a sinusoid and its filtered versions.

(a) Use the ADSP routine `chirp` to generate 500 samples of a signal $x[n]$ that consists of the sum of $\cos(0.6n\pi)$ and a chirp whose frequency varies from $F = 0$ to $F = 0.05$. Use the MATLAB command `fft` to plot its DFT magnitude. Use the ADSP routine `psdwelch` to display its power spectral density plot. Use the ADSP routine `timefreq` to display its time-frequency plot. What do the plots reveal?

(b) Design an optimal lowpass filter with a 1 dB passband edge at $F = 0.08$ and a 40-dB stopband edge at $F = 0.25$ and use the MATLAB command `filtfilt` to obtain the zero-phase filtered signal $y_1[n]$. Plot its DFT magnitude and display its PSD and time-frequency plot. What do the plots reveal? Plot $y_1[n]$. Does it look like a signal whose frequency is increasing with time? Do the results confirm that the filter performs its function?

(c) Design an optimal highpass filter with a 1-dB passband edge at $F = 0.25$ and a 40-dB stopband edge at $F = 0.08$ and use the MATLAB command `filtfilt` to obtain the zero-phase filtered signal $y_2[n]$. Plot its DFT magnitude and display its PSD and time-frequency plot. What do the plots reveal? Plot $y_1[n]$. Does it look like a sinusoid? Can you identify its period from the plot? Do the results confirm that the filter performs its function?

PGR2026-00033
Krisp Technologies EX1024 Page 774

20.34 (A Multi-Band Filter) A requirement exists for a multi-band digital FIR filter operating at 140 Hz with the following specifications:

Passband 1: from dc to 5 Hz
Maximum passband attenuation = 2 dB (from peak)
Minimum attenuation at 10 Hz = 40 dB (from peak)

Passband 2: from 30 Hz to 40 Hz
Maximum passband attenuation = 2 dB (from peak)
Minimum attenuation at 20 Hz and 50 Hz = 40 dB (from peak)

(a) Design the first stage as an odd-length optimal filter, using the ADSP routine `firpm`.
(b) Design the second stage as an odd-length half-band filter, using the ADSP routine `firhb` with a Kaiser window.
(c) Combine the two stages to obtain the impulse response of the overall filter.
(d) Plot the overall response of the filter. Verify that the attenuation specifications are met at each design frequency.

20.35 (Audio Equalizers) Many hi-fi systems are equipped with graphic equalizers to tailor the frequency response. Consider the design of a four-band graphic equalizer. The first section is to be a lowpass filter with a passband edge of 300 Hz. The next two sections are bandpass filters with passband edges of [300, 1000] Hz and [1000, 3000] Hz, respectively. The fourth section is a highpass filter with a stopband edge at 3 kHz. The sampling rate is to be 20 kHz. Implement this equalizer, using FIR filters based on window design. Repeat the design, using an optimal FIR filter. Repeat the design, using an IIR filter based on the bilinear transformation. For each design, superimpose plots of the frequency response of each section and their parallel combination. What are the differences between IIR and FIR design? Which design would you recommend?

PGR2026-00033
Krisp Technologies EX1024 Page 775

Chapter 21

MATLAB EXAMPLES

21.0 Introduction

In this last chapter, we conclude with examples that illustrate many of the principles of analog and digital signal processing, using native MATLAB commands and the ADSP toolbox routines supplied with this book. The code can be readily customized for use with other similar examples. You may wish to consult the extensive help facility within MATLAB to learn more about the MATLAB commands.

21.1 The ADSP Toolbox and Its Installation

The ADSP toolbox is a suite of MATLAB m-files (with extension .m) that resides in two directories called `adsp` and `gui` on the supplied disk. A listing of the m-files appears at the end of this chapter. The toolbox is supplied for PC compatibles but may also be used on Unix platforms (if you convert the m-files using the `dos2unix` facility), or Macintosh platforms (if you convert the m-files using the Apple file exchange facility). The ADSP toolbox requires MATLAB for its operation. It is designed to work with all (student or professional) versions of MATLAB v4 or later. Most of the nongraphics m-files will also work with the older MATLAB v3.5, but a few will not (since newer versions of MATLAB are not fully backward compatible).

Installation

The following installation procedure is generic to all platforms. To install the ADSP toolbox,

1. Create two subdirectories (preferably with the names `adsp` and `gui`).
2. Copy the `adsp` files from the disk to your `adsp` subdirectory.
3. Copy the `gui` files from the disk to your `gui` subdirectory.
4. Add the names of the two subdirectories to the MATLAB path.

Testing

To check the proper operation of the m-files in the ADSP toolbox, start MATLAB and, at the MATLAB prompt (`>>`), type

```
tour
```

The ADSP routine `tour` executes many of the m-files and prompts you to press a key to continue.

For a tour of all the GUI (graphical user interface) programs, type

```
tourgui
```


PGR2026-00033
Krisp Technologies EX1024 Page 776

The Help Facility

Most ADSP m-files are self-demonstrating if you type the filename (with no arguments) at the MATLAB prompt. For example, to seek help on the m-file `sysresp1.m`, type

```
sysresp1
```

This will bring up a help screen and demonstrate a simple example of usage.
You may also use the help facility of MATLAB by typing

```
help sysresp1
```

In this case, you will only see the help screen but no example of usage.

Bugs or Problems

If you encounter problems while installing or operating the toolbox, or wish to report any bugs or improvements, we would like to hear from you. You may contact us at the following e-mail addresses, or visit our Internet site for updates and other information:

Ashok Ambardar	Michigan Technological University
Internet:	`http://www.ee.mtu.edu/faculty/akambard.html`
e-mail:	`akambard@mtu.edu`
Brooks/Cole Publishing	511 Forest Lodge Road, Pacific Grove, CA 93950-5098,USA
Internet:	`http://www.brookscole.com`
e-mail:	`info@brookscole.com`

21.2 MATLAB Tips and Pointers

We assume you have a passing familiarity with MATLAB. However, the following tips and pointers may prove useful if you are new to MATLAB and want to get started.

21.2.1 Getting Started

When in MATLAB, you may try the following commands

```
tutor1     % Tutorial on MATLAB, part 1 (ADSP routine).
tutor2     % Tutorial on MATLAB, part 2 (ADSP routine).
doc        % Online documentation by the Mathworks Inc. (v4 and v5).
```

To exit MATLAB, type `quit` at the MATLAB prompt.

21.2.2 Some General Guidelines

1. MATLAB works in double precision. It is case sensitive. All built-in commands are in lower case.
2. Use `format long` to see results in double precision, and `format` to get back to the default display.
3. In MATLAB, the indexing starts at 1 (not 0).
4. A semicolon (;) after a statement or command suppresses on-screen display of results.
5. Comments (which follow the % sign) are not executed.
6. To abort an executing MATLAB command, or on-screen data display, use Ctrl-C.
7. Use the up/down "arrow keys" to scroll through commands and re-edit/execute them.

PGR2026-00033
Krisp Technologies EX1024 Page 777

21.2.3 Array Operations

Most operations you will perform will be elementwise operations on arrays. Be aware of the following:

```
x=0:0.01:2; % Generate an array x at intervals of 0.01
y=x .^ 2;   % Generate array of squared values; uses " .' "
z=x.';      % Transpose x from a row to column or vice versa; uses " .' "
x.*y;       % Pointwise product of equal size arrays x and y; uses " .* "
x./y;       % Pointwise division of equal size arrays x and y; uses " ./ "
l=length(x) % Returns the length of an array or vector x
p=x(3)      % Third element of x (MATLAB indexing starts with 1)
```

21.2.4 A List of Useful Commands

Some of the MATLAB commands for mathematical operations are `exp, sin, cos, tan, acos, asin, atan, pi` (for π), `eps` (for default tolerance = $(2.22)10^{-16}$), `i` or `j` (for $\sqrt{-1}$), `real, imag, sqrt` (for square root), `abs` (for absolute value or magnitude), `angle` (for the angle or phase in *radians*), `log2` and `log10` (for logarithms to the base 2 or 10), `length` (of a vector or array), and `size` (of a matrix). Other useful commands are `min, max, int, fix, find, conj, sum, cumsum, diff, prod, eval, fliplr`, and `sign`. Use the MATLAB help facility to learn more about them.

String Functions and Logical Operations

A string function (symbolic expression) is entered in single quotes, for example `x = '4*t.*exp(-2*t)'`. MATLAB displays it on screen without quotes. To use its results, or plot the expression, it must first be evaluated using `eval`. For example:

```
x = '4*t.*exp(-2*t)'     % String or symbolic expression
t=0:0.05:4;              % Pick a time range t
xe = eval(x);            % Evaluate the expression x
plot(t,xe)               % Plot
```

Logical operations are useful in many situations. Here are some examples:

```
i=find(xe>=0.7);         % Index of values in xe that equal or exceed 0.7
iz=i(1);                 % Index of first such value in i
tz=t(iz);                % Approximate time of first zero
k=find(xe==max(xe));     % Index of maximum in xe. Note the ==
tm=t(k);                 % Time(s) of maximum
```

Saving a MATLAB session

The `diary` command is a useful way of saving MATLAB statements. Type

```
diary mywork.m     % Example:  Open a file which will be called mywork.m
```

This opens a file called `mywork.m`. All subsequent MATLAB on-screen statements and displays will be included in this file. To close the file, type `diary off`. More statements may be appended to this file by issuing the commands `diary on` and `diary off`.

Frequently Asked Questions

Q. How many data points do I need to obtain a smooth curve when plotting analog signals?
A. A good rule of thumb is at least 200 points.

PGR2026-00033
Krisp Technologies EX1024 Page 778

Q. Why do I get an error while using `plot(t,x)`, with `x` and `t` previously defined?
A. Both `t` and `x` must have the same dimensions (size). To check, use `size(t)` and `size(x)`.

Q. Why do I get an error while generating $x(t) = 2^{-t}\sin(2t)$, with the array `t` previously defined?
A. You probably did not use array operations (such as `.*` or `.^`). Try `>> x=(2 .^(-t)).*sin(2*t)`

Q. How can I plot several graphs on the same plot?
A. Example: use `>>plot(t,x),hold on,plot(t1,x1),hold off` OR `>>plot(t,x,t1,x1)`

Q. How do I use `subplot` to generate 6 plots in the same window (2 rows of 3 subplots)?
A. Example: use `>>subplot(2,3,n);plot(t,x)` for the nth (count by rows) plot (`n` cycles from 1 to 6). A subplot command *precedes* each plot command and has its own xlabel, ylabel, axis, title, etc.

21.3 Graphical User Interface Programs

This book includes graphical user interface routines requiring little or no prior MATLAB skills, and designed for easy visualization of basic signal-processing concepts (some of the guis have been adapted from the book *Mastering DSP Concepts Using MATLAB*, by A. Ambardar and C. Borghesani, Prentice Hall, 1998). Their common features include a point-and-click interface, simple data entry, a help menu, plot and axis control, and cursor control. The following guis (with their function and relevant chapter listed alongside) are provided with this book and reside in the `gui` subdirectory on the supplied disk. To explore a gui, just type its filename (e.g., `ctcongui`) while in MATLAB.

```
tourgui     A Tour of All the GUIs
ctsiggui    Visualization of Analog Signals and Signal Operations (Chapter 2)
mathgui     Visualization of Special Mathematical Functions (Chapter 2)
dtsiggui    Visualization of Discrete-Time Signals and Signal Operations (Chapter 3)
randgui     Visualization of Random Signals and Histograms (Chapter 3)
ctsimgui    Simulation of the Response of Analog Systems (Chapter 4)
dtsimgui    Time-Domain Response of Discrete-Time Systems (Chapter 5)
ctcongui    Movie of Continuous-Time Convolution (Chapter 6)
dtcongui    Movie of Discrete-Time Convolution (Chapter 7)
fsgui       Fourier Series Reconstruction and Smoothing (Chapter 8)
fsrgui      System Response to Periodic Signals (Chapter 8)
fssyngui    Fourier Series Synthesis from Its Coefficients (Chapter 8)
ftprpgui    Visualization of Fourier Transform Properties (Chapter 9)
freqgui     Frequency Response Visualization for Analog Systems (Chapter 9)
modgui      Visualization of Concepts in Modulation (Chapter 10)
ctfgui      Pole-Zero Plots, Impulse Response, and Frequency Response (Chapter 11)
bodegui     Bode Plot Visualization (Chapter 12)
afdgui      Analog Filter Design (Chapter 13)
dfreqgui    Frequency Response Visualization for Discrete-Time Systems (Chapter 15)
dfftgui     The FFT of Sinusoids and Windowing (Chapter 16)
tfestgui    Transfer Function Estimation (Chapter 16)
chirpgui    Chirp Signals and Their Spectra (Chapter 16)
dtfgui      Pole-Zero Plots, Impulse Response, and Frequency Response (Chapter 17)
pzdesgui    Filter Design by Pole-Zero Placement (Chapter 18)
diirgui     IIR Filter Design (Chapter 19)
lpsiggui    Linear-Phase Sequences and Spectra (Chapter 20)
dfirgui     FIR Filter Design (Chapter 20)
printgui    Facility for Printing guis in MATLAB v4
```

PGR2026-00033
Krisp Technologies EX1024 Page 779

21.4 The ADSP Toolbox

The ADSP toolbox consists of a suite of m-files for various signal processing tasks and resides in the `adsp` directory on the supplied disk. A short description of the most useful routines is given below (support files are not listed). Examples of usage are described in the following section.

Tutorials and Demonstrations

```
TOUR      A tour of the ADSP toolbox
TUTOR1    Tutorial on MATLAB basics
TUTOR2    Tutorial on m-files and string functions
DEMOCONV  Concepts in convolution
DEMOFS    Concepts in Fourier series
DEMOSAMP  Concepts in sampling
```

Routines for Mathematical Analysis and Special Functions

```
GCD1      Greatest common divisor of integers or fractions
LCM1      Least common multiple of integers or fractions
NUMDIG    Truncates or rounds to N significant digits
SI        Sine integral sin (ax)/ax
SIMPSON   Area of a function over specified limits
POLYMAP   Converts a transfer function using polynomial transformations
BESIN     Bessel function I(x) of order n
BESINU    Spherical Bessel function i(x) of order nu
BESJN     Bessel function J(x) of order n
BESJNU    Spherical Bessel function of order nu
CHEBYFUN  Chebyshev polynomial T(x) or U(x) of order n
CX        Fresnel cosine integral C(x)
SX        Fresnel sine integral S(x)
GM        The (complete) gamma function
INVSI     Inverse of sine integral
ALOG      Antilog or dB to gain conversion
```

Signal Generation and Plotting (Chapters 2, 3)

```
DTPLOT    Plots discrete signals
AXESN     Draws axes on an existing plot
UDELTA    Discrete impulse (sample) function
URAMP     Analog or discrete ramp function
URECT     Analog or discrete rect function
USTEP     Analog or discrete step function
TRI       Analog or discrete tri function
SINC      Sinc function
CHIRP     Chirp signal
RANDIST   Random numbers with specified distributions
EVENODD   Even and odd parts of an analog or discrete signal
PEREXT    Periodic extension of a signal
ENERPWR   Energy or power in a signal
OPERATE   Time-scaling and time-shifting operations
```

PGR2026-00033
Krisp Technologies EX1024 Page 780

Time-Domain Analysis (for Chapters 4, 5)

```
SYSRESP1  Response in symbolic form (from differential/difference equation)
CTSIM     Simulation of analog systems using numerical integration
DTSIM     Simulation of discrete-time systems
```

Convolution and Correlation (for Chapters 6, 7)

```
CONVNUM   Numerical convolution
PAIRSUM   Establishes end points of convolution ranges
CONVPLOT  Movie of continuous-time convolution
CONVP     Periodic convolution
CONVMAT   Generates the circulant matrix
CORRNUM   Numerical correlation
CORRP     Periodic correlation
CLT       Demo of the central limit theorem of convolution
```

Frequency-Domain Analysis (for Chapters 8, 9, 10, 16)

```
FSERIES   Computes and plots Fourier coefficients and reconstructions
FSKERNEL  Dirichlet and Fejer kernel for Fourier series
TFPLOT    Plots magnitude/phase of a CT or DT transfer function
BODELIN   Plots asymptotic (and actual) Bode plots
TRBW      Time and bandwidth measures for lowpass filters
WBFM      Spectrum/bandwidth of wideband FM
MODSIG    Generates AM, FM, and PM signals
DEMODSIG  Demodulation of AM and FM signals
ANALYTIC  Generates the analytic signal
FFTPLOT   Computes FFT and plots its magnitude and phase
ALIAS     Computes aliased frequencies
WINFFT    Generates window functions for FFT
PSDWELCH  PSD using the Welch method
PSDTUKEY  PSD using the Tukey method
TIMEFREQ  Generates time-frequency (waterfall) spectra
```

Transform Analysis (for Chapters 11, 12, 17, 18)

```
LTR       Laplace transform
ZTR       z-transform
TF2PF     Transfer function to partial fractions
PF2TF     Partial fractions to  transfer function
ILT       Inverse Laplace transform in symbolic form
IZT       Inverse z-transform in symbolic form
IZTLONG   Inverse z-transform by long division
PLOTPZ    Pole/zero plots of CT and DT systems
SYSRESP2  CT and DT response in symbolic form (from transfer function)
SSRESP    CT and DT steady-state response in symbolic form
STABLIZE  Stabilizes an unstable H(s) or H(z)
TFMIN     Finds the minimum-phase TF from magnitude squared function
MINPHASE  Converts a nonminimum-phase TF to a minimum-phase TF
```

PGR2026-00033
Krisp Technologies EX1024 Page 781

Analog Filter Design (for Chapter 13)

```
AFD       Classical analog filter design
LPP       Design of lowpass prototypes
ATTN      dB attenuation of classical analog filters
BUTTPOLE  Plots and returns pole locations of Butterworth filters
CHEBPOLE  Pole locations of Chebyshev filter with geometric construction
CHCOEFF   Coefficients of nth order Chebyshev polynomial T(x) or U(x)
LP2AF     Analog filter transformations (LP2LP, LP2HP, LP2BP, LP2BS)
AFDBESS   Bessel lowpass filter design
BESSTF    Bessel filter transfer function of order n
DELAY     Delay of lowpass filters
```

Sampling and Quantization (for Chapter 14)

```
INTERPOL  Signal interpolation using various methods
QUANTIZ   Rounds/truncates signal values to required levels
```

IIR Filter Design (for Chapter 19)

```
DFDIIR    Design of IIR digital filters
S2ZNI     Mapping based on numerical integration/differences
S2ZMATCH  Mapping based on the matched z-transform
S2ZINVAR  Mapping based on response invariance
LP2IIR    Analog or digital prototype transformation to IIR filter
DFDPZP    Interactive filter design by pole-zero placement
BLT2ORD   Design of second-order peaking or notch filters
```

FIR Filter Design (for Chapter 20)

```
LPSIG     Amplitude spectrum of linear-phase signals
FIRWIND   Window-based FIR digital filter design
WINDOW    Generates window functions for FIR filter design
WINSPEC   Plots spectra of windows and returns figures of merit
FIRHB     Design of half-band FIR digital filters
FIRPM     Optimal (Parks-McClellan) filter design
FIRMF     Maximally flat FIR filter design
FIRDIFF   FIR differentiators
FIRHILB   FIR Hilbert transformers
TFWELCH   Transfer function estimate using Welch PSD
```

Miscellaneous (See the Disk Contents for a Full Listing)

```
ANALYTIC  Analytic signal for a real signal
ECGSIM    Generate a synthetic ECG (electrocardiogram) signal
SINC2     Sinc squared function
INVSI2    Inverse of sine-squared integral
Z2SMAP    Convert H(z) to H(s) using various mappings
```

PGR2026-00033
Krisp Technologies EX1024 Page 782

21.5 Examples of MATLAB Code

This extended section provides examples of MATLAB *script files*, using both the ADSP routines and native MATLAB commands. We remark that

1. Chapters of primary relevance are listed in parentheses in the title of each example.
2. MATLAB statements are shown in typewriter font and are *case sensitive*.
3. Statements following the % sign are comments (and are not executed).

To save space in the code that follows, we have often shown several MATLAB statements on the same line (separated by commas to display results, or semicolons to suppress the display). MATLAB statements for plotting embellishments (such as titles, axis labels, etc.) have, for the most part, been omitted. The plots have not been reproduced. The accompanying disk contains *script* m-files for each example (as `exampxx.m`, where `xx` is the two-digit example number). The files reside in the `examples` subdirectory. You may run these m-files, and add plotting commands or other embellishments to suit your own needs.

EXAMPLE 21.1 (Plotting Analog Signals (Ch. 2))
Plot $x_1(t) = u(t-1)$, $x_2(t) = r(t-1)$, $x_3(t) = \text{rect}(0.5t-1)$, $x_4(t) = \text{tri}(2t-2)$, $x_5(t) = \text{sinc}(t/2)$, and $x_6(t) = 4e^{-2t}\cos(\pi t)u(t)$ over the range $-2 \le t \le 4$.

Run `examp01.m` (in the `examples` subdirectory on disk). A partial listing of the MATLAB code follows:

```
t=-2:0.01:4;                    % CT index (2 to 4)
x1=ustep(t-1);plot(t,x1)        % (ustep is an ADSP routine)
x2=uramp(t-1);plot(t,x2)        % (uramp is an ADSP routine)
x3=urect(0.5*t-1);plot(t,x3)    % (urect is an ADSP routine)
x4=tri(2*t-2);plot(t,x4)        % (tri is an ADSP routine)
x5=sinc(t/2);plot(t,x5)
x6=4*exp(-2*t).*cos(pi*t).*ustep(t);plot(t,x6)
```

EXAMPLE 21.2 (Integration (Ch. 2))

A crude way to find the running integral $y(t) = \int_0^t x(t)\,dt$ is to approximate $x(t)$ by a staircase function, and find $y(t)$ as the cumulative sum of the areas under the rectangular strips. The MATLAB routine `cumsum` yields the cumulative sum of the function values (heights). To find the cumulative area, we multiply by the time interval t_s (the strip width).
Plot the approximate and exact running integral of $x(t) = 10e^{-t}\sin(2t)u(t)$.

Analytically, the exact running integral is $y(t) = 4 - 2e^{-t}[\sin(2t) + 2\cos(2t)]$.

Run `examp02.m` (in the `examples` subdirectory on disk). A partial listing of the MATLAB code follows:

```
ts=0.05;t=0:ts:6;                          % Choose time axis with ts=0.05
x=10*exp(-t).*sin(2*t);                    % Function x(t)
xs=cumsum(x);                              % running sum
y=ts*xs;                                   % Approximate running integral
ye=4-2*exp(-t).*(sin(2*t)+2*cos(2*t));     % Exact running integral
plot(t,x,t,y,t,ye),grid                    % Overplot
```

PGR2026-00033
Krisp Technologies EX1024 Page 783

EXAMPLE 21.3 (Derivatives (Ch. 2))
The derivative $x'(t)$ can be numerically approximated by the slope $(x[n] - x[n-1])/t_s$ where t_s is a small time step. The MATLAB routine `diff` yields the differences $x[n] - x[n-1]$. Plot the approximate and exact derivative of $x(t) = 10e^{-t}\sin(2t)u(t)$.
Note: The number of points returned by `diff` is *one less* than in $x(t)$. If the length of t and $x(t)$ is n, the length of `d=diff(x)` is $n-1$. To plot d, we can use `n=length(x);plot(t(2:n),d)` *but not* `plot(t,d)`.

The exact derivative works out to $x'(t) = 10e^{-t}[2\cos(2t) - \sin(2t)]$.

```
ts=0.05;t=0:ts:6;l=length(t)           % Choose time axis with ts=0.05
x=10*exp(-t).*sin(2*t);                % Function x(t)
z=diff(x)/ts;                          % Numerical derivative. Note: ts=0.05
ze=10*exp(-t).*(2*cos(2*t)-sin(2*t));  % Exact derivative
plot(t,x,t(2:l),z,t,ze),grid           % Plot numerical and exact result
```

EXAMPLE 21.4 (Signal Operations (Ch. 2))
The ADSP routine `operate` can be used to plot scaled/folded/shifted versions of a signal $x(t)$.
Let $x(t) = 2u(t+1) - r(t+1) + r(t-1)$.
To plot $y(t) = x(-2t+5)$, use the following commands:

```
t=-2:0.05:4;                           % Create time axis.
x=2*ustep(t+1)-uramp(t+1)+uramp(t-1);  % Create signal x.
[t1,y]=operate(t,x,-2,5);              % Generate y=x(-2t+5).
plot(t,x,t1,y)                         % Overplot
```

EXAMPLE 21.5 (Periodic Signals (Ch. 2))
A combination of sinusoids is periodic only if the ratios of the frequencies (or periods) are rational fractions. The common period equals the LCM of the individual periods. The fundamental frequency equals the GCD of the individual frequencies. The ADSP routines `lcm1` and `gcd1` allow us to find the LCM or GCD of an array of rational fractions.
What is the common period and fundamental frequency of the signal $x(t) = 2\cos(2.4\pi t) - 3\sin(5.4\pi t) + \cos(14.4\pi t - 0.2\pi)$?

The frequencies of the individual sinusoids are 1.2 Hz, 2.7 Hz, and 7.2 Hz. We express these as fractions (12/10, 27/10, and 72/10), and find the common period as their LCM.

```
[tnum, tden]=lcm1([10 10 10],[12 27 72])    % Find LCM of periods
T=tnum/tden                                 % Common period T
[fnum, fden]=gcd1([12 27 72],[10 10 10])    % Find GCD of frequencies
f0=fnum/fden                                % Fundamental frequency f0
```

EXAMPLE 21.6 (Energy and Power (Ch. 2))
A crude way to find the energy in $x(t)$ is to approximate the area of $x^2(t)$, assuming a small time step. If $x(t)$ is periodic, we can find its power by computing the energy of one period and dividing by the period.
The ADSP routine `enerpwr` computes the energy (or power) if $x(t)$ is described as a *string* expression, and does not require choice of the time step.
Let $x(t) = 6\,\mathrm{sinc}(2t)\mathrm{rect}(t)$, the central lobe of $6\,\mathrm{sinc}(2t)$ over $(-0.5 \le t \le 0.5)$. Find the energy in this signal and the signal power if $x(t)$ forms one period of a periodic signal with $T = 1.4$

PGR2026-00033
Krisp Technologies EX1024 Page 784

```
ts=0.01;                                % Select small time interval
t=-0.5:ts:0.5;                          % Create a time axis
x=6*sinc(2*t);                          % Generate x(t)
x2=x.*x;                                % Square the signal x(t)
e=sum(x2)*ts                            % Find energy e
T=1.4;p=e/T                             % And power p
ee=enerpwr('6*sinc(2*t)',[-0.5, 0.5])   % Find energy using enerpwr
p=enerpwr('6*sinc(2*t)',[-0.5, 0.5],T)  % And power using enerpwr
```

EXAMPLE 21.7 (The Impulse Function (Ch. 2))
The analog impulse function may be regarded as a tall, narrow spike that arises as a limiting form of many ordinary functions such as the rect, sinc, and Gaussian. As we decrease their width, these signals increase in height in such a way that the area remains constant.

Consider the Gaussian signal $x(t) = \frac{1}{t_0} e^{-\pi(t/t_0)^2}$. Plot $x(t)$ over $(-2, 2)$ for $t_0 = 1, 0.5, 0.1, 0.05, 0.01$, using a time step of $0.1t_0$. Do all have the same area? What can you say about the nature of $x(t)$ as $t_0 \to 0$?

```
t0=1;t1=-2:0.1*t0:2;x1=(1/t0)*exp(-pi*t1.*t1/t0/t0);
t0=0.5;t2=-2:0.1*t0:2;x2=(1/t0)*exp(-pi*t2.*t2/t0/t0);
t0=0.1;t3=-2:0.1*t0:2;x3=(1/t0)*exp(-pi*t3.*t3/t0/t0);
t0=0.05;t4=-2:0.1*t0:2;x4=(1/t0)*exp(-pi*t4.*t4/t0/t0);
t0=0.01;t5=-2:0.1*t0:2;x5=(1/t0)*exp(-pi*t5.*t5/t0/t0);
plot(t1,x1,t2,x2,t3,x3,t4,x4,t5,x5)
a1=sum(x1)*0.1;a2=sum(x2)*0.05;
a3=sum(x3)*0.01;a4=sum(x4)*0.005;
a5=sum(x5)*0.001;                       % First generate each area
a=[a1;a2;a3;a4;a5]                      % Display (all should be 1)
```

Comment: As $t_0 \to 0$, the Gaussian tends to the unit impulse $\delta(t)$.

EXAMPLE 21.8 (Plotting Discrete Signals (Ch. 3))
Plot $x_a[n] = \delta(n-3)$, $x_b[n] = u[n-3]$, $x_c[n] = r(n-3)$, $x_d[n] = \text{sinc}(n/4)$, and $x_e[n] = 4(0.8)^n \cos(0.2n\pi)u[n]$ over the range $-10 \le n \le 10$.

The MATLAB code uses the ADSP routines `udelta`, `ustep`, `uramp`, `sinc`, and `dtplot`.

```
n=-10:10;                               % DT integer index (-10 to 10)
xa=(n==3);stem(n,xa)                    % d[n-3] or
xa=udelta(n-3);dtplot(n,xa,'o')
xb=(n>=3);stem(n,xb)                    % Step u[n-3] or
xb=ustep(n-3);dtplot(n,xb,'*')
xc=(n-3).*(n>=3); stem(n,xc)            % Ramp r[n-3] or
xc=uramp(n-3);dtplot(n,xc)
xd=sinc(n/4);                           % The sinc
xe=4*(0.8 .^ n).*cos(0.2*n*pi).*(n>=0); % Damped cosine
```

PGR2026-00033
Krisp Technologies EX1024 Page 785

EXAMPLE 21.9 (Discrete-Time Complex Exponentials (Ch. 3))
Complex-valued signals must be plotted using magnitude and phase information separately or real part and imaginary part separately. Let $w[n] = 7.071e^{j\left(\frac{\pi}{9}n - \frac{\pi}{4}\right)}$, $-20 \leq n \leq 20$. Plot the real and imaginary parts, magnitude and phase, and the sum and difference of the real and imaginary parts. For the last case, derive analytic expressions for the sequences plotted and compare with your plots. Which plots allow you determine the period of $w[n]$? What is the period?

```
n=-20:20;
j=sqrt(-1);
w=7.071*exp(j*(n*pi/9-pi/4));
dtplot(n,real(w),'o'),dtplot(n,imag(w),'o')
dtplot(n,abs(w),'o'),dtplot(n,180*angle(w)/pi,'o')
dtplot(n,real(w)+imag(w),'o'),dtplot(n,real(w)-imag(w),'o')
```

EXAMPLE 21.10 (Discrete Signal Operations (Ch. 3))
Let $x[n] = r[n+6] - r[n+3] - r[n-3] + r[n-6]$. Plot $y_1[n] = x[n-4]$, $y_2[n] = x[n+4]$, $y_3[n] = x[-n-4]$, and $y_4[n] = x[-n+4]$.

```
n=-10:10;
x=uramp(n+6)-uramp(n+3)-uramp(n-3)+uramp(n-6);
[n1,y1]=operate(n,x,1,-4);
[n2,y2]=operate(n,x,1,4);
[n3,y3]=operate(n,x,-1,-4);
[n4,y4]=operate(n,x,-1,4);
dtplot(n,x),dtplot(n1,y1,'o')            % Similar plot commands for y2 etc.
```

EXAMPLE 21.11 (Discrete Signal Measures (Ch. 3))
Let $x[n] = r[n] - r[n-5] - 5u[n-10]$. Find the signal energy in $x[n]$ and sketch $x[n]$, $x_d[n] = x[n-2]$, its even part $x_e[n]$, and its odd part $x_o[n]$.
Sketch $y[n]$, the periodic extension of $x[n]$ with period $N = 7$, for three periods. What is the signal power in $y[n]$?

The MATLAB code uses the ADSP routines `uramp`, `evenodd`, `perext`, and `dtplot`.

```
n=0:15;                                  % DT index
x=uramp(n)-uramp(n-5)-5*uramp(n-10);     % Generate x[n]
e=sum(x.*x)                              % Energy in x[n]
dtplot(n,x)                              % Plot x[n] using dtplot
xd=uramp(n-2)-uramp(n-7)-5*uramp(n-12);  % Generate x[n-2]
stem(n,xd)                               % Plot x[n-2] using stem
[xe,xo,nn]=evenodd(x,n);                 % Even, odd parts, and index nn
dtplot(nn,xe)                            % Plot even part of x[n]
dtplot(nn,xo)                            % Plot odd part of x[n]
y=perext(x,7);                           % Periodic extension xpe[n]
np=0:20;dtplot(np,[y y y])               % Plot 3 periods over n=0:20
pwr=sum(y.*y)/7                          % Power in xpe[n]
```

PGR2026-00033
Krisp Technologies EX1024 Page 786

EXAMPLE 21.12 (Decimation and Interpolation (Ch. 3))
Let $h[n] = \sin(n\pi/3), \quad 0 \leq n \leq 10$. Plot $h[n]$, the decimated signal $h[3n]$, and the zero-interpolated, step-interpolated, and linearly interpolated signals, using interpolation by 3.

```
n=0:10;                  % Create DT index
h=sin(n*pi/3);           % Create signal h
yz=interpol(h,'z',3)     % Zero interpolation
yc=interpol(h,'c',3)     % Step (constant) interpolation
yl=interpol(h,'l',3)     % Linear interpolation
```

EXAMPLE 21.13 (Random Distributions (Ch. 3))
Generate 500 points each of a uniform and Gaussian (normal) random signal, plot their first 100 values, plot their histograms using 20 bins, and compute their mean and standard deviation.

The MATLAB code uses the ADSP routine `randist`.

```
xu=randist(500,'uni');      % Uniform (mean=0.5, var=0)
xr=randist(500,'nor');      % Gaussian (mean=0, var=1)
plot(xu(1:100)),            % Plot 100 samples of uniform
plot(xr(1:100)),            % Plot 100 samples of Gaussian
hist(xu,20);                % Plot histogram of uniform
hist(xr,20);                % Plot histogram of Gaussian
mu=mean(xu),mr=mean(xr)     % Display means
su=std(xu),sr=std(xr)       % Display standard deviations
```

EXAMPLE 21.14 (The Central Limit Theorem (Ch. 3))
The probability distribution of combinations of statistically independent, random phenomena often tends to a Gaussian distribution. This is the *central limit theorem*. Demonstrate the central limit theorem by generating five realizations of a uniformly distributed random signal, and plotting the histogram of the individual signals and the histogram of their sum.

```
x1=randist(500,'uni'); x2=randist(500,'uni'); % Uniform signals x1 and x2
x3=randist(500,'uni');x4=randist(500,'uni');  % And x3 and x4
x5=randist(500,'uni');x=x1+x2+x3+x4+x5;       % And x5 and their sum
hist(x1,20);                                  % Histogram of one realization x1
hist(x,20);                                   % Histogram of sum x
m1=mean(x1), m=mean(x)                        % Mean of x (should equal sum of means)
```

EXAMPLE 21.15 (Signal-to-Noise Ratio (Ch. 3))
For a noisy signal $x(t) = s(t) + An(t)$ with a signal component $s(t)$ and a noise component $An(t)$, the *signal-to-noise ratio* (SNR) is the ratio of the signal power σ_s^2 and noise power $A^2\sigma_n^2$ and defined in *decibels* (dB) as $\text{SNR} = 10 \log \frac{\sigma_s^2}{A^2\sigma_n^2}$ dB. We can adjust the SNR by varying the noise amplitude A. Use this result to generate a noisy sinusoid with a signal-to-noise ratio of 18 dB. Assume uniformly distributed noise.

We find $A = \dfrac{\sigma_s/\sigma_n}{10^{\text{SNR}/20}}$.

PGR2026-00033
Krisp Technologies EX1024 Page 787

```
t=0:0.01:2;                          % Time array
xs=sin(2*pi*t);                      % Signal
xn=randist(xs,'uni',0);              % Uniform noise (mean=0)
snr=18;                              % Desired SNR
A=std(xs)/std(xn)/(10^(snr/20));     % Compute A
x=xs+A*xn;                           % Generate noisy signal
```

Comment: You can compare the value of A obtained above with its theoretical value A_t by noting that the signal variance is 0.5, and the noise variance is $A_t^2/12$. Thus, SNR $= 18 = 20\log(6/A_t^2)$

EXAMPLE 21.16 (Signal Averaging (Ch. 3))
Extraction of signals from noise is an important signal-processing application. Averaging the results of many runs tends to average out the noise to zero and improve the signal-to-noise ratio.
Generate 16 runs (*realizations*) of a noisy signal by adding uniformly distributed random noise (with zero mean) to the signal $x(t) = \sin(40\pi t)$ and average the results. Repeat for 64 runs and compare results.

```
ts=0.001;t=0:ts:0.2;                 % Time array
x=sin(40*pi*t);                      % Pure sinusoid
xn=x+randist(x,'uni',0);             % Signal+uniform noise (mean=0)
z=0; for n=1:64;                     % Initialize z and start loop
z=z+x+randist(x,'uni',0);            % Keep summing each run
if n==16, a16=z/16; end              % Save average of 16 runs
end                                  % End of summing 64-runs
a64=z/64;                            % Average of 64 runs
plot(t,xn)                           % Plot noisy signal
plot(t,a16)                          % And 16-run average
plot(t,a64)                          % And 64-run (better)
```

Comment:
On systems with sound capability, one could create a longer signal (2 s or so) at the natural sampling rate and listen to the pure, noisy and averaged signal for audible differences, using the MATLAB command `sound`.

EXAMPLE 21.17 (Response of Analog Systems in Symbolic Form (Ch. 4))
The ADSP routine `sysresp1` finds the response of analog systems in *symbolic* form. Its syntax is

```
[yt,yzs,yzi] = sysresp1('s', RHS(or num), LHS(or den), X, IC)
```

The first argument is `'s'` in single quotes (for an analog system). The arguments `RHS` and `LHS` are coefficient arrays of the two sides of the differential equation or the numerator and denominator of the transfer function $H(s)$. The argument `X=[K a p w r]` is a five-element array corresponding to the input

$$x(t) = Ke^{-at}t^p \cos[wt + r \text{ (radians)}]u(t)$$

For M inputs, `X` is a matrix with M rows.
The last (optional) argument `IC` is the initial condition array `IC=[y(0), y'(0),..]`.
The output arguments `[yt, yzs, yzi]` are the total, zero-state, and zero-input response.
The *impulse response* is found by *omitting* the last three arguments (`NX, DX`, and `IC`).
Note: Since the output of `sysresp1` is a symbolic (or *string*) expression, it must first be evaluated (using the `eval` command) for a selected range of t before it can be plotted.

PGR2026-00033
Krisp Technologies EX1024 Page 788

Consider an analog system whose transfer function is $H(s) = \dfrac{2}{s+2}$.
Find and plot its impulse response $h(t)$ and step response $w(t)$.
Find its response $y_a(t)$ to $x(t) = 2e^{-3t}u(t)$ if $y(0) = 4$.
Find its response $y_b(t)$ to $x(t) = \cos(3t)u(t)$.

```
h=ilt('tf',2,[1 2])                    % Symbolic impulse response
h1=sysresp1('s',2,[1 2])               % Same, using sysresp1
w=sysresp1('s',2,[1 2],[1 0 0 0 0])    % Symbolic step response
ya =sysresp1('s',2,[1 2],[2 3 0 0 0],4) % Response to exponential input
yb =sysresp1('s',2,[1 2],[1 0 0 3 0])  % Response to sinusoid
t=0:0.01:5;                            % Create time index
plot(t,eval(h1))                       % Evaluate impulse response and plot
plot(t,eval(w))                        % Evaluate step response and plot
```

EXAMPLE 21.18 (Numerical Simulation of Differential Equations (Ch. 4))
The ADSP routine `ctsim` returns the numerical response of differential equations. Its syntax is

```
Y = ctsim(RHS,LHS,X,t,IC,TY)
```

Here, `RHS` and `LHS` are the coefficients of the right-hand side and left-hand side of the differential equation (or the numerator and denominator array of its transfer function), `X` is the input as a *string* function (in single quotes), `t` is a regularly spaced time vector, `IC` is the initial condition vector, and the string `TY` is the numerical integration method (`'rk1','rk2','rk4'` for Runge-Kutta methods of order 1, 2, 4, and `'sim'` for Simpson's rule). The solution is returned in `Y` at the time values in `t`.

Let $y''(t) + 4y'(t) + 3y(t) = x'(t) + 2x(t)$, where $x(t) = \sin(10t),\ 0 \le t \le \pi$, and the initial conditions are $y(0) = 1,\ y'(0) = 2$. Obtain and plot its response up to 6 s using the second-order Runge-Kutta method with a time step of 0.05 s.

```
tr = 0:0.05:6;t=tr;              % Time array
IC = [1; 2];                     % Initial condition array
RHS=[1 2]; LHS=[1 4 3];          % Coefficient arrays of LHS and RHS
X='sin(10*t).*(t>=0 & t<=pi)';   % Input as string variable
y=ctsim(RHS,LHS,X,tr,IC,'rk2');  % Response using second-order RK
plot(t,eval(X),tr,y)             % Plot input and output
```

EXAMPLE 21.19 (Response of Discrete Systems in Symbolic Form (Ch. 5))
The ADSP routine `sysresp1` finds the response of discrete systems in *symbolic* form. Its syntax is

```
[yt,yzs,yzi] = sysresp1('z', RHS(or Num), LHS(or Den), X, IC)
```

The first argument is `'z'` in single quotes (for a DT system). The arguments `RHS` and `LHS` are coefficient arrays of *equal length* of the difference equation up to the *highest* difference present on either side (*with missing coefficients set to zero*) or the numerator and denominator of the transfer function $H(z)$.
The argument `X=[K a p omega theta n0]` is a six-element array corresponding to the general input

$$x[n] = K(a)^{n-n_0}(n-n_0)^p \cos[\omega(n-n_0) + \theta(\text{radians})]$$

For M inputs, `X` is matrix with M rows.
The last argument `IC` is the initial condition array `IC=[y[-1], y[-2],..., y[-N]]`.

PGR2026-00033
Krisp Technologies EX1024 Page 789

The output arguments `[yt, yzs, yzi]` are the total, zero-state, and zero-input response.
The *impulse response* is found by *omitting* the last two arguments (`X` and `IC`).
Note: Since the output of `sysresp1` is a symbolic (or *string*) expression, it must first be evaluated (using the `eval` command) for the selected range of n before it can be plotted.

Response by Recursion
The ADSP routine `dtsim` allows the recursive solution of difference equations. Its syntax is

```
yd=dtsim(RHS, LHS, XN, IC)
```

Here `XN` is the input array $x[n]$ over an index range n.
The response `yd` has the same length as `XN`. The argument `IC` can be omitted for zero initial conditions.
For *zero initial conditions*, the MATLAB command `filter(RHS,LHS,XN)` yields identical results.

Consider the second-order system $y[n]-0.64y[n-2] = x[n]$ with $x[n] = 20(0.8)^n u[n]$, and the initial conditions $y[-1] = 5, y[-2] = 25$. Find and plot the total, zero-state, and impulse response over $0 \le n \le 30$, using *both* `dtsim` and `sysresp1`.

```
RHS=[1 0 0];LHS=[1 0 -0.64];          % RHS and LHS of difference eq
X=[20 0.8 0 0 0 0]; IC=[5;25];        % Input and initial conditions
[yt,yzs]=sysresp1('z',RHS,LHS,X,IC)   % Symbolic response
h=sysresp1('z',RHS,LHS)               % Symbolic impulse response
n=0:30;                               % DT index to evaluate sysresp1 output
dtplot(n,eval(yt),'o')                % Evaluate response and plot
dtplot(n,eval(h),'o')                 % Evaluate and plot impulse response
%
xin=20*(0.8 .^ n);                    % Input array for DTSIM
ytd=dtsim(RHS,LHS,xin,IC);            % Numerical response
yzsd=dtsim(RHS,LHS,xin);              % Numerical zero-state response
hd=dtsim(RHS,LHS,udelta(n));          % Numerical impulse response
dtplot(n,ytd,'o')                     % Plot numerical response
dtplot(n,hd,'o')                      % Plot numerical impulse response
```

EXAMPLE 21.20 (Echo and Reverb (Ch. 5))
An echo system has a simple difference equation of the form $y[n] = x[n] + \alpha x[n-N]$. Reverberations are due to multiple echoes (from the walls and other structures in a concert hall, for example). An IIR filter of the form $y[n] = x[n] + \beta y[n-M]$ simulates such an effect.

Load the audio signal `chirp.mat` into MATLAB, truncate it to 1600 samples to create signal `z`, and play back. Filter `z` through the echo system $y[n] = x[n] + \alpha x[n-N]$, where $\alpha = 0.9$ and the *integer index* N corresponds to a delay of about 0.1 s. Assume a sampling frequency of 8192 Hz. Filter `z` through the reverb system $y[n] = x[n] + \beta y[n-M]$, where $\beta = 0.4$ and $M = 800$. Listen to the filtered signal `zf2` and comment on the sound. Experiment with other values of α, β, and M and comment on your observations.

```
load chirp;                                % Load chirp signal
z=y(1:1600);sound(z)                       % Truncate and play back
d=0.1;N=fix(8192*d);                       % Index N corresponding to delay d
z1=filter([1 zeros(1,N) 0.9],1,z);         % Pass chirp through echo filter
sound(z1)                                  % Listen to filtered signal
z2=filter([1],[1 zeros(1,800) -0.4],z); % Pass chirp through reverb filter
sound(z2)                                  % Listen to filtered signal
```

PGR2026-00033
Krisp Technologies EX1024 Page 790

EXAMPLE 21.21 (Smoothing Effects of a Moving Average Filter (Ch. 5))
Consider the 20-point *moving average* FIR filter $y[n] = \frac{1}{20}\{x[n] + x[n-1] + \ldots + x[n-19]\}$. It is also called a *smoothing filter* because it tends to smooth out the rapid variations in a signal. To confirm its smoothing property, try the following:

1. Generate 200 samples of a 1-Hz sine wave sampled at 40 Hz.
2. Add some noise to generate a noisy signal.
3. Filter the noisy signal through the 20-point filter.
4. Plot each signal to display the effects of noise and smoothing.

```
n=0:199;                                % Create a DT index
x=sin(2*pi*n*0.025); x=x(:);            % 1-Hz sine sampled at 40 Hz
xn=x+0.5*randist(x,'uni');              % Add (uniform) random noise
a=[1 zeros(1,19)]; b=0.05*ones(1,20);   % Coefficient arrays of MA filter
y=filter(b, a, xn);                     % Response using filter
plot(n,x);                              % Plot input
plot(n,xn,n,y);                         % Noisy input and smoothed output
```

EXAMPLE 21.22 (Continuous Convolution (Ch. 6))
The ADSP routine `convnum` approximates the convolution of analog signals. To approximate the convolution of $x(t) = 3\cos(0.5\pi t)$ and $h(t) = 2\,\mathrm{tri}(t-1)$, use `convnum` as follows:

```
ts=0.01;                    % Choose sampling interval (time step)
tx=-1:ts:1;                 % Choose a time axis for signal x
th=0:ts:2;                  % Time axis for signal h
x=3*cos(pi*tx/2);           % Generate the signal x
h=2*tri(th-1);              % Generate the signal h
y=convnum(x,h,ts);          % Compute the convolution y
ty=-1:ts:3;                 % Create a time axis for y
plot(tx, x, th, h, ty, y);  % Overplot x, h, and y
```

Comment: You will get the same results from the MATLAB routine `conv` if you use `y=ts*conv(x,h);`. The routines `convnum` or `conv` do not tell us the start or duration of convolution. To plot results, we must know that the convolution `y` extends over $-1 \le t \le 3$.

EXAMPLE 21.23 (Discrete Convolution and Convolution Indices (Ch. 7))
An input $x[n] = \{2,\ \underset{\Downarrow}{-1},\ 3\}$ is applied to an FIR filter with impulse response $h[n] = \{1,\ 2,\ \underset{\Downarrow}{2},\ 3\}$. Find the response $y[n]$ and plot all signals.

The starting indices are $n = -1$ for $x[n]$, $n = -2$ for $h[n]$, and thus $n = -3$ for $y[n]$.

```
x=[2 -1 3];h=[1 2 2 3];                 % Signals x[n] and h[n]
y=conv(x,h)                             % Display convolution y[n]
nx=-1:1;nh=-2:1;                        % Indices of x[n] and y[n]
ns=nx(1)+nh(1);                         % Starting index of y[n]
ne=nx(length(nx))+nh(length(nh));       % Ending index of y[n]
ny=ns:ne;                               % Indices of y[n]
dtplot(nx,x)                            % Plot signal x
dtplot(nh,h)                            % Plot signal h
dtplot(ny,y)                            % Plot signal y
```

PGR2026-00033
Krisp Technologies EX1024 Page 791

EXAMPLE 21.24 (Discrete Convolution (Ch. 7))
The impulse response of a digital filter is described by $h[n] = (0.4)^n u[n]$. Evaluate and plot the response $y[n]$ of this filter to the input $x[n] = (0.8)^n u[n]$ over the range $0 \le n \le 20$.

If $x[n]$ and $h[n]$ are defined over $0 \le n \le 20$, their convolution will extend over $0 \le n \le 40$ but match the analytical results only over $0 \le n \le 20$.

```
nn=0:20;                          % DT index
x=(0.8 .^ nn);h=(0.4 .^ nn);      % Signals x[n] and h[n]
y=conv(x,h);                      % Convolution y[n]
n=0:40;                           % DT index for convolution
dtplot(n,y)                       % Plot convolution y
ya=(0.4 .^ n).*((2 .^(n+1))-1);   % Exact result
plot(n,ya-y)                      % Plot error
```

EXAMPLE 21.25 (System Response to Discrete-Time Sinusoids (Ch. 7))
We claim that the response of LTI systems to a sinusoidal input is also a sinusoid at the input frequency. Justify this statement using an input $x[n] = \cos(0.2\pi n)$ to a digital filter whose impulse response is given by $h[n] = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

The period of $x[n] = \cos(0.2\pi n)$ is $N = 10$. To find and plot the convolution result, we generate $x[n]$ for, say, $0 \le n \le 50$, and use the following MATLAB commands:

```
n=0:50;                   % Create a DT index
x=cos(0.2*pi*n);h=1:8;    % Signals x and h
y=conv(x,h);              % Response using convolution
y=y(1:length(n));         % Truncate to length of x
dtplot(n,x)               % Plot input
dtplot(n,y)               % Plot response
```

Comment: From the plots, you should observe the periodic nature of the response *except for start-up transients*. You should also be able to discern the period of both the input and output as $N = 10$.

EXAMPLE 21.26 (Convolution and Filtering (Ch. 7))
The convolution response is equivalent to the *zero-state* response obtained from the difference equation that describes the impulse response $h[n]$ and may be found using the MATLAB routine `filter`. Note that `filter` returns a result that equals the input length while the results of `conv` are longer. Over the input length, however, the results of `conv` and `filter` are identical.
The difference equation describing the digital filter of Example 21.25 may be written as

$$y[n] = x[n] + 2x[n-1] + \ldots + 8x[n-7]$$

Use this to find the response to $x[n] = \cos(0.2\pi n)$ and compare results with Example 21.25.

```
n=0:50;x=cos(0.2*pi*n);h=1:8;    % Signals x and h
a=[1 zeros(1,7)];                % Filter coefficients
yf=filter(h,a,x);                % Response using filter
yc=conv(x,h);                    % yc is longer than x and h
yt=yc(1:length(n));              % So, truncate and find error
dtplot(n,yt)                     % Plot convolution response
dtplot(n,yf)                     % Plot filter response
err=yf-yt;plot(n,err)            % Find and plot error
```

PGR2026-00033
Krisp Technologies EX1024 Page 792

EXAMPLE 21.27 (Periodic Convolution (Ch. 7))
The ADSP routine `convp` can be used to implement periodic convolution using wraparound, or the circulant matrix, or the FFT (or DFT).
Let $x_p[n] = \{1, 2, -1, 0, 2, 3\}$ and $h_p[n] = \{2, 1, 0, -1, -2, -3\}$ describe one-period segments of two periodic signals. We generate one period of the periodic convolution $y_p[n] = x_p[n] \circledast h_p[n]$, using each method.

```
x=[1 2 -1 0 2 3];       % Signal x
h=2:-1:-3;              % Signal h
ypw=convp(x,h,'w')      % Periodic convolution using wraparound
ypc=convp(x,h,'c')      % Periodic convolution using circulant matrix
ypf=convp(x,h,'f')      % Periodic convolution using FFT
```

The FFT results may show an imaginary part which is zero (to within machine roundoff).

EXAMPLE 21.28 (Normalization and Periodic Convolution (Ch. 7))
Let $x[n] = \{1,\ 2,\ -1,\ 0,\ 2,\ 3\}$ and $h[n] = \{2,\ 1,\ 0,\ -1,\ -2,\ -3\}$ describe one-period segments of two periodic signals.

1. Find their periodic convolution $y_1[n]$, using one period of $x[n]$ and $h[n]$.
2. Find their periodic convolution $y_5[n]$, using five periods of $x[n]$ and $h[n]$.
3. How is the period of $y_5[n]$ related to that of $y_1[n]$?
4. How are the convolution values of $y_5[n]$ and $y_1[n]$ related?

```
x=[1 2 -1 0 2 3];h=2:-1:-3;       % One period of x and h
y1=convp(x,h);                    % Periodic convolution (one period)
dtplot(0:length(y1)-1,y1)         % Plot y1
x5=[x x x x x];h5=[h h h h h];    % Five periods of x and h
y5=convp(x5,h5);                  % Periodic convolution (five periods)
dtplot(0:length(y5)-1,y5)         % Plot y5
scale=y5(1:length(y1))./y1        % Find ratio over one period
```

Comment: Note that $y_5[n]$ is periodic with the same period as $y_1[n]$ but five times longer and five times larger. Normalizing (dividing) $y_1[n]$ by its length $N = 6$, or $y_5[n]$ by its length $5N = 30$, we obtain *identical* results over one period.

EXAMPLE 21.29 (Regular Convolution from Periodic Convolution (Ch. 7))
Let $x[n] = \{1,\ 2,\ -1,\ 0,\ 2\}$ and $h[n] = \{2,\ 1,\ 0,\ -1,\ -2,\ -3\}$. Find their regular convolution, using zero-padding and periodic convolution.

```
x=[1 2 -1 0 2]; h=2:-1:-3;     % Signals x and h
Nx=length(x);Nh=length(h);     % Signal lengths
Ny=Nx+Nh-1;                    % Convolution length
xz=zeros(1,Ny);hz=xz;          % Initialize zero-padded signals
xz(1:Nx)=x;hz(1:Nh)=h;         % Zero-padded x and h
N=length(xz);                  % Length of xz and hz
y=convp(xz,hz)                 % Convolution using periodic conv
yr=conv(x,h)                   % Regular convolution
err=y-yr                       % Display error (should be zero)
```

PGR2026-00033
Krisp Technologies EX1024 Page 793

EXAMPLE 21.30 (Deconvolution (Ch. 7))
If $y[n] = x[n] \star h[n]$ and the response $y[n]$ and input $x[n]$ are known, we can identify the system impulse response $h[n]$ using *deconvolution*. Just as convolution is equivalent to polynomial multiplication, deconvolution is equivalent to polynomial division.
Given $y[n] = \{3, 9, 17, 21, 19, 13, 6, 2\}$ and $x[n] = \{3, 3, 2, 2\}$, identify $h[n]$.

```
x=[3 3 2 2];                % Input x
y=[3 9 17 21 19 13 6 2];    % Output y
[h,r]=deconv(y,x)           % Impulse response h and remainder r
```

Comment: For perfect identification, the remainder (after division) should be zero.

EXAMPLE 21.31 (Autocorrelation and Cross-Correlation (Ch. 7))
Consider the sequences $x[n] = n,\ 0 \le n \le 8$, and $h[n] = n,\ 0 \le n \le 3$.

1. Evaluate and plot $r_{xx}[n]$ and $r_{hh}[n]$ and find where they attain their maximum.
2. Evaluate and plot $r_{xh}[n]$ and $r_{hx}[n]$.
3. Evaluate and plot the correlation of $h[n]$ and $h[n-4]$ and find where it attains a maximum.

```
x=0:8;h=0:3;                    % Signals x and h
nx=0:8;nh=0:3;                  % DT index for x and h
rxx=conv(x,fliplr(x));          % Autocorrelation rxx
rhh=conv(h,fliplr(h));          % Autocorrelation rhh
nhh=-3:3;nxx=-8:8;              % DT index for rhh and rxx
[mx, i1] = find(rxx==max(rxx)) % Maximum rxx and its MATLAB index
[mh, i2] = find(rhh==max(rhh)) % Maximum rhh and its MATLAB index
nxxm=nxx(i1),nhhm=nhh(i2)       % DT index of max rxx and rhh
rxh=conv(x,fliplr(h));          % Cross-correlation rxh
rhx=conv(h,fliplr(x));          % Cross-correlation rhx
nxh=-3:8;nhx=-8:3;              % DT index for rxh and rhx
dtplot(nxx,rxx)                 % Plot rxx (similar statements for others)
h1=[0 0 0 0 h];n=0:7;           % Shifted h and its DT index
rhd=conv(h,fliplr(h1));         % Correlation of h and h1
nhd=-7:3;                       % DT index for rhh1
rhd2=conv(h1,fliplr(h));        % Correlation of h1 and h
nhd2=-3:7;                      % DT index for rhh2
dtplot(nhd,rhd)                 % Plot rhd
dtplot(nhd2,rhd2)               % Plot rhd2
```

Comment: Note that $r_{hd}[n] = h[n] \star \star h[n-4] = r_{hh}[n-4]$. The delay is indicated by the location of the peak in $r_{hd}[n]$. This result is the basis for target ranging by radar (or ultrasound). We send a sampled signal $x[n]$ that is reflected by the target (an airplane, say). Ideally, the received signal is just a delayed version $x[n-n_0]$. The delay $n_0 t_s = 2Rc$, where R is the target distance and c is the propagation velocity. We correlate $x[n]$ and $x[n-n_0]$ and locate the peak. This location gives n_0. From n_0, we estimate the target range R.

EXAMPLE 21.32 (Detecting Signals Buried in Noise (Ch. 7))
Correlation is an effective method of detecting signals buried in noise. Noise is essentially uncorrelated. This means that if we correlate a noisy signal with itself, the correlation will be due only to the signal (if present) and exhibit a sharp peak at $n = 0$.

PGR2026-00033
Krisp Technologies EX1024 Page 794

The spectrum of the correlated noisy signal can be used to estimate the frequency of the signal. Generate two noisy signals by adding noise to a 20-Hz sinusoid sampled at $t_s = 0.01$ s for 2 s.

1. Verify the presence of the signal by correlating the two noisy signals.
2. Estimate the frequency of the signal from the FFT spectrum of the correlation.

```
ts=0.01;t=0:ts:2;N1=length(t);  % Set up time array
x=sin(2*20*pi*t);               % Signal x
z1=2*randist(x);x1=x+z1;        % Noisy signal z1
z2=2*randist(x);x2=x+z2;        % Noisy signal z2
cc=conv(x1,fliplr(x2));         % Correlation of x1 and x2
N=length(cc);                   % Length of correlation
tc=-(N1-1):N1-1;                % Correlation index
plot(tc,cc);                    % Plot correlation
f=(0:N-1)/N/ts;                 % FFT frequency axis f
ccspec=abs(fft(cc))/N;          % FFT (magnitude) of correlation
plot(f,ccspec);                 % Plot spectrum of correlation
```

EXAMPLE 21.33 (Filtering a Chirp Signal (Ch. 7))
Digital filters can be designed and used to extract a selected range of frequencies in a signal. Frequency-domain filtering constitutes a very versatile application of signal processing.
An N-sample *chirp* signal $x[n]$ whose digital frequency *varies linearly* from F_0 to F_1 is described by

$$x[n] = \cos\left[2\pi\left(F_0 n + \frac{F_1 - F_0}{2N} n^2\right)\right], \qquad n = 0, 1, \ldots N-1$$

1. Generate 800 samples of a chirp signal `x` whose digital frequency varies from $F = 0.05$ to $F = 0.4$. Observe how the frequency of `x` varies linearly with time, using the ADSP command `timefreq(x)`.
2. Consider two filters described by

   $h_1[m] = \delta[m] - 0.4\cos(0.4m\pi)\text{sinc}(0.2m), \quad -80 \le m \le 80 \qquad h_2[m] = 0.3\text{sinc}(0.3m), \quad -80 \le m \le 80$

   Generate the response of the two filters to `xc` and use `timefreq` and `psdwelch` to plot their spectra. Identify each filter type and determine the range of digital frequencies it passes or blocks.

```
F0=0.05;F1=0.4;                        % Digital frequencies
N=800;n=0:N-1;                         % Number of points and DT index
xc=cos(2*pi*(n*F0+(F1-F0)*n.*n/2/N)); % Chirp signal
plot(n,xc)                             % Plot chirp
timefreq(xc);                          % Look at its spectrum
psdwelch(xc);                          % Look at its spectrum

n1=-80:80;                             % Index for filter coefficients
ha=0.4*cos(0.4*n1*pi).*sinc(0.2*n1);
h=udelta(n1)-ha;                       % Generate filter coefficients
yf=filter(h,1,xc);plot(n,yf)           % Filter the chirp and plot
timefreq(yf);                          % Spectrum of filtered chirp
psdwelch(yf);                          % Spectrum of filtered chirp
```

PGR2026-00033
Krisp Technologies EX1024 Page 795

```
h1=0.3*sinc(0.3*n1);                % Generate filter coefficients
yf1=filter(h1,1,xc);plot(n,yf1)     % Filter the chirp and plot
timefreq(yf1);                      % Spectrum of filtered chirp
psdwelch(yf1);                      % Spectrum of filtered chirp
```

Comment: On machines with sound capability, one could use the natural sampling rate to generate and listen to the chirp signal and the filter response for audible differences, using `sound`.

EXAMPLE 21.34 (Fourier Series (Ch. 8))
Consider a periodic signal $x_p(t)$, with period $T = 20$ ms, whose Fourier series coefficients are $X[0] = 1$ and $X[k] = j/k\pi,\ k \neq 0$. Sketch its two-sided magnitude and phase spectrum and its one-period reconstruction to 5 harmonics and 30 harmonics.

```
n=1:5;                              % Index to compute X[k]
j=sqrt(-1);xk=j./(n*pi);            % FS coefficients
k=[-fliplr(n) 0 n];                 % Harmonic index
Xk=[conj(fliplr(xk)) 1 xk];         % FS coefficients for two-sided spectrum
dtplot(k,abs(Xk))                   % Plot magnitude spectrum
dtplot(k,angle(Xk))                 % Plot phase in radians
T=0.2;t=0:0.001:T;                  % Time period and time array for one period
xp=0;                               % Initialize reconstruction
for m=1:length(k);                  % Loop for reconstruction; prompt disappears
xp=xp+Xk(m)*exp(j*2*pi*k(m)*t/T);   % Add harmonics
end                                 % End of loop; prompt reappears
xp=real(xp);                        % Force real result
plot(t,xp)                          % Plot reconstruction
```

EXAMPLE 21.35 (Fourier Series (Ch. 8))
A Fourier series describes a periodic signal in terms of a combination of harmonic signals (sines and cosines or complex exponentials) at multiples of its fundamental frequency f_0. The Fourier series coefficients and spectra may be generated using the ADSP routine `fseries`. In its simplest form, its syntax is

```
v=fseries(x, T, k)
```

Here, `T` is the width of one period, `x` is an odd-length array of n values over one period $(0,\ T)$ or a string function of t defined over $(0,\ T)$ (for example, `x='tri(t-1)'`), and `k` is an array of harmonics for intermediate reconstructions. The built-in default is `k=[1 3 5 32]`. This plots and stores results to 32 harmonics.
The matrix `v` is a 32×7 matrix whose seven columns are

$$[\text{Harmonic k},\quad a_k,\quad b_k,\quad c_k,\quad \theta_k,\quad \text{harmonic power},\quad \text{cumulative power}]$$

Consider a periodic signal $x_p(t)$ with $T = 2$ and one period over $0 \leq t \leq 2$, described by $x(t) = 3e^{-2t}$.
Find its Fourier series coefficients and compare one period of $x_p(t)$ with its reconstruction to 7 harmonics and 30 harmonics.

```
x='3*exp(-2*t)';            % Expression for signal x(t)
v=fseries(x, 2);            % Fourier series results to k=32 are in v
t=0:0.02:2;                 % Time array for one period
ak=v(:,2);bk=v(:,3);        % Trig coefficients from matrix v
sm=ak(1);                   % Start reconstruction with dc value
```

PGR2026-00033
Krisp Technologies EX1024 Page 796

```
for k=1:30;                    % Loop for reconstruction; prompt disappears
sm=sm+ak(k+1)*cos(k*pi*t)+bk(k+1)*sin(k*pi*t);
if k==7,sm7=sm;end             % Save intermediate result if k=7
end                            % End of loop; prompt reappears
plot(t,eval(x),t,sm7,t,sm)     % Overplot results
```

EXAMPLE 21.36 (System Response to Periodic Inputs (Ch. 8))
A periodic square wave $v_{\text{in}}(t)$ with a time period $T = 1$, a peak value of 4, and a duty ratio of 0.5 is applied to an RC lowpass filter with a time constant of τ. Find the Fourier series coefficients of the filter output and plot to 30 harmonics for $\tau = 0.1$ and $\tau = 1$. What output would you expect to see for *very large* τ?

```
vin='4*(t>=0 & t<0.5)+2*(t==0.5)';  % Create vin (as a string function)
fs=fseries(vin,1,30);               % Find FS coeffs to k=30 harmonics
akin=fs(:,2);bkin=fs(:,3);          % Extract a(k) and b(k)
hfilt='1./(1+j*k*2*pi*tau)';        % Filter transfer function
k=(0:30)';                          % Make k a column vector for later use
tau=0.1;hf1=eval(hfilt);            % Magnitude and phase of H(f) for tau=0.1
tau=1;hf2=eval(hfilt);              % Same for tau=1
j=sqrt(-1);
cout1=hf1.*(akin-j*bkin);           % Filter output coefficients c(k)
cout2=hf2.*(akin-j*bkin);
t=0:0.01:2;
sm1=cout1(1);sm2=cout2(1);          % Start reconstruction with dc value
for k=1:30
sm1=sm1+abs(cout1(k+1))*cos(k*2*pi*t+angle(cout1(k+1)));  % Reconstructions
sm2=sm2+abs(cout2(k+1))*cos(k*2*pi*t+angle(cout2(k+1)));
end
plot(t,sm1,t,sm2)                   % Overplot
```

Comment: For large τ, the response should be a triangular wave because the RC circuit would behave (very nearly) as an integrator.

EXAMPLE 21.37 (Analog Steady-State Response in Symbolic Form (Ch. 8))
The response to sinusoidal inputs may be found in symbolic form using the ADSP routine `ssresp`. In this routine, the sinusoidal input $K\cos(\omega t + \theta)$ is described by the three-element array $X = [K,\ \omega,\ \theta]$, where θ is in radians and ω is in rad/s. For a dc input, $\omega = 0$ and $\theta = 0$. For M inputs, `X` is a matrix with M rows.

Find the steady-state response of a system whose transfer function is $H(s) = 2/(s+2)$, if the input is given by $v_{\text{in}}(t) = 5 + 2\sin(t) + 4\cos(3t - 60^\circ)$.

```
num=2;                                % Numerator of H(s)
den=[1 2];                            % Denominator of TF
vin=[5 0 0; 2 1 -0.5*pi; 4 3 -pi/3];  % Input matrix
yss = ssresp('s', num, den, vin)      % Symbolic steady-state response
t=0:0.05:10;plot(t,eval(yss))         % Evaluate and plot response
```

PGR2026-00033
Krisp Technologies EX1024 Page 797

EXAMPLE 21.38 (The Laplace Transform (Ch. 11))
The ADSP routine `ltr` generates the Laplace transform of a *causal* signal described by

$$x(t) = Ke^{-at}t^p \cos(wt + r)u(t)$$

Its syntax is

```
[n,d]=ztr(x);
```

Here, `x` is the five-element array `x=[K a p w r]`, where `w` is in radians/sample and `r` is in radians.
For a combination of N DT signals, `x` is a matrix with N rows.
The ADSP routine `ilt` returns the inverse Laplace transform in *symbolic form*. Its syntax is

```
x=ilt('tf', n, d, ty)
```

If `ty=0` (the default), `n` and `d` are the numerator and denominator of $X(s)$ in *descending powers* of s.
If `ty=1`, the vector `d` contains the denominator *roots* (if the denominator is given in factored form).

Find the Laplace transform of $x(t) = (4e^{-t} + 3te^{-2t})u(t)$. Then, find the inverse transform of the result to recover $x(t)$. Also find the inverse Laplace transform of

$$H_1(s) = \frac{2s+4}{s^2+2s+2} \qquad H_2(s) = \frac{2s+4}{(s+1)(s^2+2s+2)} \qquad H_3(s) = \frac{2s+4}{(s+1)(s+3)}$$

```
X=[4, 1, 0, 0, 0; 3, 2, 1, 0, 0];          % Input matrix
[n,d]=ltr(X)                               % Find Laplace transform
x=ilt('tf', n, d,0)                        % Inverse transform (symbolic form)
h1=ilt('tf',[2 4],[1 2 2],0)               % Inverse transform of H1(s)
h2=ilt('tf',[2 4],conv([1 1],[1 2 2]),0) % Inverse transform of H2(s)
h3=ilt('tf',[2 4],[-1;-3],1)               % Inverse transform of H3(s)
```

Comment: Note that for $H_2(s)$, we used convolution to find the denominator coefficients, and for $H_3(s)$, we used the denominator roots directly.

EXAMPLE 21.39 (Symbolic System Response by Laplace Transforms (Ch. 11))
The ADSP routine `sysresp2` finds the response of analog systems in *symbolic* form. Its syntax is

```
[yt,yzs,yzi] = sysresp2('s', Num, Den, NX, DX, IC)
```

The first argument is `'s'` in single quotes (for an analog system). The arguments `RHS` and `LHS` are coefficient arrays of the numerator and denominator of the transfer function $H(s)$ and `NX` and `DX` are the coefficient arrays of the numerator and denominator of the Laplace transform of the input. For M inputs, `NX` and `DX` are matrices with M rows.
The last argument `IC` is the initial condition array `IC=[y(0), y'(0),..]`.
The output arguments `[yt, yzs, yzi]` are the total, zero-state, and zero-input response.
The *impulse response* is found by *omitting* the last three arguments (`NX`, `DX`, and `IC`).

Note: Since the output of `sysresp1` is a symbolic (or *string*) expression, it must first be evaluated (using the `eval` command) for a selected range of t before it can be plotted.

Consider an analog system whose transfer function is $H(s) = \dfrac{2}{s+2}$.

1. Find and plot its impulse response $h(t)$ and step response $w(t)$.
2. Find its response $y(t)$ to $x(t) = 2e^{-3t}u(t)$ if $y(0) = 4$.
3. Find its response to $x(t) = \cos(3t)u(t)$ and to $x(t) = \cos(3t)$.
4. Plot its frequency response $H(f)$ and Bode plot.
5. Find its rise time, 5% settling time, and half-power frequency.

PGR2026-00033
Krisp Technologies EX1024 Page 798

```
h=ilt('tf',2,[1 2])                    % Symbolic impulse response
h1=sysresp2('s',2,[1 2])               % Same, using sysresp2
w=sysresp2('s',2,[1 2],1,[1 0])        % Symbolic step response
y1 =sysresp2('s',2,[1 2],2,[1 3],4)    % Response for (b)
y2 =sysresp2('s',2,[1 2],[1 0],[1 0 9]) % Response for (c)
y3 =ssresp('s',2,[1 2],[1 3 0])        % Response for (c)
t=0:0.01:5;                            % Create time index
plot(t,eval(h1))                       % Evaluate expression and plot
bodelin(2,[1 2]);                      % Bode plot
trbw(1,[1 2],5)                        % Rise time, 5% settling time, bandwidth
```

EXAMPLE 21.40 (Frequency Response of Analog Systems (Ch. 12))

The MATLAB routine `freqs` allows us to evaluate the frequency response of analog systems described by differential equations or transfer functions. Plot the magnitude and phase of the following filters over the range $0 \le f \le 2$ Hz and identify the filter type from the magnitude spectrum.

$$H_A(f) = \frac{2}{2 + j2\pi f} \qquad H_B(s) = \frac{2s}{s+2} \qquad H_C(s) = \frac{s}{s^2 + s + 1}$$

```
f=0:0.01:2; w=2*pi*f;                  % Frequency array
ha=2 ./ (2+j*w);                       % Evaluate H(f) for (a)
hb=freqs([2 0],[1 2],w);               % Evaluate H(f) for (b)
hc=freqs([1 0],[1  1 1],w);            % Evaluate H(f) for (c)
plot(f,abs(ha));                       % Plot magnitude for (a)
plot(f,180*angle(ha)/pi);              % And phase (degrees)
                                       % Repeat for others
```

EXAMPLE 21.41 (Analog Filter Design (Ch. 13))

The ADSP routine `afd` allows the design of classical filters from attenuation specifications (in dB) and frequency specifications (in Hz). The transfer function is designed with *unit peak gain*. It *exactly meets* attenuation specifications at the passband edge(s) and meets/exceeds specifications elsewhere. Its syntax is

```
[na,da,np, dp]=afd(name, type, A, fp, fs, plot);
```

The *string* argument `name` may be `'bw'`, `'c1'`, `'c2'` or `'el'`.
The *string* argument `type` may be `'lp'`, `'hp'`, `'bp'` or `'bs'`.
The argument `A = [Ap As]` is a two-element array of the passband and stopband attenuation.
The argument `fp = [fp1 fp2]` specifies the passband edge(s) in Hz (each has one element for `'lp'` or `'hp'`).
The argument `fs = [fs1 fs2]` similarly specifies the stopband edge(s) in Hz.
If `plot='p'`, the results are plotted, but `plot='o'` (the default) omits plots.
The arguments `na, da` return the transfer function numerator and denominator.
The arguments `np, dp` return the lowpass prototype transfer function numerator and denominator.
Design the following:

1. A Chebyshev LPF with a 2-dB passband of 20 Hz and a 40-dB stopband edge at 40 Hz
2. A Butterworth HPF with a 2-dB passband edge at 200 Hz and a 40-dB stopband edge at 150 Hz
3. An elliptic BPF with a 3-dB passband of [5, 7] kHz and 50-dB band edges at [3, 10] kHz.

```
[n1, d1]=afd('c1', 'lp', [2 40], 20, 40)              % No plots displayed
[n2, d2]=afd('bw', 'hp', [2 40], 200, 150, 'p');      % Plots displayed
[n3, d3]=afd('el', 'bp', [3 50], [5000 7000], [3000 10000])
```

PGR2026-00033
Krisp Technologies EX1024 Page 799

EXAMPLE 21.42 (Analog Filter Design (Ch. 13))
We require a bandpass filter with $A_p = 1$ dB, $A_s = 34$ dB, a passband from 200 Hz to 300 Hz, and stopband edges at 100 Hz and 400 Hz. Design a Butterworth, Chebyshev I, Chebyshev II, and an elliptic filter to meet these specifications. Which filter requires the smallest order? Largest order?

```
Ap=1;As=34;fp=[200 300];fs=[100 400];        % Design specifications
[nbw, dbw]=afd('bw','bp',[Ap As],fp,fs);     % Butterworth filter
[nc1, dc1]=afd('c1','bp',[Ap As],fp,fs);     % Chebyshev I filter
[nc2, dc2]=afd('c2','bp',[Ap As],fp,fs);     % Chebyshev II filter
[nel, del]=afd('el','bp',[Ap As],fp,fs);     % Elliptic Filter
f=1:600;w=2*pi*f;hbw=abs(freqs(nbw,dbw,w));hc1=abs(freqs(nc1,dc1,w));
hc2=abs(freqs(nc2,dc2,w));hel=abs(freqs(nel,del,w));
plot(f,[hbw;hc1;hc2;hel]),                   % Plot linear magnitude
plot(f,20*log10([hbw;hc1;hc2;hel]))          % And dB magnitude
lbw=length(dbw)-1;                           % Butterworth order
lc1=length(dc1)-1;                           % Chebyshev I order
lc2=length(dc2)-1;                           % Chebyshev II order
lel=length(del)-1;                           % Elliptic order
disp('    BW     C1    C2    EL')
disp([lbw lc1 lc2 lel])                      % Display orders
```

Comment: You should observe that the elliptic filter requires the smallest order

EXAMPLE 21.43 (Analog Filter Design from a Specified Order (Ch. 13))
The ADSP command `[n,d]=lpp(name, n, [Ap As])` returns the transfer function of an nth-order lowpass prototype. To transform this to the required form, we use the command `[nn, dd]=lp2af(ty, n, d, fc, f0)`. It requires one band edge (`fc`) if `ty='lp'` or `ty='hp'`, and the bandwidth (`fc`) and center frequency (`f0`) if `ty='bp'` or `ty='bs'`. All frequencies are in rad/s (not Hz). Design the following:

1. A third-order Chebyshev II LPF with a 2-dB passband of 20 Hz and $A_s \geq 40$ dB
2. An eighth-order Chebyshev BPF with a 3-dB bandwidth of 1 kHz and a center frequency of 5 kHz

```
[n1, d1]=lpp('c2', 3, [2 40])                        % Lowpass prototype; order=3
[n2, d2]=lp2af('lp', n1, d1, 20*2*pi)                % Lowpass filter
[n3, d3]=lpp('c1', 4, 3);                            % Lowpass prototype; order=4
[n4, d4]=lp2af('bp', n3, d3, 4000*2*pi, 1000*2*pi)   % BPF; order=8
```

EXAMPLE 21.44 (Prototype Filter Poles and Delay (Ch. 13))
The ADSP routines `buttpole` and `chebpole` return (and plot) the pole locations of lowpass prototypes. The ADSP routine `delay` returns the delay $D(\nu)$ of Butterworth, Chebyshev I, Chebyshev II, and Bessel filters in two (numerator and denominator) arrays. Find the pole locations and plot the group delay of a third-order Chebyshev I lowpass prototype with a passband ripple of 1 dB.

```
pc=chebpole(3, 1)          % Poles
[P, Q]=delay('c1', 3, 1)   % Numerator and denominator of delay
nu=0:0.01:2;               % Frequency array
num=polyval(P,nu);         % Evaluate numerator
den=polyval(Q,nu);         % Evaluate denominator
del=num ./ den;            % Compute delay
plot(nu,del)               % Plot
```

PGR2026-00033
Krisp Technologies EX1024 Page 800

EXAMPLE 21.45 (A Multi-Band Analog Filter (Ch. 13))
Design a multi-band analog filter with the following specifications:
Passband 1: From dc to 20 Hz
Peak gain = 0 dB
Maximum passband attenuation = 1 dB (from peak)
Minimum attenuation at 40 Hz = 45 dB (from peak)
No ripples allowed in the passband or stopband
Passband 2: From 150 Hz to 200 Hz
Peak gain = 12 dB
Maximum passband ripple = 1 dB
Minimum attenuation at 100 Hz and 300 Hz = 50 dB (from peak)
Ripples permitted in the passband but not in the stopband

We use two calls to `afd` to design the filters, adjust the gain, and combine the two stages in parallel.

```
A1=[1 45];fp1=20; fs1=40;                % Filter 1 is Butterworth LP
A2=[1 50];fp2=[150 200];fs2=[100 300];   % Filter 2 is Chebyshev BP
[n1, d1]=afd('bw','lp',A1,fp1,fs1);      % Design LPF
[n2, d2]=afd('c1','bp',A2,fp2,fs2);      % Design BPF
G=10^(12/20);  n2=G*n2;                  % Multiply n2 by gain (12 dB=3.9811)
np=conv(n1,d2)+conv(n2,d1);              % Numerator of parallel combination
dp=conv(d1,d2);                          % Denominator of parallel combination
tfplot('s',np,dp,[0 400]);               % Plot frequency response (0-400 Hz)
```

EXAMPLE 21.46 (Signal Reconstruction and Resampling (Ch. 14))
Consider the signal $x(t) = \sin(2\pi ft)$. It is sampled at $S = 4$ Hz for one period, to obtain the sampled signal $x[n] = \{0, 1, 0, -1\}$. Sketch the signals that result from

1. A 16-fold interpolation of $x[n]$ using sinc interpolation.
2. Resampling $x[n]$ at 3.2 times the original sampling rate.

Since $3.2 = \frac{16}{5}$, the resampled signal of the second part can be generated by decimating the 16-fold interpolated signal of the first part by 5.

```
n=0:3;xn=[0, 1, 0, -1];          % DT index and signal samples
yi=interpol(xn,'sinc',16);       % Sinc interpolated signal
N=length(yi);ni=(0:N-1)/16;      % Index for interpolated signal
dtplot(ni,yi);                   % Plot interpolated signal
M=5;yd=yi(1:M:N);                % Decimate yi by 5
nd=ni(1:M:N);                    % DT array for resampled yd
dtplot(nd,yd)                    % Plot resampled signal
```

Comment: In theory, perfect recovery is possible for a band-limited signal using sinc interpolation. The results do not show this because signal values outside the one-period range are assumed to be zero during the interpolation.

PGR2026-00033
Krisp Technologies EX1024 Page 801

EXAMPLE 21.47 (Signal Quantization and SNR (Ch. 14))
The signal $x(t) = \sin(2\pi ft)$ is sampled with $t_s = 0.01$ s, and 200 samples are obtained, starting at $t = 0$. The sampled signal $x[n]$ is *rounded*, using a 3-bit quantizer to obtain the quantized signal $x_Q[n]$.

1. Plot the analog signal $x(t)$ and the quantized signal $x_Q[n]$.
2. Plot the histogram of the error signal $x[n] - x_Q[n]$.
3. Find the quantization SNR.

```
N=200;t=(0:N-1)*0.01;                % Time array
x=sin(2*pi*t);                       % DT sampled signal
d=max(x)-min(x);                     % Dynamic range
b=3;L=2^b;                           % Number of levels
s=d/L;                               % Step size
y=s*round(x/s);                      % Round; use y=s*floor(x/s) to truncate
plot(t,x);hold on,                   % Overplot CT signal samples
dtplot(t,y);hold off                 % And quantized signal
e=x-y;hist(e,L);                     % Error signal and its histogram
snra=10*log10(sum(x.*x)/sum(e.*e))   % Quantization SNR
```

Comment: You should observe that the error signal (`e` in the MATLAB code) is more or less uniformly distributed by examining its histogram.

EXAMPLE 21.48 (Frequency Response of Discrete Systems (Ch. 15))
The MATLAB routine `freqz` allows us to evaluate the frequency response of a discrete system described by a difference equation or transfer function. Plot the DTFT magnitude and phase of the following filters over the digital frequency range $0 \le F \le 0.5$ and identify the filter type from the magnitude spectrum.

1. $H_A(F) = \dfrac{1}{1 - 0.9e^{-j2\pi F}}$
2. $H_B(F) = \dfrac{1}{1 + 0.9e^{-j2\pi F}}$
3. $h_C[n] = \delta[n] - (0.9)^n u[n]$
4. $y[n] + 0.4y[n-1] + 0.8y[n-2] = x[n]$
5. $y[n] + 0.1y[n-1] + 0.8y[n-2] = x[n] + 0.1x[n-1] + x[n-2]$
6. $h[n] = \delta[n] - \delta[n-8]$

Note: For filter 3, we use the transfer function (after simplification)

$$H_3(F) = 1 - \frac{1}{1 - 0.9e^{-j2\pi F}} = \frac{-0.9e^{-j2\pi F}}{1 - 0.9e^{-j2\pi F}}$$

For filter 6, we use the difference equation $y[n] = x[n] - x[n-8]$.

```
F=(0:399)/800; W=2*pi*F;                          % Frequency array
h1=freqz([1 0],[1 -0.9],W);                       % Evaluate H(F) for filter 1
h2=freqz([1 0],[1 0.9],W);                        % For filter 2
h3=freqz([0 -0.9],[1 -0.9],W);                    % For filter 3
h4=freqz([1 0 0],[1 0.4 0.8],W);                  % For filter 4
h5=freqz([1 0.1 1],[1 0.1 0.8],W);                % For filter 5
h6=freqz([1 zeros(1,7) -1],[1 zeros(1,8)],W);     % For filter 6
plot(F,abs(h1));                                  % Plot magnitude for filter 1
plot(F,angle(h1));                                % And phase (radians)
                                                  % Repeat for others
```

PGR2026-00033
Krisp Technologies EX1024 Page 802

EXAMPLE 21.49 (The Time-Shift Property of the DTFT (Ch. 15))

Let $x[n] = \{3,\ 3,\ 3,\ 2,\ \underset{\Downarrow}{2},\ 2,\ 1,\ 1,\ 1\}$, $-4 \le n \le 4$. Obtain the DTFT of $x[n]$ and $y[n] = x[n-D]$, $D = 4$ over $0 \le F < 1$.

```
n=0:8; x=[3 3 3 2 2 2 1 1 1];
F=(0:199)/200; W=2*pi*F;
X=freqz(x,[zeros(1,4) 1 zeros(1,4)],W);
Y=freqz(x,[1 zeros(1,8)],W);
plot(F,abs(X));
plot(F,abs(Y));
plot(F,180*angle(X)/pi);
plot(F,180*angle(Y)/pi);
phi=unwrap(angle(Y)-angle(X));plot(F,phi)
```

Comment: The MATLAB routine `unwrap` returns the *unwrapped* phase with phase jumps of 2π removed. You should observe that $|X(F)|$ and $|Y(F)|$ are identical and the *phase difference* $\phi(F) = \angle Y(F) - \angle X(F)$ is *linear* with frequency.

EXAMPLE 21.50 (The DTFT, Periodicity, Convolution, and Multiplication (Ch. 15))

Let $x[n] = \text{tri}((n-4)/4)$, $0 \le n \le 8$. Obtain the DTFT of $x[n]$, $g[n] = x[n] \star x[n]$, and $h[n] = x^2[n]$ over $-2 \le F \le 1.99$

```
n=0:8; x=tri((n-4)/4);
F=-2:0.01:1.99; W=2*pi*F;
X=freqz(x,[1 zeros(1,8)],W);
G=freqz(conv(x,x),[1 zeros(1,16)],W);
H=freqz(x.*x,[1 zeros(1,8)],W);
plot(F,abs(X))
plot(F,abs(X).^2,F,abs(G),':')
Yp=convp(X,X)/length(X);
plot(F,abs(Yp),F,abs(H),':')
```

Comment: You should observe that $|X(F)|$ is periodic with unit period, $|X(F)|^2$ equals $|G(F)|$, and $|Yp(F)|$ equals $|H(F)|$.

EXAMPLE 21.51 (The Modulation Property of the DTFT (Ch. 15))

Let $h[n] = 1$, $0 \le n \le 19$. Obtain the DTFT of $x[n], h_2[n] = x[n]\cos(0.3n\pi)$, and $h_3[n] = x[n]\cos(2.3n\pi)$ over $-0.5 \le F \le 0.5$. How should $H(F)$ and $H_2(F)$ be related? How should $H_2(F)$ and $H_3(F)$ be related? Do your plots confirm this?

```
n=0:19;x=1+0*n;
F=-0.5:0.005:0.5; W=2*pi*F;
H1=freqz([x],[1 zeros(1,19)],W);
H2=freqz([x.*cos(0.3*n*pi)],[1 zeros(1,19)],W);
H3=freqz([x.*cos(2.3*n*pi)],[1 zeros(1,19)],W);
plot(F,abs(H1))
plot(F,abs(H2))
plot(F,abs(H3))
```

Comment: You should observe that $H_1(F)$ is the spectrum of the modulated signal, but the spectra $H_2(F)$ and $H_3(F)$ are identical.

PGR2026-00033
Krisp Technologies EX1024 Page 803

EXAMPLE 21.52 (The Gibbs Effect and Periodic Convolution (Ch. 15))
An ideal LPF with cutoff frequency F_C has the impulse response $h[n] = 2F_C \operatorname{sinc}(2F_C n)$. Symmetric truncation (about $n = 0$) leads to the *Gibbs effect* (overshoot and oscillations in the frequency response). Truncation of $h[n]$ to $|n| \leq N$ is equivalent to multiplying $h[n]$ by a rectangular *window* $w_D[n] = \text{rect}(n/2N)$. In the frequency domain, it is equivalent to the *periodic* convolution of $H(F)$ and $W_D(F)$ (the Dirichlet kernel). To eliminate overshoot and reduce the oscillations, we choose tapered windows whose DTFT sidelobes decay much faster than those of the rectangular window.
Consider an ideal filter with $F_C = 0.2$ whose impulse response $h[n]$ is truncated over $-20 \leq n \leq 20$.

1. Show that its periodic convolution with the Dirichlet kernel results in the Gibbs effect.
2. Show that the Gibbs effect is absent in its periodic convolution with the Fejer kernel.

```
n=-20:20;                          % Create a DT index
Fc=0.2; h=2*Fc*sinc(2*Fc*n);       % Create impulse response
a=[zeros(1,20) 1 zeros(1,20)];     % Denominator array of DTFT
f=(-200:199)/400;w=2*pi*f;         % Frequency array (F=-0.5 to F=-0.5)
HI=(abs(f)<=0.2);                  % Ideal filter spectrum,  FC=0.2
HT=abs(freqz(h,a,w));              % Spectrum of truncated h[n]
f1=f+0.5;                          % Frequency array (F=0 to F=1)
x=sinc(41*f1)./sinc(f1);Wd=41*x;   % Dirichlet kernel
x=sinc(20*f1)./sinc(f1);Wf=20*x.*x; % Fejer kernel
HD=convp(HI,Wd)/400;               % Periodic conv with Dirichlet kernel
HF=convp(HI,Wf)/400;               % Periodic conv with Fejer kernel
plot(f,HI)                         % Spectrum of ideal filter
plot(f,HD)                         % And windowed by Dirichlet
plot(f,HF)                         % And windowed by Fejer
plot(f,HT)                         % Spectrum of truncated h[n]
```

Comment: Since $h[n]$ is generated over $-0.5 \leq F < 0.5$, one period of the periodic convolution will cover the same range only if the Dirichlet and Fejer kernels are generated over $0 \leq F < 1$. Also note that the periodic convolution is normalized (divided) by the number of samples.

EXAMPLE 21.53 (The Effect of Smoothing Windows (Ch. 15))
Any window (other than the rectangular) does a good job of smoothing out the effects of abrupt truncation. Demonstrate this by plotting the DTFT spectrum of the windowed impulse response $h[n] = 2F_C \operatorname{sinc}(2nF_C)$, with $F_C = 0.2$, using the following windows:

1. The *Bartlett* window (or Fejer kernel, $w_B[n] = 1 - \frac{|n|}{N}, \quad |n| \leq N$)
2. The *von Hann* (*Hanning*) window ($w_V[n] = 0.5 + 0.5\cos(\pi n/N), \quad |n| \leq N$)
3. The *Hamming* window ($w_H[n] = 0.54 + 0.46\cos(\pi n/N), \quad |n| \leq N$)

```
n=-20:20;                          % Create a DT index
Fc=0.2; h=2*Fc*sinc(2*Fc*n);       % Create impulse response
a=[zeros(1,20) 1 zeros(1,20)];     % Denominator array of DTFT
f=(-200:199)/400;w=2*pi*f;         % Frequency array
HI=(abs(f)<=0.2);                  % Ideal filter spectrum,  FC=0.2
wb=1-abs(n)/20;                    % Bartlett window
wv=0.5+0.5*cos(pi*n/20);           % von Hann (Hanning) window
wh=0.54+0.46*cos(pi*n/20);         % Hamming window
```

PGR2026-00033
Krisp Technologies EX1024 Page 804

```
HB=abs(freqz(h.*wb,a,w));       % Spectrum of Bartlett windowed h[n]
HV=abs(freqz(h.*wv,a,w));       % Spectrum of von Hann windowed h[n]
HH=abs(freqz(h.*wh,a,w));       % Spectrum of Hamming windowed h[n]
plot(f,HI,f,HB,f,HV,f,HH)       % Overplot
```

EXAMPLE 21.54 (Sampling and Aliasing (Ch. 16))
Let $x(t) = \cos(200\pi t) - \cos(400\pi t)$.
Obtain samples of $x(t)$ over the interval $[0, 0.1]$ s at sampling rates of 500 Hz, 300 Hz, and 70 Hz. For each sampling rate S, derive an expression for the sampled signal $x[n]$. Compute f_1 and f_2 in the recovered signal $x_r(t) = \cos(2\pi f_1 t) - \cos(2\pi f_2 t)$. Overplot $x(t)$ and $x_r(t)$ vs. t and $x[n]$ vs. nt_s.

```
T=0.1;t=0:0.05/200:T;
x=cos(200*pi*t)-cos(400*pi*t);
S=500;tn1=0:1/S:T;
xn1=cos(200*pi*tn1)-cos(400*pi*tn1);
f1=alias(100,S);f2=alias(200,S);
xr1=cos(2*pi*f1*t)-cos(2*pi*f2*t);
plot(t,x,t,xr1),hold on
dtplot(tn1,xn1,'o'),hold off

S=300;tn2=0:1/S:T;
xn2=cos(200*pi*tn2)-cos(400*pi*tn2);
f1=alias(100,S);f2=alias(200,S);
xr2=cos(2*pi*f1*t)-cos(2*pi*f2*t);
plot(t,x,t,xr2),hold on
dtplot(tn2,xn2,'o'),hold off

S=70;tn3=0:1/S:T;
xn3=cos(200*pi*tn3)-cos(400*pi*tn3);
f1=alias(100,S);f2=alias(200,S);
xr3=cos(2*pi*f1*t)-cos(2*pi*f2*t);
plot(t,x,t,xr3),hold on
dtplot(tn3,xn3,'o'),hold off
```

Comment: You should observe that values of $x[n]$ match $x(t)$ and $x_r(t)$ for each sampling rate, but $x(t)$ and $x_r(t)$ match only when there is no aliasing.

EXAMPLE 21.55 (Sampling a Chirp Signal (Ch. 16))
Consider the chirp signal $x(t) = \cos(\pi\alpha t^2)$. The derivative of its phase $\phi(t) = \pi\alpha t^2$ yields its *instantaneous* frequency:

$$\omega(t) = \frac{d\phi(t)}{dt} = 2\pi\alpha t \ \text{ rad/s} \qquad\qquad f(t) = \alpha t \ \text{ Hz}$$

Clearly, the instantaneous frequency varies linearly with time.
Let $\alpha = 2048$. Sample $x(t)$ over $[0, 2]$ s at $S = 8192$ Hz. Generate the instantaneous frequency array `fi` of $x(t)$ vs. t over $[0, 2]$ s. Has aliasing occurred? Plot the aliased frequency against the instantaneous frequency. Repeat for $\alpha = 8192$.

PGR2026-00033
Krisp Technologies EX1024 Page 805

```
ts=1/8192;t=0:ts:2;
a1=2048;xc1=cos(pi*a1*t.*t);
fi1=a1*t;fa1=alias(fi1,8192);
plot(fi1,fa1)
a2=8192;xc2=cos(pi*a2*t.*t);
fi2=a2*t;fa2=alias(fi2,8192);
plot(fi2,fa2)
```

Comment: On machines with sound capability, you could listen to the signals using `sound` and understand how the frequency changes with time.

EXAMPLE 21.56 (Spectrum Analysis of a Sinusoid (Ch. 16))
Let $x(t) = \cos(2\pi f_0 t)$, where $f_0 = 30$ Hz. Choose a sampling frequency $S = 128$ Hz and $N = 256$ to obtain the FFT of the windowed signal, using rectangular and Hamming windows, zero-padded to $N_1 = 1024$. Plot the *normalized* FFT magnitude of the windowed signals. Which windowed signal shows a narrower mainlobe? Which shows the smaller peak sidelobe level?

```
f0=30;S=128;N=256;N1=1024;
n=0:N-1; f=(0:N1-1)*S/N1;      % DT index and frequency axis (Hz)
x=cos(2*pi*f0*n/S);            % Signal
X=abs(fft(x,N1))/N1;           % Normalized FFT for rectangular (no) window
xa=x.*window('hamm',N);        % Hamming windowed signal
XA=abs(fft(xa,N1))/N1;         % Normalized FFT of Hamming windowed signal
plot(f,X)
```

EXAMPLE 21.57 (Frequency Resolution (Ch. 16))
Let $x(t) = \cos(2\pi f_0 t) + \cos(2\pi[f_0 + \Delta f]t)$, where $f_0 = 30$ Hz, sampled at $S = 128$ Hz. Use no windowing and a Hamming window to obtain the 1024-point FFT spectrum for $\Delta f = 5$ Hz, 2 Hz, and 1 Hz if the window length is $N = 256$. Repeat for $N = 512$. For a given window (or signal) length N, which window resolves a smaller Δf?

```
df=5; N=256; N1=1024; f0=30; S=128;n=0:N-1;
x=cos(2*pi*f0*n/S)+cos(2*pi*(f0+df)*n/S);
X=abs(fft(x,N1))/N1;f=(0:N1-1)*S/N1;plot(f,X)
xb=x.*window('vonh',N);
XB=abs(fft(xb,N1))/N1;plot(f,XB)
% Repeat for df=2 and df=1

df=5; N=512; N1=1024; f0=30; S=128;n=0:N-1;
x2=cos(2*pi*f0*n/S)+cos(2*pi*(f0+df)*n/S);
X2=abs(fft(x2,N1))/N1;f=(0:N1-1)*S/N1;plot(f,X2)
xb2=x2.*window('vonh',N);
XB2=abs(fft(xb2,N1))/N1;plot(f,XB2)
% Repeat for df=2 and df=1
```

Comment: By experimenting with various values of N and zero-padded length, you should confirm that the rectangular window has the better resolution but larger sidelobes, but we cannot resolve a smaller Δf by increasing the zero-padded length N_1 *alone*.

PGR2026-00033
Krisp Technologies EX1024 Page 806

EXAMPLE 21.58 (Convolution by the FFT (Ch. 16))
The FFT can be used to perform periodic and regular convolution. For periodic convolution, we multiply the FFT of $x_p[n]$ and $h_p[n]$ pointwise and take the inverse FFT of the product. For regular convolution, we first zero-pad $x[n]$ and $h[n]$ to a length that equals the length of regular convolution before taking the FFT. Use the FFT to find

1. The periodic convolution of $x_p[n] = \{1,\ 2,\ -1,\ 0,\ 2,\ 3\}$ and $h_p[n] = \{2,\ 1,\ 0,\ -1,\ -2,\ -3\}$.
2. The regular convolution of $x[n] = \{1,\ 2,\ -1,\ 0,\ 2\}$ and $h[n] = \{2,\ 1,\ 0,\ -1,\ -2,\ -3\}$.

```
xp=[1 2 -1 0 2 3]; hp=2:-1:-3;  % One period of xp and hp
N=length(xp);                   % Length of xp and hp
YP=fft(hp).*fft(xp);            % Product of FFT of x and h
yp=ifft(YP);yp=real(yp)         % Real result of periodic convolution
Nr=N+N-1;                       % Length of regular convolution
z=zeros(1,Nr-N);                % Padding zeros
xz=[xp z];hz=[hp z];            % Zero-padded x and h
Y=fft(xz).*fft(hz);
y=real(ifft(Y))                 % Display regular conv
```

Comment: We force a real result because, for real sequences, `ifft` may show an imaginary part that is zero (to within machine roundoff).

EXAMPLE 21.59 (The z-Transform (Ch. 17))
The ADSP routine `ztr` generates the z-transform of a *causal* signal described by

$$x[n] = K(a)^{n-n_0}(n-n_0)^p \cos[(n-n_0)\Omega + \phi]u[n-n_0]$$

where $\Omega = 2\pi F$ is the digital frequency in radians/sample, and ϕ the phase in radians. Its syntax is

```
[n, d] = ztr(x)
```

The input argument `x` is the six-element array `x=[K a p w phi n0]`.
For a combination of N signals, `x` is a matrix with N rows.
The ADSP routine `izt` returns the inverse z-transform in *symbolic form*. Its syntax is

```
x=izt('tf', n, d, ty)
```

If `ty=0` (the default), n and d are the numerator and denominator of $X(z)$ in *descending powers* of z.
If `ty=1`, the vector d contains the denominator *roots* (if the denominator is given in factored form).

Find the z-transform of $x[n] = [4(0.5)^n + 3n(2)^n]u[n]$. Then, find the inverse transform of the result to recover $x(t)$. Also find the inverse Laplace transform of

$$H_1(z) = \frac{2z+4}{z^2+z+1} \qquad H_2(z) = \frac{2z+4}{(z+0.5)(z^2+z+1)} \qquad H_3(z) = \frac{2z+4}{(z-0.5)(z+0.5)}$$

```
X=[4, 0.5, 0, 0, 0, 0; 3, 2, 1, 0, 0, 0];  % Input matrix
[n, d]=ztr(X);                             % Find z-transform
x=izt('tf', n, d,0)                        % Inverse transform (symbolic form)
h1=izt('tf',[2 4],[1 1 1],0)               % Inverse transform of H1(z)
h2=izt('tf',[2 4],conv([1 0.5],[1 1 1]),0) % Inverse transform of H2(z)
h3=izt('tf',[2 4],[0.5;-0.5],1)            % Inverse transform of H3(z)
```

Comment: For $H_2(z)$ we used convolution to find the denominator coefficients. For $H_3(z)$ we used the denominator roots directly.

PGR2026-00033
Krisp Technologies EX1024 Page 807

EXAMPLE 21.60 (System Analysis by the z-Transform (Ch. 17))
The ADSP routine `sysresp2` finds the response (in symbolic form) of systems to arbitrary inputs. It requires the transfer function $H(z) = N(z)/D(z)$ and the input $X(z) = P(z)/Q(z)$ and has the syntax

```
[yt, yzs, yzi]=sysresp2('z', N, D, P, Q, IC)
```

Here, `N`, `D`, `P`, and `Q` are coefficient arrays in descending powers of z, and `IC` is a vector of initial conditions. For M inputs, `P` and `Q` are matrices with M rows.
If the input and `IC` are omitted, it returns the system impulse response (the inverse z-transform of $H(z)$).

Find the response of $H(z) = (3z + 3)/(z^2 + 5z + 6)$ to the input $x[n] = 2(0.5)^n u[n]$ subject to the initial conditions $y[-1] = 2$ and $y[-2] = 4$.

```
[Xn, Xd]=ztr([2 0.5 0 0 0 0]);                  % z-transform of input
N=[3, 3]; D=[1, 5, 6]; IC=[2; 4];               % H(z) and initial conditions
[yt, yzs, yzi]=sysresp2('z', N, D, Xn, Xd, IC)  % Complete response
h=sysresp2('z', N, D)                           % Impulse response
```

EXAMPLE 21.61 (Steady-State Response in Symbolic Form (Ch. 17))
The ADSP routine `ssresp` finds the symbolic response to discrete sinusoids. Its syntax is

```
y=sspesp('z',n,d,X)
```

Here `n` and `d` are the numerator and denominator arrays of $H(z)$ and `X` is a three-element array `X=[A, w r]` corresponding to the input $A\cos(\omega t + r(\text{radians}))$. For M inputs, `X` is a matrix with M rows.

Find the steady-state response of a system whose transfer function is $H(z) = 2z/(z - 0.4)$ to the input $v_{\text{in}}[n] = 5 + 2\sin(0.5n\pi) + 4\cos(0.25n\pi - 60°)$.

```
num=[2 0];                                  % Numerator of H(z)
den=[1 -0.4];                               % Denominator of H(z)
vin=[5 0 0; 2 pi/2 -pi/2; 4 pi/4 -pi/3];    % Input matrix
yss=ssresp('z', num, den, vin)              % Symbolic steady-state response
n=0:40;dtplot(n,eval(yss),'o')              % Evaluate and plot response
```

EXAMPLE 21.62 (Inverse Systems and Echo Cancellation (Ch. 18))
A signal $s(t)$ is corrupted by echoes. The corrupted signal $y(t) = s(t) + 0.9s(t - \tau) + 0.8s(t - 2\tau)$, with $\tau = 93.75$ ms, is sampled at $S = 8192$ Hz. The signal $y[n]$ is supplied on disk as `echosig.mat`. Load this signal and listen to it. Can you make out what is being said? Find the transfer function $H(z)$ of an echo-canceling filter that could recover $s[n]$ from $y[n]$. Filter $y[n]$ through this filter and listen to the filtered signal $s[n]$. Can you make out what is being said? Do you agree with what is being said?

The sampled signal may be written as $y[n] = x[n] + 0.8x[n - N] + 0.8x[n - 2N]$, where the integer $N = S\tau$ corresponds to a delay of τ. The inverse filter is $y[n] + 0.9y[n - N] + 0.8y[n - 2N] = x[n]$.

```
load echosig                                        % Load echo signal
sound(echosig)                                      % Listen to echo signal
N1=fix(0.09375*8192);N2=2*N1;                       % Filter delays
n=[1 zeros(1,N2)];                                  % Numerator of inverse
d=[1 zeros(1,N1-1) 0.9 zeros(1,N2-N1-1) 0.8];       % Denominator of inverse
noecho=filter(n,d,echosig);                         % Filtered signal
sound(noecho)                                       % Listen to filtered signal
```

Enjoyed this exercise? Please thank Prof. Tim Schulz (`http://www.ee.mtu.edu/faculty/schulz`).

PGR2026-00033
Krisp Technologies EX1024 Page 808

EXAMPLE 21.63 (IIR Filter Design (Ch. 19))
The ADSP routine `dfdiir` allows the design of IIR filters. It uses a direct A2D transformation for bilinear design. For all others, it first converts the analog LPP to a digital LPP followed by a D2D transformation. Its syntax is

```
[n, d]=dfdiir(name, type, map, A, SF, fp, fs, plot)
```

The *string* argument `name` may be `'bw'`, `'c1'`, `'c2'` or `'el'`.
The *string* argument `type` may be `'lp'`, `'hp'`, `'bp'` or `'bs'`.
The *string* argument `map` is the mapping: `'impulse'`, `'step'`, `'ramp'` (invariance), `'mat0'`, `'mat1'`, `'mat2'` (matched z-transform and its modifications as described in the text), and `'trap'` (bilinear).
The argument `A = [Ap As]` is a two-element array of the passband and stopband attenuation.
The argument `SF` is the sampling frequency in Hz.
The argument `fp = [fp1 fp2]` specifies the passband edge(s) in Hz (each has one element for `'lp'` or `'hp'`).
The argument `fs = fs1 fs2]` similarly specifies the stopband edge(s) in Hz.
The *string* argument `plot` is a plotting option (`'p'` plots the results; `'o'`, the default, omits plots).
The arguments `n` and `d` return the numerator and denominator of $H(z)$.

A bandpass IIR filter is to meet the following specifications:
Passband = [1.5, 3.5] kHz, stopband edges = [1, 4] kHz, $A_p = 1$ dB, $A_s = 20$ dB, and = 10 kHz.
Design a Butterworth filter, using the bilinear transformation; a Chebyshev I filter, using impulse invariance; and an Elliptic filter, using the matched z-transform

```
A=[1 20];fp=[1500 3500];
fs=[1000 4000];S=10000;
[n1, d1] = dfdiir('bw','bp','bili',A,S,fp,fs);   % Butterworth, bilinear
[n2, d2] = dfdiir('c1','bp','impu',A,S,fp,fs);   % Chebyshev, impulse invariant
[n3, d3] = dfdiir('el','bp','matc',A,S,fp,fs);   % Elliptic, matched z-transform
F=0:0.005:0.5;w=2*pi*F; f=10000*F;                % Frequency array
h1=freqz(n1,d1,w);h2=freqz(n2,d2,w);h3=freqz(n3,d3,w);
plot(f,abs([h1;h2;h3]))                          % Overplot spectra
```

EXAMPLE 21.64 (Mappings (Ch. 19))
The ADSP routines `s2zinvar`, `s2zmatch`, and `s2zni` convert an analog filter $H(s)$ to a digital filter $H(z)$, using mappings for invariance, matched z-transform or numerical algorithms. The general syntax is

```
[Nz,Dz,G]=s2zxxx(Ns, Ds, SF, ty, FM)
```

For `s2zinvar`, `ty` may be `'i'` or `'s'` or `'r'` (for impulse, step, or ramp invariance).
For `s2zmatch`, `ty` may be `'0'` or `'1'` or `'2'` (for no modifications or modified forms).
For `s2zni`, `ty` may be `'bili'` (bilinear), `'back'` (backward), etc.
`FM` is an optional match frequency at which the gains of $H(s)$ and $H(z)$ are matched.
`G` is a dc gain-matching constant such that $GH(z) = H(s)$ at dc ($z = 1$ and $s = 0$).

Convert $H(s) = \dfrac{1}{s+1}$, using impulse invariance, matched z-transform, and bilinear mappings at $S = 2$ Hz, and overplot the frequency response.

```
S=2;                                % Sampling frequency
[N1, D1]=s2zmatch(1,[1 1],S,'0',0); % Matched z-transform
```

PGR2026-00033
Krisp Technologies EX1024 Page 809

```
[N2, D2]=s2zinvar(1,[1 1],S,'i',0);  % Impulse invariance
[N3, D3]=s2zni(1,[1 1],S,'bili',0);  % Bilinear transform
F=0:0.005:0.5;w=2*pi*F;              % Frequency array
H1=freqz(N1,D1,w);                   % Frequency response of H1(z)
H2=freqz(N2,D2,w);                   % Frequency response of H2(z)
H3=freqz(N3,D3,w);                   % Frequency response of H3(z)
HA=freqs([0 1],[1 1],w*S);           % Frequency response of H(s)
f=F*S;plot(f,abs([H1;H2;H3;HA]))     % Overplot against analog frequency
```

EXAMPLE 21.65 (IIR Filter Design from a Specified Order (Ch. 19))
We start with a lowpass prototype (LPP) of the given order (using the ADSP routine `lpp`), convert it to a digital LPP (using `s2zinvar`, `s2zmatch`, or `s2zni`), and apply D2D frequency transformations (using the ADSP routine `lp2iir`) to obtain the required IIR filter. For the bilinear mapping, `lp2iir` allows direct A2D transformation of the LPP. The general syntax for `lp2iir` is

```
[n, d]=lp2iir(ty, ty1, Nlpp, Dlpp, SF, F)
```

Here, `ty = 'lp', 'hp', 'bp', or 'bs'` and `ty1='a'` for A2D or `ty1='d'` for D2D transformation. The argument `F` contains the cutoff band edge(s).

Design a sixth-order Chebyshev BPF with 3-dB edges at [2, 4] kHz and $S = 10$ kHz, using the bilinear transformation. Repeat the design, using impulse invariance.

```
[n, d]=lpp('c1', 3, 3)                    % Analog lowpass prototype (order=3)
S=10000;F=[2000 4000];FN=F/S;              % Frequency data
[NB, DB]=lp2iir('bp', 'a', n, d, 1, FN)   % Direct design for bilinear
[N2, D2]=s2zinvar(n, d, 1, 'i', 0);       % Digital LPP by impulse invariance
[NI, DI]=lp2iir('bp', 'd', N2, D2, 1, FN) % Convert to required filter
```

EXAMPLE 21.66 (Stability of IIR Filters (Ch. 19))
Roundoff error or *coefficient quantization* may make a stable IIR filter unstable! Consider the analog filter

$$H(s) = \frac{s^2 + 2s + 0.75}{s^5 + 27.5s^4 + 261.5s^3 + 1039s^2 + 1668s + 864}$$

Convert it to a digital filter $H(z)$, using the *bilinear transformation* and a sampling frequency of 100 Hz. Create the filter $H_T(z)$ by truncating the *denominator* coefficients of $H(z)$ to seven significant digits. Are the filters $H(z)$ and $H_T(z)$ stable? Are they minimum-phase?

```
n=[1 2 0.75];d=[1 27.5 261.5 1039 1668 864]; % H(s)
[nz,dz]=s2zni(n,d,100,'trap')                % H(z)
dt=numdig(dz,7);                             % Denominator of HT(z)
F=0:0.001:0.1;w=2*pi*F;                      % Frequency array
H=freqz(nz,dz,w);                            % Response of H(z)
HT=freqz(nz,dt,w);                           % Response of HT(z)
plot(F,abs([H;HT]))                          % Overplot
abs([roots(nz) roots(dz) roots(dt)])         % Check for stability
```

Comment: The plots of the two digital filters should be identical. However, some pole locations of $H_T(z)$ are outside the unit circle. Check the numerator roots for minimum phase.

PGR2026-00033
Krisp Technologies EX1024 Page 810

EXAMPLE 21.67 (Window-Based FIR Filter Design (Ch. 20))
To design a lowpass filter of length N and cutoff frequency F_c , we simply set up the impulse response and window it. For other filter types, we first convert specifications to the LPP and use frequency transformations to convert the LPP to the required filter type.
Design the following FIR filters:

1. Highpass: cutoff frequency $F_C = 0.2,\ N = 15$, Blackman window
2. Bandpass: band edges = [20, 30] Hz, $S = 100$ Hz, $N = 20$, von Hann (Hanning) window

```
Fc=0.2; n=-7:7; lp1=2*Fc*sinc(2*n*Fc);           % LPF cutoff=Fc
hw=lp1 .* window('blac', 15);                    % Windowed response
hhp=udelta(n)-hw;                                % LP2HP transformation
S=100; fp=20;fs=30;                              % BP specifications
F0=0.5*(fp+fs)/S; Fc2=0.5*(fs-fp)/S;             % Find F0 and
n=-9.5:9.5; lp2=2*Fc*sinc(2*n*Fc2);              % LPF cutoff=Fc2
hw2=lp2 .* window('vonh', 20);hw2=hw2/sum(hw2);  % Windowed, normalized response
hbp2=2*cos(2*pi*F0*n) .* hw2;                    % LP2BP transformation
```

EXAMPLE 21.68 (Automated FIR Filter Design (Ch. 20))
Window-Based Filters:
The ADSP routine `firwind` allows automated design of window-based FIR filters. Its syntax is

```
h=firwind(type, A, SF, fp, fs, window)
```

The argument `A` is a two-element attenuation array $[A_p, A_s]$.
The argument `S` is the sampling frequency.
The arguments `fp` and `fs` are the band edge(s).
The output argument `h` returns $h[n]$, the impulse response.
Control returns to the user for adjustment of filter length and the LPP cutoff if required.

Half-Band Filters:
The ADSP routine `firhb` designs half-band FIR filters. Its syntax is

```
h=firhb(type, A, SF, fp, fs, window)
```

Control returns to the user to change the filter length if required. Half-band design requires the sampling frequency to be $S = 2(f_p + f_s)$ for `'lp'` and `'hp'` filters, and $S = 4f_0$ for `'bp'` and `'bs'` filters.

Optimal Filters:
The ADSP routine `firpm` designs optimal (equiripple) FIR filters. Its syntax is

```
h=firpm(type, A, SF, fp, fs)
```

Control returns to the user to change the filter length if required.

Design a bandpass FIR filter with 2-dB band edges at [10, 40] Hz, 40-dB band edges at [10, 40] Hz, and $S = 100$ Hz. Use window-based design (with a Hamming window), half-band design (with a Kaiser window), and optimal design.

```
A=[2 40];fp=[20 30];fs=[10 40];S=100;
h1=firwind('bp', A, S, fp, fs, 'hamm')   % Window based: Hamming window
h2=firhb('bp', A, S, fp, fs, 'kais')     % Half-band design: Kaiser window
h3=firpm('bp', A, S, fp, fs)             % Optimal design
```

PGR2026-00033
Krisp Technologies EX1024 Page 811

References

Books on Digital Signal Processing

A good source of current references is the Web site of the MathWorks, Inc. (makers of MATLAB):

`http://www.mathworks.com`

Many books cover more or less the same topics as this text and at about the same level. The following list (arranged alphabetically by author) is representative.

- Ambardar, A., and C. Borghesani, *Mastering DSP Concepts Using Matlab* (Prentice Hall, 1998).
- Antoniou, A., *Digital Filters* (McGraw-Hill, 1993).
- Burrus, C. S., et al., *Computer-Based Exercises for Digital Signal Processing* (Prentice Hall, 1994).
- Cadzow, J. A., *Foundations of Digital Signal Processing and Data Analysis* (Macmillan, 1987).
- Cunningham, E. P., *Digital Filtering* (Houghton Mifflin, 1992).
- DeFatta, D. J., J. G. Lucas, and W. S. Hodgkiss, *Digital Signal Processing* (Wiley, 1988).
- Ifeachor, E. C., and B. W. Jervis, *Digital Signal Processing* (Addison-Wesley, 1993).
- Ingle, V. K., and J. G. Proakis, *Digital Signal Processing Using Matlab* (PWS, 1996).
- Jackson, L. B., *Digital Filters and Signal Processing* (Kluwer, 1995).
- Johnson, J. R., *Introduction to Digital Signal Processing* (Prentice Hall, 1989).
- Kuc, R., *Introduction to Digital Signal Processing* (McGraw-Hill, 1988).
- Ludeman, L. C., *Fundamentals of Digital Signal Processing* (Wiley, 1986).
- Mitra, S. K., *Digital Signal Processing* (Wiley, 1998).
- Oppenheim, A. V., and R. W. Schafer, *Digital Signal Processing* (Prentice Hall, 1975).
- Oppenheim, A. V., and R. W. Schafer, *Discrete-Time Signal Processing* (Prentice Hall, 1989).
- Orfanidis, S. J., *Introduction to Signal Processing* (Prentice Hall, 1996).
- Porat, B., *A Course in Digital Signal Processing* (Wiley, 1997).
- Proakis, J. G., and D. G. Manolakis, *Digital Signal Processing* (Prentice Hall, 1996).
- Rabiner, L. R., and B. Gold, *Theory and Application of Digital Signal Processing* (Prentice Hall, 1974).
- Steiglitz, K., *A Digital Signal Processing Primer* (Addison-Wesley, 1996).
- Strum, R. D., and D. E. Kirk, *First Principles of Discrete Systems and Digital Signal Processing* (Addison-Wesley, 1988).

PGR2026-00033
Krisp Technologies EX1024 Page 812

Books on Signals and Systems

Textbooks on signals and systems also cover similar topics. Here is a list of recent texts.

- Carlson, G. E., *Signal and Linear System Analysis* (Houghton Mifflin, 1992).
- Chen, C. T., *System and Signal Analysis* (Saunders, 1989).
- Gabel, R. A., and R. A. Roberts, *Signals and Linear Systems* (Wiley, 1987).
- Haykin, S., and B. Van Veen, *Signals and Systems* (Wiley, 1999).
- Kamen, E. W., and B. S. Heck, *Fundamentals of Signals and Systems* (Prentice Hall, 1997).
- Kwakernaak, H., and R. Sivan, *Modern Signals and Systems* (Prentice Hall, 1991).
- Lathi, B. P., *Signals and Systems* (Berkeley-Cambridge, 1987).
- McGillem, C. D., and G. R. Cooper, *Continuous and Discrete Signal and System Analysis* (Saunders, 1991).
- Oppenheim, A. V., et al., *Signals and Systems* (Prentice Hall, 1997).
- Phillips, C. L., and J. M. Parr, *Signals, Systems, and Transforms* (Prentice Hall, 1995).
- Siebert, W. McC., *Circuits, Signals and Systems* (MIT Press/McGraw-Hill, 1986).
- Soliman, S. S., and M. D. Srinath, *Continuous and Discrete Signals and Systems* (Prentice Hall, 1990).
- Strum, R. D., and D. E. Kirk, *Contemporary Linear Systems using* MATLAB (PWS, 1994).
- Taylor, F. J., *Principles of Signals and Systems* (McGraw-Hill, 1994).
- Ziemer, R. E., W. H. Tranter, and D. R. Fannin, *Signals and Systems: Continuous and Discrete* (Macmillan, 1993).

References on Special Topics

The following list contains more exhaustive discussions of specific topics.

Impulse Functions, Convolution, and Fourier Transforms

- Bracewell, R. N., *The Fourier Transform and Its Applications* (McGraw-Hill, 1978).
- Gaskill, J. D., *Linear Systems, Fourier Transforms and Optics* (Wiley, 1978).

Modulation and Wideband FM

- Hambley, A. R., *An Introduction to Communication Systems* (Computer Science Press, 1990).
- Fante, R. L., *Signal Analysis and Estimation* (Wiley, 1988).

Window Functions

- Geckinli, N. C., and D. Yavuz, *Discrete Fourier Transformation and Its Applications to Power Spectra Estimation* (Elsevier, 1983).
- Harris, F. J., "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform," *Proceedings of the IEEE*, vol. 66, no. 1, pp. 51–83, Jan. 1978.
- Nuttall, A. H., "Some Windows with Very Good Sidelobe Behavior," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 29, no. 1, pp. 84–91, Feb. 1981 (containing corrections to the paper by Harris).

PGR2026-00033
Krisp Technologies EX1024 Page 813

Analog Filters

- Daniels, R. W., *Approximation Methods for Electronic Filter Design* (McGraw-Hill, 1974).
- Weinberg, L., *Network Analysis and Synthesis* (McGraw-Hill, 1962).

Digital Filters and Mappings

- Kunt, M., *Digital Signal Processing* (Artech House, 1986).
- Bose, N. K., *Digital Filters* (Elsevier, 1985).
- Hamming, R. W., *Digital Filters* (Prentice Hall,1983).
- Scheider, A. M., et al., "Higher Order s-to-z Mapping Functions and Their Application in Digitizing Continuous-Time Filters," *IEEE Proceedings*, vol. 79, no. 11, pp. 1661–1674, Nov. 1991.

Applications of DSP

- Elliott, D. F. (ed.), *Handbook of Digital Signal Processing* (Academic Press, 1987).
- A good resource for audio applications on the Internet is
  `http://www.harmonycentral.com/Effects/`

Mathematical Functions

- Abramowitz, M., and I. A. Stegun (eds.), *Handbook of Mathematical Functions* (Dover, 1964).

Numerical Methods

- Press, W. H., et al., *Numerical Recipes* (Cambridge University Press, 1986).

The Life of Fourier

- Herivel, J., *Joseph Fourier—the Man and the Physicist* (Oxford, 1975).
- Carslaw, H., *Introduction to the Theory of Fourier's Series and Integrals* (Macmillan, 1921).
- Of the many Fourier-related Web pages on the Internet, we recommend
  `http://www-groups.dcs.st-and.ac.uk/ history/Mathematicians/Fourier.html`
  `http://capella.dur.ac.uk/doug/fourier.html`

PGR2026-00033
Krisp Technologies EX1024 Page 814

Index

absolutely integrable signals, 9
absolutely summable signals, 40
adaptive filters, 752
adder, 101
additivity, 69, 96
aliased frequency, 52, 449
aliasing, 52, 448, 547
allpass filters, 510, 658
almost periodic signals, 17, 50
alternation theorem, 741
amplitude distortion, 276
amplitude modulation, 300ff
    DSBSC, 300
    SSB, 310
    standard, 301
amplitude scaling, 12
amplitude shift, 12
amplitude spectrum, 208, 251
analog filters, 398ff
analog signals, 1, 8ff
analog systems, 68ff
analog-to-digital converter (ADC), 465
analog-to-digital (A2D) transformations, 696
analytic signal, 323
angle modulation, 312
    narrowband, 315
    wideband, 316
anti-aliasing filter, 465, 468
anti-causal, 39, 114
anticipating system, 75
anti-imaging filter, 465, 469
anti-symmetric (odd) signals, 42
aperture time, 466
asymptotes, 371
asymptotic stability, 345
asynchronous detection, 305
attenuation, 398
autocorrelation, 157, 187
average value, 9, 40

band edges, 276
bandpass filter, 276
bandpass signals, 300, 451
bandwidth, 5, 285
bandwidth of FM, 317
band-limited interpolation, 457, 565
band-limited signal, 269, 447
bandstop filter, 276
baseband signal, 300
Bessel filters, 432ff
Bessel functions, 317, 437
Bessel polynomials, 435
BIBO stability, 85, 114, 144, 177, 604
bilinear transformation, 691ff
bit, 460
bit reversal, 573
Black-Tukey spectrum estimate, 567
Bode plots, 370ff
break frequency, 372
butterfly, 572
Butterworth filters, 403ff

canonic realization, 638
carrier signal, 300
Carson's rule, 318
cascaded systems, 143, 176, 640
causal signal, 8, 39
causal system, 75, 100
causal (one-sided) z-transform, 613
causality, 144, 177
center frequency, 399
    geometric, 400
central limit theorem, 57, 154, 281
central moments, 29

PGR2026-00033
Krisp Technologies EX1024 Page 815

central ordinate theorems, 269, 489, 537
characteristic equation, 76, 104
Chebyshev filters, 413ff
Chebyshev II filters, 421ff
Chebyshev polynomials, 120, 412, 421
Chebyshev rational function, 427, 429
Chebyshev theorem, 413
chirp signals, 35, 64
chorusing, 662
circulant matrix, 182
circular convolution, 180
closed-loop transfer function, 383
coefficient quantization, 470, 706
coherent detection, 303
coherent gain, 556
comb filters, 512, 655
commensurate, 17
common period, 17, 40
compensating filter, 455
compensating probes, 395
compensators, 382
complex envelope, 323
complex frequency, 330, 592
conjugate reciprocal symmetry, 644
conjugate symmetry, 198, 251, 484, 537
connections, 151, 183, 250, 358, 483, 501, 542, 624
continuous-time signals, 8ff
continuous-time systems, 68ff
convergence of DTFT, 485
convergence factor, 358, 624
convergence of Fourier series, 232
convergence of Fourier transform, 271, 358
convergence of Laplace transform, 331, 358
convergence of z-transform, 624
convergence rate, 218
conversion time, 466
convolution, 130ff, 169ff
  graphical (by ranges), 138
  periodic (cyclic or circular), 150, 180
  properties, 135
convolution sum, 169
corner frequency, 372
correction factor (Bode plots), 373
correlation, 156, 187
critically damped response, 380
cross correlation, 157, 187
cutoff frequency, 276, 399
cyclic convolution, 150, 180
damping, 379
damping factor, 377
Danielson-Lanczos lemma, 572
dc value (offset), 197
dc power supplies, 224
decade, 371
decibel (dB), 58, 370
decimation, 44, 518
decimation-in-frequency FFT algorithm, 575
decimation-in-time FFT algorithm, 573
decomposition principle, 97
deconvolution, 185, 564
deemphasis, 315
delay, 29, 153, 276, 367, 379, 509
delay element, 101
delta (impulse) function, 21
demodulation, 303, 319, 325
detection, 303, 311
deterministic signals, 55
deviation ratio, 313
difference equations, 96, 103
differential equations, 76ff
differentiator (digital), 493, 749
digital audio, 470
digital audio effects, 660
digital filter realization, 101, 637
digital filters, 96, 673ff, 715ff
digital frequency, 49
digital oscillators, 651
digital resonators, 653, 740
digital signal processing (DSP), 2, 465
digital signals, 1
digital-to-analog converter (DAC), 465, 467
digital-to-digital (D2D) transformations, 694
Dirac (delta) function, 21
direct form I realization, 637
direct form II realization, 638
Dirichlet conditions (Fourier series), 233
Dirichlet conditions (Fourier transform), 271
Dirichlet kernel, 219, 506
discrete convolution, 169ff
  and polynomial multiplication, 175
  using DFT and FFT, 538, 563
discrete Fourier series (DFS), 492, 546
discrete Fourier transform (DFT), 7, 535ff
discrete spectrum, 207, 542
discrete sum, 40
discrete-time Fourier transform (DTFT), 482ff

PGR2026-00033
Krisp Technologies EX1024 Page 816

discrete-time signals, 39ff
discrete-time systems, 96ff
discriminator, 319
distortion, 276
distortionless filters, 276
Dolby noise reduction systems, 473
double-sideband suppressed carrier AM, 300
doublet, 26
down-sampler, 518
drug metabolism, 356
DSBSC AM, 300
DTFT, 482ff
DTMF receivers, 651
duality (similarity), 255
duration, 29, 153
duration-bandwidth product, 6, 284
duty ratio, 9
dynamic range, 462, 470
dynamic range processors, 472
dynamic range resolution, 559
dynamic system, 75, 100

echo, 119, 660
efficiency of modulation, 302
eigenfunctions, 151
eigenvalue, 151
eigensignal, 151, 183
elliptic filters, 427ff
elliptic functions, 428
elliptic integrals, 428
energy, 9
energy signals, 10, 40
energy spectral density, 282
envelope, 323
envelope detection, 305
equalizers (audio), 649, 700
equiripple filters, 412, 740
equivalent bandwidth, 286
equivalent duration, 287
ergodicity, 57
estimation (statistical), 57
Euler's relation, 16
Euler's algorithms, 512, 685
even symmetry, 14, 42
existence of convolution, 150
existence of Fourier series, 232
existence of Fourier transform, 271, 358
existence of Laplace transform, 331, 358
existence of the DTFT, 485
existence of the z-transform, 624
expectation, 56
extremal frequencies, 741

fast Fourier transform (FFT), 539, 571ff
feedback realization, 101, 637
feedback systems, 382ff
feedforward realization, 101, 637
Fejer kernel, 229, 506
FFT, 539, 571ff
filter, 7, 68
final value theorem, 346, 616
FIR (finite impulse response) filters, 101, 715ff
  recursive implementation, 113, 739
finite wordlength effects, 470
first-order-hold, 459
fixed system, 72
flanging, 662
flat-top sampling, 453
FM, 312ff
FM transmission, 320
FM reception, 320
folding (reversal), 12
folding frequency, 448, 543
folding index, 537
forced response, 76, 104
Fourier, J.B.J, 234ff
Fourier series, 197ff
  and system analysis, 221
  convergence, 232
  generalized, 245
  reconstruction, 216
Fourier transform, 248ff
  and system analysis, 271
  of periodic signals, 262
  properties, 252
fractional delay, 46
frequency deviation, 312
frequency domain, 4
frequency modulation (FM), 312ff
frequency multiplexing, 306
frequency resolution, 559
frequency response, 275, 367, 495, 621
  graphical interpretation, 645
  rubber sheet analogy, 646
frequency sampling filters, 736
frequency sensitivity, 313

PGR2026-00033
Krisp Technologies EX1024 Page 817

frequency transformations
    analog filters, 399
    for FIR filters, 728
    for IIR filters, 694
Fresnel integrals, 317
fundamental frequency, 17

gain, 274
gain element, 101
gain margin, 386
Gaussian function, 57, 154
generalized Fourier series, 245
generalized functions, 21
geometric center frequency, 400
geometric symmetry, 400
Gibbs effect, 218, 226, 508, 720
graphic equalizers, 700
group delay, 276, 367, 411, 421

halfband filters, 733
half-power bandwidth, 278
half-power frequency, 278, 368, 372
half-wave symmetry, 15, 201
harmonic, 197, 209
harmonic distortion, 224
harmonic signals, 16, 49
Heaviside unit step, 18
heterodyning, 307
hidden symmetry, 202
high frequency decay rate, 294, 403, 413
highpass filter, 276
Hilbert transform, 310, 321, 566
Hilbert transformers, 310, 494, 749
histograms, 58
Hölder norm, 63
homogeneity, 69, 96
Hurwitz polynomial, 345

ideal filters, 276, 504
ideal sampling, 446
ideally sampled signal, 25, 446
identity system, 642
IIR (infinite impulse response) filters, 101, 673ff
image, 52, 448
impulse approximation, 25, 48
impulse function, 21, 47
    derivatives of, 27
impulse-invariant transformation, 679
impulse response, 6, 81, 111, 271, 337, 494, 602
infinite-impulse-response (IIR) filters, 101, 673ff
in-phase component, 310
initial conditions, 69, 103
initial value theorem, 346, 616
instantaneous frequency, 35, 312
instantaneous phase, 35, 312
instantaneous power, 9
instantaneous system, 75, 100
interconnected systems, 143, 176, 357, 640
intermediate frequency (IF), 307
interpolating functions, 457
interpolation, 44, 456, 517
inverse Chebyshev (II) filters, 421ff
inverse systems, 117, 386, 642

Jacobian elliptic function, 428
jitter, 459, 591

Lanczos window, 230
Laplace transform, 330ff
leakage, 547, 550
least squares, 230, 751
left-sided signals, 8, 39
l'Hôpital's rule, 20, 205
Liapunov stability, 345
limiters 472
line spectra, 207
linear interpolation, 44, 458
linear-phase filters, 497, 715ff
linear-phase signals, 485, 644
linear operators, 69, 96
linear systems, 71, 97
linear time-invariant (LTI) systems, 73, 99
LMS algorithm, 753
lower sideband, 300
lowpass filter, 86, 277
lowpass prototype, 399

magnitude spectrum, 207, 251, 484
marginal stability, 345
mainlobe, 20
matched filters, 159
matched z-transform, 683
maximally flat filters, 403, 748
maximally flat function, 433, 443
maximum-phase system, 369
mean, 56
memory, 75
memoryless system, 75

PGR2026-00033
Krisp Technologies EX1024 Page 818

message signal, 300
minimum-phase systems, 344, 369, 642
minmax approximation, 412
mixed-phase system, 369
mixing, 307
models, 8, 68
model order reduction, 365
modulation, 257, 300ff
modulation efficiency, 302
modulation index, 301
moments, 28, 153
moving average (MA) filter, 101
multiplexing, 306
multistage decimation, 747
multistage interpolation, 744

narrowband FM, 315
natural envelope, 324
natural frequency, 339, 602
natural modes, 339, 602
natural response, 76, 104
natural sampling, 452
neper, 330
noise, 7, 284
noise gate, 472
non-anticipating system, 75
noncausal systems, 75, 100
nonrecursive filters, 101
    recursive form for, 113, 739
normal distribution, 56
norms, 63
notch filters, 657
numerical algorithms, 37, 512, 685
Nyquist frequency, 52, 448
Nyquist interval, 52, 448
Nyquist rate, 52, 448

octave, 371
odd symmetry, 14, 42
one-sided z-transform, 613ff
open loop gain, 384
operational transfer function, 100
operators, 69, 96
optimal filter design, 740
order (of a system), 70, 97
orthogonal functions, 199, 245
overdamped response, 380
overlap add method, 563
overlap save method, 563
overmodulation, 302
oversampling, 515
    and quantization, 464
overshoot, 381

Pade approximations, 711
pairwise sum (for convolution), 138
parallel connection, 143, 176, 357, 640
parametric equalizers, 702
Parks-McClellan algorithm, 742
Parseval's relation, 205, 267, 489
partial fraction expansion, 340
    for nonstandard forms, 343
partial sum, 228
passband, 276
passband ripple, 398
peak factor, 303
peak overshoot, 381
period, 9
    common, 17, 40
    principal, 51
periodic convolution, 150, 180
periodic extension, 146, 179
    and convolution, 146, 179
    and DFT, 502, 542
    and Fourier transforms, 250
    and sampling, 250, 502
periodic signals, 9, 40
periodogram, 566
phase, 4
phase delay, 16, 276, 367, 441
phase deviation, 312
phase distortion, 276
phase-locked loop (PLL), 387
phase modulation (PM), 313ff
phase sensitivity, 313
phase spectrum, 207, 251
phasing, 662
physically realizable system, 145
piecewise-continuous signals, 8
Plancherel's relation, 269
PLL (phase-locked loop), 387
plucked-string filters, 662
pole frequency, 379
pole-zero placement, 647
pole-zero plots, 339, 601
poles, 339, 600

PGR2026-00033
Krisp Technologies EX1024 Page 819

power, 9, 40
power signals, 10, 40
power spectral density (PSD), 282, 566
preemphasis, 315
pre-envelope, 323
prewarping, 691
principal period (range), 51
probability, 55
probes (for oscilloscopes), 395
processing gain, 556
projection theorem, 231
prolate spheroidal wave function, 723
proper rational function, 339
prototype transformations, 399
pseudorandom signals, 58

Q (quality factor), 377, 379
quadrantal symmetry, 370
quadrature component, 310
quantization, 460ff
quantization error, 461
quantization step size, 462
quantized signal, 1
quasiperiodic signals, 17

radix, 571
radix-2 FFT, 571
raised-cosine interpolating function, 459
ramp function, 18, 47
random signals, 55
rational fractions, 17
rational functions, 339
Rayleigh's theorem, 267
realization, 55
realization of digital filters, 101, 637
rect function, 19, 47
recursive filter, 101, 637
reflection (reversal, or folding), 12, 41
region of convergence (ROC)
    of the Laplace transform, 331
    of the z-transform, 595
regulation, 383
relaxed system, 3, 69
Remez exchange algorithm, 742
residues, 340
resolution, 559
resonators, 653, 740
response invariant transformations, 676
reverb, 119, 660
reverberation time, 510
reversal (folding, or reflection), 12, 41
RIAA equalization, 393, 714
Riemann-Lebesgue theorem, 232
Riesz-Fischer theorem, 233
right-sided signals, 8, 39
ripple, 398, 725
rise time, 4, 87, 379
RLS algorithm, 753
rms bandwidth, 287
rms duration, 287
rms quantization error, 462
rms (root-mean-square) value, 9
Routh test, 345
rubber sheet analogy, 646

sample function, 47
sampling, 446ff
    of bandpass signals, 451
    and periodic extension, 250, 502
sampling interval, 1, 39
sampling oscilloscope, 451
sampling rate (frequency), 49
sampling rate conversion, 515
sampling theorem, 2, 52, 448
scallop loss, 556
settling time, 87, 379
sgn (signum) function, 18
shelving filters, 650
shift invariance, 71, 97
shift operators, 96
sidebands, 300
sidelobes, 20
sifting property, 24
signal approximation by impulses, 25, 48
signal averaging, 58
signal-to-noise ratio (SNR), 7, 58
signum function, 18
similarity theorem, 255
simple roots, 85
Simpson's algorithm, 128, 512, 528, 710
sinc distortion, 454
sinc function, 20, 49
sinc interpolation, 456
sine integral, 227
single-sideband (SSB) AM, 310
sinusoids, 16, 49

PGR2026-00033
Krisp Technologies EX1024 Page 820

spectral image, 52, 448
spectral transformations, 399, 505, 694, 728
spectral windows, 219
spectrum, 197, 207, 251, 484
spectrum analyzers, 309
spectrum estimation, 566ff
s-plane, 339, 344
square integrable signals, 10
square summable signals, 40
SSB (single sideband) AM, 310
stability, 85, 114, 144, 177, 604
stabilization, 659
standard deviation, 56
state, 3, 68
state equations, 76, 103
state variables, 68, 103
static systems, 75, 100
stationary signals, 57
stationary systems, 72
steady-state response, 78, 104, 221, 275, 353, 500, 623
step function, 18, 47
step interpolation, 44, 458
step-invariant transformation, 676
Stirling limit, 167
stopband, 276
strength (of impulse), 22
strictly proper rational function, 339
sum by arithmetic means, 229
superheterodyne receiver, 307
superposition, 3, 69
switched periodic signals, 33, 337, 354, 617
symmetry, 14, 42,
symmetric sequences, 715
synchronous detection, 303
system classification, 70, 97
system identification, 68, 284, 568
system order, 70, 97

Thomson polynomials, 435
3-dB (half-power) frequency, 372
Tick's rule, 528
time-bandwidth product, 6, 284
time constant, 4, 86, 510
time-frequency plots, 568
time-invariant systems, 71, 97
time-limited signals, 8
time-limited/band-limited theorem, 269
time period, 9, 17
time scaling, 12
time shift, 12, 41
Toeplitz matrix, 182
tracking, 383
transfer function, 271, 337, 494, 602
transient response, 4, 78, 104
transition band (width), 398, 725
transposed realization, 638
trapezoidal rule, 513, 689
tri function, 19, 47
Tustin's rule, 689
twiddle factors, 569

uncertainty principle, 288
undamped natural frequency, 379
underdamped response, 380
undersampling, 448
uniform convergence, 232
uniform distribution, 56
unit circle, 604, 621
unitary transforms, 5, 70
upper sideband, 300
up-sampler, 44, 517

Vandermonde matrix, 570
variance, 56

waterfall plot, 568
wavelets, 569
wavetable synthesis, 119
Welch spectrum estimate, 567
white noise, 284
wide sense (marginal) stability, 345
Wiener-Khintchine theorem, 282
window-based FIR filters, 720
window functions, 19, 220, 555, 720
windowing, 219, 506, 555, 724
wraparound, 146, 179

z-plane, 592, 621
z-transform, 592ff
zero-input response (ZIR), 79, 108
zero interpolation, 44, 211, 515, 540
zero-order-hold, 458
zero-order-hold sampling, 453
zero-padding, 183, 552
zero-state response (ZSR), 79, 108
zeros, 339, 600

PGR2026-00033
Krisp Technologies EX1024 Page 821