EXHIBIT 1030

Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software
Yaakov Bar-Shalom, X. Rong Li, Thiagalingam Kirubarajan
Copyright © 2001 John Wiley & Sons, Inc.
ISBNs: 0-471-41655-X (Hardback) 0-471-22127-9 (Electronic)

Estimation with Applications To Tracking and Navigation

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 2

Estimation with Applications To Tracking and Navigation

Yaakov Bar-Shalom
X.-Rong Li
Thiagalingam Kirubarajan



A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.
New York • Chichester • Weinheim • Brisbane • Singapore • Toronto

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 3

This text is printed on acid-free paper. ♾

Copyright © 2001 by John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4744. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012, (212) 850-6011, fax (212) 850-6008, E-Mail: PERMREQ @ WILEY.COM.

For ordering and customer service, call 1-800-CALL-WILEY.

Library of Congress Cataloging-in-Publication Data

Bar-Shalom, Yaakov.
Estimation with applications to tracking and navigation / by Yaakov Bar-Shalom, X.-Rong Li, Thiagalingam Kirubarajan.
p. cm.
Includes bibliographical references and index.
ISBN 0-471-41655-X (cloth)
1. Motion control devices. 2. Remote control. 3. Telecommunication systems. 4. Robots—Control systems. 5. Process control. 6. Estimation theory. I. Li, X.-Rong. II. Kirubarajan, Thiagalingam. III. Title.

TJ214.5 .B37 2001
681'.2—dc21 2001022366

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software
Yaakov Bar-Shalom, X. Rong Li, Thiagalingam Kirubarajan
Copyright © 2001 John Wiley & Sons, Inc.
ISBNs: 0-471-41655-X (Hardback) 0-471-22127-9 (Electronic)

To Eva, Tali, Yael and Michael

YBS

To Peizhu, Helen and Linda

XRL

To Appa, Amma, Ketha, Abi and Arun

TK

Lemma 1. Make things as simple as possible but not simpler.

A. Einstein

Theorem 1. By making things absolutely clear,
people will become confused.

A Chinese fortune cookie

Corollary 1. We will make things
simple
but not too simple,
clear
but not too clear.

Lemma 2. Uncertainty is everywhere.

Theorem 2. Uncertainty cannot be conquered.

Corollary 2. Embrace it!

Paraphrased after Michael Moschen,

professonal juggler.

Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software
Yaakov Bar-Shalom, X. Rong Li, Thiagalingam Kirubarajan
Copyright © 2001 John Wiley & Sons, Inc.
ISBNs: 0-471-41655-X (Hardback) 0-471-22127-9 (Electronic)

Contents

PREFACE xvii
ACRONYMS xxi
MATHEMATICAL NOTATIONS xxii

1 INTRODUCTION 1
1.1 BACKGROUND 1
1.1.1 Estimation and Related Areas 1
1.1.2 Applications of Estimation 3
1.1.3 Preview of Estimation/Filtering 4
1.1.4 An Example of State Estimation: Vehicle Collision Avoidance 10
1.2 SCOPE OF THE TEXT 15
1.2.1 Objectives 15
1.2.2 Overview and Chapter Prerequisites 16
1.3 BRIEF REVIEW OF LINEAR ALGEBRA AND LINEAR SYSTEMS 19
1.3.1 Definitions and Notations 19
1.3.2 Some Linear Algebra Operations 20
1.3.3 Inversion and the Determinant of a Matrix 21
1.3.4 Orthogonal Projection of Vectors 23
1.3.5 The Gradient, Jacobian and Hessian 24
1.3.6 Eigenvalues, Eigenvectors, and Quadratic Forms 25
1.3.7 Continuous-Time Linear Dynamic Systems — Controllability and Observability 27
1.3.8 Discrete-Time Linear Dynamic Systems — Controllability and Observability 29
1.4 BRIEF REVIEW OF PROBABILITY THEORY 31
1.4.1 Events and the Axioms of Probability 31
1.4.2 Random Variables and Probability Density Function 33
1.4.3 Probability Mass Function 35
1.4.4 Mixed Random Variable and Mixed Probability-PDF 36
1.4.5 Expectations and Moments of a Scalar Random Variable 37
1.4.6 Joint PDF of Two Random Variables 38
1.4.7 Independent Events and Independent Random Variables 41
1.4.8 Vector-Valued Random Variables and Their Moments 41
1.4.9 Conditional Probability and PDF 44
1.4.10 The Total Probability Theorem 45
1.4.11 Bayes' Formula 47
1.4.12 Conditional Expectations and Their Smoothing Property 50
1.4.13 Gaussian Random Variables 51
1.4.14 Joint and Conditional Gaussian Random Variables 52
1.4.15 Expected Value of Quadratic and Quartic Forms 54
1.4.16 Mixture Probability Density Functions 55
1.4.17 Chi-Square Distributed Random Variables 57
1.4.18 Weighted Sum of Chi-Square Random Variables 60
1.4.19 Random Processes 61
1.4.20 Random Walk and the Wiener Process 65
1.4.21 Markov Processes 66
1.4.22 Random Sequences, Markov Sequences and Markov Chains 69

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 7

1.4.23 The Law of Large Numbers and the Central Limit Theorem 70
1.5 BRIEF REVIEW OF STATISTICS 72
1.5.1 Hypothesis Testing 72
1.5.2 Confidence Regions and Significance 74
1.5.3 Monte Carlo Runs and Comparison of Algorithms 79
1.5.4 Tables of the Chi-Square and Gaussian Distributions 82
1.6 NOTES AND PROBLEMS 85
1.6.1 Bibliographical Notes 85
1.6.2 Problems 85

2 BASIC CONCEPTS IN ESTIMATION 89
2.1 INTRODUCTION 89
2.1.1 Outline 89
2.1.2 Basic Concepts – Summary of Objectives 89
2.2 THE PROBLEM OF PARAMETER ESTIMATION 90
2.2.1 Definitions 90
2.2.2 Models for Estimation of a Parameter 91
2.3 MAXIMUM LIKELIHOOD AND MAXIMUM A POSTERIORI ESTIMATORS 92
2.3.1 Definitions of ML and MAP Estimators 92
2.3.2 MLE vs. MAP Estimator with Gaussian Prior 92
2.3.3 MAP Estimator with One-Sided Exponential Prior 94
2.3.4 MAP Estimator with Diffuse Prior 95
2.3.5 The Sufficient Statistic and the Likelihood Equation 96
2.4 LEAST SQUARES AND MINIMUM MEAN SQUARE ERROR ESTIMATION 98
2.4.1 Definitions of LS and MMSE Estimators 98
2.4.2 Some LS Estimators 100
2.4.3 MMSE vs. MAP Estimator in Gaussian Noise 100
2.5 UNBIASED ESTIMATORS 101
2.5.1 Definition 101
2.5.2 Unbiasedness of an ML and a MAP Estimator 102
2.5.3 Bias in the ML Estimation of Two Parameters 102
2.6 THE VARIANCE AND MSE OF AN ESTIMATOR 104
2.6.1 Definitions of Estimator Variances 104
2.6.2 Comparison of Variances of an ML and a MAP Estimator 105
2.6.3 The Variances of the Sample Mean and Sample Variance 106
2.6.4 Estimation of the Probability of an Event 107
2.7 CONSISTENCY AND EFFICIENCY OF ESTIMATORS 108
2.7.1 Consistency 108
2.7.2 The Cramer-Rao Lower Bound and the Fisher Information Matrix 109
2.7.3 Proof of the Cramer-Rao Lower Bound 110
2.7.4 An Example of Efficient Estimator 112
2.7.5 Large Sample Properties of the ML Estimator 113
2.8 SUMMARY 114
2.8.1 Summary of Estimators 114
2.8.2 Summary of Estimator Properties 115
2.9 NOTES AND PROBLEMS 115
2.9.1 Bibliographical Notes 115
2.9.2 Problems 116

3 LINEAR ESTIMATION IN STATIC SYSTEMS 121
3.1 INTRODUCTION 121
3.1.1 Outline 121
3.1.2 Linear Estimation in Static Systems — Summary of Objectives 121
3.2 ESTIMATION OF GAUSSIAN RANDOM VECTORS 122
3.2.1 The Conditional Mean and Covariance for Gaussian Random Vectors 122
3.2.2 Estimation of Gaussian Random Vectors — Summary 123
3.3 LINEAR MINIMUM MEAN SQUARE ERROR ESTIMATION 123
3.3.1 The Principle of Orthogonality 123

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 8

3.3.2 Linear MMSE Estimation for Vector Random Variables 127
3.3.3 Linear MMSE Estimation — Summary 129
3.4 LEAST SQUARES ESTIMATION 129
3.4.1 The Batch LS Estimation 129
3.4.2 The Recursive LS Estimator 132
3.4.3 Examples and Incorporation of Prior Information 135
3.4.4 Nonlinear LS — An Example 137
3.4.5 LS Estimation — Summary 145
3.5 POLYNOMIAL FITTING 146
3.5.1 Fitting a First-Order Polynomial to Noisy Measurements 146
3.5.2 Fitting a General Polynomial to a Set of Noisy Measurements 149
3.5.3 Mapping of the Estimates to an Arbitrary Time 152
3.5.4 Polynomial Fitting — Summary 154
3.6 GOODNESS-OF-FIT AND STATISTICAL SIGNIFICANCE OF PARAMETER ESTIMATES 154
3.6.1 Hypothesis Testing Formulation of the Problem 154
3.6.2 The Fitting Error in a Least Squares Estimation Problem 156
3.6.3 A Polynomial Fitting Example 159
3.6.4 Order Selection in Polynomial Fitting — Summary 161
3.7 USE OF LS FOR A NONLINEAR PROBLEM: BEARINGS-ONLY TARGET MOTION ANALYSIS 161
3.7.1 The Problem 161
3.7.2 Observability of the Target Parameter in Passive Localization 162
3.7.3 The Likelihood Function for Target Parameter Estimation 163
3.7.4 The Fisher Information Matrix for the Target Parameter 164
3.7.5 The Goodness-of-Fit Test 167
3.7.6 Testing for Efficiency with Monte Carlo Runs 168
3.7.7 A Localization Example 169
3.7.8 Passive Localization — Summary 169
3.8 NOTES, PROBLEMS AND A PROJECT 172
3.8.1 Bibliographical Notes 172
3.8.2 Problems 172
3.8.3 PROJECT: An Interactive Program for Bearings-Only Target Localization 176

4 LINEAR DYNAMIC SYSTEMS WITH RANDOM INPUTS 179
4.1 INTRODUCTION 179
4.1.1 Outline 179
4.1.2 Linear Stochastic Systems — Summary of Objectives 179
4.2 CONTINUOUS-TIME LINEAR STOCHASTIC DYNAMIC SYSTEMS 180
4.2.1 The Continuous-Time State-Space Model 180
4.2.2 Solution of the Continuous-Time State Equation 181
4.2.3 The State as a Markov Process 183
4.2.4 Propagation of the State's Mean and Covariance 184
4.2.5 Frequency Domain Approach 185
4.3 DISCRETE-TIME LINEAR STOCHASTIC DYNAMIC SYSTEMS 187
4.3.1 The Discrete-Time State-Space Model 187
4.3.2 Solution of the Discrete-Time State Equation 189
4.3.3 The State as a Markov Process 190
4.3.4 Propagation of the State's Mean and Covariance 191
4.3.5 Frequency Domain Approach 192
4.4 SUMMARY 195
4.4.1 Summary of State Space Representation 195
4.4.2 Summary of Prewhitening 195
4.5 NOTES AND PROBLEMS 196
4.5.1 Bibliographical Notes 196
4.5.2 Problems 196

5 STATE ESTIMATION IN DISCRETE-TIME LINEAR DYNAMIC SYSTEMS 199
5.1 INTRODUCTION 199
5.1.1 Outline 199
5.1.2 Discrete-Time Linear Estimation — Summary of Objectives 199
5.2 LINEAR ESTIMATION IN DYNAMIC SYSTEMS — THE KALMAN FILTER 200
5.2.1 The Dynamic Estimation Problem 200
5.2.2 Dynamic Estimation as a Recursive Static Estimation 202
5.2.3 Derivation of the Dynamic Estimation Algorithm 204
5.2.4 Overview of the Kalman Filter Algorithm 207
5.2.5 The Matrix Riccati Equation 211
5.2.6 Properties of the Innovations and the Likelihood Function of the System Model 213
5.2.7 The Innovations Representation 214
5.2.8 Some Orthogonality Properties 215
5.2.9 The Kalman Filter — Summary 215
5.3 EXAMPLE OF A FILTER 218
5.3.1 The Model 218
5.3.2 Results for a Kalman Filter 219
5.3.3 A Step-by-Step Demonstration of DynaEst™ 219
5.4 CONSISTENCY OF STATE ESTIMATORS 232
5.4.1 The Problem of Filter Consistency 232
5.4.2 Definition and the Statistical Tests for Filter Consistency 234
5.4.3 Examples of Filter Consistency Testing 237
5.4.4 Absolute Errors 243
5.4.5 Filter Consistency — Summary 244
5.5 INITIALIZATION OF STATE ESTIMATORS 245
5.5.1 Initialization and Consistency 245
5.5.2 Initialization in Simulations 246
5.5.3 A Practical Implementation in Tracking 247
5.5.4 Filter Initialization — Summary 248
5.6 SENSITIVITY 248
5.6.1 Model Mismatch 249
5.6.2 Reduced-Order Filters 254
5.6.3 Suboptimal Gains 256
5.6.4 Examples of Modeling Errors and Filter Approximations 256
5.7 NOTES AND PROBLEMS 261
5.7.1 Bibliographical Notes 261
5.7.2 Problems 261
5.7.3 Computer Applications 265

6 ESTIMATION FOR KINEMATIC MODELS 267
6.1 INTRODUCTION 267
6.1.1 Outline 267
6.1.2 Kinematic Models — Summary of Objectives 267
6.2 DISCRETIZED CONTINUOUS-TIME KINEMATIC MODELS 268
6.2.1 The Kinematic Models 268
6.2.2 Continuous White Noise Acceleration Model 269
6.2.3 Continuous Wiener Process Acceleration Model 270
6.3 DIRECT DISCRETE-TIME KINEMATIC MODELS 272
6.3.1 Introduction 272
6.3.2 Discrete White Noise Acceleration Model 273
6.3.3 Discrete Wiener Process Acceleration Model 274
6.3.4 Kinematic Models — Summary 275
6.4 EXPLICIT FILTERS FOR NOISELESS KINEMATIC MODELS 276
6.4.1 LS Estimation for Noiseless Kinematic Models 276
6.4.2 The KF for Noiseless Kinematic Models 276
6.5 STEADY-STATE FILTERS FOR NOISY KINEMATIC MODELS 277
6.5.1 The Problem 277
6.5.2 Derivation Methodology for the Alpha-Beta Filter 278

6.5.3 The Alpha-Beta Filter for the DWNA Model 280
6.5.4 The Alpha-Beta Filter for the Discretized CWNA Model 286
6.5.5 The Alpha-Beta-Gamma Filter for the DWPA Model 289
6.5.6 A System Design Example for Sampling Rate Selection 292
6.5.7 Alpha-Beta and Alpha-Beta-Gamma Filters — Summary 293
6.6 NOTES AND PROBLEMS 294
6.6.1 Bibliographical Notes 294
6.6.2 Problems 295

7 COMPUTATIONAL ASPECTS OF ESTIMATION 301
7.1 INTRODUCTION 301
7.1.1 Implementation of Linear Estimation 301
7.1.2 Outline 302
7.1.3 Computational Aspects — Summary of Objectives 303
7.2 THE INFORMATION FILTER 303
7.2.1 Recursions for the Information Matrices 303
7.2.2 Overview of the Information Filter Algorithm 306
7.2.3 Recursion for the Information Filter State 307
7.3 SEQUENTIAL PROCESSING OF MEASUREMENTS 308
7.3.1 Block vs. Sequential Processing 308
7.3.2 The Sequential Processing Algorithm 309
7.4 SQUARE-ROOT FILTERING 311
7.4.1 The Steps in Square-Root Filtering 311
7.4.2 The LDL′ Factorization 312
7.4.3 The Predicted State Covariance 312
7.4.4 The Filter Gain and the Updated State Covariance 314
7.4.5 Overview of the Square-Root Sequential Scalar Update Algorithm 315
7.4.6 The Gram-Schmidt Orthogonalization Procedure 315
7.5 NOTES AND PROBLEMS 317
7.5.1 Bibliographical Notes 317
7.5.2 Problems 317

8 EXTENSIONS OF DISCRETE-TIME LINEAR ESTIMATION 319
8.1 INTRODUCTION 319
8.1.1 Outline 319
8.1.2 Extensions of Estimation — Summary of Objectives 319
8.2 AUTOCORRELATED PROCESS NOISE 320
8.2.1 The Autocorrelated Process Noise Problem 320
8.2.2 An Exponentially Autocorrelated Noise 321
8.2.3 The Augmented State Equations 322
8.2.4 Estimation with Autocorrelated Process Noise — Summary 324
8.3 CROSS-CORRELATED MEASUREMENT AND PROCESS NOISE 324
8.3.1 Cross-Correlation at the Same Time 324
8.3.2 Cross-Correlation One Time Step Apart 326
8.3.3 State Estimation with Decorrelated Noise Sequences — Summary 327
8.4 AUTOCORRELATED MEASUREMENT NOISE 327
8.4.1 Whitening of the Measurement Noise 327
8.4.2 The Estimation Algorithm with the Whitened Measurement Noise 329
8.4.3 Autocorrelated Measurement Noise — Summary 330
8.5 PREDICTION 330
8.5.1 Types of Prediction 330
8.5.2 The Algorithms for the Different Types of Prediction 330
8.5.3 Prediction — Summary 332
8.6 SMOOTHING 333
8.6.1 Types of Smoothing 333
8.6.2 Fixed-Interval Smoothing 334
8.6.3 Overview of Smoothing 337
8.6.4 Smoothing — Summary 338

8.7 NOTES AND PROBLEMS 338
8.7.1 Bibliographical Notes 338
8.7.2 Problems 338

9 CONTINUOUS-TIME LINEAR STATE ESTIMATION 341
9.1 INTRODUCTION 341
9.1.1 Outline 341
9.1.2 Continuous-Time Estimation — Summary of Objectives 341
9.2 THE CONTINUOUS-TIME LINEAR STATE ESTIMATION FILTER 342
9.2.1 The Continuous-Time Estimation Problem 342
9.2.2 Connection Between Continuous - and Discrete-Time Representations and Their Noise Statistics 343
9.2.3 The Continuous-Time Filter Equations 345
9.2.4 The Continuous-Time Innovation 347
9.2.5 Asymptotic Properties of the Continuous-Time Riccati Equation 349
9.2.6 Examples of Continuous-Time Filters 351
9.2.7 Overview of the Kalman-Bucy Filter 353
9.2.8 Continuous-Time State Estimation — Summary 354
9.3 PREDICTION AND THE CONTINUOUS-DISCRETE FILTER 355
9.3.1 Prediction of the Mean and Covariance 355
9.3.2 The Various Types of Prediction 356
9.3.3 The Continuous-Discrete Filter 357
9.4 DUALITY OF ESTIMATION AND CONTROL 358
9.4.1 The Two Problems 358
9.4.2 The Solutions to the Estimation and the Control Problems 359
9.4.3 Properties of the Solutions 360
9.5 THE WIENER-HOPF PROBLEM 362
9.5.1 Formulation of the Problem 362
9.5.2 The Wiener-Hopf Equation 362
9.6 NOTES AND PROBLEMS 366
9.6.1 Bibliographical Notes 366
9.6.2 Problems 367

10 STATE ESTIMATION FOR NONLINEAR DYNAMIC SYSTEMS 371
10.1 INTRODUCTION 371
10.1.1 Outline 371
10.1.2 Nonlinear Estimation — Summary of Objectives 371
10.2 ESTIMATION IN NONLINEAR STOCHASTIC SYSTEMS 372
10.2.1 The Model 372
10.2.2 The Optimal Estimator 373
10.2.3 Proof of the Recursion of the Conditional Density of the State 374
10.2.4 Example of Linear vs. Nonlinear Estimation of a Parameter 376
10.2.5 Estimation in Nonlinear Systems with Additive Noise 379
10.2.6 Optimal Nonlinear Estimation — Summary 381
10.3 THE EXTENDED KALMAN FILTER 381
10.3.1 Approximation of the Nonlinear Estimation Problem 381
10.3.2 Derivation of the EKF 383
10.3.3 Overview of the EKF Algorithm 385
10.3.4 An Example: Tracking with an Angle-Only Sensor 387
10.3.5 The EKF — Summary 394
10.4 ERROR COMPENSATION IN LINEARIZED FILTERS 395
10.4.1 Some Heuristic Methods 395
10.4.2 An Example of Use of the Fudge Factor 396
10.4.3 An Example of Debiasing: Conversion from Polar to Cartesian 397
10.4.4 Error Compensation in Linearized Filters — Summary 402
10.5 SOME ERROR REDUCTION METHODS 404
10.5.1 Improved State Prediction 404
10.5.2 The Iterated Extended Kalman Filter 404

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 12

10.5.3 Some Error Reduction Methods — Summary 407
10.6 MAXIMUM A POSTERIORI TRAJECTORY ESTIMATION VIA DYNAMIC PROGRAMMING 407
10.6.1 The Approach 407
10.6.2 The Dynamic Programming for Trajectory Estimation 408
10.7 Nonlinear Continuous-Discrete Filter 409
10.7.1 The Model 409
10.7.2 The Fokker-Planck Equation 410
10.7.3 Example 413
10.8 NOTES, PROBLEMS AND A PROJECT 414
10.8.1 Bibliographical Notes 414
10.8.2 Problems 414
10.8.3 Project — Nonlinear Filtering with Angle-Only Measurements 419

11 ADAPTIVE ESTIMATION AND MANEUVERING TARGETS 421
11.1 INTRODUCTION 421
11.1.1 Adaptive Estimation — Outline 421
11.1.2 Adaptive Estimation — Summary of Objectives 423
11.2 ADJUSTABLE LEVEL PROCESS NOISE 424
11.2.1 Continuous Noise Level Adjustment 424
11.2.2 Process Noise with Several Discrete Levels 424
11.2.3 Adjustable Level Process Noise — Summary 426
11.3 INPUT ESTIMATION 427
11.3.1 The Model 427
11.3.2 The Innovations as a Linear Measurement of the Unknown Input 428
11.3.3 Estimation of the Unknown Input 429
11.3.4 Correction of the State Estimate 430
11.3.5 Input Estimation — Summary 431
11.4 THE VARIABLE STATE DIMENSION APPROACH 431
11.4.1 The Approach 431
11.4.2 The Maneuver Detection and Model Switching 432
11.4.3 Initialization of the Augmented Model 433
11.4.4 VSD Approach — Summary 434
11.5 A COMPARISON OF ADAPTIVE ESTIMATION METHODS FOR MANEUVERING TARGETS 435
11.5.1 The Problem 435
11.5.2 The White Noise Model with Two Levels 436
11.5.3 The IE and VSD Methods 436
11.5.4 Statistical Test for Comparison of the IE and VSD Methods 438
11.5.5 Comparison of Several Algorithms — Summary 440
11.6 THE MULTIPLE MODEL APPROACH 441
11.6.1 Formulation of the Approach 441
11.6.2 The Static Multiple Model Estimator 441
11.6.3 The Dynamic Multiple Model Estimator 444
11.6.4 The GPB1 Multiple Model Estimator for Switching Models 447
11.6.5 The GPB2 Multiple Model Estimator for Switching Models 449
11.6.6 The Interacting Multiple Model Estimator 453
11.6.7 An Example with the IMM Estimator 457
11.6.8 Use of DynaEst^{TM} to Design an IMM Estimator 460
11.6.9 The Multiple Model Approach — Summary 465
11.7 DESIGN OF AN IMM ESTIMATOR FOR ATC TRACKING 466
11.7.1 ATC Motion Models 466
11.7.2 The EKF for the Coordinated Turn Model 468
11.7.3 Selection of Models and Parameters 470
11.7.4 The ATC Scenario 471
11.7.5 Results and Discussion 472
11.8 WHEN IS AN IMM ESTIMATOR NEEDED? 476
11.8.1 Kalman Filter vs. IMM Estimator 477

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 13

11.9 USE OF EKF FOR SIMULTANEOUS STATE AND PARAMETER ESTIMATION 481
11.9.1 Augmentation of the State 481
11.9.2 An Example of Use of the EKF for Parameter Estimation 482
11.9.3 EKF for Parameter Estimation — Summary 484
11.10 NOTES, PROBLEMS, AND TERM PROJECT 484
11.10.1 Bibliographical Notes 484
11.10.2 Problems 485
11.10.3 Term Project — IMM Estimator for Air Traffic Control 488

12 INTRODUCTION TO NAVIGATION APPLICATIONS 491
12.1 INTRODUCTION 491
12.1.1 Navigation Applications — Outline 491
12.1.2 Navigation Applications — Summary of Objectives 492
12.2 COMPLEMENTARY FILTERING FOR NAVIGATION 492
12.2.1 The Operation of Complementary Filtering 492
12.2.2 The Implementation of Complementary Filtering 493
12.3 INERTIAL NAVIGATION SYSTEMS 495
12.4 MODELS FOR INERTIAL NAVIGATION SYSTEMS 496
12.4.1 State Models 496
12.4.2 Sensor Error Models 496
12.4.3 Single-Axis Models 497
12.4.4 Three-Axis Models 499
12.4.5 Coordinate Transformation 500
12.5 THE GLOBAL POSITIONING SYSTEM (GPS) 501
12.5.1 The GPS Segments 502
12.5.2 GPS Satellite Constellation 502
12.6 GPS POSITIONING 502
12.6.1 The GPS Principle 502
12.6.2 The GPS Signals 503
12.6.3 The GPS Observables 506
12.7 THE ACCURACY OF GPS POSITIONING 507
12.7.1 Dilution of Precision 507
12.7.2 GPS Positioning Accuracy 509
12.8 STATE-SPACE MODELS FOR GPS 511
12.8.1 Models for Receiver Clock State 511
12.8.2 Dynamic Models 512
12.8.3 Linearized Measurement Model 512
12.8.4 A Model for Exponentially Autocorrelated Noise 513
12.8.5 Coordinate Transformation 515
12.9 EXAMPLE: GPS NAVIGATION WITH IMM ESTIMATOR 515
12.9.1 Generation of Satellite Trajectories 516
12.9.2 Generation of Trajectories and Pseudorange Measurements 517
12.9.3 State-Space Models 518
12.9.4 Simulation Results and Discussion 520
12.9.5 Do We Need an IMM Estimator for GPS? 523
12.10 INTEGRATED NAVIGATION 523
12.10.1 Integration by Complementary Filtering 524
12.10.2 Example 525
12.10.3 Integration by Centralized Estimation Fusion 527
12.10.4 Integration by Distributed Estimation Fusion 528
12.11 NOTES AND PROBLEMS 530
12.11.1 Bibliographical Notes 530
12.11.2 Problems 530
12.11.3 Term Project — Extended Kalman Filter for GPS 533

BIBLIOGRAPHY 537

INDEX 547

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 14

Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software
Yaakov Bar-Shalom, X. Rong Li, Thiagalingam Kirubarajan
Copyright © 2001 John Wiley & Sons, Inc.
ISBNs: 0-471-41655-X (Hardback) 0-471-22127-9 (Electronic)

PREFACE

This text, which also doubles as a *set of lecture notes available in viewgraph version* (downloadable as detailed below), presents the material from a graduate-level course on *Theory and Computational Algorithms for Estimation* offered in the Department of Electrical and Computer Engineering at the University of Connecticut. This course is a standard requirement in the department's M.S. program in Information, Communication and Decision Systems and is meant for second-semester graduate students. The prerequisites are a solid knowledge of linear systems and probability theory at the first-semester graduate level. These, as well as some additional useful material from Statistics, are summarized for ready reference in Chapter 1 to enable readers to acquire the necessary background or review it. This makes the text *completely self-contained* and accessible for a person with a typical B.S.E.E. degree or equivalent. This text is a major revision and expansion of Y. Bar-Shalom and X. R. Li, **Estimation and Tracking: Principles, Techniques, and Software**, originally published by Artech House, 1993 (©YBS Publishing, 1997).

The main goal is to convey the knowledge necessary for the *design and evaluation of state estimation algorithms that operate in a stochastic environment*. These algorithms form the backbone of **information extraction systems** for **remote sensing of moving objects** as well as other similar applications. Since **information fusion** — combination of information from several sources — is an area with significant recent activity, it should be kept in mind that information extraction is a crucial step before fusion. Furthermore, since fusion should be done accounting for the existing uncertainties, it is important to quantify the accuracies of the information to be fused.

This book also provides opportunity for sharpening the students' analytical skills in addition to using software. While the greatest recent advances have been in computation, the quality of a software is (in a simplistic manner) the product of the software skills and algorithmic skills that went into its development. The algorithm development requires analytical skills, which, unfortunately, have become too scarce lately.

While the Kalman filter (KF) is the best-known state estimator, it is neither the only one nor the most versatile. The material, which covers a number of practical estimation algorithms (as many as it is felt practical for a semester-long course), leads to a major design project that is unique to this text — it combines the basics of standard estimation with the latest adaptive multiple model techniques, which use KFs as modules. The emphasis is on mapping the physical properties of the object of interest (typically called "target," whether friendly or not) and of the sensors — e.g., an aircraft in a radar-based air traffic control tracking system — into the parameters of the mathematical model, namely, the statistical characterization of the random processes describing the uncertainties of this problem. As a prelude to state estimation, parameter estimation is discussed in sufficient detail to convey the concept of "limit of extractable information" in a system.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 15

The approach is a balanced combination of mathematics — linear systems and probability theory — in order to understand how a state estimator should be designed, with the necessary tools from statistics in order to interpret the results. The use of statistical techniques has been somewhat neglected in the engineering literature pertaining to state estimation, but it is necessary for (the nontrivial task of) interpreting stochastic data and answering the question whether a design can be accepted as "good." This is particularly important for practicing engineers and is presented in sufficient detail based on our belief (and extensive experience with real systems) that it should be an integral part of advanced engineering education.[1]

The material covers the topics usually taught in control-oriented EE/systems and aeronautical engineering programs. The relevance extends to other areas dealing with control in mechanical or chemical engineering. Recently, the state estimation techniques have been gaining wider attention due to their applicability to such fields as robotics, computer vision for autonomous navigation, and image feature extraction with application to medical diagnosis. While the course is mainly directed toward the M.S. students, it is also part of the Ph.D. program at the University of Connecticut, with the intent of providing the students with the knowledge to tackle real-world problems, whether by using existing algorithms or by developing new ones.[2]

The presentation of the material stresses the algorithms, their properties and the understanding of the assumptions behind them. We do not subscribe to the philosophy of "Give me the facts and don't bother me with details." Consequently, proofs are given to the extent that they are relevant to understanding the results. This is intended to be a modest step in bridging the much talked about "gap between theory and practice" — it will illustrate to students the usefulness of state estimation for the real world and provide to engineers and scientists working in industry or laboratories a broader understanding of the algorithms used in practice. It might also avoid the situation summarized by a participant at one of the continuing education courses taught by the first author as follows: "Although I studied Kalman filters when I worked toward my Ph.D. (at one of the major U.S. universities), I did not expect that they worked with real data." This happens when, because of the theorems, the students cannot see the forest of applications.

Tuning of a KF — the choice of its design parameters — is an art. One of the contributions of this text is to make it less of a black magic technique and more of a systematic approach by connecting the filter parameters to physical system parameters, namely, the object motion uncertainty — its predictability — and the sensor measurement uncertainty — the sensor errors. This is particularly important when KFs are used as modules in an adaptive estimator, like the Interacting Multiple Model (IMM) estimator. Another contribution is providing guidelines as to what estimator should be used in specific problems, namely, when an (adaptive) IMM estimator is preferable over a KF.

The Text and Hyperlinked Viewgraph Format

The format of this text is also unique — $textgraph^{TM}$ — in that it dares to attempt to accomplish two goals in one format: to serve as a self-contained *concise text*, without excess verbosity, and, at the same time, to enable the lecturer to use the pages of this text as *viewgraphs for lectures*. To this purpose, a double-spaced version of all the pages of this text (with appropriate page breaks)

[1] The authors disclaim any responsibility for severe damage readers might suffer when falling asleep face forward while reading this book.

[2] As Ben Franklin said, the goal in his life was "to rise to an (sic) happy mediocrity and secure competency." Our objective is to provide the tools for future successful applications.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 16

is available for instructors in the form of a pdf file at http://esplab1.ee.uconn.edu/estvu.zip with the password for downloading obtainable (upon the adoption of this book as a course text) from ybs@ee.uconn.edu. Instructors can prepare transparencies for conventional projector viewgraphs, or the pdf file can be used for *direct electronic projection*. For projection from a computer, an additional feature is available: hyperlinks in the pdf file for all cross-references (equations, figures, tables, sections, bibliographical items). Starred sections (or subsections) can be skipped in lecturing at the discretion of the lecturer — by clicking on the star the hypertext display will automatically go to the end of the corresponding section, marked by a box.

Experience over a number of years shows that students with a copy of these notes in front of them can concentrate on the ideas, rather than having to copy a copious amount of equations from the board. Some of the graduate students who took the Estimation course taught using this *textgraph*™ even complained that other graduate courses had not adopted this instruction method.

This format is also a way of making lectures more efficient since it makes it possible to convey more material than through the standard lecturing technique where one has to write everything on the board. Last, but not least, the need to spend time on preparing lecture notes, as well as the main hazard of the teaching profession (the inhalation of chalk or marker dust), can be avoided. The ready availability of such class notes also makes it much easier to hold in-house courses in an industrial or government laboratory environment where senior staff — with knowledge of the material but limited time — can serve as lecturers.

It is our hope that this text will contribute to the education of *future practitioners*, as well as be of use to those already involved in the application of state estimation to real-world problems. This *textgraph*™ was written with the philosophy that engineers should be able to do more than run-of-the-mill work — they should be prepared to work at the frontiers of knowledge.[3]

A Note on the Printing Style

For the convenience of the reader, all major concepts/terms, when they are introduced or used in a particular application, are in ***italic boldface*** and indexed. Also, all the index terms appear in the text in slanted boldface.

Italics are used for emphasized ideas/properties. The main equations of important algorithms are highlighted with boxes, and the algorithms are summarized in flowcharts.

The Companion Software DynaEst™

DynaEst™ is a completely self-explanatory menu-driven interactive design and evaluation tool for Kalman filters and an adaptive Interacting Multiple Model (IMM) estimator. The package provides the following important flexibilities as a design tool:

Estimator design. The user may specify any linear Gaussian true plant (dynamic) and measurement models, as well as the plant and measurement models for the estimator. The true and the filter-assumed models can be different. Models are valid for all dimensions. Reduced-order filters can be implemented.

Selection of model parameters. The user may specify any parameter of the true model and/or the filter-assumed model(s), such as the noise means and covariances.

[3] As the ancient Romans would have said: "Orchides forum trahite, cordes et mentes veniant."

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 17

Statistical testing. Most relevant statistical tests of the Kalman filters may be performed, including the normalized estimation error squared (NEES), the normalized innovation squared (NIS), and the RMS errors.

Capability of running on external data file. Both internally generated data (simulation) or imported data can be used.

$DynaEst^{TM}$ is user friendly and has the following features:

- MATLAB based (v. 4.0 or higher, including the student edition)
- Self-explanatory menu
- Convenient user specification of parameters (on screen or in an online editable file)
- Printing on screen, to a printer, or to a file
- Automatic (and flexible) installation
- Access to the OS and MATLAB from within $DynaEst^{TM}$
- Easily readable source code transparent to the user (and thus user-modifiable)
- Knowledge of MATLAB not required (but helpful, of course)
- For convenience, the *latest version* is available for the purchaser of this text at http://esplab1.ee.uconn.edu/dynaest.zip.

$DynaEst^{TM}$ provides valuable hands-on experience on Kalman filters and the state-of-the-art Interacting Multiple Model (IMM) estimator for any person who has access to MATLAB. With the help of $DynaEst^{TM}$, the principles/concepts and techniques treated in this book can be better understood and remembered. Some practical problems of estimation and tracking can be handled by $DynaEst^{TM}$. The source code of $DynaEst^{TM}$, which is easy to read, provides a sample implementation of the Kalman filter, as well as the state-of-the-art IMM estimator.

$DynaEst^{TM}$ has been used to generate several of the examples in the text. Also, there are a number of Computer Application Exercises at the end of the chapters dealing with state estimation, designed to use $DynaEst^{TM}$.

System Requirements

- IBM PC compatible
- MATLAB (professional, education, or student version)
- An operating system capable of running MATLAB 4.0 or higher (e.g., WINDOWS 95/98/NT or SUN SOLARIS)

Acknowledgements

Many thanks to all the colleagues, friends and former students who helped in putting together the material that eventually became this book: H. Blom, L. Campo, K. C. Chang, E. Daeipour, D. L. Kleinman, A. Kumar, C. Jauffret, D. Lerro, X. D. Lin, P. B. Luh, K. R. Pattipati, C. Rago, H. M. Shertukde, P. K. Willett and S. W. Yeom.

Also, thanks to the sponsors who have supported the research that led to this book: J. G. Smith and Dr. R. Madan from ONR, Dr. N. Glassman from AFOSR, Lt. Col. Dr. J. Myers from BMDO and Dr. R. K. Baheti from NSF.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 18

NOMENCLATURE

ACRONYMS

ATC	air traffic control
cdf	cumulative (probability) distribution function
cpmf	cumulative probability mass function
CLT	central limit theorem
CRLB	Cramer-Rao lower bound
EKF	extended Kalman filter
FIM	Fisher information matrix
GPB	generalized pseudo-Bayesian
GPS	Global Positioning System
HOT	higher order terms
IE	input estimation
i.i.d.	independent identically distributed
IMM	interacting multiple model (estimator)
INS	inertial navigation system
KBF	Kalman-Bucy filter (continuous time)
KF	Kalman filter (discrete time)
LF	likelihood function
LG	linear-Gaussian (assumption)
LLN	law of large numbers
LMMSE	linear MMSE
LOS	line of sight
LS	least squares
LSE	least squares estimator (or estimate)
MAP	maximum a posteriori
ML	maximum likelihood
MLE	maximum likelihood estimator (or estimate)
MM	multiple model
MMSE	minimum mean square error
MSE	mean square error
NEES	normalized estimation error squared
NIS	normalized innovation squared
pdf	probability density function
pmf	probability mass function
RMS	root mean square
VSD	variable state dimension

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 19

MATHEMATICAL NOTATIONS

$\arg\max$	argument that maximizes
$\arg\min$	argument that minimizes
χ^2_n	chi-square distribution with n degrees of freedom [Subsection 1.4.17]
$\chi^2_n(Q)$	$100Q$ percentile point of chi-square distribution with n degrees of freedom [Subsection 1.5.4]
$\text{col}(\cdot)$	column vector [Subsection 1.3.1]
cov	covariance
$\delta(\cdot)$	Dirac (continuous) delta function (impulse function)
δ_{ij}	Kronecker (discrete) delta function
$\text{diag}(\cdot)$	diagonal or block-diagonal matrix [Subsection 1.3.3, Subsection 3.4.1]
$E[\cdot]$	expectation
$\mathcal{G}(Q)$	$100Q$ percentile point of standard normal (Gaussian) distribution [Subsection 1.5.4]
Λ	likelihood function
$\mu(\cdot)$	pmf
$\mathcal{N}(x; \bar{x}, \sigma^2)$	pdf of a normal (Gaussian) random variable x with mean $\bar{x}$ and variance σ^2 [Subsection 1.4.13]
$\sim (\bar{x}, \sigma^2)$	a random variable with mean $\bar{x}$ and variance σ^2 [Subsection 1.4.13]
$\mathcal{N}(\bar{x}, \sigma^2)$	normal (Gaussian) distribution with mean $\bar{x}$ and variance σ^2 [Subsection 1.4.13]
$\mathcal{N}(x; \bar{x}, P)$	pdf of a normal (Gaussian) random vector x with mean $\bar{x}$ and covariance P [Subsection 1.4.13]
n_x	dimension of the vector x
$O(\cdot)$	order of magnitude of
$p[\cdot]$	pdf [Subsection 1.4.2]
$p(\cdot)$	pdf [Subsection 1.4.2]
$P(k\|j)$	conditional covariance matrix of state at time k given observations through time j
P_{xz}	covariance (matrix) of x and z
$P_{x\|z}$	conditional covariance (matrix) of x given z
$P[\cdot, \cdot]$	mixed joint probability-pdf
$P\{\cdot\}$	probability of an event
$\tilde{P}$	the spread of the means term [Subsection 1.4.16]
P_D	probability of detection [Subsection 1.5.1]
P_{FA}	false alarm probability [Subsection 1.5.1]
σ_θ	standard deviation of θ
σ^2_θ	variance of θ
$\text{tr}(\cdot)$	trace (of a matrix) [Subsection 1.3.2]
$\mathcal{U}(a, b)$	uniform distribution over the interval $[a, b]$
$\mathcal{U}(x; a, b)$	pdf of random variable x uniformly distributed over the interval $[a, b]$ [Subsection 1.4.5]
var	variance
Z^k	the sequence $z(j)$, $j = 1, \ldots, k$ [Subsection 2.2.1]
0	scalar zero, zero (null) vector, zero (null) matrix
$\emptyset$	empty set, impossible event [Subsection 1.4.1]

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 20

$\mathbf{1}$	unity, or a (column) vector with all elements unity
$1(\cdot)$	unit step function
$'$	transposition (of a matrix or a vector) [Subsection 1.3.1]
$(\cdot)^*$	complex conjugate and transposed [Subsection 1.4.23]
$\bar{x}$	expected value of x [Subsection 1.4.5],
$\hat{x}$	estimate of x [Subsection 2.2.1],
$\tilde{x}$	error corresponding to the estimate of x [Subsection 2.2.1],
$\bar{A}$	complement of set (or event) A [Subsection 1.4.1]
$\forall$	for all
$(\cdot\|\cdot)$	conditioning (for probabilities, estimates or covariances)
$\|\cdot\|$	determinant (of a matrix)
$\|\cdot\|$	magnitude (of a scalar)
$\|\|\cdot\|\|$	norm of a vector [Subsection 1.3.4]
$\|\|\cdot\|\|$	norm of a matrix [Subsection 1.3.6]
$\|\|\cdot\|\|$	norm of a random variable [Subsection 3.3.1]
$\perp$	orthogonal vectors [Subsection 1.3.4]
$\perp$	orthogonal random variables [Subsection 3.3.1]
$f(x)\|_{x=0}$	function evaluated at $x = 0$
$[a, b]$	closed interval between points a and b (includes the endpoints)
(a, b)	open interval between points a and b (without the endpoints)
$[a, b)$	semi-open (semi-closed) interval between points a and b
$[a_{ij}]$	a matrix whose component at i-row j-column is a_{ij} [Subsection 1.3.1]
$x_{[t_0,t_1]}$	the function $x(t)$ in the corresponding interval [Subsection 4.2.1]
$\{x(j)\}_{j=j_0}^{j_1}$	the sequence $x(j)$, $j = j_0, \ldots, j_1$ [Subsection 4.3.3]
$\{x(j), j = j_0, \ldots, j_1\}$	the set $\{x(j_0), x(j_0 + 1), \ldots, x(j_1)\}$
$\langle \cdot, \cdot \rangle$	inner product of vectors [Subsection 1.3.2]
$\langle \cdot, \cdot \rangle$	inner product of random variables [Subsection 3.3.1]
$\wedge$	the smallest of two variables [Subsection 10.2.4]
$\vee$	the largest of two variables [Subsection 10.2.4]
$\gg$	much larger than
$\triangleq$	equal by definition
$\equiv$	identically equal
$\approx$	approximately equal
$\sim$	distributed as [Subsection 1.4.13]
$\Longrightarrow$	implies
$\Longleftrightarrow$	if and only if — implies and is implied by
$\exists$	there exists
$\ni$	such that
$\in$	element of
$\cup$	logical "OR" operation (set union) [Subsection 1.4.1]
$\cap$	logical "AND" operation (set intersection) [Subsection 1.4.1]
∇_x	gradient with respect to (the vector) x

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 21

Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software
Yaakov Bar-Shalom, X. Rong Li, Thiagalingam Kirubarajan
Copyright © 2001 John Wiley & Sons, Inc.
ISBNs: 0-471-41655-X (Hardback) 0-471-22127-9 (Electronic)

Chapter 1

INTRODUCTION

1.1 BACKGROUND

1.1.1 Estimation and Related Areas

Estimation is the process of inferring the value of a quantity of interest from indirect, inaccurate and uncertain observations.

The purpose of estimation can be:

- Determination of planet orbit parameters — probably the first estimation problem — studied by Laplace, Legendre, and Gauss
- Statistical inference
- Determination of the position and velocity of an aircraft in an air traffic control system — tracking
- Application of control to a plant in the presence of uncertainty (noise and/or unknown parameters) — parameter identification, state estimation and stochastic control
- Determination of model parameters for predicting the state of a physical system or forecasting economic or other variables — system identification
- Determination of the characteristics of a transmitted message from noise-corrupted observation of the received signal — communication theory
- Determination of some parameters or characteristics of a signal or image — signal/image processing

More rigorously, ***estimation*** can be viewed as the *process of selecting a point from a continuous space* — the "best estimate."

Decision can be viewed as the *selection of one out of a set of discrete alternatives* — the "best choice" from a discrete space. However, one can talk about estimation in a discrete-valued case with the possibility of not making a choice but obtaining some conditional probabilities of the various alternatives. This information can be used without making "hard decisions."

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 22

Estimation and decision can, therefore, be seen to be overlapping and techniques from both areas are used simultaneously in many practical problems.

Tracking is the estimation of the state of a moving object based on *remote measurements*. This is done using one or more sensors at fixed locations or on moving platforms.

At first sight, tracking might seem to be a special case of estimation. However, it is wider in scope: Not only does it use all the tools from estimation, but it also requires extensive use of statistical decision theory when some of the practical problems (data association — "Which is my measurement?" — see, e.g., [Bar-Shalom95]) are considered.

Filtering is the estimation of the (current) state of a dynamic system — the reason for the use of the word "filter" is that the process for obtaining the "best estimate" from noisy data amounts to "filtering out" the noise. The term filtering is thus used in the sense of eliminating an undesired signal, which, in this case, is the noise.

In control systems, signals are also filtered to obtain the estimate of the state of the (noisy) dynamic system, needed by the controller. Filtering of signals is very commonly used in signal processing — in the frequency domain as well as in the spatial domain. The latter is done, for example, to select signals coming from a certain direction. ***Navigation*** is the estimation of the state of the platform ("own ship") on which the sensor is located.

An ***optimal estimator*** is a computational algorithm that processes observations (measurements) to yield an estimate of a variable of interest, which optimizes a certain criterion.

In view of the above discussion, estimation can be viewed as a scheme for ***information extraction and enhancement***: Based on measurements (observations), we want to maximize our knowledge about a parameter, a state, a signal, an image, and so on.

In general, one can classify the variable that is to be estimated into the following two categories:

- A parameter — a time-invariant quantity (a scalar, a vector, or a matrix)
- The state of a dynamic system (usually a vector), which evolves in time according to a stochastic equation

Consequently, one has two classes of estimators:

- Parameter estimators
- State estimators

While some engineering texts deal mainly (or exclusively) with state estimation, which is more general than parameter estimation, a balanced approach is taken here. The reason it is important to understand both is that some results — e.g., convergence and, more importantly, ***information limits*** — are more readily available/usable for parameter estimators than for state estimators. This

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 23

understanding can be helpful in choosing the type of mathematical model used for solving practical problems.

Figure 1.1.1-1 presents a concise block diagram that illustrates state estimation. In this figure the first two blocks are "black boxes" — there is no access to variables inside them. The only variables to which the estimator has access are the *measurements*, which are affected by the error sources in the form of "measurement noise."

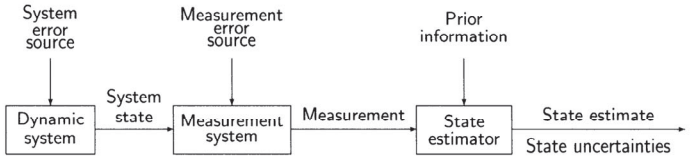


Figure 1.1.1-1: State estimation — information extraction and enhancement.

The estimator uses knowledge about the following:

- The evolution of the variable (the system dynamics)
- The measurement model
- The probabilistic characterization of the various random factors (disturbances)
- The prior information

Optimal Estimation — Advantages and Disadvantages

The advantage of an optimal estimator is that it makes the best utilization of the data and the knowledge of the system and the disturbances.

The disadvantages, like for any optimal technique, are that it is possibly sensitive to modeling errors and might be computationally expensive. In view of this, it is very important to have a clear understanding of the assumptions under which an algorithm is optimal and how they relate to the real world.

1.1.2 Applications of Estimation

State estimation, in combination with decision theory, has a great variety of applications. A partial list of the areas where this has found use is as follows:

- Tracking/surveillance and trajectory determination
- Navigation
- Control systems
  - guidance
  - attitude control

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 24

  - sensor pointing
  - steel making
  - chemical, nuclear, and industrial processes, etc.
- Power systems
- Failure detection
- Signal processing
- Image processing
- Communications
- Biomedical engineering
- Operations research
- Mapping via remote sensing
- Geophysical problems
- Fluid flow rate measurement
- Econometric systems
  - macroeconomic models
  - microeconomic models
- Demographic systems

1.1.3 Preview of Estimation/Filtering

This subsection will present a bird's-eye view of estimation, from the the Least Squares technique originated by Gauss and Legendre for parameter estimation in astronomical studies (motion of heavenly bodies) to Kalman filtering for state estimation. All this will be done without mathematics, to illustrate the concepts. [1]

Estimation is the process of inferring the values of some parameters that characterize, e.g., the motion of a heavenly body (the six parameters of planetary motion — 3 position and 3 velocity — evolving in a predictable manner) or the state of an object, e.g., an aircraft (evolving in a not completely predictable manner).

Gauss made the following (philosophical) observations on the (physical) observations that could be made on the planetary motion:

- If the observations were absolutely correct, the parameters could be determined with perfect accuracy from a minimum number of observations (n observations for n parameters).
- Then subsequent observations would confirm, but not correct, the values obtained for the parameters.

[1] The alternative title for this subsection could have been "Kalman Filtering for Dummies." This material could be used by readers eager to show their spouses how they make a living; however, we assume no responsibility for the possible ensuing marital problems.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 25

- But, since the observations are only approximations of the truth, we should combine more observations than the minimum to determine more accurately the unknown quantities.
- Thus, starting with approximate knowledge, we can correct it with subsequent observations so as to satisfy all the observations in the most accurate manner possible.

The above (philosophical) observations lead to the following implications:

- A basic description (model) of the "system" is available with some unknown parameters to be estimated.
- Redundant data are required to reduce the effect of measurement errors.
- To satisfy all the observations in the most accurate manner possible, the "residuals" (differences between the observed values and the values predicted from the estimates) should be as small as posible.
- Inaccuracy of observations necessitates probabilistic modeling.
- Combination of the initial knowledge and the subsequent observations leads to the recursive algorithm concept.

Estimation theory provides a systematic approach to deal with the discrepancies in different measurements on the same object.

There are a number of estimation problems that will be briefly outlined below.

The Least Squares (Gauss-Legendre) Problem

If the observations consist of

- linear combinations of the parameters of interest, in the presence of additive errors (noise) assumed to be zero-mean Gaussian random variables,

then the

- minimization of the sum of the squares of the residuals (errors) in fitting the data namely, the least squares (LS) of the difference between the model and the observations,

is equivalent to the

- maximization of the likelihood function of the parameters (the probability density function of the measurements conditioned on the unknown parameters).

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 26

The Wiener-Hopf Problem

The variables to be estimated are not constant but time-varying: The objective is to estimate a random process.

The process to be estimated, $x(t)$, is assumed stationary (typically, zero-mean) and with known autocorrelation. The cross-correlation between this process and the observations $z(t)$ is assumed also known.

The solution that yields the estimate $\hat{x}(t)$ of the process, obtained in the frequency domain as a transfer function $W_o(s)$ via Fourier transforms, so as to minimize a mean square error is the Wiener filter. This is illustrated in Fig. 1.1.3-1.

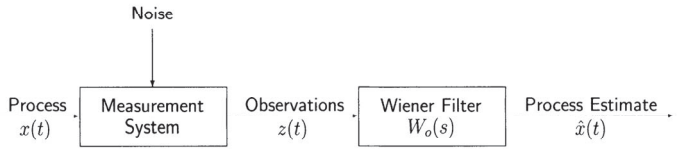


Figure 1.1.3-1: The Wiener filter — linear time-invariant system with transfer function $W_o(s)$.

The Kalman-Bucy Problem

The object of the estimation is a random process, which is not necessarily stationary. Its characterization is given by a linear differential (or difference) equation with known coefficients — the state equation of a dynamic system. Random disturbances affect the evolution of the process (process noise). The process is observed in the presence of measurement errors (measurement noise).

The solution obtained (in the time domain) so as to minimize a mean square error is the Kalman-Bucy or Kalman filter.

Solution to the Least Squares Estimation Problem

The standard (batch) solution to the Least Squares Estimation (LSE) problem consists of the following:

- Given k measurements (each an m-vector) of a constant parameter (an n-vector), the LS estimate of the unknown parameter is obtained by processing simultaneously the entire measurement set (a km-vector).
- To obtain the new estimate of the parameter when one gets a new measurement (the $(k+1)$th m-vector), one reprocesses the entire data, a $(k+1)m$-vector, i.e., a new problem of *larger size* is solved.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 27

The batch LSE solution can be rewritten (if the noise components in the k sets of measurements are uncorrelated) as a *recursion*

$$(\text{estimate})_{k+1} = (\text{estimate})_k + (\text{weighting})_{k+1} \times (\text{residual})_{k+1}$$

where k is the time index, $(\cdot)_k$ indicates the value at time k and

$$(\text{residual})_{k+1} = (\text{measurement})_{k+1} - (\text{predicted measurement})_{k+1} \text{ based on } (\text{estimate})_k$$

The size of this problem is *the same* for every k.

The accuracy of the estimate also obeys a recursion; that is, the "precision" of the estimate (the inverse of the parameter mean square error — MSE) evolves according to

$$(\text{precision})_{k+1} = (\text{precision})_k + (\text{information about the parameter in measurement})_{k+1}$$

Thus the "precision" increases (the MSE or variance decreases) strictly if the measurements contain information about the parameters — if the parameters are *observable*.

If the parameters are constant, as $k \to \infty$ one accumulates an unlimited amount of information and the error in the estimates decreases to zero — the LSE converges to the true value. As the MSE in the estimate decreases, so does the weighting in the recursion: as the confidence in the estimates grows, new data are weighted less and less.

Estimation of the State of a Linear Stochastic Dynamic System

A realistic model for the state of a system has typically

- an imperfect initial estimate
- an imperfectly predictable evolution (a stochastic disturbance affects its dynamics — the process noise)

Similarly, the measurement model is as follows:

- Only some state components (or combinations of them) are observed.
- These observations are inaccurate — corrupted by measurement noise.

The evolution model for the state (discrete time dynamic equation) is as follows:

$$(\text{state})_{k+1} = (\text{linear function of the state})_k + (\text{process noise})_k$$

where k is the time index and $(\cdot)_k$ indicates the value at time k.

The observation process (measurement equation) is

$$(\text{measurement})_k = (\text{linear function of the state})_k + (\text{measurement noise})_k$$

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 28

The two noise sequences are assumed to have known means (typically zero) and variances (covariance matrices in the mutidimensional case), uncorrelated in time and from each other.

The (discrete time) Kalman filter (KF) computes the best estimate of the current state (in the minimum mean square error sense) based on the measurements up to the current time, denoted as $(\text{state estimate})_k$. The KF recursion is as follows:

$$(\text{state estimate})_{k+1} = (\text{predicted state})_{k+1^-} + (\text{weighting})_{k+1} \times (\text{residual})_{k+1}$$

where $k+1^-$ denotes "just before time $k+1$" and

$$(\text{residual})_{k+1} = (\text{measurement})_{k+1} - (\text{predicted measurement})_{k+1^-}$$

The LSE recursion is remarkably similar:

$$(\text{parameter estimate})_{k+1} = (\text{parameter estimate})_k + (\text{weighting})_{k+1} \times (\text{residual})_{k+1}$$

The accuracy of the estimate (precision — inverse of the variance) is as follows:

$$(\text{precision})_{k+1} = (\text{precision})_{k+1^-} + (\text{information about the state in measurement})_{k+1}$$

The filter gain reflects the relative accuracy of the predicted state vs. the new observation. The new (updated) state estimate is the optimal combination of

- the entire past data (with the predicted state being the sufficient statistic) and
- the latest measurement

The sequence of the operations in the KF is outlined in Fig. 1.1.3-2. The calculations involving the state estimate use the measurements and have to be done in real time. However, the calculations involving the variances (covariance matrices in the mutidimensional case) are data-independent and can be done off-line (before the data are obtained). The variance of the estimates (their "precision") can be precomputed using the statistical description (variances) of the noise components entering into the system and the initial state variance.

Illustration for Tracking a Moving Point

Figure 1.1.3-3 illustrates graphically the state estimation process for a point moving with a nearly constant velocity along the x axis. The state vector consists of position x and velocity $\dot{x}$, with estimates $\hat{x}$ and $\hat{\dot{x}}$, respectively. The

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 29

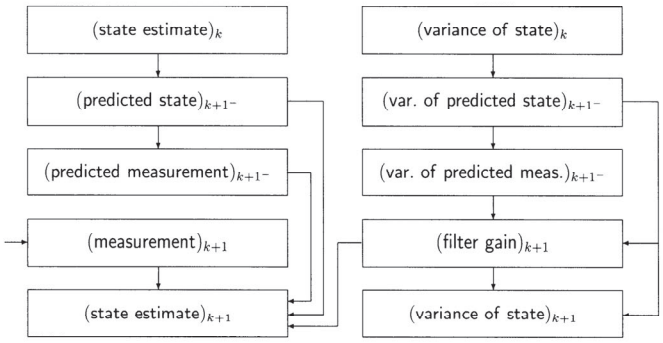


Figure 1.1.3-2: Sequence of operations in the KF.

covariance matrix of the state is denoted by P and it indicates the uncertainty region around the estimated state where the true state lies with high probability.

The measurement is

$$z_{k+1} = x_{k+1} + (\text{measurement noise})_{k+1}$$

The new position estimate will be between the observed and predicted values:

$$z_{k+1} < \hat{x}_{k+1} < \hat{x}_{k+1^-}$$

Assuming that at time $k+1$ the measured position is behind the predicted position, that is,

$$z_{k+1} - \hat{x}_{k+1^-} < 0$$

the new velocity estimate will be less then the predicted one:

$$\hat{\dot{x}}_{k+1} < \hat{\dot{x}}_{k+1^-}$$

Under stationarity (and state observability) conditions, the variance of the estimates converges to a steady-state value, as illustrated in Fig. 1.1.3-4, where the prediction and updated variances are shown.

Remarks

The KF is a stochastic counterpart of the (deterministic) state observer.

When the covariance is in steady state, the KF gain becomes a constant — such filters (for "nearly constant velocity" motion models) are known as α-β filters.

The Wiener filter is the frequency domain representation of the steady-state KF.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 30

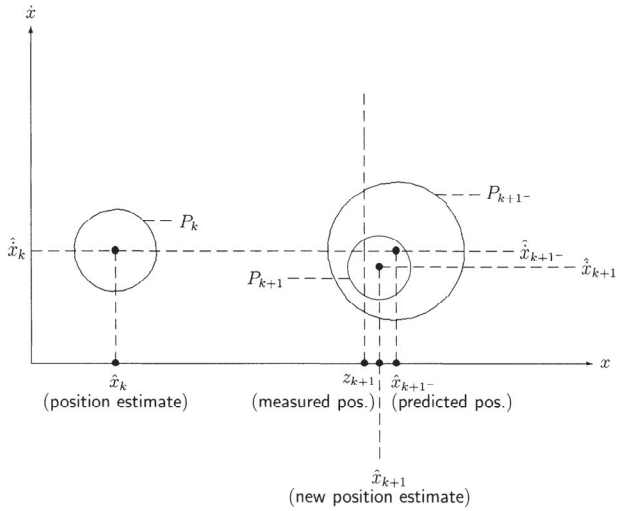


Figure 1.1.3-3: Illustration of the state estimation process.

The sampling interval (time between successive measurements) does not have to be constant in the KF.

Nonstationary disturbances (e.g., measurement noise with time-varying intensity) can be accomodated easily by the KF algorithm.

1.1.4 An Example of State Estimation: Vehicle Collision Avoidance

This subsection introduces the problem of state estimation for a dynamic system by considering the estimation of distance and speed between two vehicles as it arises in a collision avoidance system.

An algorithm is presented that estimates the distance (range) between two vehicles and their relative speed (range rate), based on imperfect (noisy) measurements of the distance between them. This algorithm is a simplified version of the Kalman filter, known as the α-β filter. It can also be viewed as a state observer [Chen84], which, based on observations of one state component, reconstructs (estimates) the entire state vector.

Define the state vector consisting of range r and range rate $\dot{r}$ as

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} r \\ \dot{r} \end{bmatrix} \tag{1.1.4-1}$$

Assuming a constant speed (constant range-rate), the state equation is, for

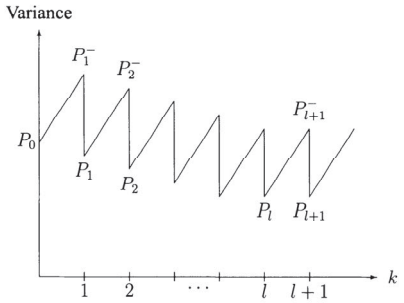


Figure 1.1.3-4: Evolution of the variance of the estimate with time.

discrete time with sampling period T,

$$x(k) = \begin{bmatrix} r(k) \\ \dot{r}(k) \end{bmatrix} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} x(k-1) = Fx(k-1) \tag{1.1.4-2}$$

where the system matrix is

$$F = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \tag{1.1.4-3}$$

The measurement is

$$y(k) = x_1(k) = r(k) = [1 \;\; 0]x(k) = Hx(k) \tag{1.1.4-4}$$

where the measurement matrix is

$$H = [1 \;\; 0] \tag{1.1.4-5}$$

In truth, the measurement is a "corrupted " (noisy) version of (1.1.4-4), i.e.,

$$y(k) = Hx(k) + w(k) \tag{1.1.4-6}$$

where $w(k)$ is the (range) measurement error.

Similarly, the state equation should reflect an acceleration, denoted as u, namely,

$$x(k) = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} x(k-1) + \begin{bmatrix} T^2/2 \\ T \end{bmatrix} u(k-1) \tag{1.1.4-7}$$

Using (1.1.4-2) instead of (1.1.4-7), since we do not know u, as the "motion model" for the estimation problem leads to motion modeling errors.

The Estimator

In view of the fact that there are both measurement errors and motion modeling errors, we shall calculate the "estimate" $\hat{x}(k)$ of the state vector $x(k)$ via the

following algorithm, which is a simplified form of the Kalman filter, known as the α-β filter (see Chapters 5 and 6).

$$\hat{x}(k) - \bar{x}(k) + W[y(k) - \bar{y}(k)] \tag{1.1.4-8}$$

where $\bar{x}(k)$ is the "predicted state" at time k from $k-1$

$$\begin{aligned} \bar{x}(k) = F\hat{x}(k-1) &= \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \hat{x}(k-1) = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{r}(k-1) \\ \hat{\dot{r}}(k-1) \end{bmatrix} \\ &= \begin{bmatrix} \hat{r}(k-1) + T\hat{\dot{r}}(k-1) \\ \hat{\dot{r}}(k-1) \end{bmatrix} \end{aligned} \tag{1.1.4-9}$$

and $y(k)$ is the "predicted measurement" at k from $k-1$

$$\bar{y}(k) = H\bar{x}(k) = [1 \;\; 0]\bar{x}(k) = \hat{r}(k-1) + T\hat{\dot{r}}(k-1) \tag{1.1.4-10}$$

The difference between the measurement (observed range) and its predicted value

$$y(k) - \bar{y}(k) = y(k) - H\bar{x}(k) \tag{1.1.4-11}$$

is the "residual," which is multiplied (weighted) by the "filter gain"

$$W = \begin{bmatrix} \alpha \\ \beta/T \end{bmatrix} \tag{1.1.4-12}$$

The residual multiplied by the gain is used in (1.1.4-8) to "correct" or "update" the predicted state to obtain the new estimate $\hat{x}(k)$. The filter (1.1.4-8) can be seen as a ***predictor-corrector algorithm***.

The (dimensionless) constants α and β are such that $0 < \alpha < 1$ and $0 < \beta < 2$. The filter gain coefficients α and β define the α-β filter (1.1.4-8).

The breakdown of the state estimation equation (1.1.4-8) for its two components r and $\dot{r}$ is

$$\hat{r}(k) = \hat{r}(k-1) + T\hat{\dot{r}}(k-1) + \alpha[y(k) - \bar{y}(k)] \tag{1.1.4-13}$$

$$\hat{\dot{r}}(k) = \hat{\dot{r}}(k-1) + \frac{\beta}{T}[y(k) - \bar{y}(k)] \tag{1.1.4-14}$$

The intuitive interpretation of (1.1.4-13) is as follows:

- If the predicted range is larger than the measured one, the residual will be negative and a negative correction is made with gain $\alpha < 1$ to obtain the new estimated range.
- If the predicted range is smaller than the measured one, the residual will be positive, and a positive correction will be made.

Similarly, the range rate (1.1.4-14) is corrected downward for a negative residual and upward for a positive one. Also note that the residual has the dimension of length and, when multiplied by β/T, its dimension becomes length/time, i.e., velocity.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 33

A Deterministic Error Analysis

Assuming a constant acceleration (or deceleration) u, the state equation is, based on (1.1.4-7), given by

$$x(k) = Fx(k-1) + Gu \tag{1.1.4-15}$$

where

$$G = \begin{bmatrix} T^2/2 \\ T \end{bmatrix} \tag{1.1.4-16}$$

reflects the effect of the acceleration u on the range ($uT^2/2$) and on the range rate (uT).

Assuming the measurements to be given by (1.1.4-4), i.e., neglecting measurement errors, the evaluation of the effect of the acceleration on the errors in the range and range-rate estimates is carried out next.

The error is defined as the difference between the actual value and the estimate, namely,

$$\tilde{x}(k) - x(k) - \hat{x}(k) \tag{1.1.4-17}$$

From (1.1.4-8), using (1.1.4-9) and (1.1.4-10), the estimate is given by

$$\hat{x}(k) = F\hat{x}(k-1) + W[y(k) - HF\hat{x}(k-1)] \tag{1.1.4-18}$$

Substituting (1.1.4-4) for $y(k)$ in the above yields

$$\hat{x}(k) = F\hat{x}(k-1) + WH[x(k) - F\hat{x}(k-1)] \tag{1.1.4-19}$$

Substituting (1.1.4-15) for $x(k)$ in the above yields

$$\hat{x}(k) - F\hat{x}(k-1) + WH[Fx(k-1) + Gu - F\hat{x}(k-1)] \tag{1.1.4-20}$$

or, using notation (1.1.4-17),

$$\hat{x}(k) = F\hat{x}(k-1) + WHF\tilde{x}(k-1) + WHGu \tag{1.1.4-21}$$

Subtracting (1.1.4-21) from (1.1.4-15) yields

$$x(k) - \hat{x}(k) = Fx(k-1) + Gu - F\hat{x}(k-1) - WHF\tilde{x}(k-1) - WHGu \tag{1.1.4-22}$$

which can be rewritten as

$$\tilde{x}(k) = [I - WH]F\tilde{x}(k-1) + [I - WH]Gu \tag{1.1.4-23}$$

Note that the dynamic equation obeyed by the error $\tilde{x}$ is a state equation with system matrix

$$F = [I - WH]F = \begin{bmatrix} 1-\alpha & (1-\alpha)T \\ -\frac{\beta}{T} & 1-\beta \end{bmatrix} \tag{1.1.4-24}$$

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 34

and input with gain

$$\tilde{G} = [I - WH]G = \begin{bmatrix} (1-\alpha)T^2/2 \\ (1-\frac{\beta}{2})T \end{bmatrix} \tag{1.1.4-25}$$

i.e.,

$$\tilde{x}(k) = \tilde{F}\tilde{x}(k-1) + \tilde{G}u \tag{1.1.4-26}$$

It can be shown that the inequalities indicated earlier for α and β guarantee that the matrix $\tilde{F}$ above is stable, i.e., it has eigenvalues inside the unit circle. In other words, for a constant "disturbance" u, the error $\tilde{x}(k)$ converges to a *steady-state* value $\tilde{x}_{ss}$, which is given by the expression

$$\tilde{x}_{ss} = (I - \tilde{F})^{-1}\tilde{G}u = \begin{bmatrix} \frac{1-\alpha}{\beta}T^2 \\ \frac{2\alpha-\beta}{2\beta}T \end{bmatrix} u \tag{1.1.4-27}$$

For $u = 5\,\text{m/s}^2$, $T = 0.1\,\text{s}$, $\alpha = 0.75$, $\beta = 0.5$, one obtains $\tilde{r}_{ss} = 0.39\,\text{m}$, $\dot{\tilde{r}}_{ss} = 0.5\,\text{m/s}$.

Frequency Domain View of Estimation: The Bandwidth of a Filter

Based on (1.1.4-23), one has the following z-domain transfer function from the input to the error $\tilde{x}$:

$$\tilde{H}(z) = \frac{\tilde{X}(z)}{U(z)} = (zI - \tilde{F})^{-1}\tilde{G} \tag{1.1.4-28}$$

With the above numbers,

$$\tilde{F} = \begin{bmatrix} 0.25 & 0.025 \\ -5 & 0.5 \end{bmatrix} \qquad \tilde{G} = \begin{bmatrix} 0.005 \\ 0.1 \end{bmatrix} \tag{1.1.4-29}$$

the transfer function from the input to the position component error is

$$\tilde{H}_1(z) = \frac{\tilde{X}_1(z)}{U(z)} = \frac{0.005z}{z^2 - 0.75z + 0.25} \tag{1.1.4-30}$$

The two poles of $\tilde{H}_1(z)$ have magnitude 1/2, i.e., this is a stable transfer function.

Figure 1.1.4-1 shows the plot of the magnitude of the above transfer function for the range of frequencies $[0, \frac{1}{2T}]$ (the upper limit is the point for which the Nyquist frequency is the sampling frequency). It can be clearly seen that this is a low-pass filter with a bandwidth (based on the half-power point) of approximately 2 Hz.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 35

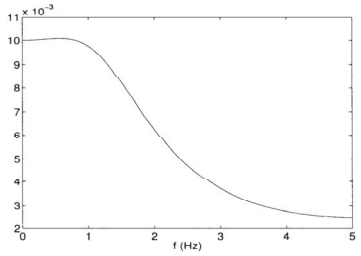


Figure 1.1.4-1: The frequency response of the state estimation filter.

1.2 SCOPE OF THE TEXT

1.2.1 Objectives

The objectives of this text are to present the fundamentals of *state estimation theory* and the tools for the design of state-of-the-art algorithms for *target tracking and navigation*.

The text covers the basic concepts and estimation techniques for static and dynamic systems, linear and nonlinear, as well as adaptive estimation. This constitutes a one-semester graduate course in estimation theory in an electrical/systems engineering program.

Special emphasis is given to the statistical tools that can be used for the *interpretation of the output from stochastic systems*. These are key tools for the assessment of the performance of state estimation/tracking filters — in simulations as well as in real-time implementation.

The discussion deals mainly with discrete time estimation algorithms, which are natural for digital computer implementation. Also, typically, sensors provide observations in discrete time. The basic state estimation algorithm — the Kalman filter — is presented in discrete as well as in continuous time.

The use of the estimation algorithms is illustrated on kinematic motion models because they reveal all the major issues and, in particular, the subtleties encountered in estimation. Also this serves as an introduction to tracking.

Guidelines for the *design of tracking filters*, namely,

- selection of the filter design parameters
- selection of the type of filter — nonadaptive (single model based) vs. adaptive (multiple model based)

are given and illustrated in several examples. Most of the examples are run with the companion software DynaEst^{TM}; also, a step-by-step demonstration of how to use DynaEst^{TM} is provided.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 36

Prerequisite Material

The *linear algebra tools* that form the backbone of the state space analysis techniques used in estimation/tracking are presented in Section 1.3 together with some useful results from linear systems. Matrix notation will be used throughout this text — every quantity should be considered a matrix, unless otherwise stated.

A review of the *tools from probability theory and random processes* [2] is presented in Section 1.4.

Section 1.5 reviews *statistical hypothesis testing* and presents some useful statistical tables for the chi-square and the normal (Gaussian) distributions. Particular attention is given to the concept of statistical significance, which, while common in the statistics literature, has not received much attention in the engineering literature.

These review sections cover all the background material needed for the text and the basic concepts are illustrated by examples. This might be used as a (crash) minicourse to bring advanced undergraduate students up to the level necessary to understand the main material of the text.

1.2.2 Overview and Chapter Prerequisites

In the following an overview of the text is given with the prerequisite to each chapter.

Review of Background Techniques (remainder of Chapter 1)

- Linear algebra
- Probability theory and stochastic processes
- Statistics

Basic Concepts in Estimation — Chapter 2 (prerequisite: 1)

- Estimation techniques (ML, MAP, LS and MMSE)
- Properties of estimators
- Limit of existing information in the data

Linear Estimation in Static Systems — Chapter 3 (prerequisite: 2)

- Techniques (LMMSE, recursive LS and prelude to Kalman filtering)
- Application to polynomial fitting

Linear Dynamic Systems with Random Inputs — Chapter 4 (prerequisite: 1)

- Models — continuous and discrete time state space models
- Response of dynamic systems to random inputs

[2] Probability demonstrates both our knowledge as well as our ignorance.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 37

State Estimation in Discrete-Time Linear Dynamic Systems — Chapter 5 (prerequisites: 3, 4)

- The Kalman filter and its properties
- Use of the companion software DynaEst™ to design and evaluate a KF
- Consistency of state estimators
- Initialization — in simulations and in practice

Estimation for Kinematic Models — Chapter 6 (prerequisite: 5)

- Types of kinematic models
- Explicit filters and the target maneuvering index
- Filter design and system performance prediction
- Sampling interval selection (system design)

Computational Aspects of Estimation — Chapter 7 (prerequisite: 5)

- The Information filter
- Sequential update
- Square-root filtering

Extensions of Discrete-Time Linear Estimation — Chapter 8 (prerequisite: 5)

- Correlated process noise
- Cross-correlated process and measurement noise sequences
- Correlated measurement noise
- Prediction
- Smoothing

Continuous-Time Linear State Estimation — Chapter 9 (prerequisite: 5)

- The Kalman-Bucy filter and its properties
- Prediction
- Duality between estimation and control
- The Wiener-Hopf problem

State Estimation for Nonlinear Dynamic Systems — Chapter 10 (prerequisite: 5)

- Optimal estimation
- Suboptimal estimator — the extended Kalman filter (EKF)
- Practical issues in implementation
- Trajectory estimation via dynamic programming

Adaptive Estimation and Maneuvering Targets — Chapter 11 (prerequisite: 5)

- White and colored noise models
- Input estimation
- Variable state dimension filtering

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 38

- Comparison of several methods
- The multiple model approach — static and dynamic, the Interacting Multiple Model (IMM) estimator
- Use of the companion software DynaEst™ to design and evaluate an IMM estimator
- An air traffic control (ATC) tracking algorithm design example with the IMM estimator
- Estimator selection — nonadaptive vs. adaptive
- Use of the EKF for parameter estimation

Estimation for Navigation Systems — Chapter 12 (prerequisites: 5 and, for the last part, 11)

- Inertial navigation systems (INS)
- Global Positioning System (GPS)
- Combination of INS and GPS
- Use of the KF and IMM estimator for GPS navigation

At the end of each chapter, a number of problems that enhance the understanding of the theory and connect the theoretical material to the real world are given.

The sequence of prerequisites is illustrated in Fig. 1.2.2-1.

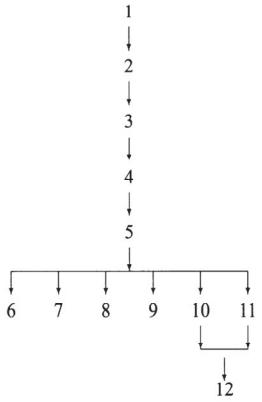


Figure 1.2.2-1: Sequence of chapter prerequisites.

Two major examples of realistic systems are presented:

- A bearings-only target localization from a moving platform based on the Maximum Likelihood technique
- An ATC problem for tracking maneuvering aircraft using an adaptive estimation technique (the IMM estimator)

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 39

As part of the problems at the end of Chapter 3, a project is given to develop an interactive program for the first example listed above.

A major term project that involves a realistic ATC tracking problem (with nonlinear measurements from a radar — range and azimuth) is given at the end of Chapter 11. The sequence of prerequisites for this is Chapters 1–6, Section 10.3 and Section 11.6.

An alternative major term project that involves a realistic ATC navigation problem with GPS is given at the end of Chapter 11. The sequence of prerequisites for this is Chapters 1–6, Section 10.3 and Sections 12.6–12.8.

1.3 BRIEF REVIEW OF LINEAR ALGEBRA AND LINEAR SYSTEMS

1.3.1 Definitions and Notations

A ***matrix*** of dimension $n \times m$ is the two-dimensional array

$$A = [a_{ij}] = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix} \tag{1.3.1-1}$$

The first dimension is the number of rows, and the second dimension is the number of columns. The elements of A will be denoted as a_{ij} or as A_{ij}.

An n-dimensional ***vector*** is the one-dimensional array ($n \times 1$ matrix)

$$a = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \triangleq \text{col}(a_i) \tag{1.3.1-2}$$

In this text all vectors will be ***column vectors***.

Transposition of a matrix or vector will be denoted by an apostrophe. The *transpose* of the matrix (1.3.1-1) is

$$A' = [a_{ji}] = \begin{bmatrix} a_{11} & \cdots & a_{n1} \\ \vdots & \ddots & \vdots \\ a_{1m} & \cdots & a_{nm} \end{bmatrix} \tag{1.3.1-3}$$

With this notation one can write the column vector (1.3.1-2) as

$$a = [a_1 \ \cdots \ a_n]' \tag{1.3.1-4}$$

The ***row vector*** obtained from transposing (1.3.1-2) is

$$a' = [a_1 \ \cdots \ a_n] \tag{1.3.1-5}$$

A (square) matrix is said to be ***symmetric*** if

$$A = A' \tag{1.3.1-6}$$

or

$$a_{ij} = a_{ji} \qquad \forall i, j \tag{1.3.1-7}$$

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 40

1.3.2 Some Linear Algebra Operations

Addition of matrices and the multiplication of a matrix by a scalar are defined as follows. With α and β scalars, the matrix

$$C = \alpha A + \beta B \tag{1.3.2-1}$$

has elements given by

$$c_{ij} = \alpha a_{ij} + \beta b_{ij} \qquad i = 1, \ldots, n; \;\; j = 1, \ldots, m \tag{1.3.2-2}$$

where all the matrices have the same dimension $n \times m$.

The product of two matrices

$$C = AB \tag{1.3.2-3}$$

has elements

$$c_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj} \qquad i = 1, \ldots, n; \;\; j = 1, \ldots, p \tag{1.3.2-4}$$

where A is $n \times m$, B is $m \times p$, and the result C is $n \times p$. The matrix product is, in general, not commutative.

The ***transpose of a product*** is

$$C' = (AB)' = B'A' \tag{1.3.2-5}$$

Thus, if

$$Ab = c \tag{1.3.2-6}$$

where A is $n \times m$, b is $m \times 1$ (i.e., an m-vector) and c is $n \times 1$ (n-vector), then

$$c' = b'A' \tag{1.3.2-7}$$

where c' is $1 \times n$ (n-vector in row form), etc. The ***inner product*** of two (real) n-vectors in a Euclidean space is

$$\langle a, b \rangle \triangleq a'b = \sum_{i=1}^{n} a_i b_i \tag{1.3.2-8}$$

The inner product has a counterpart in probability theory where it is used for random variables, which can be looked upon as vectors in a suitable space. This has applications in linear minimum mean square error estimation.

The ***outer product*** of two vectors is the matrix

$$ab' = C = [c_{ij}] \qquad \text{with} \qquad c_{ij} = a_i b_j \tag{1.3.2-9}$$

The ***trace*** of an $n \times n$ matrix is the sum of its diagonal elements

$$\text{tr}(A) \triangleq \sum_{i=1}^{n} a_{ii} = \text{tr}(A') \tag{1.3.2-10}$$

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 41

It can be easily shown[3] that if A is $m \times n$ and B is $n \times m$, then

$$\text{tr}(AB) = \text{tr}(BA) \tag{1.3.2-11}$$

For example, if a and b are both n-vectors, then

$$a'b = \text{tr}(a'b) = \text{tr}(ba') \tag{1.3.2-12}$$

The first equality above is due to the fact that the trace of a scalar is itself.

It can be shown[4] that for matrices of suitable dimension,

$$\text{tr}(ABC) = \text{tr}(BCA) = \text{tr}(CAB) \tag{1.3.2-13}$$

that is, the trace is *invariant under circular permutations* in its argument. A (square) matrix is ***idempotent*** if and only if[5] for all positive integers n

$$A^n = A \tag{1.3.2-14}$$

The ***square root*** of the square matrix A is the (in general, nonunique) matrix $A^{1/2}$, such that

$$A^{1/2}A^{1/2} = A \tag{1.3.2-15}$$

A symmetric matrix A can also be decomposed as

$$A = CC' = LDL' \tag{1.3.2-16}$$

where C is ***lower triangular*** (with zeros above the diagonal), L is ***unit lower triangular*** (with zeros above the diagonal and units on the diagonal) and D is ***diagonal***. The first decomposition above is the ***Cholesky factorization*** whereas the last form above is known as the ***LDL′ factorization***, or, since $L' = U$ is a unit upper triangular matrix, the ***LDU factorization***.

1.3.3 Inversion and the Determinant of a Matrix

The ***inverse*** A^{-1} of the $n \times n$ matrix A is such that

$$A^{-1}A = I \tag{1.3.3-1}$$

where

$$I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \triangleq \text{diag}(1) \tag{1.3.3-2}$$

[3] Also known in mathematical circles as the ICBES argument.
[4] Also known in mathematical circles as the ICBS argument.
[5] The standard shorthand for "if and only if" in mathematical texts is "***iff***."

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 42

is the ***identity matrix*** of the same dimension as A.

The inverse is given by the expression

$$A^{-1} = \frac{1}{|A|} C' \tag{1.3.3-3}$$

where C, called the ***adjugate*** of A, is the matrix of cofactors of A and $|A|$ is the ***determinant*** of A. The cofactor (i, j) of A is the determinant of the matrix A without its row i and column j, multiplied by $(-1)^{i+j}$.

The inverse of a (square) matrix exists if and only if its columns a_i, $i = 1, \ldots, n$ (or its rows) are ***linearly independent***, that is,

$$\sum_{i=1}^{n} \alpha_i a_i = 0 \qquad \Longleftrightarrow \qquad \alpha_i = 0 \quad i = 1, \ldots, n \tag{1.3.3-4}$$

where 0 denotes the zero vector (of dimension n). This is equivalent to the determinant in (1.3.3-3) being nonzero.

An invertible matrix is also called ***nonsingular***.

The determinant of an $n \times n$ matrix multiplied by a scalar is

$$|\alpha A| = \alpha^n |A| \tag{1.3.3-5}$$

This is useful in writing the probability density function of Gaussian random vector.

Inversion of a Partitioned Matrix

The inverse of the (nonsingular) $n \times n$ ***partitioned matrix***

$$\begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}^{-1} = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} \tag{1.3.3-6}$$

where P_{11} is $n_1 \times n_1$, P_{12} is $n_1 \times n_2$, P_{21} is $n_2 \times n_1$, P_{22} is $n_2 \times n_2$ and $n_1 + n_2 = n$, has the partitions

$$V_{11} = P_{11}^{-1} + P_{11}^{-1} P_{12} V_{22} P_{21} P_{11}^{-1} = (P_{11} - P_{12} P_{22}^{-1} P_{21})^{-1} \tag{1.3.3-7}$$

$$V_{12} = -P_{11}^{-1} P_{12} V_{22} = -V_{11} P_{12} P_{22}^{-1} \tag{1.3.3-8}$$

$$V_{21} = -V_{22} P_{21} P_{11}^{-1} = -P_{22}^{-1} P_{21} V_{11} \tag{1.3.3-9}$$

$$V_{22} = P_{22}^{-1} + P_{22}^{-1} P_{21} V_{11} P_{12} P_{22}^{-1} = (P_{22} - P_{21} P_{11}^{-1})^{-1} P_{12} \tag{1.3.3-10}$$

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 43

The Matrix Inversion Lemma

Another useful result is the following identity known as the ***matrix inversion lemma***,

$$(P^{-1} + H'R^{-1}H)^{-1} = P - PH'(HPH' + R)^{-1}HP \tag{1.3.3-11}$$

where P is $n \times n$, H is $m \times n$, and R is $m \times m$.

An alternative version of the above is

$$(A + BCB')^{-1} = A^{-1} - A^{-1}B(B'A^{-1}B + C^{-1})^{-1}B'A^{-1} \tag{1.3.3-12}$$

It can be shown that (1.3.3-6) to (1.3.3-10) and (1.3.3-11) hold by verifying that the corresponding multiplications will yield the identity matrix.

These results have direct application in the derivation of the recursive form of the least squares estimation of parameters as well as in linear estimation for dynamic systems — they yield various forms of the Kalman filter.

1.3.4 Orthogonal Projection of Vectors

The inner product of a vector with itself

$$\langle a, a \rangle \triangleq \|a\|^2 \tag{1.3.4-1}$$

is the *squared* l_2 ***norm*** of this vector. This applies for the inner product defined in (1.3.2-8) or for any other properly defined inner product. The l_2 norm is the ***magnitude*** of the vector or its ***length***.

The ***Schwarz inequality*** states the following relationship between the magnitude of the inner product of two vectors and their norms

$$|\langle a, b \rangle| \leq \|a\| \, \|b\| \tag{1.3.4-2}$$

Two vectors are ***orthogonal***, which is denoted as

$$a \perp b \tag{1.3.4-3}$$

if

$$\langle a, b \rangle = 0 \tag{1.3.4-4}$$

The ***orthogonal projection*** of the vector a on b is

$$\Pi_b(a) = \frac{\langle a, b \rangle}{\|b\|^2} \, b \tag{1.3.4-5}$$

It can be easily shown that the difference vector between a and its orthogonal projection on b is orthogonal on b, that is,

$$[a - \Pi_b(a)] \perp b \tag{1.3.4-6}$$

Figure 1.3.4-1 illustrates this.

The orthogonal projection is used to obtain the linear minimum mean square error (LMMSE) estimates of random variables.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 44

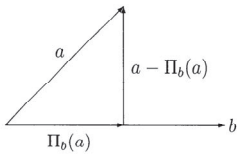


Figure 1.3.4-1: Orthogonal projection of vector a on vector b.

1.3.5 The Gradient, Jacobian and Hessian

The ***gradient*** operator with respect to the n-vector x is defined as

$$\nabla_x = \left[\frac{\partial}{\partial x_1} \cdots \frac{\partial}{\partial x_n}\right]' \tag{1.3.5-1}$$

The gradient of an m-dimensional vector-valued function $f(x)$ is

$$\nabla_x f(x)' = \begin{bmatrix} \dfrac{\partial}{\partial x_1} \\ \vdots \\ \dfrac{\partial}{\partial x_n} \end{bmatrix} [f_1(x) \cdots f_m(x)] = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_m}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_1}{\partial x_n} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{bmatrix} \tag{1.3.5-2}$$

The transpose of the above is the ***Jacobian matrix***, an $m \times n$ matrix in this case

$$f_x(x) \triangleq \frac{\partial f}{\partial x} = [\nabla_x f(x)']' = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{bmatrix} \tag{1.3.5-3}$$

The dimensions m and n are usually (but not necessarily) equal.

It can be easily shown that

$$\nabla_x x' = I \tag{1.3.5-4}$$

and, if A is a symmetric matrix,

$$\nabla_x (x'Ax) - 2Ax \tag{1.3.5-5}$$

The ***Hessian*** of the scalar function $\phi(x)$ with respect to the n-vector x is

$$\phi_{xx}(x) \triangleq \frac{\partial^2 \phi(x)}{\partial x^2} = \nabla_x \nabla_x' \phi(x) = \begin{bmatrix} \dfrac{\partial^2 \phi}{\partial x_1 \partial x_1} & \cdots & \dfrac{\partial^2 \phi}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 \phi}{\partial x_n \partial x_1} & \cdots & \dfrac{\partial^2 \phi}{\partial x_n \partial x_n} \end{bmatrix} \tag{1.3.5-6}$$

which is, obviously, a symmetric $n \times n$ matrix.

These results are used in the linear estimation of the state of nonlinear systems via series expansion.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 45

1.3.6 Eigenvalues, Eigenvectors, and Quadratic Forms

Eigenvalues and Eigenvectors

The ***eigenvalues*** of an $n \times n$ matrix A are the scalars λ_i such that

$$Au_i = \lambda_i u_i \qquad i = 1, \ldots, n \tag{1.3.6-1}$$

where the vectors u_i are the corresponding ***eigenvectors***. If the eigenvalues are distinct, then there are n eigenvectors; otherwise the number of the linearly independent eigenvectors might be less than n.

A matrix is ***nonsingular*** if and only if all its eigenvalues are nonzero. This follows from the fact that its determinant is the product of its eigenvalues

$$|A| = \prod_{i=1}^{n} \lambda_i \tag{1.3.6-2}$$

The ***rank*** of a (not necessarily square) matrix is equal to the number of its linearly independent rows or columns. For a square $(n \times n)$ matrix, this equals the number of its nonzero eigenvalues.

A nonsingular matrix must have ***full rank***. In the case of a full rank $n \times n$ matrix, its columns (and rows) span the n-dimensional space.

The eigenvalues of a real matrix can be real or complex; however, a *symmetric* real matrix has only real eigenvalues.

It can be shown that the trace of a matrix equals the sum of its eigenvalues

$$\text{tr}(A) = \sum_{i=1}^{n} \lambda_i \tag{1.3.6-3}$$

An ***idempotent*** matrix, defined in (1.3.2-14), that is, with the property

$$AA = A \tag{1.3.6-4}$$

has eigenvalues that are *either zero or unity*.

Quadratic Forms

The (scalar) function of the real vector x

$$q = x'Ax \tag{1.3.6-5}$$

is called a ***quadratic form***. It can be easily shown that, without loss of generality, in a quadratic form the matrix A can be considered symmetric.

A ***positive (semi)definite quadratic form*** is one which is positive (nonnegative) for all nonzero vectors x.

A ***positive (semi)definite matrix*** is one for which the quadratic form (1.3.6-5) is positive (semi)definite. This can be summarized as follows:

$$A > 0 \qquad \Longleftrightarrow \qquad x'Ax > 0 \quad \forall x \neq 0 \tag{1.3.6-6}$$

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 46

$$A \geq 0 \qquad \Longleftrightarrow \qquad x'Ax \geq 0 \quad \forall x \neq 0 \tag{1.3.6-7}$$

If the matrix A is positive definite, the expression on the right hand side of (1.3.6-5) can also be called the ***squared norm with respect to*** A of the vector x and denoted sometimes as $\|x\|_A^2$.

A matrix is positive (semi)definite if and only if all its eigenvalues are positive (nonnegative).

The ***inequality of two matrices*** is defined as follows: the matrix A is smaller (not larger) than the matrix B if and only if the difference $B - A$ is positive (semi)definite. The outer product of a vector with itself, aa', called a ***dyad***, is a positive semidefinite matrix (it has rank one).

The sum of dyads of n-vectors a_i with positive weights

$$\tilde{A} = \sum_{i=1}^{m} \alpha_i a_i a_i' \qquad \alpha_i > 0 \tag{1.3.6-8}$$

is (at least) positive semidefinite. If the vectors $a_1, \ldots, a_m$ span the n-dimensional space, in which case one needs $m \geq n$, then (1.3.6-8) is positive definite. The ***spectral representation*** of an $n \times n$ matrix that has n eigenvectors is

$$A = \sum_{i=1}^{n} \lambda_i u_i v_i' \tag{1.3.6-9}$$

where λ_i are its eigenvalues, u_i are its eigenvectors, and v_i, $i = 1, \ldots, n$, is the ***reciprocal basis***, that is,

$$u_i' v_j = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{1.3.6-10}$$

The vectors v_i are the columns of the inverse of the matrix consisting of the eigenvectors u_i' as rows.

If the $n \times n$ real matrix A is *symmetric* with nonzero eigenvalues, it will always have n eigenvectors that are *orthogonal*. Furthermore, in this case,

$$v_i = u_i \tag{1.3.6-11}$$

Thus, the spectral representation of a *symmetric* real $n \times n$ matrix A is

$$A = \sum_{i=1}^{n} \lambda_i u_i u_i' \tag{1.3.6-12}$$

The ***condition number*** of a positive definite symmetric matrix is (usually taken as) the common logarithm of the ratio of its largest to its smallest eigenvalue:

$$\kappa(A) \triangleq \log_{10} \frac{\lambda_{max}}{\lambda_{min}} \tag{1.3.6-13}$$

Large condition numbers indicate near-singularity (e.g., $\kappa > 6$ for a 32-bit computer indicates an ill-conditioned matrix).

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 47

The (induced) ***norm of a matrix*** is

$$\|A\| \triangleq \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} \tag{1.3.6-14}$$

A consequence of the above and the Schwarz inequality is

$$\|Ax\| \leq \|A\| \, \|x\| \tag{1.3.6-15}$$

Also

$$\|AB\| \leq \|A\| \, \|B\| \tag{1.3.6-16}$$

1.3.7 Continuous-Time Linear Dynamic Systems — Controllability and Observability

The State-Space Representation

The state space representation of ***continuous-time linear systems*** is

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) \tag{1.3.7-1}$$

where

$x(t)$ is the state vector of dimension n_x,
$u(t)$ is the input (control) vector of dimension n_u,
$A(t)$ is an $n_x \times n_x$ matrix, called the ***system matrix***,
$B(t)$ is an $n_x \times n_u$ matrix, called the (continuous-time) ***input gain***.

Equation (1.3.7-1) is known as the ***dynamic equation*** or the ***plant equation***.

The output of the system is, in general, a vector

$$z(t) = C(t)x(t) \tag{1.3.7-2}$$

of dimension n_z and C a known $n_z \times n_x$ matrix, called the ***measurement matrix***.

Equation (1.3.7-2) is known as the ***output equation*** or the ***measurement equation***.

Given the initial condition $x(t_0)$ and the input function denoted as

$$u_{[t_0,t]} = \{u(\tau), t_0 \leq \tau \leq t\} \tag{1.3.7-3}$$

one can compute the future output at any time $t > t_0$

$$z(t) = z[x(t_0), u_{[t_0,t]}, t, t_0] \tag{1.3.7-4}$$

The ***state*** (of a deterministic system) is defined as the smallest dimension vector that *completely summarizes the past of the system.*

A continuous-time system is ***linear time-invariant*** if it is described by (1.3.7-1) and (1.3.7-2) with $A(t) = A$, $B(t) = B$, and $C(t) = C$.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 48

Controllability

A continuous-time (deterministic) system is ***completely controllable*** if, given an *arbitrary destination point* in the state space, there is an input function that will bring the system from any initial state to this point in a finite time.

For a *linear time-invariant system*, the controllability condition is that the pair $\{A, B\}$ is controllable, that is, the ***controllability matrix***

$$\mathcal{Q}_C \triangleq \begin{bmatrix} B & AB & \cdots & A^{n_x-1}B \end{bmatrix} \tag{1.3.7-5}$$

has ***full rank***, which in this case is n_x, the lower of its two dimensions (the other is $n_x n_u$).

Observability

A continuous-time (deterministic) system is ***completely observable*** if its initial state can be *fully and uniquely* recovered from its output, observed over a finite time interval, and the knowledge of the input.

Note that since the system is deterministic, knowledge of the initial state is equivalent to knowledge of the state at any time. Thus, using the output (1.3.7-2), which is in general a vector of dimension $n_z < n_x$, and the input, one can then recover the state perfectly.

For a *linear time-invariant system*, the observability condition is that the pair $\{A, C\}$ is observable; that is, the ***observability matrix***

$$\mathcal{Q}_O \triangleq \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n_x-1} \end{bmatrix} \tag{1.3.7-6}$$

has full rank n_x.

Example

Consider the system (double integrator)

$$\ddot{\xi}(t) = u(t) \tag{1.3.7-7}$$

With the state vector consisting of the position ξ and velocity $\dot{\xi}$

$$x = [\xi \;\; \dot{\xi}]' \tag{1.3.7-8}$$

the state space representation of this system is

$$\dot{x}(t) = Ax(t) + Bu(t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t) \tag{1.3.7-9}$$

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 49

The controllability matrix is

$$\mathcal{Q}_C = [B \;\; AB] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{1.3.7-10}$$

which has full rank $n_x = 2$, i.e., the state is *completely controllable*. While the input u enters directly only into the velocity component of the state, it does have the capability of controlling both state components. If the observation is

$$z(t) = Cx(t) = [1 \;\; 0]x(t) \tag{1.3.7-11}$$

i.e., the position ξ, the observability matrix is

$$\mathcal{Q}_O = \begin{bmatrix} C \\ CA \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{1.3.7-12}$$

and it has full rank $n_x = 2$, i.e., the state is *completely observable*.

If the observation is

$$z(t) = Cx(t) = [0 \;\; 1]x(t) \tag{1.3.7-13}$$

i.e., the velocity $\dot{\xi}$, the observability matrix is

$$\mathcal{Q}_O = \begin{bmatrix} C \\ CA \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \tag{1.3.7-14}$$

and it has rank $1 < n_x = 2$, i.e., the state is *not completely observable*. The reason is that, while the position observations (1.3.7-11) provide information about the velocity, the velocity observations (1.3.7-13) do not provide any position information.

1.3.8 Discrete-Time Linear Dynamic Systems — Controllability and Observability

The State-Space Representation

Deterministic linear dynamic systems can be described in discrete time by a state equation of the form

$$x(k+1) = F(k)x(k) + G(k)u(k) \tag{1.3.8-1}$$

where

$x(k)$ is the state of the system, a vector of dimension n_x,
$F(k)$ is the transition matrix ($n_x \times n_x$),
$u(k)$ is the input (control), a vector of dimension n_u,
$G(k)$ is the input gain ($n_x \times n_u$ matrix),

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 50

all at time k.

The output equation is

$$z(k) = H(k)x(k) \tag{1.3.8-2}$$

where

$z(k)$ is the output vector (observation or measurement), of dimension n_z,
$H(k)$ is the measurement matrix ($n_z \times n_x$).

The ***state*** of the system is the smallest dimension vector that completely summarizes the past of the system. Then the output at time $j > k$ can be determined fully from $x(k)$ and the intervening inputs

$$z(j) = z[j, x(k), u(k), \ldots, u(j-1)] \tag{1.3.8-3}$$

A discrete-time system is ***linear time-invariant*** if it is described by (1.3.8-1) and (1.3.8-2) with $F(k) = F$, $G(k) = G$, and $H(k) = H$.

Controllability

A discrete-time (deterministic) system is ***completely controllable*** if, given an *arbitrary destination point* in the state space, there is an input sequence that will bring the system from any initial state to this point in a finite number of steps.

For a *linear time-invariant system*, the controllability condition is that the pair $\{F, G\}$ is controllable; that is, the ***controllability matrix***

$$\mathcal{Q}_C \triangleq \begin{bmatrix} G & FG & \cdots & F^{n_x-1}G \end{bmatrix} \tag{1.3.8-4}$$

has full rank n_x.

Observability

A (deterministic) system is ***completely observable*** if its initial state can be *fully and uniquely* recovered from a finite number of observations of its output and the knowledge of its input.

Note that since the system is deterministic, knowledge of the initial state is equivalent to knowledge of the state at any time. Thus, using the input and the output (1.3.8-2), which is in general a vector of dimension $n_z < n_x$, one can, *in the absence of noise*, recover the state perfectly.

For a *linear time-invariant* system, the observability condition is that the pair $\{F, H\}$ is observable; that is, the ***observability matrix***

$$\mathcal{Q}_O \triangleq \begin{bmatrix} H \\ HF \\ \vdots \\ HF^{n_x-1} \end{bmatrix} \tag{1.3.8-5}$$

has full rank n_x.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 51

1.4 BRIEF REVIEW OF PROBABILITY THEORY

1.4.1 Events and the Axioms of Probability

Consider an "experiment," or, in general, a process with random outcomes. An ***event*** is a collection (set) of such outcomes — it is said to have occurred if the outcome is one of the elements of this set.

Denote by A an event in such an experiment (e.g., "even" or "5" in a die rolling experiment). Let S be the ***sure event*** in the experiment (e.g., "any number between 1 and 6" in the die rolling). Then the ***probability*** of an event is a number (***measure***) that satisfies the following three ***axioms of probability***:

1. It is nonnegative

$$P\{A\} \geq 0 \qquad \forall A \tag{1.4.1-1}$$

2. It is unity for the sure event

$$P\{S\} = 1 \tag{1.4.1-2}$$

3. It is additive over the union of ***mutually exclusive*** events; that is, if the events A and B have no common elements (their set intersection is $\emptyset$, the empty set)

$$A \cap B \triangleq \{A \text{ and } B\} \triangleq \{A, B\} = \emptyset \tag{1.4.1-3}$$

then their union (logical "or") has probability

$$P\{A \cup B\} \triangleq P\{A \text{ or } B\} \triangleq P\{A + B\} = P\{A\} + P\{B\} \tag{1.4.1-4}$$

From the above it follows that

$$P\{A\} = 1 - P\{\bar{A}\} \leq 1 \tag{1.4.1-5}$$

where the overbar denotes the ***complementary event***.

Since $\emptyset$ is the complement of S, one has

$$P\{\emptyset\} = 0 \tag{1.4.1-6}$$

The event $\emptyset$ is called the ***impossible event***.

Extension

The extended version of Axiom 3 such that the probability is additive over the union of an *infinite number of mutually exclusive events* — ***countable additivity*** — is necessary when an experiment has an infinite number of outcomes. This is a key point in the definition of continuous-valued random variables.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 52

Remarks

Relative-frequency or ***measure-of-belief*** interpretation of probability are alternative, even though not rigorous, ways of introducing the concept of probability.

In practice one can use the relative frequency or measure of belief to assign, based on intuition (or engineering common sense), probabilities to certain events. This can be of major importance in certain engineering systems.

Note that probability is a scalar quantity that is dimensionless (in a physical sense). This is in contrast to, for instance, a probability density function (pdf) which, while a scalar, has in general a *physical dimension*.

Example

For a die-rolling experiment, the space of outcomes is

$$S = \{f_1, \ldots, f_6\} \tag{1.4.1-7}$$

where f_i denotes "face i", $i = 1, \ldots, 6$.

Events are subsets of S, for example,

$$F_1 = \{f_1\} \tag{1.4.1-8}$$

$$\{i = \text{even}\} = \{f_2, f_4, f_6\} \tag{1.4.1-9}$$

$$\{2 < i \leq 4\} = \{f_3, f_4\} \tag{1.4.1-10}$$

The sure event

$$\{\text{any face}\} = S \tag{1.4.1-11}$$

is denoted by the same symbol as the space of outcomes since it is the set of all possible outcomes.

An impossible event is, for example,

$$\{i \leq 1 \text{ and even}\} = \emptyset \tag{1.4.1-12}$$

This is clearly an empty subset of S.

Probabilities are assigned (postulated) for each element of S (each outcome). For a finite number of outcomes this *granular approach* yields, using Axiom 3, the probabilities of all the subsets of S, i.e., the events.

Postulating a fair die — each outcome (face) is equally probable — one has from Axioms 2 and 3,

$$1 = P\{S\} = P\left\{\bigcup_{i=1}^{6} F_i\right\} = \sum_{i=1}^{6} P\{F_i\} = 6P\{F_i\} \tag{1.4.1-13}$$

which yields

$$P\{F_i\} = 1/6 \qquad i = 1, \ldots, 6 \tag{1.4.1-14}$$

From this it follows with Axiom 3 that

$$P\{i = \text{even}\} = P\{F_2\} + P\{F_4\} + P\{F_6\} = 1/2 \tag{1.4.1-15}$$

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 53

1.4.2 Random Variables and Probability Density Function

A scalar ***random variable*** is a (real-valued) function that assumes a certain *value* according to the outcome of a *random experiment*. The value taken by a random variable is called its ***realization***.

The ***probability density function (pdf)*** of a scalar continuous-valued random variable x at $x = \xi$ is

$$p_x(\xi) = \lim_{d\xi \to 0} \frac{P\{\xi - d\xi < x \leq \xi\}}{d\xi} \geq 0 \tag{1.4.2-1}$$

where $P\{\cdot\}$ is the probability of the event $\{\cdot\}$.

The more common notation

$$p_x(x) = p(x) \tag{1.4.2-2}$$

where the argument defines the function, is used instead of (1.4.2-1). Also the term ***density*** is sometimes used instead of *pdf*.

From (1.4.2-1) and Axiom 3 from Subsection 1.4.1 it follows that

$$P\{\eta < x \leq \xi\} = \int_{\eta}^{\xi} p(x)\, dx \tag{1.4.2-3}$$

The function

$$P_x(\xi) = P\{x \leq \xi\} = \int_{-\infty}^{\xi} p(x)\, dx \tag{1.4.2-4}$$

is called the ***cumulative probability distribution function (cdf)*** of x at ξ. This is usually referred to as ***distribution***.

Since the event $\{x \leq \infty\}$ is the "sure" event, one has

$$P\{x \leq \infty\} = \int_{-\infty}^{\infty} p(x)\, dx = 1 \tag{1.4.2-5}$$

A pdf has to have the ***normalization property*** (1.4.2-5) that its total ***probability mass*** is unity — otherwise it is not a ***proper density***.

The relationship between the density and the cumulative distribution is, from (1.4.2-4),

$$p(x) = \frac{d}{d\xi} P_x(\xi)\Big|_{\xi = x} \tag{1.4.2-6}$$

if the derivative exists.

An outline of the rigorous way of introducing the pdf is as follows:

1. First define the "basic" events $\{x \leq \xi\}$.
2. Show that from these events one can obtain via (a countable number of) set operations all the events of interest (e.g., $\{a < x \leq \xi\}$, $\{x = a\}$).
3. Define the probabilities of the basic events, from which the probabilities of the other events of interest can be computed using the extended version of the third axiom of probability.
4. Finally, the pdf follows from (1.4.2-6) under suitable conditions of differentiability; note that (1.4.2-6) is equivalent to (1.4.2-1).

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 54

Remarks

Note from (1.4.2-1) that with the numerator being dimensionless, the pdf of the random variable x has as its *physical dimension* the inverse of the physical dimension of x. (See also problem 1-5.)

The event $\{x = a\}$ has, for a continuous-valued random variable, probability zero (even though it is *not impossible*!). This follows from (1.4.2-3) when the interval length tends to zero, in which case, as long as the density is finite, the integral is zero. Finite density means "no point masses" — this is discussed in the next section.

Improper densities — which do not integrate to unity (actually, their integral is not defined) — can be used, however, in certain circumstances.

Example

A **uniformly distributed random variable** in the interval $[a,\ b]$, symbolically denoted as

$$x \sim \mathcal{U}(a, b) \tag{1.4.2-7}$$

has the pdf

$$p(x) = \mathcal{U}(x; a, b) \triangleq \begin{cases} \dfrac{1}{b-a} & x \in [a, b] \\ 0 & \text{elsewhere} \end{cases} \tag{1.4.2-8}$$

Its cdf is

$$P_x(\xi) = \begin{cases} 0 & \xi < a \\ \dfrac{\xi - a}{b-a} & \xi \in [a, b] \\ 1 & \xi > b \end{cases} \tag{1.4.2-9}$$

A **Gaussian (normal) random variable** with parameters μ, σ^2 (the mean and variance, respectively, discussed later) is symbolically denoted as

$$x \sim \mathcal{N}(\mu, \sigma^2) \tag{1.4.2-10}$$

Its pdf is

$$p(x) = \mathcal{N}(x; \mu, \sigma^2) \triangleq \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{1.4.2-11}$$

and the cdf is

$$\begin{aligned} P_x(\xi) &= \int_{-\infty}^{\xi} \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = \int_{-\infty}^{\frac{\xi-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{\eta^2}{2}}\, d\eta \\ &= \int_{-\infty}^{\frac{\xi-\mu}{\sigma}} \mathcal{N}(\eta; 0, 1)\, d\eta \triangleq \mathcal{G}\left(\frac{\xi - \mu}{\sigma}\right) \end{aligned} \tag{1.4.2-12}$$

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 55

where $\mathcal{G}$ is the **cumulative standard Gaussian distribution** (with mean $\mu = 0$ and standard deviation $\sigma = 1$). Its argument above is the *number of standard deviations away from the mean*. This integral has been extensively tabulated (see, e.g., Table 1.5.4-1).

Note

The symbols $\mathcal{U}$ and $\mathcal{N}$ with two arguments indicate "*distributed as*," while with three arguments they denote the corresponding pdf.

1.4.3 Probability Mass Function

The ***probability mass function (pmf)*** of a scalar random variable x, which can take values in the set $\{\xi_i, i = 1, \ldots, n\}$ (i.e., it is discrete-valued), is

$$\mu_x(\zeta_i) = P\{x = \zeta_i\} = \mu_i \qquad i = 1, \ldots, n \tag{1.4.3-1}$$

where μ_i are the ***point masses***.

Similarly to (1.4.2-5), the requirement for a ***proper pmf*** is

$$\sum_{i=1}^{n} \mu_i = 1 \tag{1.4.3-2}$$

Using the ***Dirac (impulse) delta function*** defined by

$$\delta(x) = 0 \qquad \forall x \neq 0 \tag{1.4.3-3}$$

and

$$\int_{-\infty}^{\infty} \delta(x)dx = 1 \tag{1.4.3-4}$$

one can write a pdf corresponding to (1.4.3-1) as

$$p(x) = \sum_{i=1}^{n} \mu_i \delta(x - \xi_i) \tag{1.4.3-5}$$

Note that the above satisfies the ***normalization property*** (1.4.2-5) of a pdf in that it integrates to unity.

The distribution corresponding to the above density, called ***cumulative probability mass function (cpmf)***, has jumps at ξ_i — it is a staircase function: Its derivative is zero everywhere except at the jumps where it is an impulse function.

The expression of the cpmf can be written, in terms of the ***unit step function*** $1(\cdot)$, as

$$P\{x \leq \xi\} = \sum_{i=1}^{n} \mu_i \, 1(\xi - \xi_i) \tag{1.4.3-6}$$

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 56

Note

For a random variable with no point mass at ξ, one has

$$P\{x \leq \xi\} = P\{x < \xi\} \tag{1.4.3-7}$$

Example

The ***Poisson pmf*** with rate λ

$$P\{x = n\} = e^{-\lambda T}\frac{(\lambda T)^n}{n!} \qquad n = 0, 1, \ldots \tag{1.4.3-8}$$

describes the number of random points in an interval T can also be written as

$$p(x) = \sum_{n=0}^{\infty} e^{-\lambda T}\frac{(\lambda T)^n}{n!}\delta(x - n) \tag{1.4.3-9}$$

1.4.4 Mixed Random Variable and Mixed Probability-PDF

A ***mixed random variable*** or a ***hybrid random variable*** x is one which can take values in a continuous set X as well as over a discrete set of points $\{\xi_i, i = 1, \ldots, n\}$. Such a random variable has a pdf of the form

$$p(x) = p_c(x) + \sum_{i=1}^{n} \mu_i \delta(x - \xi_i) \tag{1.4.4-1}$$

where $p_c(x)$ is the continuous part of the pdf and μ_i are the point masses.

Then

$$\int_{-\infty}^{\infty} p(x)dx = \int_{x\in X} p_c(x)dx + \sum_{i=1}^{n} \mu_i = 1 \tag{1.4.4-2}$$

The ***joint probability-pdf*** *of an event and a random variable* is defined as

$$P_{A,x}[A, \xi] = \lim_{d\xi \to 0} \frac{P\{A, \xi - d\xi < x \leq \xi\}}{d\xi} \tag{1.4.4-3}$$

and is denoted without subscripts as $P[A, x]$ where the arguments define the function.

The following notations will be observed[6] in the sequel:

$$P\{\cdot\} = \text{probability of an event} \tag{1.4.4-4}$$

$$p(\cdot) \text{ or } p[\cdot] - \text{pdf} \tag{1.4.4-5}$$

$$\mu(\cdot) = \text{pmf} \tag{1.4.4-6}$$

$$P[\cdot, \cdot] = \text{mixed (joint) probability-pdf} \tag{1.4.4-7}$$

Some of the uncertainties in estimation problems are naturally modeled as mixed random variables.

[6]Most of the time. Some of the time we will be guilty of using "dynamic notation" — the same variable keeps being denoted by new symbols, and the same symbol is used for different variables.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 57

1.4.5 Expectations and Moments of a Scalar Random Variable

The ***expected value*** of a scalar random variable, also called its ***mean***, ***average***, or ***first moment***, is

$$E[x] = \int_{-\infty}^{\infty} xp(x)\,dx \triangleq \bar{x} \tag{1.4.5-1}$$

The nth ***moment*** is

$$E[x^n] = \int_{-\infty}^{\infty} x^n p(x)\,dx \tag{1.4.5-2}$$

The ***second central moment*** or ***variance*** is

$$\text{var}(x) \triangleq E[(x-\bar{x})^2] = \int_{-\infty}^{\infty} (x-\bar{x})^2 p(x)\,dx = E[x^2] - (\bar{x})^2 \triangleq \sigma_x^2 \tag{1.4.5-3}$$

The square root σ of the variance is called the ***standard deviation***.

A random variable x with mean $\bar{x}$ and variance σ^2 will be denoted as

$$x \sim [\bar{x}, \sigma^2] \tag{1.4.5-4}$$

Note that this is similar to (1.4.2-10) except that the distribution is not specified.

The second moment, which is the ***mean square (MS)*** value, is equal to the square of the mean plus the variance

$$E[x^2] = [E(x)]^2 + \text{var}(x) = \bar{x}^2 + \sigma_x^2 \tag{1.4.5-5}$$

For a zero-mean random variable, the standard deviation is the ***root mean square (RMS)*** value.[7]

The expected value of a function $g(x)$ of the random variable x is

$$E[g(x)] = \int_{-\infty}^{\infty} g(x)p(x)\,dx \tag{1.4.5-6}$$

that is, the (Lebesgue) integral with respect to the measure $p(x)\,dx = dP(x)$.

Examples

If x is uniformly distributed in the interval $[a, b]$, that is,

$$p(x) = \mathcal{U}(x; a, b) \triangleq \begin{cases} \dfrac{1}{b-a} & x \in [a, b] \\ 0 & \text{elsewhere} \end{cases} \tag{1.4.5-7}$$

then one has

$$E[x] = \frac{b+a}{2} \qquad \text{var}(x) = \frac{(b-a)^2}{12} \tag{1.4.5-8}$$

[7] Not to be confused with the RMS value of a waveform.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 58

(See also problem 1-6.)

For a Gaussian random variable with parameters μ and σ^2 (see 1.4.2-10) the mean is

$$\begin{aligned} E[x] &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \int_{-\infty}^{\infty} (x-\mu) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \mu \end{aligned} \tag{1.4.5-9}$$

where in the last line above the first integral is zero due to symmetry and the second is unity due to the pdf normalization property.

The variance is

$$\begin{aligned} \text{var}(x) &= E[(x-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{x^2}{2\sigma^2}} dx \end{aligned} \tag{1.4.5-10}$$

The last integral above can be evaluated by differentiating w.r.t. σ the identity

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} dx = \sigma \tag{1.4.5-11}$$

which follows from the pdf normalization property. Thus

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma^3} x^2 e^{-\frac{x^2}{2\sigma^2}} dx = 1 \tag{1.4.5-12}$$

Using this in (1.4.5-10) yields

$$\text{var}(x) = \sigma^2 \tag{1.4.5-13}$$

The mean of the Poisson random variable x from (1.4.3-9) is

$$E[x] = \int_{\infty}^{\infty} x \sum_{n=0}^{\infty} e^{-\lambda T} \frac{(\lambda T)^n}{n!} \delta(x-n)\, dx = \sum_{n=0}^{\infty} n\, e^{-\lambda T} \frac{(\lambda T)^n}{n!} = \lambda T \tag{1.4.5-14}$$

1.4.6 Joint PDF of Two Random Variables

The ***joint pdf*** of two random variables x and y is defined in terms of the probability of the following joint event, denoted with the set intersection symbol (see Fig. 1.4.6-1), namely,

$$p_{x,y}(\xi,\eta) \triangleq \lim_{d\xi\to 0, d\eta\to 0} \frac{P\{\{\xi - d\xi < x \le \xi\} \cap \{\eta - d\eta < y \le \eta\}\}}{d\xi d\eta} \tag{1.4.6-1}$$

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 59

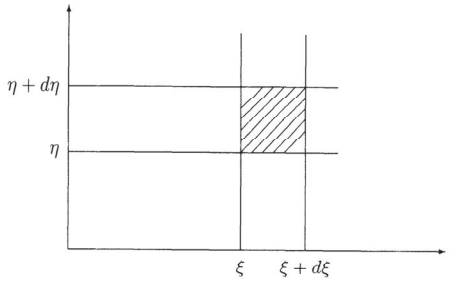


Figure 1.4.6-1: Event for definition of a joint pdf.

Integrating the joint pdf of two random variables over one of the variables yields the pdf of the other random variable

$$\int_{-\infty}^{\infty} p_{x,y}(\xi, \eta)\, d\eta = p_x(\xi) \tag{1.4.6-2}$$

or, using the simpler notation (1.4.2-2),

$$\int_{-\infty}^{\infty} p(x, y)\, dy = p(x) \tag{1.4.6-3}$$

The resulting pdf, which pertains to a single random variable (introduced in Subsection 1.4.2), is also called ***marginal pdf*** or ***marginal density***.

Similarly to (1.4.2-4), the ***joint cdf*** is

$$P_{x,y}(\xi, \eta) = P\{x \le \xi, y \le \eta\} = \int_{x=-\infty}^{\xi} \int_{y=-\infty}^{\eta} p_{x,y}(x, y)\, dxdy \tag{1.4.6-4}$$

Covariance and Correlation Coefficient

The ***covariance of two scalar random variables*** x_1 and x_2 with means $\bar{x}_1$ and $\bar{x}_2$, respectively, is

$$\begin{aligned} \operatorname{cov}(x_1, x_2) &\triangleq E[(x_1 - \bar{x}_1)(x_2 - \bar{x}_2)] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - \bar{x}_1)(x_2 - \bar{x}_2) p(x_1, x_2)\, dx_1 dx_2 \triangleq \sigma^2_{x_1 x_2} \end{aligned} \tag{1.4.6-5}$$

The ***correlation coefficient*** of these two random variables is the normalized quantity

$$\rho_{12} \triangleq \frac{\sigma^2_{x_1 x_2}}{\sigma_{x_1} \sigma_{x_2}} \tag{1.4.6-6}$$

where σ_{x_i} is the standard deviation of x_i.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 60

Due to the normalization, the *magnitude of the correlation coefficient* of any two random variables obeys the following inequality:

$$|\rho_{12}| \leq 1 \tag{1.4.6-7}$$

It can be shown that this is a consequence of the Schwarz inequality (1.3.4-2) for random variables (where the inner product is the covariance and the norm squared is the variance — this is discussed in Section 3.3).

Two random variables whose correlation coefficient is zero are said to be ***uncorrelated***. Clearly, the uncorrelatedness of two random variables is equivalent to their covariance being zero. Furthermore, the random variables x_1 and x_2 are uncorrelated if and only if

$$E[x_1 x_2] = E[x_1]E[x_2] \tag{1.4.6-8}$$

At the other extreme, if the correlation coefficient of two random variables has magnitude unity, then it can be shown that they are linearly dependent (i.e., one is a linear function of the other). Two random variables x_1 and x_2 are said to be ***linearly dependent*** if

$$a_1 x_1 + a_2 x_2 = 0 \tag{1.4.6-9}$$

for some $a = [a_1 \;\; a_2]' \neq 0$.

Example

Consider the random variables x and y with means $\bar{x}$, $\bar{y}$ and variances σ_x^2, σ_y^2, respectively, and uncorrelated from each other (i.e., $\text{cov}(x, y) = 0$).

Let

$$z = ax + by \tag{1.4.6-10}$$

The mean of z is

$$\bar{z} = E[z] = E[ax + by] = aE[x] + bE[y] = a\bar{x} + b\bar{y} \tag{1.4.6-11}$$

while its variance is

$$\begin{aligned} \text{var}(z) &= E[(z - \bar{z})^2] = E[(ax + by - a\bar{x} - b\bar{y})^2] \\ &= E\{[a\,(x - \bar{x}) + b\,(y - \bar{y})]^2\} \\ &= a^2\,E[(x - \bar{x})^2] + b^2\,E[(y - \bar{y})^2] + 2ab\,E[(x - \bar{x})(y - \bar{y})] \\ &= a^2\sigma_x^2 + b^2\sigma_y^2 \end{aligned} \tag{1.4.6-12}$$

since the cross-term — the covariance of x and y — is zero.

The covariance of z and x is

$$\begin{aligned} E[(z - \bar{z})(x - \bar{x})] &= E\{[a\,(x - \bar{x}) + b\,(y - \bar{y})](x - \bar{x})\} \\ &= a\sigma_x^2 + b\,\text{cov}(x, y) = a\sigma_x^2 \end{aligned} \tag{1.4.6-13}$$

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 61

1.4.7 Independent Events and Independent Random Variables

Two events are ***independent*** if the probability of their joint event equals the product of their marginal probabilities

$$P\{A \cap B\} \triangleq P\{A, B\} = P\{A\}P\{B\} \tag{1.4.7-1}$$

A set of n events A_i, $i = 1, \ldots, n$ are independent if the (joint) probability of their intersection is equal to the product of the corresponding (marginal) event probabilities

$$P\left\{\bigcap_{i=1}^{n} A_i\right\} = \prod_{i=1}^{n} P\{A_i\} \tag{1.4.7-2}$$

and the same property holds also for any subset of these events.

Similarly, n random variables are ***independent*** if their joint pdf equals the product of their corresponding marginal densities

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i) \tag{1.4.7-3}$$

A set of random variables is called ***independent, identically distributed (i.i.d.)*** if (1.4.7-3) holds and their marginal distributions (or densities) are identical.

The pdf of the sum of two independent random variables

$$y = x_1 + x_2 \tag{1.4.7-4}$$

is the ***convolution*** of their marginal densities

$$p(y) = \int p_{x_1}(y - x)p_{x_2}(x)\, dx \tag{1.4.7-5}$$

(See also problem 1-1.)

1.4.8 Vector-Valued Random Variables and Their Moments

The pdf of the ***vector-valued random variable***

$$x = [x_1 \ \cdots \ x_n]' \tag{1.4.8-1}$$

at

$$\xi = [\xi_1 \ \cdots \ \xi_n]' \tag{1.4.8-2}$$

is defined as the ***joint density*** of its components

$$p_{x_1,\ldots,x_n}(\xi_1, \ldots, \xi_n) \triangleq p_x(\xi) \triangleq \lim_{d\xi_1 \to 0, \ldots, d\xi_n \to 0} \frac{P\{\cap_{i=1}^{n}\{\xi_i - d\xi_i < x_i \leq \xi_i\}\}}{d\xi_1 \cdots d\xi_n} \tag{1.4.8-3}$$

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 62

where the set intersection symbol $\cap$ is used to denote a joint event

$$A \cap B \triangleq \{A \text{ and } B\} \triangleq \{A, B\} \tag{1.4.8-4}$$

The shorter notation $p(x)$ will be used if it does not lead to ambiguities.

The mean of the vector x is the result of the n-fold integration

$$E[x] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \xi\, p_x(\xi)\, d\xi_1 \cdots d\xi_n = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x p(x)\, dx_1 \cdots dx_n \triangleq \bar{x} \tag{1.4.8-5}$$

The ***covariance matrix*** of the n-vector x is obtained from the n-fold integration (written with the short notation and with the limits omitted)

$$\text{cov}(x) \triangleq E[(x - \bar{x})(x - \bar{x})'] = \int (x - \bar{x})(x - \bar{x})' p(x)\, dx \triangleq P_{xx} \tag{1.4.8-6}$$

where $dx \triangleq \prod dx_i$. Note that the covariance matrix is a *symmetric* $n \times n$ matrix.

The diagonal elements of the *covariance matrix* are the *variances* of the components of x, while the off-diagonal elements are the *(scalar) covariances* between its components, as in (1.4.6-5).

The *covariance matrix is positive definite* (and thus nonsingular) unless there is a linear dependence among the components of x. If there is such a dependence, the covariance matrix is positive semidefinite.

Note

Covariance matrices will be denoted by the symbol P, with a subscript or argument (in parentheses); the designation of a probability will be $P\{\cdot\}$ where braces are used for the event it pertains to, as in (1.4.4-4).

Characteristic Function and Moments

The ***characteristic function*** of a vector random variable is defined as the n-fold integral (with the limits, which are, as in (1.4.8-5), omitted)

$$M_x(s) = E[e^{s'x}] = \int e^{s'x} p(x)\, dx \tag{1.4.8-7}$$

which is the (n-dimensional) Fourier transform of the pdf with argument the (purely imaginary) vector s of the same dimension n as that of x.

The first moment of x can be obtained from the characteristic function as

$$E[x] = \nabla_s M_x(s)|_{s=0} \tag{1.4.8-8}$$

that is, its gradient is evaluated at $s = 0$, where ∇_s is the (column) gradient operator

$$\nabla_s = \left[\frac{\partial}{\partial s_1} \cdots \frac{\partial}{\partial s_n}\right]' \tag{1.4.8-9}$$

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 63

Similarly,

$$E[xx'] = \nabla_s \nabla_s' M_x(s)|_{s=0} \tag{1.4.8-10}$$

and so forth.

Due to this property, the characteristic function is also called the ***moment-generating function***.

Example

The characteristic function of the scalar Gaussian random variable

$$x \sim \mathcal{N}(\mu, \sigma^2) \tag{1.4.8-11}$$

is

$$M_x(s) = e^{s\mu + \frac{1}{2}s^2\sigma^2} \tag{1.4.8-12}$$

To find the moments of x it is convenient to deal with the zero-mean random variable

$$y = x - \mu \tag{1.4.8-13}$$

The moments of x follow easily from the moments of y, according to

$$E[x^n] = E[(y + \mu)^n] \tag{1.4.8-14}$$

The characteristic function of y is

$$M_y(s) = e^{\frac{1}{2}s^2\sigma^2} \tag{1.4.8-15}$$

and

$$\frac{d}{ds} M_y(s) = s\sigma^2 \, e^{\frac{1}{2}s^2\sigma^2} \tag{1.4.8-16}$$

Evaluating the above at $s = 0$ yields the first moment (mean) of y as 0.

Similarly,

$$\frac{d^2}{ds^2} M_y(s) = [\sigma^2 + (s\sigma^2)^2] \, e^{\frac{1}{2}s^2\sigma^2} \tag{1.4.8-17}$$

which, at $s = 0$, yields

$$E[y^2] = E[(x - \mu)^2] = \text{var}(x) = \sigma^2 \tag{1.4.8-18}$$

The third derivative of the characteristic function of y is

$$\frac{d^3}{ds^3} M_y(s) = [2s\sigma^4 + (\sigma^2 + s^2\sigma^4)s\sigma^2] \, e^{\frac{1}{2}s^2\sigma^2} = [3s\sigma^4 + s^3\sigma^6] \, e^{\frac{1}{2}s^2\sigma^2} \tag{1.4.8-19}$$

which yields zero as the third moment of y.

The fourth moment of y is given by

$$\frac{d^4}{ds^4} M_y(s) = [3\sigma^4 + 3s^2\sigma^6 + (3s\sigma^4 + s^3\sigma^6)s\sigma^2] \, e^{\frac{1}{2}s^2\sigma^2} \tag{1.4.8-20}$$

at $s = 0$ as

$$E[y^4] = E[(x - \mu)^4] = 3\sigma^4 \tag{1.4.8-21}$$

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 64

1.4.9 Conditional Probability and PDF

The ***conditional probability*** of an event A given B is defined as

$$P\{A|B\} = \frac{P\{A, B\}}{P\{B\}} \tag{1.4.9-1}$$

For independent events, the above becomes the unconditional probability.

Similarly, the ***conditional pdf*** of one random variable given another random variable is, using the simpler notation as in (1.4.2-2), given by

$$p(x|y) = \frac{p(x, y)}{p(y)} \tag{1.4.9-2}$$

For an event conditioned on a random variable, one has

$$P\{A|x\} = \frac{P[A, x]}{p(x)} \tag{1.4.9-3}$$

The conditional pdf of a random variable x, given an event A, is

$$p(x|A) = \frac{P[A, x]}{P\{A\}} \tag{1.4.9-4}$$

where the numerator above is a mixed probability-pdf, as defined in (1.4.4-3).

Example

If the conditioning event is

$$A = \{x \leq a\} \tag{1.4.9-5}$$

then one gets the ***truncated pdf***

$$p(x|x \leq a) = \frac{P[x, x \leq a]}{P\{x \leq a\}} = \begin{cases} \dfrac{p(x)}{P\{x \leq a\}} & \text{if } x \leq a \\ 0 & \text{elsewhere} \end{cases} \tag{1.4.9-6}$$

which is the original one restricted to $x \leq a$ and suitably *renormalized*.

If $z = x + y$ and the conditional pdf of y given x is $p_{y|x}(\cdot)$, then

$$p(z|x) = p_{y|x}(z - x) \tag{1.4.9-7}$$

If, in addition, x and y are independent, then, with $p_y(\cdot)$ denoting the pdf of y, one has

$$p(z|x) = p_y(z - x) \tag{1.4.9-8}$$

(See also problem 1-1.)

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 65

Assume $x \sim \mathcal{N}(\mu_x, \sigma_x^2)$, $y \sim \mathcal{N}(\mu_y, \sigma_y^2)$ and independent. Then, with $z = x + y$,

$$\begin{aligned} p(z|x) &= p_y(z-x) = \mathcal{N}(z-x; \mu_y, \sigma_y^2) \\ &= \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{(z-x-\mu_y)^2}{2\sigma_y^2}} = \mathcal{N}(z; x+\mu_y, \sigma_y^2) \end{aligned} \tag{1.4.9-9}$$

Note that the above can be viewed, due to the special form of the Gaussian density, as the pdf of $z - x$ with mean μ_y or the pdf of z with mean $x + \mu_y$.

If $x \sim \mathcal{N}(\mu, \sigma^2)$, then

$$p(x|\,|x-\mu| \le a) = \begin{cases} \frac{1}{P\{|x-\mu|\le a\}} \mathcal{N}(x; \mu, \sigma^2) & |x| \le a \\ 0 & \text{elsewhere} \end{cases} \tag{1.4.9-10}$$

is a ***truncated Gaussian***.

The normalization constant above is the probability that $x - \mu$ lies in the closed interval[8] $[-a, a]$, which is

$$\begin{aligned} P\{|x-\mu| \le a\} &= P\{x-\mu \in [-a,a]\} = P\{x \in [\mu-a, \mu+a]\} \\ &= P\{x \in (-\infty, \mu+a]\} - P\{x \in (-\infty, \mu-a)\} \\ &= \mathcal{G}\left(\frac{(\mu+a)-\mu}{\sigma}\right) - \mathcal{G}\left(\frac{(\mu-a)-\mu}{\sigma}\right) \\ &= \mathcal{G}\left(\frac{a}{\sigma}\right) - \mathcal{G}\left(\frac{-a}{\sigma}\right) = 1 - 2\mathcal{G}\left(\frac{-a}{\sigma}\right) \end{aligned} \tag{1.4.9-11}$$

where $\mathcal{G}$ is the cumulative standard Gaussian distribution and the last equality follows from the symmetry of the Gaussian density. The last term above is the tail probability to the left of $(-a)/\sigma$ or to the right of $(a)/\sigma$.

1.4.10 The Total Probability Theorem

Let the events B_i, $i = 1, \ldots, n$, be ***mutually exclusive***, that is,

$$P\{B_i, B_j\} = 0 \qquad \forall i \ne j \tag{1.4.10-1}$$

and ***exhaustive***, that is,

$$\sum_{i=1}^{n} P\{B_i\} = 1 \tag{1.4.10-2}$$

Such a set of events is a *partition of the space of outcomes*. Then, the ***total probability theorem*** states that for any event A, its probability can be decomposed in terms of conditional probabilities as follows:

$$P\{A\} = \sum_{i=1}^{n} P\{A, B_i\} = \sum_{i=1}^{n} P\{A|B_i\}P\{B_i\} \tag{1.4.10-3}$$

[8] A bracket indicates that the interval is closed at the corresponding end, i.e., it includes the endpoint. A parenthesis indicates that the interval is open at the corresponding end, i.e., it excludes the endpoint.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 66

The key to the above is the mutual exclusiveness and exhaustiveness of the set of events B_i, $i = 1, \ldots, n$.

Figure 1.4.10-1 illustrates this with the set theory counterpart, namely,

$$A = \bigcup_{i=1}^{n} (A \cap B_i) \tag{1.4.10-4}$$

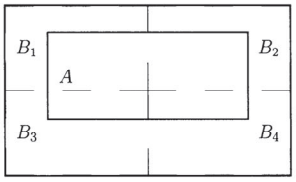


Figure 1.4.10-1: The total probability theorem (set theory counterpart).

The version of the total probability theorem for random variables is

$$p(x) = \int_{-\infty}^{\infty} p(x, y)\, dy = \int_{-\infty}^{\infty} p(x|y)p(y)\, dy \tag{1.4.10-5}$$

Note that this is a combination of the definition of the conditional pdf (1.4.9-2) and property (1.4.6-3) that the joint pdf integrated out with respect to one random variable yields the marginal pdf of the other variable.

For mixed event/random variable situations, one has

$$p(x) = \sum_{i=1}^{n} p(x|B_i)P\{B_i\} \tag{1.4.10-6}$$

(this is illustrated in an example in Subsection 1.4.14) and

$$P\{A\} = \int_{-\infty}^{\infty} P\{A|x\}p(x)\, dx = E[P\{A|x\}] \tag{1.4.10-7}$$

where the expectation operator averages over x.

An additional *common* conditioning in *all* the probabilities is permissible, for instance,

$$P\{A|C\} = \sum_{i=1}^{n} P\{A, B_i|C\} = \sum_{i=1}^{n} P\{A|B_i, C\}P\{B_i|C\} \tag{1.4.10-8}$$

or

$$P\{A|y\} = \int_{-\infty}^{\infty} P\{A|x, y\}p(x|y)\, dx \tag{1.4.10-9}$$

Similarly,

$$p(x|z) = \int_{-\infty}^{\infty} p(x, y|z)\, dy = \int_{-\infty}^{\infty} p(x|y, z)p(y|z)\, dy \tag{1.4.10-10}$$

The total probability theorem is the primary tool in obtaining the state estimate in the presence of extraneous uncertainties (e.g., model uncertainties or measurements of uncertain origin).

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 67

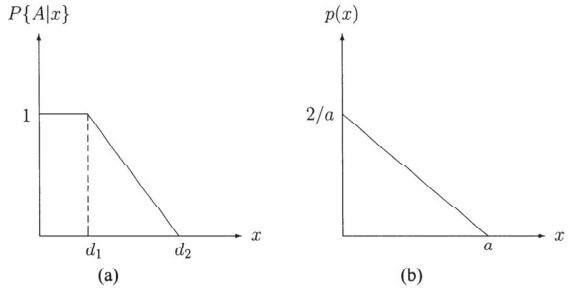


Figure 1.4.10-2: Success probability and miss distance pdf.

Example

The probability of event A, which denotes success in a hunting target shooting, can be modeled as a function of the "miss distance" x by the conditional probability (see Fig. 1.4.10-2a)

$$P\{A|x\} = \begin{cases} 1 & x \le d_1 \\ \dfrac{x - d_2}{d_1 - d_2} & d_1 < x \le d_2 \\ 0 & x > d_2 \end{cases} \tag{1.4.10-11}$$

The miss distance pdf can be taken as (see Fig. 1.4.10-2b)

$$p(x) = \begin{cases} \dfrac{2(a-x)}{a^2} & 0 < x \le a \\ 0 & \text{otherwise} \end{cases} \tag{1.4.10-12}$$

Then, assuming $d_2 < a$, the (unconditional) probability of success is

$$\begin{aligned} P\{A\} &= \int_{-\infty}^{\infty} P\{A|x\}p(x)\,dx = \int_0^{d_1} p(x)\,dx + \int_{d_1}^{d_2} \frac{x - d_2}{d_1 - d_2} p(x)\,dx \\ &= P\{x \le d_1\} + \int_{d_1}^{d_2} \frac{x - d_2}{d_1 - d_2} \frac{2(a-x)}{a^2}\,dx \\ &= \frac{d_1(2a - d_1)}{a^2} + \frac{(d_2 - d_1)}{3a^2}(3a - 2d_1 - d_2) \end{aligned} \tag{1.4.10-13}$$

1.4.11 Bayes' Formula

The probability of an event B_i conditioned on event A can be expressed in terms of the reverse conditioning as follows:

$$P\{B_i|A\} = \frac{P\{A|B_i\}P\{B_i\}}{P\{A\}} \tag{1.4.11-1}$$

This is known as ***Bayes' formula*** or ***Bayes' theorem*** and is also referred to as ***Bayes' rule***.[9]

The conditional probability of B_i is sometimes referred to as ***posterior probability***, while the unconditional one is referred to as ***prior probability***.

Using the total probability theorem (1.4.10-3) for the denominator in the above, one also has (if B_j, $j = 1, \ldots, n$ are mutually exclusive and exhaustive)

$$P\{B_i|A\} = \frac{P\{A|B_i\}P\{B_i\}}{\sum_{j=1}^{n} P\{A|B_j\}P\{B_j\}} = \frac{1}{c}P\{A|B_i\}P\{B_i\} \tag{1.4.11-2}$$

where now the denominator c appears clearly as the ***normalizing constant***, which guarantees that

$$\sum_{i=1}^{n} P\{B_i|A\} = 1 \tag{1.4.11-3}$$

Bayes' formula for random variables is written as

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} = \frac{p(y|x)p(x)}{\int p(y|x)p(x)\,dx} \tag{1.4.11-4}$$

In this case the unconditional pdf $p(x)$ is also called the ***prior pdf*** and the conditional pdf $p(x|y)$ is also called the ***posterior pdf***.

For a mixed case

$$P\{B_i|x\} = \frac{p(x|B_i)P\{B_i\}}{p(x)} = \frac{p(x|B_i)P\{B_i\}}{\sum_{j=1}^{n} p(x|B_j)P\{B_j\}} \tag{1.4.11-5}$$

where the denominator is again the normalizing factor.

When there are several conditioning random variables or events, Bayes' formula can be used to "switch" only some of them. For example,

$$P\{B_i|x, y\} = \frac{p(x|B_i, y)P\{B_i|y\}}{p(x|y)} = \frac{p(x|B_i, y)P\{B_i|y\}}{\sum_{j=1}^{n} p(x|B_j, y)P\{B_j|y\}} \tag{1.4.11-6}$$

These equations are the key tools in state estimation for ***hybrid systems*** — systems with continuous as well as discrete uncertainties.

Remarks

The prior $p(x)$ reflects the (possibly subjective[10]) initial ***degree of belief***, which, when combined with the ***evidence from the data*** — the ***likelihood function*** of x, namely, $p(y|x)$ — yields the posterior $p(x|y)$.

[9] ***Bayes' decision rule*** (or ***Bayes' criterion*** or ***Bayes' principle***) in a decision problem is to minimize the expected value of a cost function, called ***Bayes' risk***.

[10] Similarly to the well-known fact that the entire population of the world is divided into two groups (those who believe this and those who don't), the people dealing with probability are divided into two groups: proud Bayesians (who believe in priors) and closet Bayesians.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 69

Bayes' postulate states that in the absence of prior knowledge, a ***uniform prior pdf*** should be chosen. While this is reasonable for discrete valued random variables, for continuous valued random variables this would imply a nonuniform pdf on any nonlinear function of this random variable and this has generated a lot of controversy among statisticians.

The use of a uniform prior pdf *over an infinite interval*, called ***diffuse*** or ***noninformative***, also called ***Dirichlet prior*** is discussed in more detail in Subsection 2.3.4.

Example

The "noise" in a device, denoted as x and modeled as a random variable, is observed. If the device is operating according to its specifications, then one has event B_1, defined as

$$B_1 = \{x \sim \mathcal{N}(0, \sigma_1^2)\} \tag{1.4.11-7}$$

The defective device event is assumed for simplicity defined as

$$B_2 = \{x \sim \mathcal{N}(0, \sigma_2^2)\} \tag{1.4.11-8}$$

The conditional probability that the device is defective given the observation x is

$$P\{B_2|x\} = \frac{p(x|B_2)P\{B_2\}}{p(x|B_1)P\{B_1\} + p(x|B_2)P\{B_2\}} \tag{1.4.11-9}$$

Using the explicit expressions of the Gaussian densities defining the events B_1 and B_2, the above can be written, after cancellations, as

$$P\{B_2|x\} = \frac{1}{1 + \frac{\sigma_2}{\sigma_1} e^{-\frac{x^2}{2\sigma_1^2} + \frac{x^2}{2\sigma_2^2}} \frac{P\{B_1\}}{P\{B_2\}}} \tag{1.4.11-10}$$

Assume that a priori the two events are equiprobable

$$P\{B_1\} = \frac{1}{2} \qquad P\{B_2\} = 1 - P\{B_1\} = \frac{1}{2} \tag{1.4.11-11}$$

If $\sigma_2 = 4\sigma_1$ and the observation obtained is $x = \sigma_2$, then

$$P\{B_2|x\} = \frac{1}{1 + 4\, e^{-8+0.5}} \approx 0.998 \tag{1.4.11-12}$$

If the prior is different, namely, $P\{B_2\} = 10^{-2}$, then

$$P\{B_2|x\} = \frac{1}{1 + 396\, e^{-8+0.5}} \approx 0.82 \tag{1.4.11-13}$$

i.e., regardless of the prior, the "evidence from the data" indicates a very high probability that the device is defective.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 70

1.4.12 Conditional Expectations and Their Smoothing Property

The ***conditional expectation*** is defined similarly to the unconditional expectation (1.4.5-1) but *with respect to a conditional pdf*, that is,

$$E[x|y] = \int_{\infty}^{\infty} xp(x|y)\,dx \tag{1.4.12-1}$$

Similarly, for a function of the random variable x (and possibly y), one has

$$E[g(x,y)|y] = \int_{-\infty}^{\infty} g(x,y)p(x|y)\,dx \tag{1.4.12-2}$$

Note that (1.4.12-1) and (1.4.12-2) are functions of the conditioning argument y.

The ***smoothing property of the expectations*** states that the expected value of a conditional expected value is the (unconditional) expected value:

$$\begin{aligned} E\{E[x|y]\} &= \int_{y=-\infty}^{\infty} \left[\int_{x=-\infty}^{\infty} xp(x|y)\,dx\right] p(y)\,dy \\ &= \int_{x=-\infty}^{\infty} x\left[\int_{y=-\infty}^{\infty} p(x,y)\,dy\right] dx \\ &= \int_{x=-\infty}^{\infty} xp(x)\,dx = E[x] \end{aligned} \tag{1.4.12-3}$$

In the first line of the above, the inside expectation is a function of y, which is "averaged out" (integrated over) by the outside expectation.

Equation (1.4.12-3) is also called the ***law of iterated expectations***, summarized as

$$\boxed{E[E[x|y]] = E[x]} \tag{1.4.12-4}$$

The same property holds when the conditioning is on events or mixed random variables.

This is used in the evaluation of the performance of certain estimation algorithms operating in the presence of several types of uncertainties — it allows us to handle different types of uncertainties sequentially rather than simultaneously.

Example

To illustrate the smoothing property of expectations, consider the conditional mean

$$E(x|z) = \bar{x} + P_{xz}P_{zz}^{-1}(z - \bar{z}) \tag{1.4.12-5}$$

where x and z are jointly Gaussian (this is derived later in (1.4.14-17)); overbar denotes the unconditional expected values.

Applying the unconditional expectation operator on the above, where the only random variable on the right-hand side is z, with mean $\bar{z}$, yields

$$E[E(x|z)] = E[\bar{x}] + P_{xz}P_{zz}^{-1}\,E(z - \bar{z}) = \bar{x} = E[x] \tag{1.4.12-6}$$

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 71

1.4.13 Gaussian Random Variables

The pdf of a (scalar) ***Gaussian*** or ***normal random variable*** is

$$p(x) = \mathcal{N}(x; \bar{x}, \sigma^2) \triangleq \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} \tag{1.4.13-1}$$

where $\mathcal{N}(\cdot)$ denotes the ***normal pdf (density)*** with argument x, mean $\bar{x}$, and variance σ^2. The first two moments, which fully characterize a Gaussian random variable, are referred to as its ***statistics***.

Another notation equivalent to the above is

$$\boxed{x \sim \mathcal{N}(\bar{x}, \sigma^2)} \tag{1.4.13-2}$$

which states that x is *normally distributed* with the corresponding mean and variance. Note the different meanings of $\mathcal{N}$ in (1.4.13-1) and (1.4.13-2) — they are, however, specified by their arguments: three in (1.4.13-1) vs. two in (1.4.13-2).

A ***vector-valued Gaussian*** random variable has the density

$$\boxed{\mathcal{N}(x; \bar{x}, P) \triangleq |2\pi P|^{-1/2} e^{-\frac{1}{2}(x-\bar{x})' P^{-1}(x-\bar{x})}} \tag{1.4.13-3}$$

where

$$\bar{x} = E[x] \tag{1.4.13-4}$$

$$P = E[(x - \bar{x})(x - \bar{x})'] \tag{1.4.13-5}$$

are, respectively, the mean and covariance matrix of the vector x.

The determinant in (1.4.13-3) has been written with the factor 2π inside it by making use of (1.3.3-5). This avoids the need to indicate the dimension of the vector x in (1.4.13-3).

The components of a Gaussian distributed vector are said to be ***jointly Gaussian.***

If the covariance matrix P is diagonal — that is, the components of the *Gaussian* random vector x are *uncorrelated* — then they are also *independent* because their joint pdf equals the product of the marginals.

A very important property of Gaussian random variables is that they *stay Gaussian under linear transformations*. Consequently, the sum of (jointly) Gaussian random variables is also Gaussian. This property is a major reason that the Gaussian model is very commonly used in estimation.

Example

If x_1 and x_2 are random vectors of the same dimension and jointly Gaussian distributed, that is,

$$p\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \mathcal{N}\left\{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}; \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix}, \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}\right\} \tag{1.4.13-6}$$

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 72

then

$$x = x_1 + x_2 \tag{1.4.13-7}$$

is Gaussian

$$p(x) = \mathcal{N}(x; \bar{x}, P) \tag{1.4.13-8}$$

where

$$\bar{x} = \bar{x}_1 + \bar{x}_2 \tag{1.4.13-9}$$

$$P = P_{11} + P_{12} + P_{21} + P_{22} \tag{1.4.13-10}$$

Symbolically, one can write, with notation (1.4.13-2),

$$\mathcal{N}(\bar{x}_1, P_{11}) + \mathcal{N}(\bar{x}_2, P_{22}) = \mathcal{N}(\bar{x}_1 + \bar{x}_2, P_{11} + P_{12} + P_{21} + P_{22}) \tag{1.4.13-11}$$

If x_1 and x_2 independent, the pdf of their sum $x = x_1 + x_2$ conditioned on x_1 is

$$\begin{aligned} p(x|x_1) &= p_{x_2}(x - x_1) \\ &= \mathcal{N}(x - x_1; \bar{x}_2, P_{22}) = |2\pi P_{22}|^{-1/2}\, e^{-\frac{1}{2}(x-x_1-\bar{x}_2)' P_{22}^{-1}(x-x_1-\bar{x}_2)} \\ &= \mathcal{N}(x; x_1 + \bar{x}_2, P_{22}) \end{aligned} \tag{1.4.13-12}$$

Note that in (1.4.13-12) x_1 acts as a constant — it shifts the mean — since it is given in the conditioning.

1.4.14 Joint and Conditional Gaussian Random Variables

Two random vectors x and z are ***jointly Gaussian*** if the ***stacked vector***

$$y = \begin{bmatrix} x \\ z \end{bmatrix} \tag{1.4.14-1}$$

is Gaussian, that is,

$$p(x, z) = p(y) = \mathcal{N}(y; y, P_{yy}) \tag{1.4.14-2}$$

The mean and covariance matrix of y in terms of those of x and z are

$$\bar{y} = \begin{bmatrix} \bar{x} \\ \bar{z} \end{bmatrix} \tag{1.4.14-3}$$

$$P_{yy} = \begin{bmatrix} P_{xx} & P_{xz} \\ P_{zx} & P_{zz} \end{bmatrix} \tag{1.4.14-4}$$

where

$$P_{xx} = \text{cov}(x) = E[(x - \bar{x})(x - \bar{x})'] \tag{1.4.14-5}$$

$$P_{zz} = \text{cov}(z) = E[(z - \bar{z})(z - \bar{z})'] \tag{1.4.14-6}$$

$$P_{xz} = \text{cov}(x, z) = E[(x - \bar{x})(z - \bar{z})'] = P_{zx}' \tag{1.4.14-7}$$

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 73

are the blocks of the ***partitioned covariance matrix*** (1.4.14-4).

The conditional pdf of x given z when they are jointly Gaussian is

$$p(x|z) = \frac{p(x,z)}{p(z)} = \frac{|2\pi P_{yy}|^{-1/2}\, e^{-\frac{1}{2}(y-\bar{y})' P_{yy}^{-1}(y-\bar{y})}}{|2\pi P_{zz}|^{-1/2}\, e^{-\frac{1}{2}(z-\bar{z})' P_{zz}^{-1}(z-\bar{z})}} \tag{1.4.14-8}$$

Advantage will be taken of the fact that the above is an exponential whose exponent is the difference of the exponents of the numerator and denominator.

The following notations will be used

$$\xi \triangleq x - \bar{x} \tag{1.4.14-9}$$

$$\zeta \triangleq z - \bar{z} \tag{1.4.14-10}$$

Using the new variables ξ and ζ, the problem for the *nonzero mean random variables* x and y is reduced to that of *zero-mean variables*.

The exponent on the right-hand side of the conditional density (1.4.14-8) is (after multiplication by -2) the quadratic form

$$\begin{aligned} q &= \begin{bmatrix} \xi \\ \zeta \end{bmatrix}' \begin{bmatrix} P_{xx} & P_{xz} \\ P_{zx} & P_{zz} \end{bmatrix}^{-1} \begin{bmatrix} \xi \\ \zeta \end{bmatrix} - \zeta' P_{zz}^{-1} \zeta \\ &= \begin{bmatrix} \xi \\ \zeta \end{bmatrix}' \begin{bmatrix} T_{xx} & T_{xz} \\ T_{zx} & T_{zz} \end{bmatrix} \begin{bmatrix} \xi \\ \zeta \end{bmatrix} - \zeta' P_{zz}^{-1} \zeta \end{aligned} \tag{1.4.14-11}$$

The relationships between the partitions of the inverse of the covariance matrix and the partitions of the original matrix are, using (1.3.3-7) to (1.3.3-10), given by

$$T_{xx}^{-1} = P_{xx} - P_{xz} P_{zz}^{-1} P_{zx} \tag{1.4.14-12}$$

$$P_{zz}^{-1} = T_{zz} - T_{zx} T_{xx}^{-1} T_{xz} \tag{1.4.14-13}$$

$$T_{xx}^{-1} T_{xz} = -P_{xz} P_{zz}^{-1} \tag{1.4.14-14}$$

The exponent (1.4.14-11) can be rewritten as

$$\begin{aligned} q &= \xi' T_{xx} \xi + \xi' T_{xz} \zeta + \zeta' T_{zx} \xi + \zeta' T_{zz} \zeta - \zeta' P_{zz}^{-1} \zeta \\ &= (\xi + T_{xx}^{-1} T_{xz} \zeta)' T_{xx} (\xi + T_{xx}^{-1} T_{xz} \zeta) + \zeta' (T_{zz} - T_{zx} T_{xx}^{-1} T_{xz}) \zeta - \zeta' P_{zz}^{-1} \zeta \\ &= (\xi + T_{xx}^{-1} T_{xz} \zeta)' T_{xx} (\xi + T_{xx}^{-1} T_{xz} \zeta) \end{aligned} \tag{1.4.14-15}$$

where use has been made of (1.4.14-13). The above procedure is called ***completion of the squares*** (actually, of the quadratic forms).

The result is a quadratic form in x, meaning that the conditional pdf of x given z is also Gaussian. This can be seen as follows: In view of (1.4.14-9), (1.4.14-10), and (1.4.14-14), the expression on the right-hand side of (1.4.14-15) is a quadratic form in

$$\xi + T_{xx}^{-1} T_{xz} \zeta = x - \bar{x} - P_{xz} P_{zz}^{-1} (z - \bar{z}) \tag{1.4.14-16}$$

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 74

From this one can recognize the conditional mean of x given z as

$$\boxed{E(x|z) \triangleq \hat{x} = \bar{x} + P_{xz}P_{zz}^{-1}(z - \bar{z})} \tag{1.4.14-17}$$

The corresponding conditional covariance is

$$\boxed{\text{cov}(x|z) \triangleq P_{xx|z} = T_{xx}^{-1} = P_{xx} - P_{xz}P_{zz}^{-1}P_{zx}} \tag{1.4.14-18}$$

Note that the conditional mean (1.4.14-17) is linear in the observation z and that the covariance (1.4.14-18) is independent of the observation.

The above are the ***fundamental equations of linear estimation.***

1.4.15 Expected Value of Quadratic and Quartic Forms

Consider a vector-valued random variable x with mean $\bar{x} = 0$ and covariance matrix P. Then, the ***expected value of a quadratic form*** with this random vector can be written as

$$E[x'Ax] = E[\text{tr}(x'Ax)] = E[\text{tr}(Axx')] = \text{tr}[AE(xx')] = \text{tr}(AP) \tag{1.4.15-1}$$

The same result can be obtained for a Gaussian random vector using its characteristic function as follows.

If

$$x \sim \mathcal{N}(\bar{x}, P) \tag{1.4.15-2}$$

then its characteristic (or moment generating) function is

$$M_x(s) = E[e^{s'x}] = e^{\frac{1}{2}s'Ps + s'\bar{x}} \tag{1.4.15-3}$$

Using the properties of the gradient, one can write

$$\begin{aligned} E[x'Ax] &= E[(\nabla_s\, e^{s'x})'Ax]|_{s=0} = E[\nabla_s'Ax\, e^{s'x}]|_{s=0} \\ &= \nabla_s'A\, E[x\, e^{s'x}]|_{s=0} = \nabla_s'A\, E[\nabla_s\, e^{s'x}]|_{s=0} \\ &= \nabla_s'A\, \nabla_s E[e^{s'x}]|_{s=0} = \nabla_s'A\, \nabla_s M_x(s)|_{s=0} \end{aligned} \tag{1.4.15-4}$$

Using the characteristic function (1.4.15-3) with $\bar{x} = 0$ in the above yields (1.4.15-1) after the evaluation of the gradient. This technique, even though not the simplest for a quadratic form, can be used conveniently to evaluate the expected value of a quartic form, which is needed for the covariance of quadratic forms.

Note that the expected value of a quadratic form (1.4.15-1) holds regardless of the distribution of x.

For the ***quartic form*** to be considered next, the result will rely on the fact that x is Gaussian.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 75

Analogously to (1.4.15-4), one has the ***expected value of a quartic form*** written as follows:

$$E[x'Axx'Bx] = \nabla'_s A\ \nabla_s \nabla'_s B\ \nabla_s M_x(s)|_{s=0} \tag{1.4.15-5}$$

Using (1.4.15-3), it can be shown that after some computations the final result is

$$E[x'Axx'Bx] = \text{tr}(AP)\text{tr}(BP) + 2\text{tr}(APBP) \tag{1.4.15-6}$$

The scalar version of the above for $A = B = 1$ and $P = \sigma^2$ is the well-known expression

$$E[x^4] = 3\sigma^4 \tag{1.4.15-7}$$

The ***covariance of two quadratic forms*** is, using (1.4.15-6),

$$\begin{aligned} E[[x'Ax - E(x'Ax)][x'Bx - E(x'Bx)]] &= E[x'Axx'Bx] - E[x'Ax]E[x'Bx] \\ &= 2\text{tr}(APBP) \end{aligned} \tag{1.4.15-8}$$

These results are used in the state estimation for nonlinear systems where nonlinear functions (system dynamics or measurement equation) are approximated by a series expansion of up to second order.

1.4.16 Mixture Probability Density Functions

A ***mixture*** pdf is a ***weighted sum of pdfs*** with the weights summing up to unity.

A ***Gaussian mixture*** is a pdf consisting of a weighted sum of Gaussian densities

$$p(x) = \sum_{j=1}^{n} p_j \mathcal{N}(x; \bar{x}_j, P_j) \tag{1.4.16-1}$$

where, obviously,

$$\sum_{j=1}^{n} p_j = 1 \tag{1.4.16-2}$$

Denote by A_j the event that x is Gaussian distributed with mean $\bar{x}_j$ and covariance P_j, that is,

$$A_j \triangleq \{x \sim \mathcal{N}(\bar{x}_j, P_j)\} \tag{1.4.16-3}$$

With A_j, $j = 1, \ldots, n$, mutually exclusive and exhaustive, and

$$P\{A_j\} = p_j \tag{1.4.16-4}$$

one can rewrite (1.4.16-1) with the Total Probability Theorem as

$$p(x) = \sum_{j=1}^{n} p(x|A_j) P\{A_j\} \tag{1.4.16-5}$$

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 76

The Moments of a Mixture

The mean of a mixture is easily seen to be

$$\boxed{\bar{x} = \sum_{j=1}^{n} p_j \bar{x}_j} \tag{1.4.16-6}$$

that is, the weighted sum of the means of the component densities regardless of the latter's nature.

The covariance of this mixture is

$$\begin{aligned} E[(x-\bar{x})(x-\bar{x})'] &= \sum_{j=1}^{n} E[(x-\bar{x})(x-\bar{x})'|A_j]p_j \\ &= \sum E[(x-\bar{x}_j+\bar{x}_j-\bar{x})(x-\bar{x}_j+\bar{x}_j-\bar{x})'|A_j]p_j \\ &= \sum p_j E[(x-\bar{x}_j)(x-\bar{x}_j)'|A_j] + \sum(\bar{x}_j-\bar{x})(\bar{x}_j-\bar{x})'p_j \end{aligned} \tag{1.4.16-7}$$

which can be written as

$$\boxed{E[(x-\bar{x})(x-\bar{x})'] = \sum_{j=1}^{n} p_j P_j + \tilde{P}} \tag{1.4.16-8}$$

In the above

$$\tilde{P} \triangleq \sum(\bar{x}_j - \bar{x})(\bar{x}_j - \bar{x})'p_j \tag{1.4.16-9}$$

is the ***spread of the means term***.

An alternative expression for the above is

$$\begin{aligned} \tilde{P} &= \sum \bar{x}_j \bar{x}_j' p_j - \bar{x}\sum \bar{x}_j' p_j - \sum \bar{x}_j p_j \bar{x}' + \bar{x}\bar{x}'\sum p_j \\ &= \sum p_j \bar{x}_j \bar{x}_j' - \bar{x}\bar{x}' \end{aligned} \tag{1.4.16-10}$$

which yields another form for (1.4.16-8)

$$E[(x-\bar{x})(x-\bar{x})'] = \sum_{j=1}^{n} p_j P_j + \sum_{j=1}^{n} p_j \bar{x}_j \bar{x}_j' - \bar{x}\bar{x}' \tag{1.4.16-11}$$

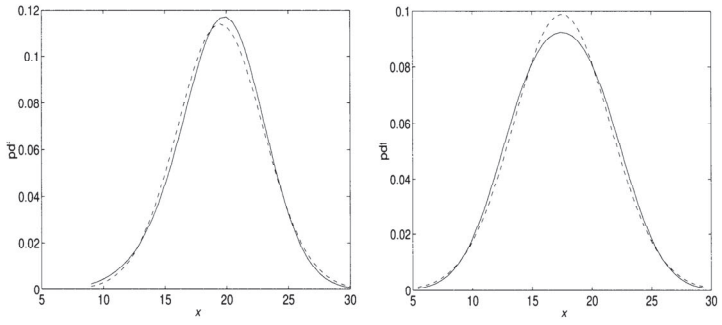
The spread of the means term $\tilde{P}$, defined in (1.4.16-9), is a sum of dyads with positive weightings, and thus it follows from Subsection 1.3.6 that $\tilde{P}$ is *positive semidefinite*.

Note that (1.4.16-6) and (1.4.16-8), which will be referred to as the ***mixture equations***, hold even if the densities in the mixture are *not Gaussian*.

Approximation of a Mixture

A mixture pdf can be approximated by a single Gaussian pdf with momcnts equal to those of the mixture, given by (1.4.16-6) and (1.4.16-8) — this is called ***moment matching***.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 77



(a) $p_1 = 0.1, \quad p_2 = 0.9$ (b) $p_1 = 0.5, \quad p_2 = 0.5$

Figure 1.4.16-1: The exact Gaussian mixture pdf (solid line) and the corresponding moment-matched Gaussian (dashed line) for $\bar{x}_1 = 15, \bar{x}_2 = 20, \sigma_1^2 = \sigma_2^2 = 10$.

Figures 1.4.16-1 and 1.4.16-2 present comparisons between the exact pdf of Gaussian mixtures and the corresponding *moment-matched Gaussian* for several values of the parameters entering into the mixture.

In Fig. 1.4.16-1 the difference of the means is 5 while in Fig. 1.4.16-2 it is 10; in both cases the (common) standard deviation of the mixture components is $\sqrt{10}$.

The resulting "umbrella" Gaussian appears to be close to the exact pdf of the mixture as long as its components are not too far apart; namely, the distance between the means of the componenets is up to about *two standard deviations*.

This approximate condition is met for the cases depicted in Fig. 1.4.16-1, where the match appears good, and it is not met for the cases in Fig. 1.4.16-2, where the match does not appear good.

These results are used in obtaining recursive filtering algorithms for systems with hybrid uncertainties — ***hybrid systems***. Systems falling into this category are those described by multiple models and/or with measurements of uncertain origin. (See also problem 1-4.)

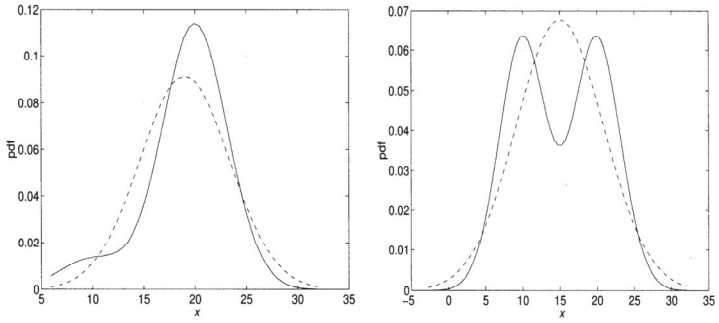
1.4.17 Chi-Square Distributed Random Variables

If the n-dimensional random vector x is Gaussian, with mean $\bar{x}$ and covariance P, then the (scalar) random variable defined by the quadratic form

$$q = (x - \bar{x})' P^{-1} (x - \bar{x}) \tag{1.4.17-1}$$

can be shown to be the sum of the squares of n independent zero-mean, unity-variance Gaussian random variables. Such a random variable is said to have a

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 78



(a) $p_1 = 0.1, \ p_2 = 0.9$ (b) $p_1 = 0.5, \ p_2 = 0.5$

Figure 1.4.16-2: The exact Gaussian mixture pdf (solid line) and the corresponding moment-matched Gaussian (dashed line) for $\bar{x}_1 = 10, \bar{x}_2 = 20, \sigma_1^2 = \sigma_2^2 = 10$.

chi-square distribution with n degrees of freedom (the meaning of degrees of freedom is discussed in Subsection 3.6.2).

This can be seen as follows. Let

$$u \triangleq P^{-1/2}(x - \bar{x}) \tag{1.4.17-2}$$

Then u is Gaussian with

$$E[u] = 0 \tag{1.4.17-3}$$

$$E[uu'] = P^{-1/2}E[x - \bar{x})(x - \bar{x})']P^{-1/2} = P^{-1/2}PP^{-1/2} = I \tag{1.4.17-4}$$

where I is the identity matrix. Therefore, since the covariance matrix of u is diagonal, its components are uncorrelated and, in view of the fact that they are jointly Gaussian, they are also independent.

Thus

$$q = u'u = \sum_{i=1}^{n} u_i^2 \tag{1.4.17-5}$$

where, from (1.4.17-3) and (1.4.17-4), u_i are now ***standard Gaussian random variables*** (zero mean and unity variance)

$$u_i \sim \mathcal{N}(0, 1) \tag{1.4.17-6}$$

Therefore q is chi-square distributed with n degrees of freedom, which can be written as

$$q \sim \chi_n^2 \tag{1.4.17-7}$$

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 79

The mean and variance of the χ^2_n random variable q are

$$E[q] = E\Big[\sum_{i=1}^{n} u_i^2\Big] = n \tag{1.4.17-8}$$

$$\begin{aligned} \text{var}(q) &= E\Big[\sum_{i=1}^{n}(u_i^2 - 1)\Big]^2 = \sum_{i=1}^{n} E[(u_i^2 - 1)^2] \\ &= \sum_{i=1}^{n}\big(E[u_i^4] - 2E[u_i^2] + 1\big) = \sum_{i=1}^{n}(3 - 2 + 1) = 2n \end{aligned} \tag{1.4.17-9}$$

where the fact that the cross-terms are zero mean has been used together with the expression of the fourth moment of a Gaussian variable, as given in (1.4.15-7).

The pdf of q — the chi-square density with n degrees of freedom — is

$$p(q) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} q^{\frac{n-2}{2}}\, e^{-\frac{q}{2}} \qquad q \geq 0 \tag{1.4.17-10}$$

where Γ is the gamma function, with the following useful properties:

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \qquad \Gamma(1) = 1 \qquad \Gamma(m+1) = m\Gamma(m) \tag{1.4.17-11}$$

The square root of q has the ***chi distribution***. The "chi-distributed" random variables with $n = 2$ and 3 degrees of freedom (and multiplied by an arbitrary constant) have the ***Rayleigh distribution*** and ***Maxwell distribution***, respectively.

Given the *independent* random variables

$$q_1 \sim \chi^2_{n_1} \qquad q_2 \sim \chi^2_{n_2} \tag{1.4.17-12}$$

then it can be easily shown that their sum

$$q_3 = q_1 + q_2 \tag{1.4.17-13}$$

is chi-square distributed with

$$n_3 = n_1 + n_2 \tag{1.4.17-14}$$

degrees of freedom. Symbolically, one can write

$$\chi^2_{n_1} + \chi^2_{n_2} = \chi^2_{n_1+n_2} \tag{1.4.17-15}$$

The chi-square distribution is often used to check state estimators for "consistency" — that is, whether their actual errors are consistent with the variances calculated by the estimator.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 80

1.4.18 Weighted Sum of Chi-Square Random Variables

Consider the following independent and identically distributed random variables

$$x_i \sim \chi^2_m \qquad i = 1, \ldots, n \tag{1.4.18-1}$$

with m degrees of freedom chi-square distribution.

Then the ***weighted sum of chi-square variables*** with weights a_i,

$$y_n \triangleq \sum_{i=1}^{n} a_i x_i \tag{1.4.18-2}$$

has a distribution which is not chi-square and is very complicated. Its mean is

$$E[y_n] = m \sum_{i=1}^{n} a_i \tag{1.4.18-3}$$

and, from (1.4.17-9), its variance is

$$\text{var}(y_n) = 2m \sum_{i=1}^{n} a_i^2 \tag{1.4.18-4}$$

The pdf of (1.4.18-2) can be approximated by ***moment matching*** — equating its first two moments to those of another random variable. In this case, the latter is chosen as a "scaled" chi-square with a number of degrees of freedom n', to be determined together with the scaling factor.

To find a random variable whose moments are matched to those of (1.4.18-2), let

$$v \sim \chi^2_{n'} \tag{1.4.18-5}$$

Then the mean and variance of a "scaled" version of v,

$$w \triangleq cv \tag{1.4.18-6}$$

are cn' and $2c^2n'$, respectively. Equating these two moments to (1.4.18-3) and (1.4.18-4) yields the following equations for c and n':

$$m \sum_{i=1}^{n} a_i = cn' \tag{1.4.18-7}$$

$$2m \sum_{i=1}^{n} a_i^2 = 2c^2 n' \tag{1.4.18-8}$$

The solution of these equations yields the scaling factor as

$$c = \frac{\sum_{i=1}^{n} a_i^2}{\sum_{i=1}^{n} a_i} \tag{1.4.18-9}$$

and the number of degrees of freedom of v as

$$n' = \frac{m \left(\sum_{i=1}^{n} a_i\right)^2}{\sum_{i-1}^{n} a_i^2} \tag{1.4.18-10}$$

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 81

Therefore, the distribution of y_n, defined in (1.4.18-2) is approximately

$$y_n \sim \frac{\sum_{i=1}^n a_i^2}{\sum_{i=1}^n a_i} \chi_{n'}^2 \tag{1.4.18-11}$$

Note that the χ_n^2 density was defined only for integer n; a noninteger n makes it into a case of the gamma density. In practice, one can interpolate from the chi-square tables.

Fading Memory Average

Consider the case where the weights in (1.4.18-2) are exponential, that is,

$$a_i = \alpha^{n-i} \tag{1.4.18-12}$$

with $0 < \alpha < 1$. Then the weighted sum

$$z_n \triangleq \sum_{i=1}^n \alpha^{n-i} x_i \tag{1.4.18-13}$$

is called the ***fading memory sum*** of the variables x_i.

Then, for $n \gg 1$ one obtains

$$z_n \sim \frac{1}{1+\alpha} \chi_{n'}^2 \tag{1.4.18-14}$$

and

$$n' = m \frac{1+\alpha}{1-\alpha} \tag{1.4.18-15}$$

When (1.4.18-13) is normalized by dividing it with the sum of the coefficients, it becomes the ***fading memory average*** or ***exponentially discounted average***.

These results are used in real-time monitoring of state estimator performance.

1.4.19 Random Processes

A scalar random variable is a (real) number x determined by the outcome ω of a random experiment

$$x = x(\omega) \tag{1.4.19-1}$$

A (scalar) ***random process*** or a ***stochastic process*** is a *function of time* determined by the outcome of a random experiment

$$x(t) = x(t, \omega) \tag{1.4.19-2}$$

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 82

This is a family or ***ensemble*** of functions of time, in general different for each outcome ω.

The ***mean*** or ***ensemble average*** of the random process is

$$\bar{x}(t) = E[x(t)] = \int_{-\infty}^{\infty} \xi \, p_{x(t)}(\xi) \, d\xi \tag{1.4.19-3}$$

while its ***autocorrelation*** is defined for a real-valued (scalar) process as

$$R(t_1, t_2) \triangleq E[x(t_1)x(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \zeta\eta \, p_{x(t_1),x(t_2)}(\zeta, \eta) \, d\zeta d\eta \tag{1.4.19-4}$$

The ***autocovariance*** of this random process is

$$\begin{aligned} V(t_1, t_2) &\triangleq E[[x(t_1) - \bar{x}(t_1)][x(t_2) - \bar{x}(t_2)]] \\ &= R(t_1, t_2) - \bar{x}(t_1)\bar{x}(t_2) \end{aligned} \tag{1.4.19-5}$$

Note that the autocorrelation of a random process is an (unnormalized) *noncentral moment* — the mean is not subtracted — while the correlation of two random variables is the normalized joint *central moment* (1.4.6-6). If the process is zero mean, the distinction between central and noncentral moments disappears.

For a ***vector-valued random process***, (1.4.19-4) and (1.4.19-5) contain the corresponding outer products.

Stationarity

A random process whose mean is time invariant and whose autocorrelation is of the form

$$R(t_1, t_2) = R(t_1 - t_2) = R(\tau) \tag{1.4.19-6}$$

where

$$\tau = t_1 - t_2 \tag{1.4.19-7}$$

that is, its first two moments are invariant with respect to a shift of the time axis, is called ***wide sense stationary*** or, less rigorously, ***stationary***.

As can be seen from (1.4.19-4), for a real-valued process the autocorrelation is symmetric in its two time arguments

$$R(t_1, t_2) = R(t_2, t_1) \tag{1.4.19-8}$$

Thus

$$R(\tau) = R(-\tau) \tag{1.4.19-9}$$

The ***power spectrum*** or ***power spectral density*** of a *stationary* random process is the Fourier transform (denoted by $\mathcal{F}$) of its autocorrelation

$$S(\omega) = \mathcal{F}\{R(\tau)\} = \int_{-\infty}^{\infty} e^{-j\omega\tau} R(\tau) \, d\tau \tag{1.4.19-10}$$

where ω denotes the (angular) frequency.

Strict stationarity requires that all pdfs (rather than only moments up to second order) be invariant with respect to time shift. In practice, wide-sense stationarity is about all one can assume (and hope for).

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 83

Example

Consider the scalar dynamic system

$$\dot{x}(t) = -ax(t) + n(t) \tag{1.4.19-11}$$

with $a > 0$, i.e., a stable system. The input $n(t)$ is a zero-mean white noise with autocorrelation (same as its autocovariance, since it is zero-mean)

$$E[n(t_1)n(t_2)] = S_0\delta(t_1 - t_2) \tag{1.4.19-12}$$

Assuming that the system in steady state (i.e., the initial condition died out), we obtain

$$x(t) = \int_{-\infty}^{t} h(t,\tau)n(\tau)\,d\tau \tag{1.4.19-13}$$

where the impulse response of the system considered is

$$h(t,\tau) = e^{-a(t-\tau)} \tag{1.4.19-14}$$

The mean of the output $x(t)$ is

$$E[x(t)] = E\left[\int_{-\infty}^{t} h(t,\tau)n(\tau)\,d\tau\right] = \int_{-\infty}^{t} h(t,\tau)E[n(\tau)]\,d\tau = 0 \tag{1.4.19-15}$$

The autocorrelation of $x(t)$ is

$$E[x(t_1)x(t_2)] = E\left[\int_{-\infty}^{t_1} h(t_1,\tau_1)n(\tau_1)\,d\tau_1 \int_{-\infty}^{t_2} h(t_2,\tau_2)n(\tau_2)\,d\tau_2\right] \tag{1.4.19-16}$$

Assuming $t_1 < t_2$, the above can be written as

$$\begin{aligned} E[x(t_1)x(t_2)] &= E\left\{\int_{-\infty}^{t_1} h(t_1,\tau_1)n(\tau_1)\,d\tau_1 \right. \\ &\quad \left. \cdot \left[\int_{-\infty}^{t_1} h(t_2,\tau_2)n(\tau_2)\,d\tau_2 + \int_{t_1}^{t_2} h(t_2,\tau_2)n(\tau_2)\,d\tau_2\right]\right\} \end{aligned} \tag{1.4.19-17}$$

In view of the fact that the first and last integral above are over nonoverlapping intervals, the expected value of their product will be zero due to the whiteness of $n(t)$, whose autocorrelation is zero at any two different times.

Thus one can write

$$\begin{aligned} E[x(t_1)x(t_2)] &= E\left\{\int_{-\infty}^{t_1} e^{-a(t_1-\tau_1)}n(\tau_1)\,d\tau_1 \int_{-\infty}^{t_1} e^{-a(t_2-\tau_2)}n(\tau_2)\,d\tau_2\right\} \\ &= e^{a(t_1-t_2)}E\left\{\int_{-\infty}^{t_1} e^{-a(t_1-\tau_1)}n(\tau_1)\,d\tau_1 \int_{-\infty}^{t_1} e^{-a(t_1-\tau_2)}n(\tau_2)\,d\tau_2\right\} \\ &= e^{a(t_1-t_2)}E[x(t_1)^2] \end{aligned} \tag{1.4.19-18}$$

If $t_1 > t_2$ the result would be the same except that they are switched. Therefore,

$$E[x(t_1)x(t_2)] = e^{-a|t_1-t_2|}E[x(t_1)^2] \tag{1.4.19-19}$$

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 84

i.e., the autocorrelation decreases exponentially from its peak, which is the (steady-state) mean square value of $x(t)$ — its instantaneous power.

The mean square value of $x(t)$ is

$$\begin{aligned} E[x(t)^2)] &= E\left\{\int_{-\infty}^{t} e^{-a(t-\tau_1)}n(\tau_1)\,d\tau_1 \int_{-\infty}^{t} e^{-a(t-\tau_2)}n(\tau_2)\,d\tau_2\right\} \\ &= \int_{-\infty}^{t}\int_{-\infty}^{t} e^{-a(2t-\tau_1-\tau_2)}E[n(\tau_1)n(\tau_2)]\,d\tau_1\,d\tau_2 \\ &= e^{-2at}\int_{-\infty}^{t}\int_{-\infty}^{t} e^{a(\tau_1+\tau_2)}S_0\delta(\tau_1-\tau_2)\,d\tau_1\,d\tau_2 \\ &= e^{-2at}S_0\int_{-\infty}^{t} e^{2a\tau_1}\,d\tau_1 = e^{-2at}\frac{S_0}{2a}\left[e^{2at}-0\right] = \frac{S_0}{2a} \end{aligned} \quad (1.4.19\text{-}20)$$

Thus

$$R_{xx}(\tau) = E[x(t)x(t+\tau)] = \frac{S_0}{2a}e^{-a|\tau|} \quad (1.4.19\text{-}21)$$

The Fourier transform of the above, using the unit step function $1(\cdot)$ to rewrite the two-sided exponential, yields the spectrum of x as

$$\begin{aligned} S_{xx}(\omega) &= \mathcal{F}\{R_{xx}(\tau)\} = \frac{S_0}{2a}\mathcal{F}\{e^{a\tau}1(-\tau)+e^{-a\tau}1(\tau)\} \\ &= \frac{S_0}{2a}\left[\frac{1}{a-j\omega}+\frac{1}{a+j\omega}\right] = S_0\frac{1}{a^2+\omega^2} \end{aligned} \quad (1.4.19\text{-}22)$$

Since this stable system acts as a low-pass filter, its effect on the spectrum of the input (which, being a white noise, has a flat spectrum) is to attenuate the high frequencies.

Ergodicity

If a stationary random process is ***ergodic***, then ***time averages*** (of some functions of the process) are equal to the corresponding expected values — ***ensemble averages*** — given by (1.4.19-3), that is,

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} x(t)dt = \bar{x} \quad (1.4.19\text{-}23)$$

Time averages are used in real-time performance monitoring of state estimation filters.

White Noise

A (not necessarily stationary) random process whose autocovariance is zero for any two different times is called ***white noise***. In this case

$$V(t_1,t_2) = \sigma^2(t_1)\delta(t_1-t_2) = S_0(t_1)\delta(t_1-t_2) \quad (1.4.19\text{-}24)$$

where $\sigma^2(t_1)$ is its "instantaneous variance" and $\delta(\cdot)$ is the ***Dirac (impulse) delta function***. The above property is the ***wide-sense whiteness*** as opposed to the ***strict-sense whiteness***, which is defined by *independence* rather than *uncorrelatedness* as in (1.4.19-24).

A ***stationary zero-mean white*** process has the autocorrelation

$$R(\tau) = E[x(t+\tau)x(t)] = S_0\delta(\tau) \tag{1.4.19-25}$$

The impulse function autocorrelation (1.4.19-25) leads to a power spectral density S_0 constant across the frequency spectrum.

It is convenient in some applications to consider a ***nonstationary zero-mean white*** process $x(t)$. The autocorrelation of such a process is

$$E[x(t_1)x(t_2)] = S_0(t_1)\delta(t_1 - t_2) \tag{1.4.19-26}$$

where $S_0(t)$ is, with some abuse of language, the "time-varying spectral density" or the "instantaneous variance." The term ***instantaneous variance*** is in fact *incorrect* in this case since the variance is actually *infinite*.[11]

Note on Whiteness

In the sequel, when a result requires whiteness, it usually requires strict-sense whiteness. In practice, however, one has only moments (typically up to second order) and thus only wide-sense whiteness can be assumed. Then the result is of practical value only when the strict sense whiteness is replaced by wide-sense whiteness, although in this case the result is only approximately true in theory.

1.4.20 Random Walk and the Wiener Process

The ***Wiener random process*** (or Wiener-Levy or Brownian motion) is a limiting form of the ***random walk***: the sum of independent steps of size $s \to 0$, equiprobable in each direction, taken at intervals $\Delta \to 0$ such that

$$\frac{s}{\sqrt{\Delta}} \to \sqrt{\alpha} \tag{1.4.20-1}$$

where α is a constant. This yields a stochastic process $\mathbf{w}(t)$ with the following pdf [assuming $\mathbf{w}(0) = 0$],

$$p[\mathbf{w}(t)] = \mathcal{N}[\mathbf{w}(t); 0, \alpha t] \tag{1.4.20-2}$$

that is, normal, zero mean, and with variance αt.

[11] Also, the physical dimension of σ^2 is the square of the dimension of x, *multiplied by time*, since δ has as dimension the inverse of time.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 86

Note that the Wiener process is *nonstationary*. It relates to the zero-mean white noise, denoted here as $n(t)$, as follows

$$\mathbf{w}(t) = \int_0^t n(\tau)\, d\tau \tag{1.4.20-3}$$

where

$$E[n(t_1)n(t_2)] = \alpha\delta(t_1 \quad t_2) \tag{1.4.20-4}$$

Another way of writing (1.4.20-3) is

$$d\mathbf{w}(t) = n(t)\, dt \tag{1.4.20-5}$$

which shows that the Wiener process is an ***independent increment process***. Formally, the white noise is the derivative of the Wiener process; however, this is not rigorous, since the Wiener process is ***nowhere differentiable*** — its derivative has infinite variance.

The autocorrelation of the Wiener process is

$$E[\mathbf{w}(t_1)\mathbf{w}(t_2)] - \alpha\ \min(t_1, t_2) \tag{1.4.20-6}$$

The white noise and the Wiener process are used to model unknown inputs (maneuvers) in state estimation/tracking.

Note

Since the Wiener process is not stationary, it does not have a power spectral density.

1.4.21 Markov Processes

Markov processes are defined by the following ***Markov property***

$$p[x(t)|x(\tau), \tau \le t_1] = p[x(t)|x(t_1)] \qquad \forall t > t_1 \tag{1.4.21-1}$$

that is, the past up to any t_1 is *fully characterized* by the value of the process at t_1.

An equivalent statement to the above is:

"The future is independent of the past if the present is known."

The Wiener process is Markov. This follows from the fact that it is the integral of white noise

$$\mathbf{w}(t) = \mathbf{w}(t_1) + \int_{t_1}^t n(\tau)\, d\tau \tag{1.4.21-2}$$

and $n(\tau)$, $\tau \in [t_1, t]$ is independent of $\mathbf{w}(t_1)$.

Furthermore, the state $x(t)$ of a (possibly time-varying) dynamic system driven by white noise $n(t)$,

$$\dot{x}(t) = f[t, x(t), n(t)] \tag{1.4.21-3}$$

is a Markov process. In general, both $x(t)$ and $n(t)$ are vector-valued random processes.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 87

Markov Processes with Rational Spectra

Given a *linear time-invariant dynamic system excited by a stationary white noise* $n(t)$ with mean zero

$$\dot{x}(t) = Ax(t) + Bn(t) \tag{1.4.21-4}$$

its state is (in steady state — if the system is stable) a ***stationary Markov process*** with spectrum (power spectral density)

$$S(\omega) = H(j\omega)QH(j\omega)^* \tag{1.4.21-5}$$

In the above the asterisk denotes complex conjugate transpose and

$$H(j\omega) = (j\omega I - A)^{-1}B \tag{1.4.21-6}$$

is the ***transfer function matrix*** (from the input to the state) of the system (1.4.21-4) and the autocorrelation of the input (in general, a matrix) is

$$R_n(t_1, t_2) = E[n(t_1)n(t_2)'] = Q\delta(t_1 - t_2) \tag{1.4.21-7}$$

The matrix Q is sometimes called (nonrigorously) the covariance of $n(t)$; actually it is its *power spectral density*.

Note that (1.4.21-5) is a ***rational spectrum*** — a ratio of polynomials.

Conversely, one has the following result:

> *Every Markov process with a rational spectrum can be represented as a linear time-invariant system excited by white noise.*

The models used in state estimation are Markov processes — linear or nonlinear systems driven by white noise. These systems can be time-invariant or time-varying and the noise can be stationary or nonstationary. If a time-invariant system is driven by a stationary noise that is not white, but has a rational spectrum, then it has to undergo ***prewhitening***: It is the output of a subsystem driven by white noise. The transfer function of this subsystem is obtained by ***spectral factorization*** — decomposition of the output spectrum in the form (1.4.21-5). For more details on spectral factorization, see, e.g., [Poor88].

Prewhitening

The prewhitening is illustrated in Figs. 1.4.21-1 and 1.4.21-2. In Fig. 1.4.21-1 the system $\mathcal{S}$ with state $x(t)$ is driven by the ***autocorrelated noise*** (also called ***colored noise***) $n(t)$. Since $n(t)$ is not white, $x(t)$ is not a Markov process.

Assume that $n(t)$ can be represented as the output of a system $\mathcal{S}_0$ with white noise input $v(t)$. With the state of the ***prewhitening system*** or ***shaping filter*** $\mathcal{S}_0$ denoted as $x_0(t)$, the augmented state of the composite system shown in Fig. 1.4.21-2,

$$y(t) \triangleq \begin{bmatrix} x(t) \\ x_0(t) \end{bmatrix} \tag{1.4.21-8}$$

is then a Markov process.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 88

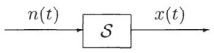


Figure 1.4.21-1: A system driven by autocorrelated noise.

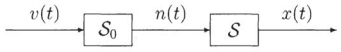


Figure 1.4.21-2: The same system augmented to be driven by white noise.

Example of Spectral Factorization

Consider the random process $x(t)$ with the spectrum

$$S_{xx}(\omega) = S_0 \frac{1}{a^2 + \omega^2} \tag{1.4.21-9}$$

This can be factorized as follows. Consider the random process $x(t)$ with the spectrum

$$S_{xx}(\omega) = \frac{1}{a + j\omega} S_0 \frac{1}{a - j\omega} \tag{1.4.21-10}$$

In the above, one can identify the *causal* transfer function

$$H(\omega) = \frac{1}{a + j\omega} \tag{1.4.21-11}$$

as the transfer function of the system

$$\dot{x}(t) = -ax(t) + n(t) \tag{1.4.21-12}$$

driven by the white noise $n(t)$.

If $n(t)$ is zero-mean with spectrum S_0, then the above system is the *prewhitening system* for the autocorrelated process $x(t)$.

Note that in the course of the factorization, the roles of H and its complex conjugate H^* can be switched. It is, however, necessary to choose H such that it represents a stable and causal system. Namely, the impulse response, given by the inverse Fourier transform, denoted by $\mathcal{F}^{-1}$, of the transfer function, i.e.,

$$\mathcal{F}^{-1}\left\{\frac{1}{a + j\omega}\right\} = e^{-at}1(t) \tag{1.4.21-13}$$

is ***causal***. The impulse response corresponding to the other factor, namely,

$$\mathcal{F}^{-1}\left\{\frac{1}{a - j\omega}\right\} = e^{at}1(-t) \tag{1.4.21-14}$$

is ***noncausal*** (cannot be implemented by a physical system) and thus cannot serve as H.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 89

1.4.22 Random Sequences, Markov Sequences and Markov Chains

A ***random sequence***, or a ***discrete-time stochastic process***, is a time-indexed sequence of random variables

$$X^k = \{x(j)\}_{j=1}^k \qquad k = 1, 2, \ldots \tag{1.4.22-1}$$

Similarly to the continuous-time definition of the Markov property, a random sequence is Markov if

$$p[x(k)|X^j] = p[x(k)|x(j)] \qquad \forall k > j \tag{1.4.22-2}$$

The (real-valued) *zero-mean sequence* $v(j)$, $j = 1, \ldots,$ is a ***discrete-time white noise*** (a ***white sequence***) if

$$E[v(k)v(j)] = q(k)\delta_{kj} \tag{1.4.22-3}$$

where the ***Kronecker delta function***

$$\delta_{kj} = \begin{cases} 1 & \text{if } k = j \\ 0 & \text{if } k \neq j \end{cases} \tag{1.4.22-4}$$

is used and $q(k)$ denotes its variance. If $q(k) = q$, that is, the variance is time-invariant, then this is a ***stationary white sequence***.

The sequence with property (1.4.22-3) is actually only uncorrelated, or weakly independent — strong independence requires that (1.4.7-3) be satisfied. However, in practice one does not have pdfs but only moments up to second order, and thus the usual assumption is the ***weak independence*** (***uncorrelatedness***) indicated in (1.4.22-3).

The state $x(k)$ of a dynamic system excited by white noise $v(k)$

$$x(k+1) = f[k, x(k), v(k)] \tag{1.4.22-5}$$

is a ***discrete-time Markov process*** or ***Markov sequence***. In general, both $x(k)$ and $v(k)$ are vector-valued.

The state of a linear dynamic system excited by white Gaussian noise

$$x(k+1) = F(k)x(k) + v(k) \tag{1.4.22-6}$$

is a ***Gauss-Markov sequence***.

The reasons for this are as follows: Because of the linearity, $x(k)$ is Gaussian (assuming the initial condition is Gaussian); and because of the whiteness of the driving ("process") noise, it is Markov.

A special case of (1.4.22-6), for a scalar x, is

$$x(k+1) = x(k) + v(k) \tag{1.4.22-7}$$

in which case x becomes the *integral (sum)* of the white noise sequence terms, and is called a ***discrete-time Wiener process***.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 90

Markov Chains

A ***Markov chain*** is a special case of a Markov sequence, in which the state space is discrete and finite:

$$x(k) \in \{x_i,\ i = 1, \ldots, n\} \tag{1.4.22-8}$$

Its characterization is given in full by the ***transition (jump) probabilities***

$$P\{x(k) = x_j \mid x(k-1) = x_i\} \triangleq \pi_{ij} \qquad i, j = 1, \ldots, n \tag{1.4.22-9}$$

and the initial probabilities.

Define the vector

$$\mu(k) \triangleq [\mu_1(k), \ldots, \mu_n(k)]' \tag{1.4.22-10}$$

where the components are the probabilities of the chain being in state i

$$\mu_i(k) \triangleq P\{x(k) = x_i\} \tag{1.4.22-11}$$

This vector describes the pmf of the state of the chain.

The evolution in time of (1.4.22-11) is then given by

$$\mu_i(k+1) = \sum_{j=1}^{n} \pi_{ji}\mu_j(k) \qquad i = 1, \ldots, n \tag{1.4.22-12}$$

It can be easily shown that the above can be written in vector form with notation (1.4.22-10) as

$$\mu(k+1) = \Pi'\mu(k) \tag{1.4.22-13}$$

where

$$\Pi = [\pi_{ij}] \tag{1.4.22-14}$$

is the ***transition matrix of the Markov chain.***

1.4.23 The Law of Large Numbers and the Central Limit Theorem

The Law of Large Numbers

The ***law of large numbers (LLN)*** states loosely that the sum of a large number of random variables tends, under some fairly nonrestrictive conditions, to its expected value. One of the versions of the LLN is the following.

Given a stationary sequence of random variables x_i, $i = 1, \ldots$, with means

$$E[x_i] - \bar{x} \tag{1.4.23-1}$$

and covariances

$$E[(x_i - \bar{x}_i)(x_j - \bar{x}_j)] = \sigma^2\rho(i-j) \tag{1.4.23-2}$$

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 91

where the correlation coefficients are such that

$$\lim_{|i-j|\to\infty} \rho(i-j) = 0 \tag{1.4.23-3}$$

that is, this sequence is *asymptotically uncorrelated.*

If the correlation coefficients in (1.4.23-3) tend to zero "sufficiently fast" (e.g., exponential decay), then the ***sample average***

$$y_n \triangleq \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1.4.23-4}$$

converges, as $n \to \infty$, in the mean square sense to its expected value

$$\bar{y}_n = \frac{1}{n}\sum_{i=1}^{n} \bar{x} = \bar{x} \tag{1.4.23-5}$$

In other words, the variance of y_n in (1.4.23-4) tends to zero as n increases.

If the stronger condition of uncorrelatedness holds, then the convergence of y_n given by (1.4.23-4) to $\bar{y}_n$ given by (1.4.23-5) is quite obvious — it can be easily shown that in this case the mean square value of their difference is σ^2/n.

The Central Limit Theorem

The ***central limit theorem (CLT)*** states that if the sequence x_i, $i = 1, \ldots$, consists of *independent* random variables, then under some reasonably mild conditions the pdf of the sum

$$z_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i \tag{1.4.23-6}$$

will tend to a Gaussian pdf as $n \to \infty$.

If these random variables are ***independent and identically distributed (i.i.d.)*** zero-mean and with variance σ^2, then the z_n tends to $\mathcal{N}(0, \sigma^2)$.

Since y_n defined in (1.4.23-4) is a scaled version of z_n, the central limit theorem holds for y_n as well. Thus, if the random variables x_i are i.i.d., then for large n,

$$y_n \sim \mathcal{N}(\bar{y}_n, \sigma^2_{y_n}) \tag{1.4.23-7}$$

where

$$\bar{y}_n = \bar{x} \tag{1.4.23-8}$$

$$\sigma^2_{y_n} = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{\sigma^2}{n} \tag{1.4.23-9}$$

are the mean and variance of y_n.

The independence requirement can be weakened — there are versions of the central limit theorem that allow dependence, as long as the random variables become "asymptotically independent," in a manner somewhat similar to the asymptotic uncorrelatedness property (1.4.23-3).

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 92

Note on the CLT

The CLT has a very important role in characterizing many real-world sources of uncertainty: It is used as the justification/excuse to make the omnipresent Gaussian assumption. For example, the thermal noise in electronic devices, as the sum of many "small contributions," is indeed close to Gaussian.

1.5 BRIEF REVIEW OF STATISTICS

1.5.1 Hypothesis Testing

Consider the ***hypothesis testing*** problem between the following "simple" hypotheses — each is defined by a point (a certain value of a parameter).

The ***null hypothesis*** is

$$H_0: \quad \theta = \theta_0 \tag{1.5.1-1}$$

while the ***alternate hypothesis*** is

$$H_1: \quad \theta = \theta_1 \tag{1.5.1-2}$$

where θ is a certain parameter whose value is equal to a certain value θ_0 (θ_1) under H_0 (H_1).

The ***type I error*** probability is defined as

$$P_{e_I} \triangleq P\{\text{“}H_1\text{”}|H_0\} \triangleq P\{\text{accept } H_1|H_0 \text{ true}\} \tag{1.5.1-3}$$

while the ***type II error*** probability is

$$P_{e_{II}} \triangleq P\{\text{“}H_0\text{”}|H_1\} \triangleq P\{\text{accept } H_0|H_1 \text{ true}\} \tag{1.5.1-4}$$

In signal detection, if H_0 stands for "signal equal to zero" (i.e., absent) and H_1 stands for "signal present," then the type I error is a ***false alarm*** while the type II error is a ***miss***.

The ***power of the test*** is

$$\pi \triangleq P\{\text{“}H_1\text{”}|H_1\} = 1 - P_{e_{II}} \tag{1.5.1-5}$$

and it measures the test's capability to discern H_1 when it is true — the ***detection probability***.

The decision as to which hypothesis to accept is made based on a set of observations, Z, whose pdfs conditioned on H_0 and H_1 are known.

According to the ***Neyman-Pearson Lemma***, the optimal decision, in the sense of *minimizing the probability of type II error* (or maximizing the power of the test), *subject to a given (maximum) probability of type I error* is as follows.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 93

The test, based on the ***likelihood ratio***, is

$$\Lambda(H_1, H_0) = \frac{p(Z|H_1)}{p(Z|H_0)} \underset{\text{"}H_0\text{"}}{\overset{\text{"}H_1\text{"}}{\gtrless}} \Lambda_0 \tag{1.5.1-6}$$

that is, "accept H_1" if Λ exceeds the threshold Λ_0 and "accept H_0" if Λ is below this threshold. The threshold Λ_0 is such that

$$P\{\Lambda(H_1, H_0) > \Lambda_0 | H_0\} = P_{e_I} \tag{1.5.1-7}$$

The pdf of the observations, $p(Z|H_j)$, is called the ***likelihood function*** of H_j.

Example

Consider the test between the hypotheses (1.5.1-1) and (1.5.1-2) with $\theta_0 = 0$ and $\theta_1 > 0$, based on the single observation

$$z \sim \mathcal{N}(\theta, \sigma^2) \tag{1.5.1-8}$$

that is, we are testing whether the mean of z is zero or positive.

The likelihood ratio is

$$\Lambda(H_1, H_0) = \frac{p(z|H_1)}{p(z|H_0)} = e^{-\frac{(z-\theta_1)^2 - z^2}{2\sigma^2}} = e^{\frac{2z\theta_1 - \theta_1^2}{2\sigma^2}} \tag{1.5.1-9}$$

Instead of comparing (1.5.1-9) to a threshold, it is convenient to take its logarithm, incorporate the various constants into the threshold, and then determine it.

Since $\theta_1 > 0$, comparing (1.5.1-9) to a threshold is equivalent to

$$z \underset{\text{"}H_0\text{"}}{\overset{\text{"}H_1\text{"}}{\gtrless}} \lambda_0 \tag{1.5.1-10}$$

where λ_0 is the new threshold to be determined.

In view of (1.5.1-8), the pdf of the observation under H_0 is

$$p(z|H_0) = \mathcal{N}(z; 0, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \; e^{-\frac{z^2}{2\sigma^2}} \tag{1.5.1-11}$$

The threshold λ_0 for ***false alarm probability*** $P_{FA} = \alpha$ follows from

$$P_{FA} \triangleq P_{e_I} = P\{z > \lambda_0 | H_0\} = \int_{\lambda_0}^{\infty} \mathcal{N}(z; 0, \sigma^2) dz = \alpha \tag{1.5.1-12}$$

The ***tail mass*** or ***tail probability*** α from (1.5.1-12) is shaded in Fig. 1.5.1-1.

Under H_0 there is only a small probability α that z falls in the shaded tail region of the likelihood function of H_0; if it does fall there, then it is deemed

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 94

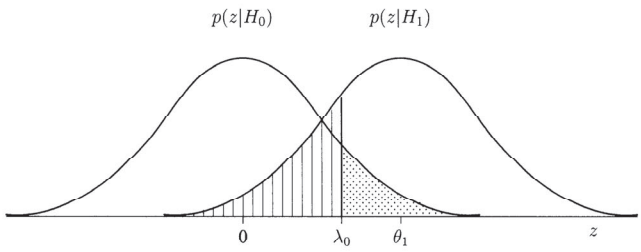


Figure 1.5.1-1: Test between two simple hypotheses.

unlikely that H_0 is true, and it is rejected. In this case it is said that the ***level of significance*** of H_0 is low: α or less.[12] The latter holds if $z = z_1 > \lambda_0$ because such an observation belongs to an even smaller tail mass.

The type II error probability — the ***miss probability*** — of the test (1.5.1-10) is

$$P_{e_{II}} \triangleq P\{z < \lambda_0 | H_1\} = \int_{-\infty}^{\lambda_0} \mathcal{N}(z; \theta_1, \sigma^2) dz \tag{1.5.1-13}$$

This corresponds to the striped area of Fig. 1.5.1-1.

The power of this test — the ***probability of detection*** — also denoted as P_D is

$$P_D \triangleq \pi = P\{z > \lambda_0 | H_1\} = \int_{\lambda_0}^{\infty} \mathcal{N}(z; \theta_1, \sigma^2) dz = 1 - P_{e_{II}} \tag{1.5.1-14}$$

and, obviously, it depends on θ_1 — the farther it is from 0, the more powerful the test will be.

Example

For $\alpha = 5\%$, the threshold is obtained (see Subsection 1.5.4) based on the one-sided tail of the Gaussian pdf as $\lambda_0 = 1.64\sigma$. Table 1.5.1-1 shows the power of this test as a function of the separation between the two means (in units of σ, the common standard deviation for both hypotheses).

1.5.2 Confidence Regions and Significance

Assume one desires to test the null hypothesis

$$H_0: \quad \theta = 0 \tag{1.5.2-1}$$

[12] Equivalently, one can say that the ***confidence level*** for H_1 is high ($1 - \alpha$ or more).

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 95

Table 1.5.1-1: Power of the test between simple hypotheses as a function of their separation (threshold set for $\alpha = 5\%$).

θ_1/σ	2	3	4	5
Power π	0.641	0.913	0.991	0.9996

versus the ***one-sided alternative hypothesis***

$$H_1' : \quad \theta > 0 \tag{1.5.2-2}$$

based upon a set of observations

$$Z = \{z_i, i = 1, \ldots, n\} \tag{1.5.2-3}$$

where

$$z_i = \theta + w_i \tag{1.5.2-4}$$

and the "noises" w_i are independent and identically distributed:

$$w_i \sim \mathcal{N}(0, \sigma^2) \tag{1.5.2-5}$$

Note that H_1' is a ***composite hypothesis*** — it is defined by more than one point (an interval in this case).

The test is subject to

$$P_{e_I} = \alpha \tag{1.5.2-6}$$

The likelihood ratio for hypothesis H_1' versus H_0 is

$$\Lambda(H_1', H_0) = \frac{p(Z|\theta)}{p(Z|\theta = 0)} = e^{\frac{1}{2\sigma^2}\sum_{i=1}^{n}[z_i^2 - (z_i - \theta)^2]} = e^{\frac{1}{2\sigma^2}\sum_{i=1}^{n}(2z_i\theta - \theta^2)} \tag{1.5.2-7}$$

The comparison of the above to a threshold is equivalent to comparing the sample mean of the observations to another threshold:

$$\bar{z} \triangleq \frac{1}{n}\sum_{i=1}^{n} z_i \underset{H_0}{\overset{H_1'}{\gtrless}} \lambda_1 \tag{1.5.2-8}$$

Equation (1.5.2-8) follows by taking the logarithm of (1.5.2-7) and lumping θ and σ into the (yet undetermined) threshold; the division by n is done for the convenience of normalization — this yields z in the same range regardless of the number of samples n.

For this problem $\bar{z}$ is the ***test statistic*** — the function of the observations used in the test.

The threshold λ_1 is obtained by noting that

$$p(\bar{z}|H_0) = \mathcal{N}(\bar{z}; 0, \frac{\sigma^2}{n}) \tag{1.5.2-9}$$

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 96

and putting the condition

$$P\{\text{“}H_1'\text{”}|H_0\} = P\{\bar{z} > \lambda_1|H_0\} = 1 - \int_{-\infty}^{\lambda_1} \mathcal{N}(\bar{z}; 0, \frac{\sigma^2}{n})d\bar{z} = \alpha \quad (1.5.2\text{-}10)$$

If the one-sided alternate hypothesis (1.5.2-2) is replaced by the ***two-sided alternate hypothesis***, denoted as H_1,

$$H_1: \quad \theta \neq 0 \quad (1.5.2\text{-}11)$$

then it can be shown that (1.5.2-8) is to be replaced by

$$|\bar{z}| \underset{\text{“}H_0\text{”}}{\overset{\text{“}H_1\text{”}}{\gtrless}} \lambda \quad (1.5.2\text{-}12)$$

The threshold λ is obtained by putting the condition

$$P\{\text{“}H_1\text{”}|H_0\} = P\{|\bar{z}| > \lambda|H_0\} = 1 - \int_{-\lambda}^{\lambda} \mathcal{N}(\bar{z}; 0, \frac{\sigma^2}{n})d\bar{z} = \alpha \quad (1.5.2\text{-}13)$$

In other words, λ is such that the ***acceptance region*** for H_0, which is the interval $[-\lambda, \lambda]$, contains $1 - \alpha$ probability mass for the pdf (1.5.2-9).

For $\alpha = 0.05$, one obtains from tables of the normal distribution (see Subsection 1.5.4)

$$\lambda = 1.96\frac{\sigma}{\sqrt{n}} \quad (1.5.2\text{-}14)$$

The power of the test depends on the specific value of θ under the alternate hypothesis H_1.

The sample mean $\bar{z}$ is also the ***maximum likelihood estimate*** (and the ***least squares estimate***) of the unknown parameter θ, i.e.,

$$\bar{z} = \arg\max_{\theta} p(Z|\theta) = \arg\min_{\theta} \sum_{i=1}^{n} (z_i - \theta)^2 \quad (1.5.2\text{-}15)$$

since $p(Z|\theta)$ is the likelihood function of the parameter θ.

Figure 1.5.2-1 illustrates the acceptance region for H_0 — the *two-sided* 95% ***probability region*** for $\bar{z}$ under hypothesis H_0 — the interval $[-\lambda, \lambda]$. The shaded areas represent the tails on the two sides of $p(\bar{z}|H_0)$, whose total probability mass is $\alpha = 0.05$; this is the region of rejection of H_0 (acceptance of H_1).

Since the difference between the sample mean $\bar{z}$ and the true mean θ (whatever its value) is

$$\bar{z} - \theta \sim \mathcal{N}(0, \frac{\sigma^2}{n}) \quad (1.5.2\text{-}16)$$

one can say that the true mean lies within the interval $[\bar{z} - \lambda, \bar{z} + \lambda]$ with “confidence” $1 - \alpha$ (since θ is not a random variable, no probabilistic statement about it can be made). In view of this, such an interval is called the ***confidence region*** for θ.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 97

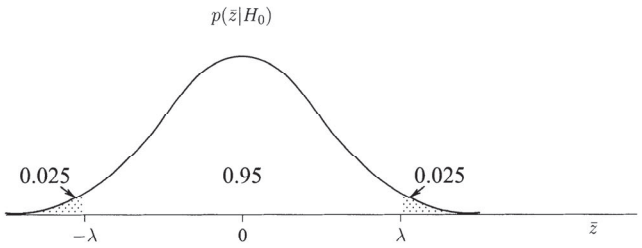


Figure 1.5.2-1: The test between hypotheses H_0 and H_1.

Another way of interpreting test (1.5.2-12) is the following: If the null hypothesis value $\theta = 0$ falls within the confidence region $[\bar{z} - \lambda, \bar{z} + \lambda]$, then H_0 is accepted. This is equivalent to $\bar{z}$ falling within the acceptance region $[-\lambda, \lambda]$.

If the estimate $\hat{\theta} = \bar{z}$ of θ falls outside the interval $[-\lambda, \lambda]$, then it is said to be a ***significant estimate*** — hypothesis H_1 is accepted because H_0 is ***insignificant***.

Note that while $\bar{z}$ is the ***point estimate*** of θ, the confidence region around z can be seen as an ***interval estimate*** of the parameter θ.

Figure 1.5.2-2 illustrates these concepts.

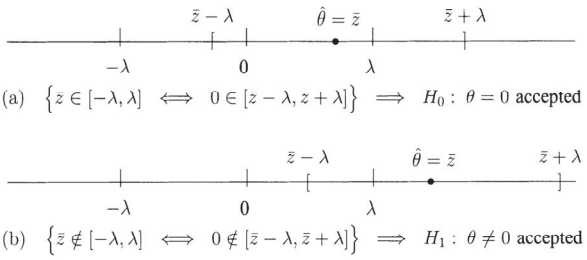


Figure 1.5.2-2: Parameter estimate. (a) $\hat{\theta}$ insignificant; (b) $\hat{\theta}$ significant.

Thus the estimate of a parameter that might be zero is accepted as *significant* (at a "level" of 5% for the null hypothesis) if

$$\frac{|\hat{\theta}|}{\sigma_{\hat{\theta}}} > 1.96 \tag{1.5.2-17}$$

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 98

where the threshold has been determined under the normal assumption and

$$\sigma_{\hat{\theta}} = \frac{\sigma}{\sqrt{n}} \tag{1.5.2-18}$$

is the standard deviation of the estimate, or its ***standard error***.

Defining the ***estimation error*** as

$$\tilde{\theta} \triangleq \theta - \hat{\theta} \tag{1.5.2-19}$$

one can see that

$$\begin{aligned} p(\hat{\theta}|\theta) &= \mathcal{N}(\hat{\theta}; \theta, \sigma_{\hat{\theta}}^2) - \mathcal{N}(\hat{\theta} - \theta; 0, \sigma_{\hat{\theta}}^2) \\ &= \mathcal{N}(\tilde{\theta}; 0, \sigma_{\hat{\theta}}^2) \triangleq p(\tilde{\theta}) = \mathcal{N}(\tilde{\theta}; 0, \sigma_{\tilde{\theta}}^2) \end{aligned} \tag{1.5.2-20}$$

Therefore (1.5.2-18), which is the ***standard deviation of the estimate*** (about the true mean), $\sigma_{\hat{\theta}}$, is the same as $\sigma_{\tilde{\theta}}$, the ***standard deviation of the estimation error*** (which is zero mean), that is,

$$\sigma_{\hat{\theta}} = \sigma_{\tilde{\theta}} \tag{1.5.2-21}$$

The (statistical) significance of parameter estimates is used in choosing the order of models to describe system equations (e.g., for target motion). (See also problem 1-2.)

Example

A parameter θ is estimated based on n measurements, each with an additive error with standard deviation $\sigma = 10$. Assume the estimate obtained from $n = 10$ measurements is $\hat{\theta} = 2$. The significance statistic (at a level of 5%) is in this case

$$\frac{|\hat{\theta}|}{\sigma_{\hat{\theta}}} = \frac{|\hat{\theta}|}{\sigma/\sqrt{n}} = \frac{2}{10/\sqrt{10}} = 0.63 \ll 1.96 \tag{1.5.2-22}$$

i.e., based on these $n = 10$ measurements, this estimate is clearly not statistically significant — the true value of the parameter can be assumed as zero. The threshold above corresponds to 5% level of significance (5% probability that the estimate is "due to noise only," or 95% confidence that it is nonzero).

Assume next that the observations are repeated until one has $n = 100$ measurements and that the estimate is still $\hat{\theta} = 2$. The significance statistic is now

$$\frac{|\hat{\theta}|}{\sigma/\sqrt{n}} = \frac{2}{10/\sqrt{100}} = 2 > 1.96 \tag{1.5.2-23}$$

i.e., based on these $n = 100$ measurements, this estimate is statistically significant, even though only marginally.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 99

Note that it is somewhat arbitrary to use a 5% level of significance level for the threshold for acceptance/rejection of the null hypothesis. A marginally significant parameter can become not significant if the above level is lowered. However, while there is some subjectivity in the selection of the level for the test, this only affects marginal cases. For example, if $\hat{\theta} = 12$, then this value would be accepted even from $n = 10$ measurements since the test statistic is

$$\frac{|\hat{\theta}|}{\sigma/\sqrt{n}} = \frac{12}{10/\sqrt{10}} = 3.8 \tag{1.5.2-24}$$

i.e., large enough to exceed even the 1% level threshold, which is 3.38.

1.5.3 Monte Carlo Runs and Comparison of Algorithms

The performance of an estimation (or control) algorithm is usually quantified by the expected value of a "cost" function

$$J = E[C] \tag{1.5.3-1}$$

If C is, for instance, the squared error in the estimation of a certain variable, then J is the corresponding mean square error.

In many situations, the performance of an algorithm of interest cannot be evaluated analytically.[13] In such a case, ***Monte Carlo simulations (runs)*** are made to obtain an estimate of J from a sample average of *independent* realizations C_i, $i = 1, \ldots, N$, of the cost C. The larger the number of such runs, the smaller is the variability (error) of the resulting estimate. Also, a larger number of runs increases the power of the hypothesis testing used in comparing different algorithms.

The estimate of the performance from N independent runs is the ***sample average*** (or ***sample mean***) of the N realizations of the cost

$$\bar{C} = \frac{1}{N}\sum_{i=1}^{N} C_i \tag{1.5.3-2}$$

with the associated standard deviation — the ***standard error***

$$\sigma_{\bar{C}} = \sqrt{\frac{1}{N^2}\sum_{i=1}^{N}(C_i - \bar{C})^2} \tag{1.5.3-3}$$

The above follows from the fact that the variance of the sample mean $\bar{C}$ is the variance of C divided by N; since this variance is not known, it is replaced by the sample variance, which has another N in the denominator. This is the reason for having N^2 in the denominator of (1.5.3-3).

[13] Those who can, do. Those who cannot, simulate. Those who cannot simulate, supervise.

Ex.-1030 - ADB Safegate et al.
ADB Safegate et al. v. Honeywell International Inc. – PGR2023-00053
Page 100