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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

Meta 1018
p. 2

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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)

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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)

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

Meta 1018
p. 6

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

Meta 1018
p. 7

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Meta 1018
p. 8

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

Meta 1018
p. 9

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

Meta 1018
p. 10

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

Meta 1018
p. 11

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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™ 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

Meta 1018
p. 12

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Meta 1018
p. 13

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)

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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.

Meta 1018
p. 14

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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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 $textgraph^{TM}$ 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 $textgraph^{TM}$ 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^{TM}

$DynaEst^{TM}$ 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."

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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™ 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*™
- 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™ 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*™, 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*™. The source code of *DynaEst*™, which is easy to read, provides a sample implementation of the Kalman filter, as well as the state-of-the-art IMM estimator.

DynaEst™ 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*™.

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.

Meta 1018
p. 17

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

Meta 1018
p. 18

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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]

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$\mathbf{1}$	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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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)

Index

a posteriori, 91
a priori, 91
acceleration increment, 274
acceleration increment over a sampling period, 289
accceleration, longitudinal and lateral, 418
accelerometer, 495
acceptable estimation error, 233
acceptance interval, 235
acceptance region, 76
acceptance regions, 240
actual MSE, 233
adaptive bandwidth, 479
adaptive estimation algorithms, 421
adjoint, 359
adjugate, 22
affine transformation, 362
air traffic control (ATC), 423, 466
algebraic matrix Riccati equation, 211, 350
algebraic Riccati equation, 211
algorithm comparison, 86
alpha filter, 296
alpha filter and suboptimality, 297
alpha filter, bias and total MS error, 297
alpha-beta, 278
alpha-beta filter, 282, 286, 287
alpha-beta filter design and evaluation, 296
alpha-beta filter gain, 278
alpha-beta filter, bias and total MS error, 298
alpha-beta-gamma, 278
alpha-beta-gamma filter, 289
alternate hypothesis, 72
alternative derivation of the Kalman filter gain, 263
Altitude estimation from slant range, 118
amplitude estimation of a sinusoid, 368
angle-only measurements, 137
angular rate, 183
angular rate estimation, 416
arbitrary gain, 262
artificial noise, 243
artificial process noise, 212, 395, 482
asymptotically efficient, 113
asymptotically Gaussian, 113
asymptotically unbiased, 101, 113
ATC, 423
augmented state, 481
augmented state equation, 322
autocorrelated measurement noise, 327
autocorrelated noise, 67
autocorrelated process noise, 320
autocorrelation, 62
autocovariance, 62
autocovariance of the state of a discrete time system, 197
autocovariance of the state of a system 197
average, 37
average covariance associated with the optimal estimate, 378
average NEES, 234, 238
average NIS, 236, 238
average, fading memory, 61
axioms of probability, 31
azimuth, 397

backward recursion, 334
bandwidth of an alpha filter, 298
bandwidth of the filter, 207
bank of filters, 442
base state, 444, 481
batch estimator, 130
Bayes' criterion, 48
Bayes' decision rule, 48
Bayes' formula, 48, 91, 114
Bayes' formula for random variables, 48
Bayes' postulate, 49
Bayes' principle, 48
Bayes' risk, 48
Bayes' rule, 48
Bayes' theorem, 48
Bayesian approach, 91, 114
bearing, 397
bearings-only target localization, 176
best linear state estimator, 207
best linear unbiased estimator, 126
bias and total MS error in the alpha filter, 297
bias and total MS error in the alpha-beta filter, 298
bias in the measurements, 263
bias significance, 400
block processing, 308

C/A code, 504
CA model, 272, 277
calculation of the merging probabilities, 451
calculation of the mixing probabilities, 455
carrier phase measurement, 507

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Cartesian basis vector, 383
cascaded filtering, 528
causal, 68
cdf, 33
central limit theorem (CLT), 71
centralized estimation, 527
Chapman-Kolmogorov equation, 87, 374
characteristic function, 42
Chebyshev inequality, 85
chi distribution, 59
chi-square distribution with n degrees of freedom, 58
chi-square test, 235
Cholesky factor, 211
Cholesky factorization, 21, 311
closed-loop system matrix, 361
CLT, 71
colored noise, 67, 320
column vectors, 19
complementary distribution, 82
complementary event, 31
complementary filtering, 492
completely controllable, 28, 30
completely observable, 28, 30
completion of the squares, 53
composite hypothesis, 75
computational requirements, 210
condition number, 26, 302
conditional covariance matrix, 123, 201
conditional expectation, 50
conditional mean, 99, 123
conditional MSE, 104
conditional pdf, 44
conditional pdf of the sum of two Gaussian random variables, 88
conditional probability, 44
conditional variance, 104
conditional variance versus unconditional variance, 118
confidence level, 74
confidence region, 76, 166
consistency checking, 261
consistency criteria of a filter, 234
consistency evaluation, 233
consistency in the mean square sense, 108
consistency testing, 244, 387
consistency tests, 237
consistent, 234
consistent estimator, 108, 115
constant acceleration (CA) model, 272, 277
constant acceleration motion, 151
Constant altitude trajectory estimation from range-only observations, 175
constant angular rate, 183
constant jerk motion, 151
constant speed, 183
constant velocity (CV) model, 270, 274, 276
constant velocity motion, 150
constant velocity object, 269
continuation, track, 153
continuous noise level adjustment, 422, 424
continuous time state estimate, 346
continuous time state estimation filter, 346
continuous-discrete filter, 347, 358
continuous-time EKF, 416
continuous-time innovation, 347
continuous-time linear stochastic systems, 180
continuous-time linear systems, 27
continuous-time nonlinear systems, 404
controllability condition, 481
controllability matrix, 28, 30
convergence in mean square, 108
converted measurement, 390
convolution, 41
coordinate transformation from polar to Cartesian, 415
coordinated turn, 183
coordinated turn (CT), 467
coordinated turn transition matrix, 198
correction term, 134
correlated noises, 116
correlation between current estimate and smoothed value, 339
correlation coefficient, 39
correlation matrices, 250
cost function, 99
cost functional, 358
cost weighting matrices, 359
countable additivity, 31
course RMS errors, 243
covariance associated with the estimate, 201
covariance matching, 233
covariance matrices, filter-calculated, 249
covariance matrix, 42, 128
covariance matrix of the LS estimator, 131
covariance of the residual, 133, 134, 173
covariance of the state versus covariance of the estimation error., 262
covariance of two quadratic forms, 55
covariance of two scalar random variables, 39
covariance prediction equation, 301
covariance propagation equation, 191
covariance update, 206
covariance update and gain calculation, 314
covariance update equation, 302
cpmf, 35
Cramer-Rao lower bound (CRLB), 109, 164
CRLB, 109, 115, 164
CRLB as the actual covariance, 169
CRLB for a two-model parameter estimation problem, 485
cross-correlated noise sequences, 325
cross-covariance, 122, 362
CT, 467
CT model, 467
cubic fit, 151
cumulative probability distribution function (cdf), 33

Meta 1018
p. 22

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

cumulative probability mass function (cpmf), 35
cumulative standard Gaussian distribution, 35
curse of dimensionality, 380
CV model, 270, 274, 276
CWNA model, 286

data fusion, 524
debiasing threshold, 488
decision, 1
decision-free, 466
decorrelation time, 321
degree of belief, 48
degrees of freedom, 158
delta function, 35, 69
delta range, 507
density, 33
derivative of the inverse of a matrix, 367
derivative with respect to a matrix, 87
design and evaluation, 296
design parameter, 267
detection probability, 72
determinant, 22
DFT, 193
DGPS, 509
diagonal matrix, 21
differenced measurement, 327
Differential GPS (DGPS), 509
diffuse, 49, 117, 137
diffuse pdf, 95
diffuse prior, 302, 307
diffuse uniform prior pdf, 95
diffusion process, 180, 410
dilution of precision (DOP), 507
dimension of a pdf, 86
dimensionality, curse of, 380
Dirac (impulse) delta function, 35, 65
direct discrete time model, 189
Dirichlet prior, 49
discounting, 396
discrete time gain, 187
discrete white noise acceleration model, 273
discrete white noise acceleration (DWNA) model, 280
discrete Wiener process acceleration model, 274
discrete-time (difference) matrix Riccati equation, 211
discrete-time Fourier transform (DTFT), 193
discrete-time linear stochastic systems, 188
discrete-time Markov process, 69
discrete-time measurement, 345
discrete-time measurement equation, 188
discrete-time process noise, 187
discrete-time stochastic process, 69
discrete-time systems, 187
discrete-time transfer function, 193
discrete-time white noise, 69
discrete-time Wiener process, 69
discrete-valued parameter, 116
discretized continuous-time white noise acceleration (CWNA) model, 286
distance estimation in sailing, 419
distributed estimation fusion, 528
distribution, 33
distribution-free probability region, 85
divergence, 232
divergence of the filter, 385
DOP, 507
downcrossing threshold, 426
drift, 180, 410
DTFT, 193
dual problems, 360, 361
duality, 305
DWNA model, 273, 280
DWPA model, 274
dyad, 26
dyad of a state estimate — the expected value, 339
DynaEst, 219, 460
dynamic equation, 27, 180
dynamic estimation problem, 91
dynamic MM estimator, 441
dynamic programming, 407
dynamic range capability, 479

effective window length, 425, 432
efficiency check, 165
efficient, 109
efficient estimator, 115
eigenvalues, 25
eigenvectors, 25
EKF, 371, 382
EKF prediction equations for a system with unknown parameters, 486
EKF simulation for state and parameter estimation, 486
ellipsoid, g-sigma, 166
ensemble, 62
ensemble average, 62
ensemble averages, 64, 237
ergodic, 64
error state of an inertial navigation system, 497
estimability, 166
estimate, 90, 122
estimate and covariance combination, 457
estimate of a parameter, 114
estimate of the state, 201
estimation, 1
estimation criteria that lead to the conditional mean, 117
estimation error, 78, 90, 101, 125, 201, 233
estimation from indirect observations, 175
estimation of the probability of an event, 107
estimation with correlated measurements, 119
estimation with correlated noises, 172
estimation, dynamic, 91
estimation, static, 91
estimator, 90
estimator design parameters, 446
estimator of a parameter, 114

Meta 1018
p. 23

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

estimator optimality, 199, 233
evaluation of estimator optimality, 199, 233
evaluation of the CRLB, 113
event, 31
evidence from the data, 48, 91
exhaustive, 45
expected value, 37
expected value of a quadratic form, 54
expected value of a quartic form, 55
exponential pdf, 94
exponentially autocorrelated acceleration model, 271
exponentially autocorrelated noise, 321
exponentially decaying autocorrelation, 186
exponentially discounted average, 61, 192, 425
extended Kalman filter (EKF), 371, 382
extrapolation, 152

factorization, 186, 193
factorization of the covariance matrix, 303
factorization, Cholesky, 21
factorization, LDL′, 21
factorization, LDU, 21
factorized prediction covariance, 313
fading memory, 396
fading memory average, 61, 192, 425, 432, 486
fading memory average and effective window length, 197
fading memory sum, 61, 192
false alarm, 72
false alarm probability, 73
fault detection, 487
filter design parameter, 267
filter for CWPA model without initial estimates, 318
filter for DWNA model without initial estimates, 318
filter for DWPA model without initial estimates, 318
filter gain, 204, 205, 346
filter stability, 212
filter tuning, 243
filter-calculated covariance, 233
filter-calculated covariance matrices, 249
filter-calculated covariances, 207
filtering, 2
FIM, 110, 115, 164
finite impulse response (FIR), 194
finite-sample consistency, 234
FIR, 194
first moment, 37
first-order polynomial fitting, 247
Fisher approach, 91
Fisher information, 109
Fisher information matrix, 115
Fisher information matrix (FIM), 110, 164
Fisher information matrix in bearings-only target localization, 173
Fisher information matrix in the LS problem, 173
fitting error, 155–157, 161, 167
fixed interval prediction, 330, 356
fixed lead prediction, 330
fixed lead prediction evolution in continuous time, 368
fixed point prediction, 330
fixed point prediction — the mean and covariance, 338
fixed point prediction evolution in continuous time, 367
fixed point prediction prior moments, 368
fixed-interval prediction, 332
fixed-interval smoothing, 333, 334
fixed-lag smoothing, 333
fixed-lead prediction, 332, 356
fixed-point prediction, 356
fixed-point smoothing, 333
Fokker-Planck (F-P) equation, 412
Fokker-Planck equation, 410
forward diffusion equation, 410
fourth joint moment of Gaussian random variables, 87
frequency estimation of a sinusoid, 368
fudge factor, 396, 424
full rank, 25, 28
function in the interval, 181
fundamental equations of linear estimation, 54, 128
fusion equations, 529
fusion of estimates, 264

gain, 134
gain coefficients, 289
gate, 153, 154
gate probability, 166
Gating, 263
Gauss-Markov sequence, 69
Gaussian, 51
Gaussian assumption as the worst case, 128
Gaussian mixture, 55, 445
Gaussian random variable, 34
GDOP, 145, 508, 531
generalized pseudo-Bayesian (GPB), 423, 446
generalized pseudo-Bayesian approach of first order (GPB1), 447
generalized pseudo-Bayesian estimator of second order (GPB2), 449
generation of a random sequence with exponential autocorrelation, 317
generation of correlated random variables, 317
Geometric Dilution of Precision, 145
Geometric Dilution of Precision (GDOP) in Global Positioning System (GPS), 417
Geometric Dilution of Precision in GPS, 531
Geometric DOP (GDOP) matrix, 508
gimbaled system, 495
Global Positioning System, 176, 417
good observability, 167
goodness-of-fit, 155, 156
GPB, 423, 446
GPB1, 447
GPB2, 449

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

GPS, 176, 417, 501
GPS-aided INS, 524
gradient, 24
Gram-Schmidt orthogonalization, 313, 317
guidance, 491
gyroscope, 495

Hessian, 24
hidden Markov model, 445
higher-order polynomial fitting, 247
higher-order terms, 383
homogeneous Markov chain, 445
hybrid, 441
hybrid random variable, 36
hybrid system with i.i.d. modal state sequence, 487
hybrid systems, 48, 57, 423
hypothesis testing, 72
hypothesis testing with correlated noise, 87

i.i.d., 41, 71
i.i.d. modal state sequence in a hybrid system, 487
idempotent, 21, 25
identity matrix, 22
IE, 436
iff, 21
ILS estimator, 138
IMM, 423, 446
IMM estimator, 453
implementation of the linear estimation techniques, 301
imported measurements, 223
impossible event, 31
improper densities, 34
improper pdf, 95
incompletely observed Markov process, 376
incorrect gain, 206
incorrect prior variance, 173
independence of measurements, 198
independence versus conditional independence, 85
independent, 41
independent increment process, 66
independent runs, 237
independent, identically distributed (i.i.d.), 41, 71
inequality of two matrices, 26
inertial navigation system (INS), 495
inertial navigation system, gimbaled, 495
inertial navigation system, strapdown, 495
information, 94, 132
information extraction and enhancement, 2
information filter, 248, 302, 304
information filter state, 307
information in the data, insufficient, 165
information limits, 2
information matrix, 304, 349
information matrix filter, 206
information set, 373
information state, 181, 373, 375
information, existing, 110
information, extracted, 110
information, new, 146
information, old, 145
initial covariance, 202
initial estimate, 202
initial information, 202
initialization, 134, 147, 245, 262
inner product, 20, 124
innovation, 205
innovation covariance, 205, 384
innovation sequence, 213
innovations representation, 214
input estimate, 430
input estimation, 422
input estimation (IE), 436
input gain, 27, 180
input vector, 180
INS-aided GPS, 524
INS/GPS integration by distributed estimation fusion, 528
insignificant, 77
instantaneous variance, 65, 321
insufficient information in the data, 165
integrated navigation, 524
integration of INS and GPS, 524
intensity of the white noise, 342
interacting multiple model (IMM), 423, 446
interacting multiple model (IMM) estimator, 453
interaction, 454
interpolating polynomial, 182
interval estimate, 77
inverse, 21
inversion lemma, 23
inversion, partitioned matrix, 22
iterated extended Kalman filter, 406
Iterated Least Squares estimator, 138
Ito differential equation, 343, 347

Jacobian matrix, 24
jerk, 151, 271
joint cdf, 39
joint density, 41
joint pdf, 38
joint probability-pdf, 36
jointly Gaussian, 51, 52
Joseph form covariance update, 206, 302, 310
jump probabilities, 70
jump process, mode, 444
jump-linear systems, 444

Kalman filter, 199
Kalman filter (KF), 202
Kalman filter gain, 263
Kalman filter with diferent measurements, 265
Kalman filter with diferent sampling intervals, 265
Kalman filter with nonzero noise means, 263
Kalman-Bucy filter, 341, 346
Kalman-Bucy filter with non-zero-mean noises, 369

Meta 1018
p. 25

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

KF, 202
KF as a lowpass filter, 369
kinematic state models, 268
known acceleration objects, 299
known input, 200
Kolmogorov forward equation, 410, 412
Kronecker delta function, 69

Laplacian pdf, 88, 117
large-sample properties of the ML estimator, 113
law of iterated expectations, 50
law of large numbers (LLN), 70
LDL′ factorization, 21, 311, 312
LDU factorization, 21
least mean square (LMS), 115, 129
least squares (LS), 129
least squares (LS) method, 98
least squares estimate, 76
least squares estimator (LSE), 98
left half-plane (LHP), 364
left-continuous, 444
length of a vector, 23
level of significance, 74
LF, 91
LG, 201
likelihood equation, 92, 98
likelihood equations, 102
likelihood function, 48, 73, 97, 114, 374
likelihood function (LF), 91
likelihood function of mode, 442
likelihood function of the system model, 214
likelihood ratio, 73
limiting process of an autocorrelated noise to white noise, 338
linear dependence, 124
linear Gaussian (LG) assumption, 201
linear least squares (LS), 129
linear LS problem, 98
linear MMSE, 124
linear MMSE estimator, 125, 129
linear MMSE estimator for the multidimensional case, 127
linear space, 124
linear time-invariant system, 27, 30
linear transformation of random variables, 87
linear-quadratic control problem (LQ), 358
linearly dependent, 40
linearly independent, 22
lines of sight (LOS), 507
LLN, 70
LMMSE estimation for vectors, 174
LMMSE estimator, 176
LMMSE state estimator, 207
LMS, 115, 129
log-likelihood function, 98, 164
log-normal random variable, 86
LOS, 507
loss of positive definiteness, 301
loss of symmetry, 301
low probability events, 86
lower triangular matrix, 21, 311
lowpass filter, 369
LQ, 358
LS, 98, 114, 129
LS estimator, 130
LSE, 98
lumped estimate, 447
Lyapunov differential equation, 347
Lyapunov equation, 185

MA, 194
magnitude of a vector, 23
Mahalanobis distance, 166
maneuver, 466
maneuver detection, 430
maneuver onset, 433
maneuver onset time, 428
maneuver termination, 433
maneuvering index, 267, 281, 287
maneuvering index threshold, 479, 488
maneuvering model, 432
MAP, 92
MAP estimate, 114
MAP estimate with two-sided exponential (Laplacian) prior, 117
MAP smoothing technique, 409
MAP with Gaussian prior — vector version, 118
margin of error, 108
marginal density, 39
marginal observability, 167
marginal pdf, 39
Markov chain, 70
Markov chain Monte Carlo methods, 414
Markov inequality, 85
Markov process, 184, 190
Markov processes, 66
Markov property, 66
Markov sequence, 69, 190
matched, 237
matrix, 19
matrix inversion lemma, 23
matrix MSE, 128
matrix Riccati differential equation, 346, 349
maximization, multistage, 407
maximum a posteriori (MAP), 92
maximum likelihood estimate, 76
maximum likelihood estimator (MLE), 92
maximum likelihood method, 92
Maxwell distribution, 59
MCMC methods, 414
mean, 37, 62
mean and covariance of the state of a linear system, 196
mean sojourn time, 460
mean sojourn time, 471
mean square (MS), 37

Meta 1018
p. 26

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

mean square error (MSE), 104
mean square error (MSE) matrix, 140
mean-square value matrix, 250
measure, 31
measure of belief, 32
measurement, 122
measurement equation, 27, 180
Measurement error variance, 119
measurement matrix, 27, 180, 188
measurement noise, 180, 188, 200
measurement prediction covariance, 203, 205, 384
measurement prediction error, 205
measurement residual, 205
measurement update, 208
measurement-dependent accuracy, 377
measurements, 3
measurements, imported, 223
measurements, simulated, 223
median, 117
merging, 452
merging probabilities, 451
minimum magnitude error estimate, 117
minimum mean square error (MMSE), 123
minimum mean square error (MMSE) estimator, 99
minimum variance (MV), 115, 129
mismatched, 237
mismatched filter, 240
miss, 72
miss probability, 74
mixed coordinate filter, 390
mixed pmf/pdf model — the variance, 338
mixed pmf/pdf of the acceleration model, 339
mixed random variable, 36
mixing, 454, 455
mixing probabilities, 454, 455
mixture, 55, 86
mixture equations, 56
ML estimate, 114
MLE, 92
MLE from correlated measurements, 118
MM, 423, 441
MM estimator, dynamic, 441
MM estimator, static, 441
MMSE, 114, 123
MMSE estimation with correlated noises, 172
MMSE estimator, 99, 123
MMSE with exponential prior, 118
Mobile sensor platform location error, 531
Mobile sensor platform location error reduction, 532
modal state, 444
modal trajectory, 371, 407
mode, 93, 100, 407, 441
mode (model) switching, 444
mode conditioned covariances, 442
mode conditioned estimate, 449
mode conditioned state estimate, 442
mode conditioned state estimates, 442
mode estimates, 442
mode history, 444
mode jump, 443
mode jump process, 444
mode jumping, 441
mode likelihood function, 442, 447
mode matched filter, 442
mode observation in a hybrid system, 487
mode probabilities, 452
mode probability update, 448, 456
mode probability updating, 452
mode transition probabilities, 445, 446, 452
mode-conditioned state estimate, 447
mode-matched filtering, 448, 451, 456
model, 441
modes, 423
modified process noise covariance, 396
moment, 37
moment matching, 56, 60
moment-generating function, 43
moments of the white noise response of a scalar linear system, 196
Monte Carlo runs, 168, 439
Monte Carlo runs for algorithm comparison, 86
Monte Carlo runs for low probability events, 86
Monte Carlo simulations, 168
Monte Carlo simulations (runs), 79, 234
motion uncertainty, 281
moving average (MA), 194, 425, 433
moving sum, 425
MS, 37
MSE, 104, 233
MSE matrices of predicted and updated state, 250
MSE matrix, 140, 384
MSE, unconditional, 104
MSV matrix, 250
multiframe signal processing, 409
multilateration, 137
multiple model (MM), 423
multiple model (MM) approach, 441
multiscan signal processing, 409
multisensor, 309
multistage maximization, 407
mutually exclusive, 31, 45
MV, 115, 129

navigation, 2, 491
Navigation Satellite Timing and Ranging/Global Positioning System — Navstar/GPS, 501
NCA model, 272
NCV model, 270, 274
nearly constant acceleration (NCA) model, 272
nearly constant velocity model, 270, 274
nearly coordinated turn model, 467
NEES, 165, 234, 238
Neyman-Pearson Lemma, 72
NIS, 236
NLS, 137, 164
NMEE, 235

Meta 1018
p. 27

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

noise, 179
Noise correlated to prior error, 173
noise gain, 180, 189, 272
noise level switching, 422, 424
noise only measurement, 174
noise reduction, 438
noise reduction factor, 282, 458
noiseless kinematic model, 276
nominal trajectory, 385
non-Bayesian, 91
non-Bayesian approach, 91, 114
noncausal, 68, 364
noninformative, 49, 137
noninformative pdf, 95
noninformative prior, 302, 307
nonlinear, 164
nonlinear least squares, 161, 164, 176
Nonlinear Least Squares estimator, 137
nonlinear LS, 130
nonlinear LS problem, 98
nonmaneuvering filter, 427
nonsingular, 22, 25
nonstationary, 191
nonstationary white process noise, 184
nonstationary zero-mean white, 65
norm, 23, 124
norm of a matrix, 27
norm of the residual, 156
normal (Gaussian), 83
normal pdf (density), 51
normal random variable, 34, 51
normalization constant, 91
normalization property, 33, 35
normalized (state) estimation error squared (NEES), 234
normalized error squared (NES), 401
normalized estimation error squared (NEES), 165
normalized innovation squared, 424
normalized innovation squared (NIS), 236
normalized mean estimation error (NMEE), 235
normalized sum of the squares of the residuals, 167
normalizing constant, 48
normally (Gaussian) distributed, 122
nowhere differentiable, 66
null hypothesis, 72
numerical integration, 404

observability Gramian, 349
observability matrix, 28, 30, 349
observability of a linear dynamic system, 266
observability, marginal (poor), 132
observability, parameter, 165
observable, 132
observation, 122
observation uncertainty, 281
one sigma region, 153
one-sided alternative hypothesis, 75
one-sided exponential pdf, 94
one-sided probability region, 82, 83
one-step prediction, 152
operating regime, 441
operator, 375
optimal cost, 359
optimal cost matrix, 359
optimal estimator, 2
optimal gain, 207
optimal MMSE state estimator, 207
optimal nonlinear continuous-discrete filter, 410
optimal nonlinear state estimator, 372
optimistic, 245
Ornstein-Uhlenbeck process, 186, 321
orthogonal, 23, 124
orthogonal projection, 23, 125
orthogonality, 363
orthogonality of estimation error to previous estimates, 263
orthogonality of the innovations to the state prediction, 263
orthogonality properties, 215
orthonormal vectors, 158
outer product, 20
output equation, 27, 180
overfitting, 154

parabolic fit, 151
parameter, 90
parameter estimate significance test, 155
parameter observability, 165
partitioned covariance matrix, 53
partitioned matrix inversion, 22
passive localization, 161
passive localization with direction cosine measurements, 173
passive ranging, 161
pdf, 33
percentile point, 82
perversity of inanimate objects, 88, 96
pessimistic, 245
phase estimation of a sinusoid, 368
physical dimension, 344
piecewise constant acceleration, 273
piecewise constant white acceleration model, 273
piecewise constant Wiener process acceleration model, 274, 289
plant equation, 27, 180
plant noise, 180
pmf, 35
point estimate, 77
point masses, 35
point prediction, 374
Poisson pmf, 36
polar coordinates, 397
polar-to-Cartesian conversion bias, 416
polynomial filters, 268
polynomial fitting, 146, 149
polynomial model, 269

Meta 1018
p. 28

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

polynomial models, 268
position and velocity estimation for known acceleration objects, 299
position confidence region, 166
position estimate for an arbitrary time, 166
position estimation, 175
position estimation improvement, 282
position gain, 281
position RMS errors, 243
position/velocity measurement, 297
positive (semi)definite matrix, 25
positive (semi)definite quadratic form, 25
posterior, 91
posterior pdf, 48
posterior probability, 48
power of the test, 72
power spectral density, 62
power spectral density (spectrum), 193
power spectral density matrix of the output, 186
power spectrum, 62, 186
Precision Positioning Service (PPS), 504
predicted measurement, 203, 205, 384
predicted state, 202, 204, 355, 383
predicted state covariance, 203
predicted value, 134
predicted value of the state, 201
prediction, 152, 330
prediction covariance, 356
prediction with unequal sampling times and variances, 176
predictor-corrector algorithm, 12
prewhitening, 67
prewhitening of a colored noise, 339
prewhitening of a second-order process, 339
prewhitening system, 67, 186, 193, 320, 321
principle of orthogonality, 124, 125
prior, 91
prior information, 136
prior pdf, 48
prior probability, 48
probability, 31
probability concentration ellipsoid, 166
probability concentration region, 235
probability density function (pdf), 33
probability estimation for an event, 107
probability mass, 33
probability mass function (pmf), 35
probability matrix, 88
probability of detection, 74
probability region, 76
probability region, distribution-free, 85
probability region, Gaussian, 83
probability region, two-sided, 156
process noise, 180, 200, 212
process noise and initialization, 263
process noise intensity choice, 270, 272
process noise rescaling, 295
propagation equation of the covariance, 185
propagation equation of the mean, 184, 191
proper density, 33
proper pdf, 95
proper pmf, 35
pseudo-noise, 212, 243, 395, 482
pseudorange equation, 506
public opinion polls, 119

quadratic form, 25
quadratic form with non-zero-mean random variables, 88
quadratic observations, 175
quartic form, 54
quiescent model, 432

radar cross-section estimation, 418
random number generator testing, 262
random process, 61
random sequence, 69
random variable, 33
random walk, 65
range, 397
range: Horizontal and Slant, 417
rank, 25
rational function, 186, 193
rational spectrum, 67, 186
Rayleigh distribution, 59
real-time consistency tests, 237, 240
realization, 33, 91
reciprocal basis, 26
recursive LS estimator, 134
recursive parameter estimate updating equation, 134
reference time, 146, 149, 150
regime, operating, 441
region, one sigma, 153
regression, 146
relative frequency, 32
relinearization of the measurement equation, 405
rescaling of the variance of the process noise, 275
residual, 134
residual covariance, 384
retrodicted state, 201
retrodiction, 333
revisit rate, 267, 292
revisits, 225
Riccati differential equation, 346
Riccati differential equation solution for the inverse covariance, 367
Riccati equation, 211
right half-plane (RHP), 364
RMS error from Monte Carlo runs, 243
RMS speed and velocity errors in Monte Carlo simulations, 417
RMS value of a random variable, 37
root mean square value of a random variable, 37
round-off errors, 301
row vector, 19

SA, 506

Meta 1018
p. 29

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

sample autocorrelation, 236, 238
sample average, 71, 79, 100
sample correlation, 262
sample mean, 79, 100, 135, 439
sample performance differences, 81, 439
sample variance, 103
sampling frequency, 267
sampling period change, 295
sampling rate, 292
scalar updates, 308
scaling factor, 424
scenario, 246
Schwarz inequality, 23, 111
second central moment, 37
second-order kinematic model, 269
Selective Availability, 506
sensitivity of the filter to an incorrect gain, 206
sensitivity to incorrect prior variance, 173
sensitivity to process noise variance, 264
sensor bias and target location estimation, 264
sensor platform location error, 531
sensor platform location error reduction, 532
sequence of models, 444
sequential maximization, 407
sequential Monte Carlo methods, 414
sequential processing, 308
sequential updating, 303
shaping filter, 67, 186, 193, 320, 339
short-term average, 345
sigma, g-ellipsoid, 166
signal interpolation for measurement generation, 414
significance test, parameter estimate, 155
significant estimate, 77
simulated kinematic trajectory behavior, 295
simulated kinematic trajectory variability, 295
simulated measurements, 223
single run, 237
sliding window, 425
slowly varying parameters, 482
SMC methods, 414
smoothed covariance, 336
smoothed state, 336
smoothed state covariance, 337
smoothed value of the state, 201
smoother gain, 336
smoothing, 333
smoothing compared with estimation, 339
smoothing property of the expectations, 50
smoothing residual, 337
soft switching, 458, 466
sojourn time, 460, 471
sojourn time in a state for a Markov chain, 487
spatial window, 154
spectral factorization, 67
spectral factorization for a discrete-time system, 198
spectral factorization for prewhitening (shaping), 197
spectral representation, 26
speed, 183
speed and course estimates, 416
speed and velocity errors in Monte Carlo simulations, 417
speed RMS errors, 243
spread of the means, 443
spread of the means term, 56
square root, 21
square root filtering, 303, 311
square root of a matrix, 311
squared norm of the estimation error, 233
squared norm with respect to a matrix, 26
stable, 193, 351
stacked measurement matrix, 130
stacked vector, 52, 122, 130, 481
standard, 83
standard deviation, 37
standard deviation associated with the estimator, 105
standard deviation improvement versus variance improvement, 415
standard deviation of the estimate, 78
standard deviation of the estimation error, 78, 105
standard error, 78, 79, 105, 106
standard error of the estimate, 106
standard error of the sample variance, 106
standard Gaussian distribution, 35
standard Gaussian random variables, 58
Standard Positioning Service (SPS), 504
state, 27, 30, 181
state augmentation, 320
state equation for fading memory average, 198
state estimate and covariance combination, 449, 452
state estimation error, 347
state estimation errors' autocorrelation, 263
state MSE matrix for continuous-time filter with arbitrary gain, 368
state prediction covariance, 203, 204, 384
state prediction error, 204
state prediction in a linear time-invariant discrete time system, 198
state prediction pdf, 374
state space representation, 187
state transition matrix, 181
state transition pdf, 375
state update, 206
state vector, 180
state-space representation, 180
static estimation problem, 91
static MM estimator, 441
stationary, 62
stationary Markov process, 67
stationary white sequence, 69
stationary zero-mean white, 65
statistical characterization of the disturbances, 233
statistically insignificant, 155, 161
statistics, 51
steady-state covariance, 211, 350
steady-state filter, 350, 352

Meta 1018
p. 30

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

steady-state filter existence for a kinematic model, 296
steady-state filters for noisy kinematic models, 278
steady-state gain, 211
stochastic observability, 166
stochastic process, 61
straight line fit, 150
strapdown system, 495
strict stationarity, 62
strict-sense whiteness, 65
subspace, 124
sufficient statistic, 96
superefficiency, 118
sure event, 31
switching models, 441
switching, mode, 444
symmetric, 19
system design, 267, 292
system identification, 421
system matrix, 27, 180

tail mass, 73
tail probability, 73
target localization, bearings-only, 176
target maneuvering index, 267, 281
target motion analysis, 161
target tracking, 491
target tracking index, 281
test statistic, 75
third-order model, 271
three-point initialization, 264
tightly coupled, 527
time average autocorrelation, 237
time average normalized innovation squared, 237, 240
time averages, 64, 237
time update, 208
time-average autocorrelation, 240
time-invariant, 190
time-invariant system, 27, 30
time-varying discrete-time system, 188
total probability theorem, 45
trace, 20
track before detect, 409
track continuation, 153
tracking, 2
tracking index, 281
trade-off, 292
transfer function matrix, 67
transition matrix, 187
transition matrix of the Markov chain, 70
transition probabilities, 70
transpose of a product, 20
transposition, 19
triangular matrix, 311
triangulation, 137
truncated Gaussian, 45
truncated pdf, 44
tuning of the filter, 482
turn, 183
two-model parameter estimation problem with a prior, 486
two-point differencing, 247
two-sided alternate hypothesis, 76
two-sided probability region, 83, 156, 168
type I error, 72
type II error, 72

UM, 466
unbiased, 101
unbiased estimator, 115
unbiasedness, 233, 363
unbiasedness property, 126
unconditional MSE, 104
uncorrelated, 40
uncorrelatedness, 69
underfitting, 154
uniform motion (UM), 466
uniform prior pdf, 49
uniformly distributed random variable, 34
unimodal, 101
unit lower triangular matrix, 21, 311
unit step function, 35
unknown input, 427
unknown input as a random process, 422
unnormalized errors, 243
unobservable, 162, 167
upcrossing threshold, 426
update gain, 133
update with an arbitrary gain, 262
updated covariance, 206
updated state, 203
updated state covariance, 203
updated state covariance conditioned on prior data, 338
updated state estimate, 203, 205
upper tail, 82
upper tail probability, 82
upper triangular matrix, 311

validity limit of the standard conversion, 400
variability reduction, 235
variable state dimension, 422
variable state dimension (VSD), 431, 436
variance, 37
variance estimate with the smallest MSE, 117
variance of the estimator, 104
variance of the sample mean, 106
variance of the sample variance, 106
variance, conditional, 104
variance/MSE of an estimator, 115
vector, 19
vector Taylor series expansion, 383
vector-valued Gaussian, 51
vector-valued random process, 62
vector-valued random variable, 41
vectors in an abstract vector space, 124

Meta 1018
p. 31

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

velocity estimation, 174
velocity estimation improvement, 282
velocity gain coefficient, 281
velocity increment over a sampling period, 436
velocity RMS errors, 243
Viterbi algorithm, 409
VSD, 431, 436

weak independence, 69
weighted sum of chi-square variables, 60
weighted sum of pdfs, 55
weighting matrix, 127
WGS-84, 515
white noise, 64
white noise acceleration, 268
white noise acceleration model, 269
white noise jerk (WNJ) model, 271
white sequence, 69
whiteness test, 236
Wide Area Augmentation System (WAAS), 509
wide sense stationary, 62
wide-sense whiteness, 65
Wiener filter, 364, 365
Wiener filter as the steady-state Kalman-Bucy filter, 369
Wiener process acceleration, 268
Wiener process acceleration (WPA) model, 271
Wiener process increments, 88
Wiener random process, 65
Wiener-Hopf equation, 363
Wiener-Hopf problem, 362
window length for fading memory average, 197
window, spatial, 154
WNA model, 269
World Geodetic System of 1984 (WGS-84), 515
WPA model, 271

zero-mean random variables, 124

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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)

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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."

Meta 1018
p. 33

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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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."

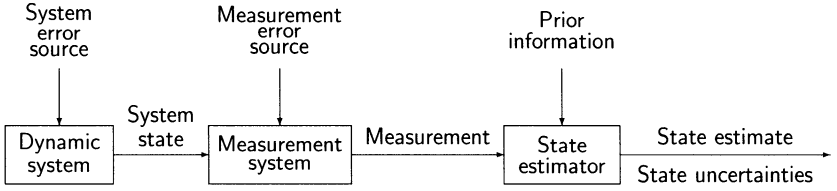


Figure 1.1.1-1: State estimation — information extraction and enhancement.

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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

  - 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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

- 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).

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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.

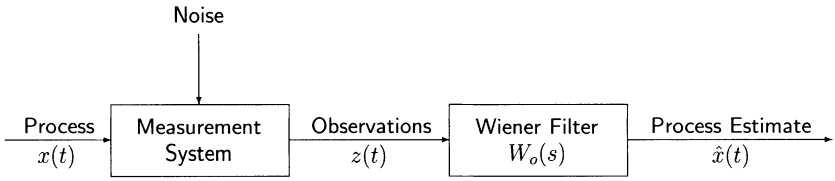


Figure 1.1.3-1: The Wiener filter — linear time-invariant system with transfer function $W_o(s)$.

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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

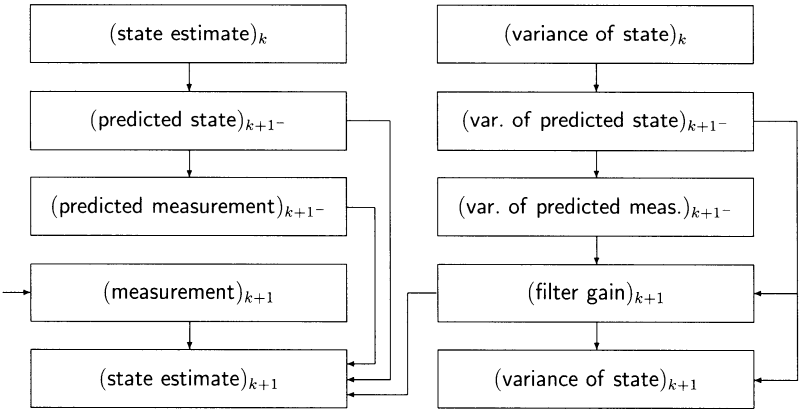


Figure 1.1.3-2: Sequence of operations in the KF.

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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

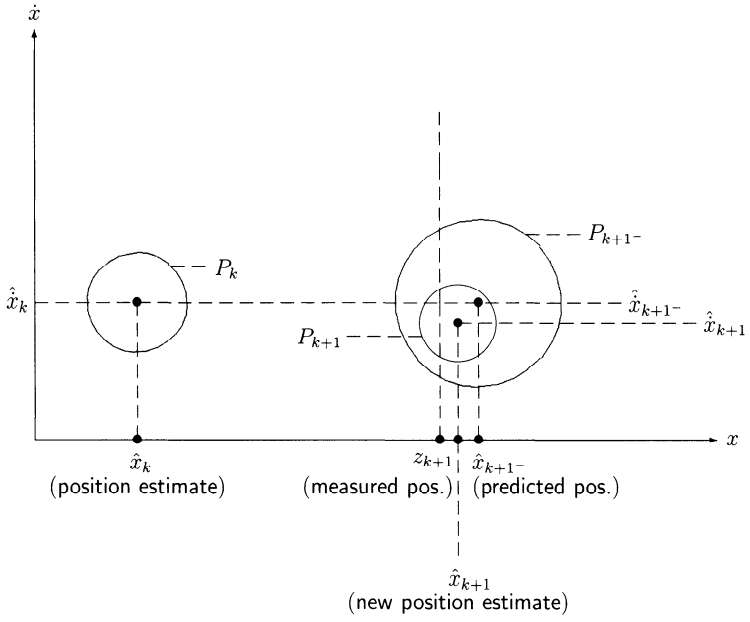


Figure 1.1.3-3: Illustration of the state estimation process.

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

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} r \\ \dot{r} \end{bmatrix} \qquad (1.1.4\text{-}1)$$

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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

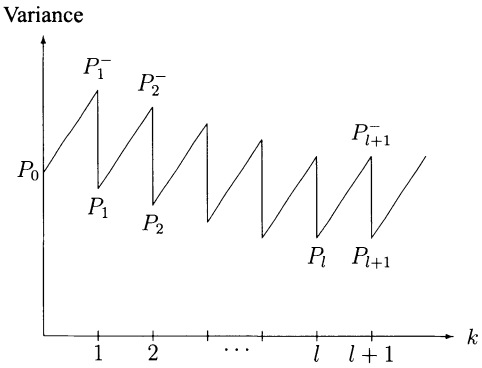


Figure 1.1.3-4: Evolution of the variance of the estimate with time.

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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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 $\bar{y}(k)$ 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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

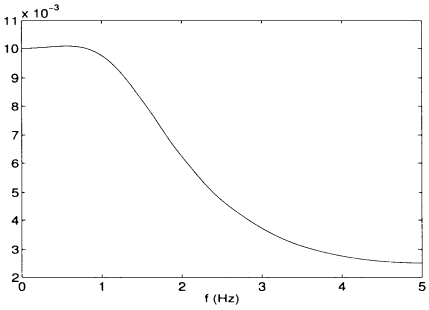


Figure 1.1.4-1: The frequency response of the state estimation filter.

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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

- Comparison of several methods
- The multiple model approach — static and dynamic, the Interacting Multiple Model (IMM) estimator
- Use of the companion software DynaEst^{TM} 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.

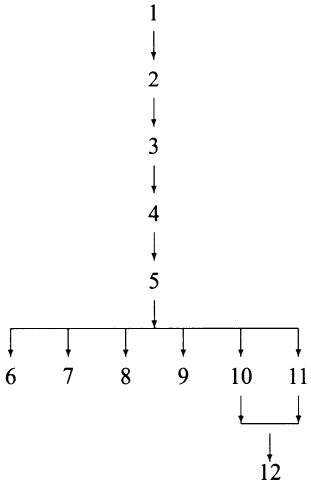


Figure 1.2.2-1: Sequence of chapter prerequisites.

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)

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

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

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

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

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

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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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***."

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

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

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

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

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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

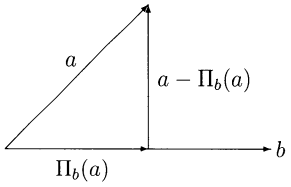


Figure 1.3.4-1: Orthogonal projection of vector a on vector b.

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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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:

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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

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).

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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 $C(t) = C$.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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),

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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).

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

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

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

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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(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,

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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

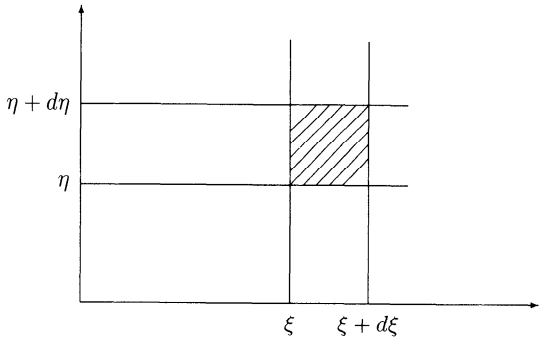


Figure 1.4.6-1: Event for definition of a joint pdf.

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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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.)

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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,

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

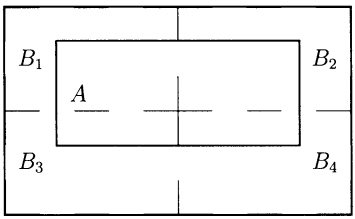


Figure 1.4.10-1: The total probability theorem (set theory counterpart).

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).

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

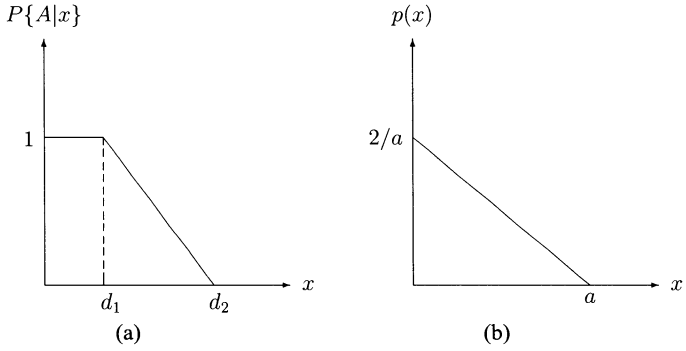


Figure 1.4.10-2: Success probability and miss distance pdf.

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:

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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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)

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

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

$$\sum_{i=1}^{n} P\{B_i|A\} = 1 \qquad (1.4.11\text{-}3)$$

Bayes' formula for random variables is written as

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

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

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

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,

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

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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

$$E[E(x|z)] = E[\bar{x}] + P_{xz}P_{zz}^{-1}\,E(z - \bar{z}) = \bar{x} = E[x] \tag{1.4.12-6}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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,

$$p\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \mathcal{N}\left\{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}; \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix}, \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}\right\} \tag{1.4.13-6}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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,

$$p(x, z) = p(y) = \mathcal{N}(y; \bar{y}, P_{yy}) \tag{1.4.14-2}$$

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

$$P_{xx} = \text{cov}(x) = E[(x - \bar{x})(x - \bar{x})'] \tag{1.4.14-5}$$

$$P_{zz} = \text{cov}(z) = E[(z - \bar{z})(z - \bar{z})'] \tag{1.4.14-6}$$

$$P_{xz} = \text{cov}(x, z) = E[(x - \bar{x})(z - \bar{z})'] = P_{zx}' \tag{1.4.14-7}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

$$\xi + T_{xx}^{-1} T_{xz} \zeta = x - \bar{x} - P_{xz} P_{zz}^{-1} (z - \bar{z}) \tag{1.4.14-16}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

$$p(x) = \sum_{j=1}^{n} p(x|A_j)P\{A_j\} \tag{1.4.16-5}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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)

$$E[(x-\bar{x})(x-\bar{x})'] = \sum_{j=1}^{n} p_j P_j + \sum_{j=1}^{n} p_j\bar{x}_j\bar{x}_j' - \bar{x}\bar{x}' \tag{1.4.16-11}$$

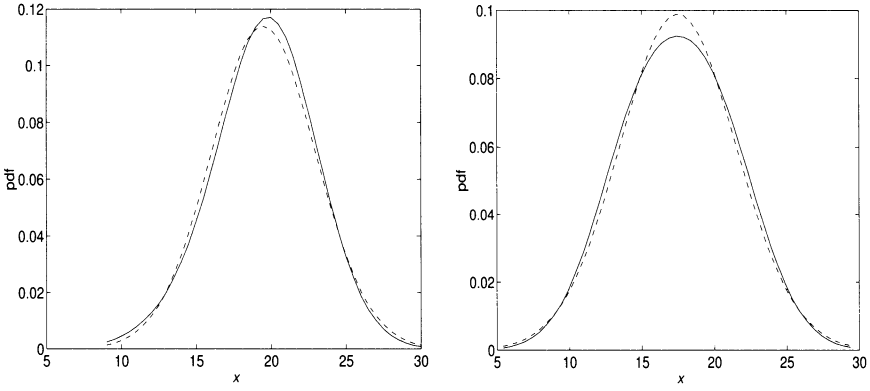
The spread of the means term $\tilde{P}$, defined in (1.4.16-9), is a sum of dyads with positive weightings, and thus it follows from Subsection 1.3.6 that $\tilde{P}$ is *positive semidefinite*.

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 moments equal to those of the mixture, given by (1.4.16-6) and (1.4.16-8) — this is called ***moment matching***.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



(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.)

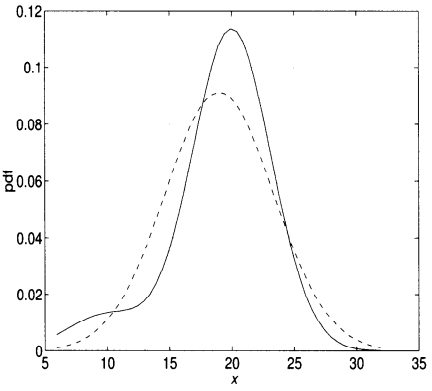
1.4.17 Chi-Square Distributed Random Variables

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

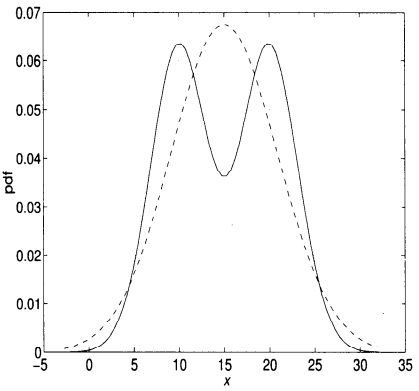
$$q = (x - \bar{x})' P^{-1} (x - \bar{x}) \qquad (1.4.17\text{-}1)$$

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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



(a) $p_1 = 0.1, \quad p_2 = 0.9$



(b) $p_1 = 0.5, \quad p_2 = 0.5$

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

$$q \sim \chi_n^2 \tag{1.4.17-7}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

$$n' = \frac{m \left(\sum_{i=1}^{n} a_i\right)^2}{\sum_{i=1}^{n} a_i^2} \tag{1.4.18-10}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Therefore, the distribution of y_n, defined in (1.4.18-2) is approximately

$$y_n \sim \frac{\sum_{i=1}^{n} a_i^2}{\sum_{i=1}^{n} a_i} \chi^2_{n'} \tag{1.4.18-11}$$

Note that the χ^2_n 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

$$z_n \sim \frac{1}{1+\alpha} \chi^2_{n'} \tag{1.4.18-14}$$

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

$$x(t) = x(t, \omega) \tag{1.4.19-2}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

$$R(t_1, t_2) \triangleq E[x(t_1)x(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \xi\eta \, p_{x(t_1),x(t_2)}(\xi, \eta) \, d\xi d\eta \tag{1.4.19-4}$$

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).

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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,

$$E[x(t_1)x(t_2)] = e^{-a|t_1-t_2|}E[x(t_1)^2] \tag{1.4.19-19}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

$$V(t_1,t_2) = \sigma^2(t_1)\delta(t_1-t_2) = S_0(t_1)\delta(t_1-t_2) \quad (1.4.19\text{-}24)$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

$$E[n(t_1)n(t_2)] = \alpha\delta(t_1 - t_2) \tag{1.4.20-4}$$

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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

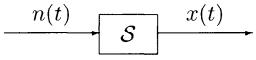


Figure 1.4.21-1: A system driven by autocorrelated noise.

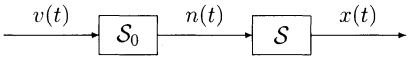


Figure 1.4.21-2: The same system augmented to be driven by white noise.

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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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***.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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***

$$P\{x(k) = x_j \,|\, x(k-1) = x_i\} \triangleq \pi_{ij} \qquad i, j = 1, \ldots, n \tag{1.4.22-9}$$

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

$$E[(x_i - \bar{x}_i)(x_j - \bar{x}_j)] = \sigma^2\rho(i-j) \tag{1.4.23-2}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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).

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

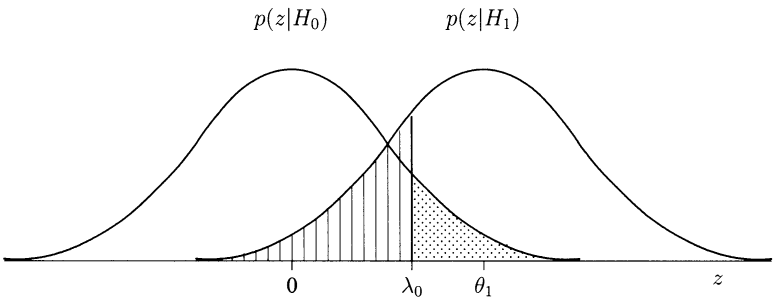


Figure 1.5.1-1: Test between two simple hypotheses.

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

$$P_{e_{II}} \triangleq P\{z < \lambda_0 | H_1\} = \int_{-\infty}^{\lambda_0} \mathcal{N}(z; \theta_1, \sigma^2) dz \qquad (1.5.1\text{-}13)$$

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

$$P_D \triangleq \pi = P\{z > \lambda_0 | H_1\} = \int_{\lambda_0}^{\infty} \mathcal{N}(z; \theta_1, \sigma^2) dz = 1 - P_{e_{II}} \qquad (1.5.1\text{-}14)$$

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

$$H_0: \quad \theta = 0 \qquad (1.5.2\text{-}1)$$

[12] Equivalently, one can say that the ***confidence level*** for H_1 is high ($1 - \alpha$ or more).

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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:

$$\bar{z} \triangleq \frac{1}{n}\sum_{i=1}^{n} z_i \underset{\text{"}H_0\text{"}}{\overset{\text{"}H_1'\text{"}}{\gtrless}} \lambda_1 \tag{1.5.2-8}$$

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 $\bar{z}$ 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

$$p(\bar{z}|H_0) = \mathcal{N}(\bar{z}; 0, \frac{\sigma^2}{n}) \tag{1.5.2-9}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

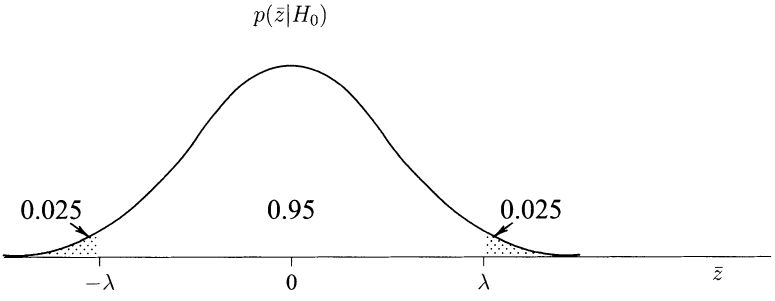


Figure 1.5.2-1: The test between hypotheses H_0 and H_1.

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 $\bar{z}$ can be seen as an ***interval estimate*** of the parameter θ.

Figure 1.5.2-2 illustrates these concepts.

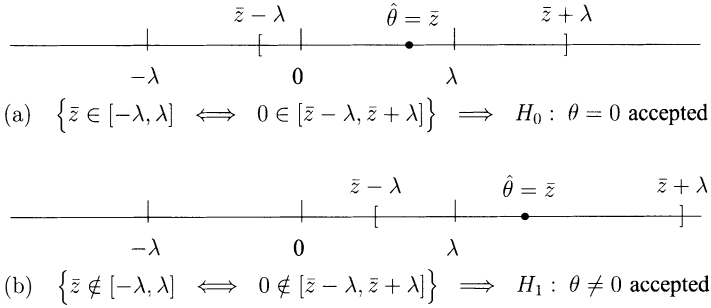


Figure 1.5.2-2: Parameter estimate. (a) $\hat{\theta}$ insignificant; (b) $\hat{\theta}$ significant.

Thus the estimate of a parameter that might be zero is accepted as *significant* (at a "level" of 5% for the null hypothesis) if

$$\frac{|\hat{\theta}|}{\sigma_{\hat{\theta}}} > 1.96 \qquad (1.5.2\text{-}17)$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

These can be used to obtain a confidence region for the performance assuming that its distribution is, in view of the central limit theorem, approximately normal. Since the pdf of (1.5.3-2) is in general skewed, confidence regions based on the normal assumption might not be accurate unless N is very large — possibly of the order of thousands.

Algorithm Comparison as a Hypothesis Testing Problem

When two algorithms are compared based on simulations, this can be formulated as a hypothesis testing problem as follows.

Assume one has the sample mean of the performance (e.g., MSE, or some other cost to be minimized) for algorithm j from N independent runs

$$\bar{C}^{(j)} = \frac{1}{N}\sum_{i=1}^{N} C_i^{(j)} \qquad j = 1, 2 \tag{1.5.3-4}$$

where $C_i^{(j)}$ is the performance of algorithm j in run i and $C_i^{(j)}$ is independent of $C_k^{(j)}$, $\forall k \neq i$. Each pair of runs i uses the *same random variables for the two algorithms* $j = 1, 2$. As shown in problem 1-3, it is *beneficial* to use the same random variables for both algorithms in the same run.

Some Simple Comparison Techniques

Assume that from the simulations one has

$$\bar{C}^{(1)} < \bar{C}^{(2)} \tag{1.5.3-5}$$

This does not necessarily imply that algorithm 1 is better than algorithm 2. Any such statement must be qualified by a probability α of error of type I.

Comparison technique #0, based on (1.5.3-5) without any statistical analysis of the individual run outcomes, is clearly *naive*.

Comparison technique #1 — the simplest (but still simplistic) statistical approach — consists of the following:

1. Calculate confidence regions around the two sample means (based on (1.5.3-2) and (1.5.3-3)).
2. If they do not overlap, declare that the algorithm with the smaller sample mean is the superior one.

The last step above implicitly assumes the two sample means to be uncorrelated. Note that this is *incorrect*, since each pair of runs uses the same random numbers.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The Optimal Comparison Technique

Since $C_i^{(1)}$ is correlated with $C_i^{(2)}$ (because they use the same random variables), the optimal test, which will be based on *independent samples*, is as follows.

The hypothesis testing problem is

$$H_0: \Delta = J^{(2)} - J^{(1)} \leq 0 \quad \text{(algorithm 1 not better than 2)} \tag{1.5.3-6}$$

versus

$$H_1: \Delta = J^{(2)} - J^{(1)} > 0 \quad \text{(algorithm 1 better than 2)} \tag{1.5.3-7}$$

subject to

$$P\{\text{accept } H_1 | H_0 \text{ true}\} = \alpha \quad \text{(level of significance (of hypothesis } H_0\text{))} \tag{1.5.3-8}$$

where

$$J^{(j)} = E[C^{(j)}] \qquad j = 1, 2 \tag{1.5.3-9}$$

are the *true expected values of the cost functions (true performance).*

The decision whether H_1 ("1" better than "2") should be accepted is made based upon the ***sample performance differences***

$$\boxed{\Delta_i = C_i^{(2)} - C_i^{(1)}} \tag{1.5.3-10}$$

Note that Δ_i is independent of Δ_k, $\forall k \neq i$.

Then, H_1 is accepted if

$$\boxed{\mu \triangleq \frac{\bar{\Delta}}{\sigma_{\bar{\Delta}}} > \mu_0} \tag{1.5.3-11}$$

where

$$\boxed{\bar{\Delta} = \frac{1}{N} \sum_{i=1}^{N} \Delta_i} \tag{1.5.3-12}$$

and

$$\boxed{\sigma_{\bar{\Delta}} = \sqrt{\frac{1}{N^2} \sum_{i=1}^{N} (\Delta_i - \bar{\Delta})^2}} \tag{1.5.3-13}$$

are the *sample mean* of the differences (1.5.3-10) and the *standard error* of this sample mean, respectively.

Assuming the error in $\bar{\Delta}$ to be normal, the threshold μ_0 is based on the *upper tail* of the normal density: $\mu_0 = 1.64$ for $\alpha = 5\%$, $\mu_0 = 2.33$ for $\alpha = 1\%$, and so on (see Subsection 1.5.4). This follows from the fact that the test is for *positive mean (H_1) versus zero or negative mean* and H_1 is accepted if and only if the sample mean (1.5.3-12) is positive and statistically significant (1.5.3-11); α is the probability of accepting H_1 (positive mean) when the true mean is zero.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

This procedure is summarized below:

$C_1^{(1)}$	$C_1^{(2)}$	Δ_1
$\vdots$	$\vdots$	$\vdots$
$C_N^{(1)}$	$C_N^{(2)}$	Δ_N
$\bar{C}^{(1)}$	$\bar{C}^{(2)}$	$\bar{\Delta}$, $\sigma_{\bar{\Delta}}$

Remarks

The optimal algorithm comparison is based on the significance test of the *sample mean of the differences* (the mean of the terms in the last column above) instead of the naive comparison of the performance estimates from the first two columns (technique #0) or using two confidence regions around these estimates (technique #1).

The applicability of the CLT is much more realistic on the average difference (1.5.3-12) than on the average outcomes (1.5.3-4) since the differences are much less skewed (and independent from run to run).

The fact that the same random numbers are used for the two algorithms leads to a *positive correlation* between the two sample means, and the optimal test presented takes advantage of this. (See also problem 1-3.)

Section 11.5 presents a major example of the application of this optimal comparison technique.

1.5.4 Tables of the Chi-Square and Gaussian Distributions

Table 1.5.4-1 presents the points x on the chi-square distribution for a given ***upper tail probability***

$$Q = P\{y > x\} \tag{1.5.4-1}$$

where

$$y \sim \chi_n^2 \tag{1.5.4-2}$$

and n is the number of degrees of freedom. This tabulated function is also known as the ***complementary distribution***.

An alternative way of writing (1.5.4-1) is

$$x(1-Q) \triangleq \chi_n^2(1-Q) \tag{1.5.4-3}$$

which indicates that to the left of the point x the probability mass is $1-Q$. This is the $100(1-Q)$ ***percentile point***.

The 95% probability region for a χ_2^2 variable can be taken as the ***one-sided probability region*** (cutting off the 5% ***upper tail***)

$$[0, \chi_2^2(0.95)] = [0, 5.99] \tag{1.5.4-4}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

or the ***two-sided probability region*** (cutting off both 2.5% tails)

$$[\chi^2_2(0.025), \chi^2_2(0.975)] = [0.05, 7.38] \tag{1.5.4-5}$$

For a χ^2_{100} variable, the two-sided 95% probability region is

$$[\chi^2_{100}(0.025), \chi^2_{100}(0.975)] = [74, 130] \tag{1.5.4-6}$$

Note the skewedness of the chi-square distribution: The above two-sided regions are not symmetric about the corresponding means (2 and 100, respectively, in the above two equations).

For degrees of freedom above 100, the following approximation of the points on the chi-square distribution can be used:

$$\chi^2_n(1-Q) = \frac{1}{2}\left[\mathcal{G}(1-Q) + \sqrt{2n-1}\right]^2 \tag{1.5.4-7}$$

where $\mathcal{G}$ is defined in (1.5.4-11). For example, $\chi^2_{400}(0.025) = 346, \chi^2_{400}(0.975) = 457$.

The last line of Table 1.5.4-1 shows the points x on the ***standard*** (zero mean and unity variance) ***normal (Gaussian)*** distribution for the same tail probabilities. In this case

$$y \sim \mathcal{N}(0,1) \tag{1.5.4-8}$$

and, with

$$Q = P\{y > x\} \tag{1.5.4-9}$$

these points will be denoted as

$$x(1-Q) \triangleq \mathcal{G}(1-Q) \tag{1.5.4-10}$$

Thus, with this notation the 95% ***two-sided probability region*** for an $\mathcal{N}(0,1)$ random variable is

$$[\mathcal{G}(0.025), \mathcal{G}(0.975)] = [-1.96, 1.96] \tag{1.5.4-11}$$

In terms of the cumulative distribution function (cdf) of a standard Gaussian random variable

$$\begin{aligned} P_{\mathcal{G}}(x) &\triangleq P\{y \le x\} \\ &= \int_{-\infty}^{x} \mathcal{N}(y;0,1)\,dy \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}}\,dy \end{aligned} \tag{1.5.4-12}$$

one can write (1.5.4-11) as

$$P_{\mathcal{G}}(1.96) = 0.975 \tag{1.5.4-13}$$

$$P_{\mathcal{G}}(-1.96) = 0.025 \tag{1.5.4-14}$$

Note that $P_{\mathcal{G}}$ defined in (1.5.4-12) is the inverse function of $\mathcal{G}$ defined in (1.5.4-10), that is, $\mathcal{G}$ is the inverse cdf.

The 95% ***one-sided probability region*** for a standard Gaussian random variable is given by its 95% point:

$$(-\infty, \mathcal{G}(0.95)] = (-\infty, 1.64] \tag{1.5.4-15}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 1.5.4-1: Tail probabilities of the chi-square and normal densities.

$n \backslash Q$	0.99	0.975	0.95	0.90	0.75	0.5	0.25	0.10	0.05	0.025	0.01	5E-3	1E-3
1	2E-4	.001	.003	.016	.102	.455	1.32	2.71	3.84	5.02	6.63	7.88	10.8
2	.020	.051	.103	.211	.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6	13.8
3	.115	.216	.352	.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8	16.3
4	.297	.484	.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9	18.5
5	.554	.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7	20.5
6	.872	1.24	1.64	2.20	3.35	5.35	7.84	10.6	12.6	14.4	16.8	18.5	22.5
7	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.1	18.5	20.3	24.3
8	1.65	2.18	2.73	3.49	5.07	7.34	10.2	13.4	15.5	17.5	20.1	22.0	26.1
9	2.09	2.70	3.33	4.17	5.90	8.34	11.4	14.7	17.0	19.0	21.7	23.6	27.9
10	2.56	3.25	3.94	4.87	6.74	9.34	12.5	16.0	18.3	20.5	23.2	25.2	29.6
11	3.05	3.82	4.57	5.58	7.58	10.3	13.7	17.3	19.7	22.0	24.7	26.8	31.3
12	3.57	4.40	5.23	6.30	8.44	11.3	14.8	18.5	21.0	23.3	26.2	28.3	32.9
13	4.11	5.01	5.90	7.04	9.30	12.3	16.0	19.8	22.4	24.7	27.7	29.8	34.5
14	4.66	5.63	6.57	7.79	10.2	13.3	17.1	21.1	23.7	26.1	29.1	31.3	36.1
15	5.23	6.26	7.26	8.55	11.0	14.3	18.2	22.3	25.0	27.5	30.6	32.8	37.7
16	5.81	6.91	7.96	9.31	11.9	15.3	19.4	23.5	26.3	28.8	32.0	34.3	39.3
17	6.41	7.56	8.67	10.1	12.8	16.3	20.5	24.8	27.6	30.2	33.4	35.7	40.8
18	7.01	8.23	9.40	10.9	13.7	17.3	21.6	26.0	28.9	31.5	34.8	37.2	42.3
19	7.63	8.91	10.1	11.7	14.6	18.3	22.7	27.2	30.1	32.9	36.2	38.6	43.8
20	8.26	9.60	10.9	12.4	15.5	19.3	23.8	28.4	31.4	34.2	37.6	40.0	45.3
25	11.5	13.1	14.6	16.5	19.9	24.3	29.3	34.4	37.7	40.6	44.3	46.9	52.6
30	15.0	16.8	18.5	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	53.7	59.7
40	22.2	24.4	26.5	29.1	33.7	39.3	45.6	51.8	55.8	59.3	63.7	66.8	73.4
50	29.7	32.4	34.8	37.7	43.0	49.3	56.3	63.2	67.5	71.4	76.2	79.5	86.7
60	37.5	40.5	43.2	46.5	52.3	59.3	67.0	74.4	79.1	83.3	88.4	92.0	99.6
70	45.4	48.8	51.7	55.3	61.7	69.3	77.6	85.5	90.5	95.0	100	104	112
80	53.5	57.2	60.4	64.2	71.1	79.3	88.1	96.6	102	107	112	116	125
90	61.8	65.6	69.1	73.3	80.6	89.3	98.6	108	113	118	124	128	137
100	70.1	74.2	77.9	82.4	90.1	99.3	109	118	124	130	136	140	149
150	112	118	123	128	138	149	161	173	180	186	193	198	209
200	157	163	168	175	186	199	213	226	234	241	249	255	268
300	246	254	261	269	283	299	316	332	341	350	360	367	381
400	337	346	355	364	381	399	419	437	448	457	469	478	495
500	429	440	449	460	478	499	521	541	553	563	576	585	603
800	710	724	735	749	773	799	827	852	867	880	896	907	929
$\mathcal{G}()$	-2.33	-1.96	-1.64	-1.28	-.675	0	.675	1.28	1.64	1.96	2.33	2.58	3.09

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Some Distribution-free Probability Regions

The ***Chebyshev inequality*** provides a ***distribution-free probability region.*** For a random variable x with mean μ and variance σ^2, one has, for any $\epsilon > 0$,

$$P\{|x - \mu| \geq \epsilon\} \leq \frac{\sigma^2}{\epsilon^2} \tag{1.5.4-16}$$

regardless of the distribution of x.

For a non-negative random variable x (i.e., $p(x) = 0 \ \forall x < 0$) with mean μ, the ***Markov inequality*** states that, for any $\alpha > 0$,

$$P\{x \geq \alpha\} \leq \frac{\mu}{\alpha} \tag{1.5.4-17}$$

1.6 NOTES AND PROBLEMS

1.6.1 Bibliographical Notes

An excellent historical survey of the developments of estimation theory, starting from the least squares technique of Gauss and Legendre, the maximum likelihood technique of R. A. Fisher, the Wiener filtering for signals in noise, and the emergence of digital computer oriented recursive algorithms developed by Swerling, Kalman, Bucy, and others, can be found in [Sorenson85].

Other texts on the same topic as this one are [Meditch69, Sage71, Gelb74, Anderson79, Lewis86, Maybeck79, Maybeck82, Grewal93, Brown97].

For background material, further references on linear algebra as well as linear systems are [Zadeh63, Fortmann77, Chen84]. The existing literature on this topic consists of tens of books, and most of them are equally good. A succinct summary of this material can be found in [Gelb74]. The probabilistic tools are covered in [Papoulis84] and succinctly in [Gelb74]. Spectral factorization is discussed in [Poor88]. The statistical tools (hypothesis testing) can be found, for example, in [Melsa78, Johnston72], among others.

For the more specialized target tracking topics — in particular, tracking with uncertain origin observations and multitarget tracking — the reader is referred to [Blackman86, Bar-Shalom88]. The texts [Bar-Shalom95] and [Blackman99] are the most recent compilations of algorithms in this area, with suitable explanations and design guidelines.

1.6.2 Problems

1-1 Independence versus conditional independence. Consider x and y independent Gaussian random variables with zero mean and unity variance.

1. For an arbitrary new random variable does the following hold

$$p(x, y|z) = p(x|z)p(y|z) \quad ?$$

2. Let $z = x + y$. Write the explicit expression of the pdf $p(x, y|z)$. Use the Dirac delta function if necessary.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

1-2 Monte Carlo runs for low probability events. Given an experiment with binary outcome x, assume your colleague did a theoretical calculation that indicates that

$$P\{x = 1\} = \hat{p} = 10^{-4}$$

However, you do not believe this colleague and plan to conduct N Monte-Carlo (i.i.d.) repetitions of the experiment. The outcomes are x_i, $i = 1, \ldots, N$. Estimate, using a (say, 95%) confidence region, the number N of experiments you have to carry out to confirm or refute your colleague's theoretical calculation with, e.g., 10% tolerance.

1-3 Monte Carlo runs for algorithm comparison. Two estimation algorithms are to be compared based on Monte Carlo runs. Assume their average performances (MSE) were obtained as $\bar{C}^{(1)} = 5$ and $\bar{C}^{(2)} = 8$ from $N = 100$ runs.

1. Can one say that algorithm 1 is superior to algorithm 2?

2. Assume the standard deviations associated with these sample means are

$$\sigma_{\bar{C}^{(j)}} = \left[\frac{1}{N^2}\sum_{i=1}^{N}\left[C_i^{(j)} - \bar{C}^{(j)}\right]^2\right]^{1/2} = 1 \qquad j = 1, 2$$

Using the above information answer question 1.

3. Denote the correlation coefficient between the outcomes of the runs with the two algorithms as ρ, that is,

$$E\left[(C_i^{(1)} - J^{(1)})(C_i^{(2)} - J^{(2)})\right] = \rho\sigma_{C^{(1)}}\sigma_{C^{(2)}}$$

Show that if $\rho = 0.5$ and the variances are as in item 2, the optimal statistical test will yield a conclusive answer.

1-4 Covariance of a mixture. Prove that an equivalent expression of the "spread of the means" term (1.4.16-9) in the covariance of a mixture is

$$\tilde{P} = \sum_i \sum_{j<i} p_i p_j (\bar{x}_i - \bar{x}_j)(\bar{x}_i - \bar{x}_j)'$$

where the summation is over all the pairs of indices without repetition.

1-5 Dimension of a pdf.

1. Find the physical dimension of the pdf of the random vector x when x is
   a. The position of a point in an n-dimensional Euclidean space.
   b. The state of a constant (unknown) acceleration point moving along, say, the ξ-axis.
2. Find the units of the pdf of a velocity measured in furlongs/fortnight.
3. Can one compare the pdf of a random variable whose dimension is length to the joint pdf of two such random variables?

1-6 Log-normal random variable. Given the random variable $y \sim \mathcal{N}(\mu, \sigma^2)$, then $x \triangleq e^y$ is called a log-normal random variable. Find its mean and variance.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

1-7 Moments of linear transformation of random variables. Given the random variables x and y of dimensions n_x and n_y, with means $\bar{x}$ and $\bar{y}$, respectively, and with covariances P_{xx}, P_{yy}, P_{xy}:

1. Find the mean and covariance of the n_z-dimensional vector

$$z = Ax + By + c$$

where A, and B are matrices of appropriate dimensions.

2. Indicate the dimensions of A, B, and c.

1-8 Chapman-Kolmogorov equation. Prove that the following equation holds for a (discrete time) Markov process

$$\int p[x(k)|x(k-1)]p[x(k-1)|x(k-2)]\,dx(k-1) = p[x(k)|x(k-2)]$$

1-9 Hypothesis testing with correlated noise. Consider the hypotheses

$$H_0 : \theta = 0 \qquad H_1 : \theta \neq 0$$

and the observations

$$z_i = \theta + w_i \qquad i = 1, \ldots, n$$

with w_i zero-mean jointly Gaussian but not independent. Denoting

$$w = [w_1 \ \cdots \ w_n]'$$

one has the covariance matrix (assumed given)

$$E[ww'] = P$$

For the above

1. Specify the optimal hypothesis test for false alarm probability α.
2. Solve explicitly for $n = 2$, $P = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$ and $\alpha = 1\%$.

1-10 Partial derivative with respect to a matrix. The partial derivative of a scalar q with respect to the matrix $A = [a_{ij}]$ is defined as

$$\frac{\partial q}{\partial A} \triangleq \left[\frac{\partial q}{\partial a_{ij}}\right]$$

Prove that

1. For B symmetric, $\frac{\partial}{\partial A}\text{tr}[ABA'] = 2AB$.
2. For B not symmetric, $\frac{\partial}{\partial A}\text{tr}[AB] = B'$.

1-11 Fourth joint moment of Gaussian random variables. Show that, if the zero-mean scalar random variables x_i, $i = 1, \ldots, 4$, are jointly Gaussian, then

$$E[x_1x_2x_3x_4] = E[x_1x_2]E[x_3x_4] + E[x_1x_3]E[x_2x_4] + E[x_1x_4]E[x_2x_3]$$

Hint: Use the characteristic function (1.4.15-3) and take a suitable fourth derivative of it.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

1-12 Wiener process increments.

1. Using (1.4.20-3) show that

$$E\left[[d\mathbf{w}(t)]^2\right] = \alpha\, dt \qquad \text{and} \qquad d\mathbf{w}(t) \sim \mathcal{N}(0, \alpha\, dt)$$

2. Using the above, show that the limit of $\frac{d\mathbf{w}(t)}{dt}$ tends to ∞. *Hint*: Find the "order of magnitude" of $d\mathbf{w}(t)$ in terms of dt (a zero-mean random variable is of the order of its standard deviation).

1-13 Conditional pdf of the sum of two Gaussian random variables. Find the pdf (1.4.13-12) in the general case (1.4.13-6).

1-14 Probability matrix. Find $\sum_{j=1}^{n} \pi_{ij}$ for (1.4.22-9).

1-15 Laplacian pdf. Given the random variable x with Laplacian (two-sided exponential) pdf

$$p(x) = c\, e^{-a|x|} \qquad a > 0 \qquad -\infty < x < \infty$$

1. Find the normalization constant c, the mean and the variance of x.

2. Repeat the above if it is known that $|x| \leq b$.

1-16 Moments of a quadratic form with non-zero-mean random variables. Consider the random variables x and y with means $\bar{x}$ and $\bar{y}$, respectively, and with covariances P_{xx}, P_{yy}, P_{xy}. Evaluate $E[x'Ay]$.

1-17 The law of perversity of inanimate objects. A buttered toast falls off a table. According to the "law of perversity of inanimate objects," the probability of it falling on the floor with the buttered side down is greater than that with the buttered side up.

The motion (rotation and fall) of the toast can be modeled as follows.

Starting with the buttered side facing up, it acquires a rotational motion with a random angular rate $\dot{\theta} \sim U[\dot{\theta}_1, \dot{\theta}_2]$, assumed to stay constant until it hits the floor. The side which will hit the floor depends on the angle it rotated.

1. Indicate the interval(s) of the rotation angle θ for which the buttered side will hit the floor.

2. Assuming a constant acceleration fall from the typical height of a table, h, with acceleration $g = 10$ m/s^2, find the time the toast has to rotate.

3. Find the pdf of the angle θ it rotated until it hit the floor.

4. Give the expression of the probability that the buttered side will be facing down when hitting the floor.

5. Evaluate the above for $h = 0.8$ m, $\dot{\theta}_1 = 200^0$/s, $\dot{\theta}_2 = 300^0$/s.

6. Reevaluate the above for the land of the Brobdingnagians, in which the size of people (and their furniture) is scaled up by a factor of 9 compared to the above.

Meta 1018
p. 120

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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)

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Chapter 2

BASIC CONCEPTS IN ESTIMATION

2.1 INTRODUCTION

2.1.1 Outline

This chapter introduces some of the basic techniques of estimation that provide the foundation for state estimation and its applications like tracking, navigation, etc.

The problem of parameter estimation is defined in Section 2.2, where the two most commonly used models for unknown parameters (nonrandom and random) are also described. Section 2.3 deals with the maximum likelihood (ML) and the maximum a posteriori (MAP) estimators. The least squares (LS) and the minimum mean square error (MMSE) estimators are presented in Section 2.4.

The remaining sections deal with various "measures of quality" of estimators. Section 2.5 discusses unbiasedness and Section 2.6 discusses the variances of estimators. The consistency of estimators is discussed in Section 2.7, together with "information limit" results: the Cramer-Rao lower bound, the Fisher information, and estimator efficiency.

2.1.2 Basic Concepts – Summary of Objectives

Distinguish between

- Random parameters
- Nonrandom parameters

Define the following estimates

- Maximum likelihood
- Maximum a posteriori
- Least squares

- Minimum mean square error

Present "measures of quality" of estimators and "information limit" results

- Unbiasedness
- Variance
- Consistency
- The Cramer-Rao lower bound, Fisher information
- Efficiency

2.2 THE PROBLEM OF PARAMETER ESTIMATION

2.2.1 Definitions

The term ***parameter*** is used to designate a quantity (scalar or vector valued) that is assumed to be *time invariant.* If it does change with time, it can be designated (with a slight abuse of language) as a "time-varying parameter," but its time variation must be "slow" compared to the state variables of a system. *State estimation*, which is for *dynamic systems*, is covered starting with Chapter 5.

The problem of estimating a (time invariant) parameter x is the following. Given the measurements

$$z(j) = h[j, x, w(j)] \qquad j = 1, \ldots, k \tag{2.2.1-1}$$

made in the presence of the disturbances (noises) $w(j)$, find a function of the k observations

$$\boxed{\hat{x}(k) \triangleq \hat{x}[k, Z^k]} \tag{2.2.1-2}$$

where these observations are denoted compactly as

$$Z^k \triangleq \{z(j)\}_{j=1}^{k} \tag{2.2.1-3}$$

that estimates the value of x in some sense.

The function (2.2.1-2) is called the ***estimator***. The value of this function is the ***estimate***. These terms, while not the same, will be used (sometimes) interchangeably.

The ***estimation error*** corresponding to the estimate $\hat{x}$ is

$$\boxed{\tilde{x} \triangleq x - \hat{x}} \tag{2.2.1-4}$$

An alternate notation instead of (2.2.1-2) that will be used when k is fixed (and, therefore, can be omitted) is

$$\hat{x}(Z) \triangleq \hat{x}[k, Z^k] \tag{2.2.1-5}$$

where Z is the set of observations.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Remark

Parameter estimation can be viewed as a ***static estimation problem***, while state estimation can be viewed as a ***dynamic estimation problem***.

2.2.2 Models for Estimation of a Parameter

There are two models one can use in the estimation of a (time invariant) parameter:

1. Nonrandom ("unknown constant"): There is an unknown true value x_0. This is also called the ***non-Bayesian*** or ***Fisher approach***.
2. Random: The parameter is a random variable with a *prior* (or *a priori*) pdf $p(x)$ — a *realization* (see Subsection 1.4.2) of x according to $p(x)$ is assumed to have occurred; this value then stays constant during the measurement process. This is also called the ***Bayesian approach***.

The Bayesian Approach

In the ***Bayesian approach***, one starts with the ***prior*** pdf of the parameter from which one can obtain its ***posterior*** pdf (or ***a posteriori*** pdf) using ***Bayes' formula***:

$$\boxed{p(x|Z) = \frac{p(Z|x)p(x)}{p(Z)} = \frac{1}{c}p(Z|x)p(x)} \tag{2.2.2-1}$$

where c is the ***normalization constant***, which does not depend on x.

The posterior pdf can be used in several ways to estimate x.

The Non-Bayesian (Likelihood Function) Approach

In contrast to the above, in the ***non-Bayesian approach*** there is no prior pdf associated with the parameter and thus one cannot define a posterior pdf for it.

In this case, one has the *pdf of the measurements conditioned on the parameter*, called the ***likelihood function (LF)*** of the parameter

$$\boxed{\Lambda_Z(x) \triangleq p(Z|x)} \tag{2.2.2-2}$$

or

$$\Lambda_k(x) \triangleq p(Z^k|x) \tag{2.2.2-3}$$

as a measure of how "likely" a parameter value is given the obtained observations. The likelihood function serves as a measure of the ***evidence from the data***.

The use of the LF (2.2.2-2) and a similar usage of the posterior pdf (2.2.2-1) are discussed in the next section.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.3 MAXIMUM LIKELIHOOD AND MAXIMUM A POSTERIORI ESTIMATORS

2.3.1 Definitions of ML and MAP Estimators

Maximum Likelihood Estimator

A common method of estimating nonrandom parameters is the ***maximum likelihood method*** that maximizes the likelihood function (2.2.2-2). This yields the ***maximum likelihood estimator (MLE)***

$$\boxed{\hat{x}^{\mathrm{ML}}(Z) = \arg\max_x \Lambda_Z(x) = \arg\max_x p(Z|x)} \tag{2.3.1-1}$$

Note that, while x is an unknown constant, $\hat{x}^{\mathrm{ML}}(Z)$, being a function of the set of random observations Z, is a random variable.

The MLE is the solution of the ***likelihood equation***

$$\frac{d\Lambda_Z(x)}{dx} = \frac{dp(Z|x)}{dx} = 0 \tag{2.3.1-2}$$

Maximum A Posteriori Estimator

The corresponding estimate for a random parameter is the ***maximum a posteriori (MAP)*** estimator, which follows from the maximization of the posterior pdf (2.2.2-1):

$$\boxed{\hat{x}^{\mathrm{MAP}}(Z) = \arg\max_x p(x|Z) = \arg\max_x [p(Z|x)p(x)]} \tag{2.3.1-3}$$

The last equality above follows from the fact that, when using Bayes' formula (2.2.2-1), the normalization constant is irrelevant for the maximization.

The MAP estimate, which depends on the observations Z, and through them on the realization of x is, obviously, a random variable.

2.3.2 MLE vs. MAP Estimator with Gaussian Prior

Consider the single measurement

$$z = x + w \tag{2.3.2-1}$$

of the unknown parameter x in the presence of the additive measurement noise w, assumed to be a normally (Gaussian) distributed random variable with mean zero and variance σ^2, that is,

$$w \sim \mathcal{N}(0, \sigma^2) \tag{2.3.2-2}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

First assume that x is an unknown constant (no prior information about it is available). The likelihood function of x (denoted here without a subscript, for simplicity) is

$$\Lambda(x) = p(z|x) = \mathcal{N}(z; x, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-x)^2}{2\sigma^2}} \tag{2.3.2-3}$$

Then

$$\hat{x}^{\text{ML}} = \arg\max_x \Lambda(x) = z \tag{2.3.2-4}$$

since the peak or ***mode*** of (2.3.2-3) occurs at $x = z$.

Next assume that the prior information about the parameter is that x is Gaussian with mean $\bar{x}$ and variance σ_0^2, that is,

$$p(x) = \mathcal{N}(x; \bar{x}, \sigma_0^2) \tag{2.3.2-5}$$

It is also assumed that x is independent of w.

Then the posterior pdf of x conditioned on the observation z is

$$p(x|z) = \frac{p(z|x)p(x)}{p(z)} = \frac{1}{c} e^{-\frac{(z-x)^2}{2\sigma^2} - \frac{(x-\bar{x})^2}{2\sigma_0^2}} \tag{2.3.2-6}$$

where

$$c = 2\pi\sigma\sigma_0 p(z) \tag{2.3.2-7}$$

is the normalization constant independent of x. This normalization constant, which guarantees that the pdf integrates to unity, is given explicitly next, after rearranging the exponent in (2.3.2-6).

After rearranging the exponent in the above by completing the squares in x, it can be easily shown in a manner similar to the one used in Subsection 1.4.14 that the posterior pdf of x is

$$p(x|z) = \mathcal{N}[x; \xi(z), \sigma_1^2] = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{[x-\xi(z)]^2}{2\sigma_1^2}} \tag{2.3.2-8}$$

i.e., Gaussian, where

$$\xi(z) \triangleq \frac{\sigma^2}{\sigma_0^2 + \sigma^2}\bar{x} + \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2} z = \bar{x} + \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}(z - \bar{x}) \tag{2.3.2-9}$$

and (the "parallel resistors formula")

$$\sigma_1^2 \triangleq \frac{\sigma_0^2 \sigma^2}{\sigma_0^2 + \sigma^2} \tag{2.3.2-10}$$

The maximization of (2.3.2-8) with respect to x yields immediately

$$\hat{x}^{\text{MAP}} = \xi(z) \tag{2.3.2-11}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

that is, $\xi(z)$ given by (2.3.2-9) is the *maximum a posteriori estimator* for the random parameter x with the prior pdf (2.3.2-5).

Note that the MAP estimator (2.3.2-9) for this (purely Gaussian) problem is a weighted combination of

1. z, the MLE, which is the peak (or mode) of the likelihood function;
2. $\bar{x}$, which is the peak of the prior pdf of the parameter to be estimated.

Equation (2.3.2-9) can be rewritten as follows:

$$\begin{aligned}\hat{x}^{\text{MAP}} &= (\sigma_0^{-2}+\sigma^{-2})^{-1}\sigma_0^{-2}\bar{x} + (\sigma_0^{-2}+\sigma^{-2})^{-1}\sigma^{-2}z \\ &= (\sigma_0^{-2}+\sigma^{-2})^{-1}\left[\frac{\bar{x}}{\sigma_0^2}+\frac{z}{\sigma^2}\right] \qquad (2.3.2\text{-}12)\end{aligned}$$

which indicates that the weightings of the prior mean and the measurement are *inversely proportional to their variances.*

Similarly, (2.3.2-10) can be rewritten as follows:

$$\sigma_1^{-2} = \sigma_0^{-2} + \sigma^{-2} \qquad (2.3.2\text{-}13)$$

which shows that the *inverse variances* (also called ***information*** — this will be discussed in more detail in Subsection 3.4.2) are *additive.* This additivity property of information holds in general when the information sources are *independent.* (See also problem 2-8.)

2.3.3 MAP Estimator with One-Sided Exponential Prior

Consider the same problem as before except that the prior pdf of x is a ***one-sided exponential pdf***

$$p(x) = ae^{-ax} \qquad x \geq 0 \qquad (2.3.3\text{-}1)$$

This can model, for instance, the arrival time in a stochastic process where the number of arrivals is Poisson distributed.

The ML estimate is the same as before in (2.3.2-4), that is,

$$\hat{x}^{\text{ML}} = z \qquad (2.3.3\text{-}2)$$

The posterior pdf of x is now

$$p(x|z) = c(z)e^{-\frac{(z-x)^2}{2\sigma^2}-ax} \qquad x \geq 0 \qquad (2.3.3\text{-}3)$$

Since the exponent is quadratic in x, the above posterior pdf is Gaussian but truncated due to the fact that x cannot be negative as modeled by the prior given in (2.3.3-1).

In view of the fact that it cannot be negative, the maximizing argument of (2.3.3-3) is given by

$$\hat{x}^{\text{MAP}} = \max(z - \sigma^2 a, 0) \qquad (2.3.3\text{-}4)$$

Note that the MAP estimate (2.3.3-4) in this case will always be smaller than the MLE (2.3.3-2) as long as the latter is not negative because the prior (2.3.3-1) attaches higher probability to smaller values of x.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.3.4 MAP Estimator with Diffuse Prior

While $\hat{x}^{\mathrm{ML}}$ is based on a non-Bayesian approach and $\hat{x}^{\mathrm{MAP}}$ is based on the Bayesian approach, the latter will coincide with the former for a certain prior pdf, called a ***diffuse pdf***.

This can be seen by rewriting the denominator in Bayes' formula

$$p(x|Z) = \frac{p(Z|x)p(x)}{p(Z)} \tag{2.3.4-1}$$

with the total probability theorem

$$p(Z) = \int_{-\infty}^{\infty} p(Z|x)p(x)\,dx \tag{2.3.4-2}$$

and assuming a ***diffuse uniform prior pdf*** for the parameter

$$p(x) = \epsilon \qquad \text{for } |x| < \frac{1}{2\epsilon} \tag{2.3.4-3}$$

over a "sufficiently large" region of length $1/\epsilon$ where $\epsilon > 0$ but small. Using (2.3.4-3) in (2.3.4-2) yields

$$p(Z) = \epsilon \int_{-\infty}^{\infty} p(Z|x)dx = \epsilon g(Z) \tag{2.3.4-4}$$

where g does not depend on x.

Then, inserting (2.3.4-4) into Bayes' formula (2.3.4-1) yields

$$p(x|Z) = \frac{p(Z|x)\epsilon}{\epsilon g(Z)} = \frac{p(Z|x)}{g(Z)} = \frac{1}{c}p(Z|x) \tag{2.3.4-5}$$

since $\epsilon \neq 0$.

This diffuse pdf is also called ***improper pdf*** because as $\epsilon \to 0$, it does not integrate to unity as a ***proper pdf*** does. Another name for it is ***noninformative pdf*** because it carries no information about the parameter: uniform distribution over an infinite interval at the limit.

A diffuse prior causes the posterior pdf of x to be proportional to its likelihood function and, thus, the MAP estimate to coincide with the MLE.

Bayesian vs. Non-Bayesian Philosophies

The non-Bayesian MLE is, in view of the above discussion, nothing but the Bayesian MAP estimate with complete prior ignorance, reflected by the diffuse prior (2.3.4-3). *This provides a philosophically unifying view of the Bayesian and non-Bayesian approaches to estimation.*[1]

[1]Thus one can say that statisticians can be divided into two categories: Bayesians and closet Bayesians. This is a particular case of the following general taxonomy theorem: People can be divided into two categories. (*Proof*: The categories are those who believe that people can be divided into two categories and those who do not.)

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Remark

In spite of the fact that the diffuse prior (2.3.4-3) does not integrate to unity and has no moments (i.e., it is *not a proper pdf*), the posterior pdf of x will be, in general, proper.

Example

Consider the problem of Subsection 2.3.2 where the prior (2.3.2-5) is made diffuse by making $\sigma_0 \to \infty$. A Gaussian pdf with very large variance becomes flat and at the limit looks like a uniform pdf over the whole real line.

When $\sigma_0 \to \infty$ it can be seen from (2.3.2-9) that

$$\lim_{\sigma_0 \to \infty} \xi(z) = z \tag{2.3.4-6}$$

that is, $\hat{x}^{\text{MAP}}$ coincides with $\hat{x}^{\text{ML}}$. This occurs regardless of the value of $\bar{x}$, which becomes irrelevant when $\sigma_0 \to \infty$ (i.e., this is a *noninformative prior*). Thus the non-Bayesian approach can be seen as a degenerate case of the Bayesian approach.

The Philosophical Meaning of the Prior

The prior pdf assumed in a problem is in many cases the subjective assessment of phenomena. The uniform prior assumes Nature as "indifferent." In game theory, Nature is assumed to be opposed to our interests. While neither of these two extreme points of view is correct, it is useful to keep in mind the well-known ***principle of perversity of inanimate objects.***[2]

2.3.5 The Sufficient Statistic and the Likelihood Equation

If the likelihood function of a parameter can be decomposed as follows

$$\Lambda(x) \triangleq p(Z|x) = f_1[g(Z), x] f_2(Z) \tag{2.3.5-1}$$

then it is clear that the maximum likelihood estimate of x depends only on the function $g(Z)$, called the ***sufficient statistic***, rather than on the entire data set Z.

The sufficient statistic *summarizes the information about x contained in the entire data.*

[2] As mentioned by Richard Bellman [Bellman61], this has been established by a number of experiments. The most conclusive of these involved dropping a piece of buttered toast on a rug. In 79.3% of the trials the toast fell buttered side down; for a mathematical proof of this principle, which does not hold in the land of the Brobdingnagians, see problem 1-17. For an extensive discussion on priors, see [Raiffa72].

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Example

Consider the scalar measurements

$$z(j) = x + w(j) \qquad j = 1, \ldots, k \tag{2.3.5-2}$$

If the noise components $w(j)$, $j = 1, \ldots, k$, are independent and identically distributed zero-mean Gaussian random variables with variance σ^2, that is,

$$w(j) \sim \mathcal{N}(0, \sigma^2) \tag{2.3.5-3}$$

then

$$z(j) \sim \mathcal{N}(x, \sigma^2) \tag{2.3.5-4}$$

and, conditioned on x, the observations $z(j)$ are mutually independent.

Thus, the ***likelihood function*** of x in terms of

$$Z^k \triangleq \{z(j), j = 1, \ldots, k\} \tag{2.3.5-5}$$

is then

$$\begin{aligned} \Lambda_k(x) &\triangleq p(Z^k|x) \triangleq p[z(1), \ldots, z(k)|x] \\ &= \prod_{j=1}^{k} \mathcal{N}[z(j); x, \sigma^2] = ce^{-\frac{1}{2\sigma^2}\sum_{j=1}^{k}[z(j)-x]^2} \end{aligned} \tag{2.3.5-6}$$

The likelihood function (2.3.5-6) can be rewritten into the product of two functions as in (2.3.5-1) as follows:

$$\begin{aligned} \Lambda_k(x) &= c\, e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{k} z(j)^2 + \frac{1}{2\sigma^2} 2\sum_{j=1}^{k} z(j)x - \frac{1}{2\sigma^2}kx^2} \\ &= c\, e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{k} z(j)^2} e^{-\frac{1}{2\sigma^2}kx\left[x - \frac{2}{k}\sum_{j=1}^{k} z(j)\right]} \\ &\triangleq f_2(Z) f_1[g(Z), x] \end{aligned} \tag{2.3.5-7}$$

where

$$f_2(Z) \triangleq ce^{-\frac{1}{2\sigma^2}\sum_{j=1}^{k} z(j)^2} \tag{2.3.5-8}$$

$$f_1[g(Z), x] \triangleq e^{-\frac{1}{2\sigma^2}kx[x - 2\bar{z}]} \tag{2.3.5-9}$$

$$g(Z) \triangleq \frac{1}{k}\sum_{j=1}^{k} z(j) \triangleq \bar{z} \tag{2.3.5-10}$$

Thus, according to the definition (2.3.5-1), $\bar{z}$ is the *sufficient statistic* for estimating x.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The Likelihood Equation

To maximize the likelihood function (2.3.5-7), one sets its derivative with respect to x to zero. Since f_2 is independent of x, the ***likelihood equation*** is

$$\frac{d\Lambda_k(x)}{dx} = 0 \qquad \Longleftrightarrow \qquad \frac{df_1[g(Z), x]}{dx} = 0 \tag{2.3.5-11}$$

Since the logarithm is a monotonic transformation, equivalently one can use the derivative of the ***log-likelihood function***

$$\frac{d \ln \Lambda_k(x)}{dx} = \frac{d \ln f_1[g(Z), x]}{dx} = -\frac{k}{2\sigma^2} 2(x - \bar{z}) = 0 \tag{2.3.5-12}$$

which yields

$$\hat{x}^{\mathrm{ML}} = \bar{z} \tag{2.3.5-13}$$

The concept of sufficient statistic carries over to the MAP procedure in a completely analogous manner.

2.4 LEAST SQUARES AND MINIMUM MEAN SQUARE ERROR ESTIMATION

2.4.1 Definitions of LS and MMSE Estimators

The LS Estimator

Another common estimation procedure for nonrandom parameters is the ***least squares (LS) method***. Given the (scalar and nonlinear) measurements

$$z(j) = h(j, x) + w(j) \qquad j = 1, \ldots, k \tag{2.4.1-1}$$

the ***least squares estimator (LSE)*** of x is, with notation (2.2.1-2),

$$\boxed{\hat{x}^{\mathrm{LS}}(k) = \arg\min_x \left\{ \sum_{j=1}^{k} [z(j) - h(j, x)]^2 \right\}} \tag{2.4.1-2}$$

This is the ***nonlinear LS problem*** — if the function h is linear in x, then one has the ***linear LS problem***. The linear LS problem is considered in more detail for the vector case in Section 3.4.

The criterion in (2.4.1-2) makes no assumptions about the "measurement errors" or "noises" $w(j)$. If these are independent and identically distributed zero-mean Gaussian random variables, that is,

$$w(j) \sim \mathcal{N}(0, \sigma^2) \tag{2.4.1-3}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

then the LSE (2.4.1-2) coincides with the MLE under these assumptions. In this case,

$$z(j) \sim \mathcal{N}[h(j,x), \sigma^2] \qquad j = 1, \ldots, k \tag{2.4.1-4}$$

The likelihood function of x is then

$$\begin{aligned} \Lambda_k(x) &\triangleq p(Z^k|x) \triangleq p[z(1), \ldots, z(k)|x] \\ &= \prod_{j=1}^{k} \mathcal{N}[z(j); h(j,x), \sigma^2] = ce^{-\frac{1}{2\sigma^2}\sum_{j=1}^{k}[z(j)-h(j,x)]^2} \end{aligned} \tag{2.4.1-5}$$

and the minimization (2.4.1-2) is equivalent to the maximization of (2.4.1-5); that is, *the LS method is a "disguised" ML approach.*

The MMSE Estimator

For random parameters, the counterpart of the above is the ***minimum mean square error (MMSE) estimator***

$$\hat{x}^{\text{MMSE}}(Z) = \arg\min_{\hat{x}} E[(\hat{x} - x)^2|Z] \tag{2.4.1-6}$$

The solution to (2.4.1-6) is the ***conditional mean*** of x

$$\boxed{\hat{x}^{\text{MMSE}}(Z) = E[x|Z] \triangleq \int_{-\infty}^{\infty} xp(x|Z)\,dx} \tag{2.4.1-7}$$

where the expectation is with respect to the conditional pdf (2.2.2-1).

The above follows by setting the derivative of (2.4.1-6) with respect to $\hat{x}$ to zero:

$$\frac{d}{d\hat{x}}E[(\hat{x} - x)^2|Z] = E[2(\hat{x} - x)|Z] = 2(\hat{x} - E[x|Z]) = 0 \tag{2.4.1-8}$$

For vector random variables, (2.4.1-7) is obtained similarly by setting the gradient of the mean of the squared norm of the error to zero, that is

$$\nabla_{\hat{x}}E[(\hat{x} - x)'(\hat{x} - x)|Z] = 2(\hat{x} - E[x|Z]) = 0 \tag{2.4.1-9}$$

from which (2.4.1-7) follows immediately.

Remarks

1. With x being an unknown constant (nonrandom) and the noises in (2.4.1-1) modeled as random (not necessarily Gaussian), the LSE is a random variable.
2. The MMSE estimate (2.4.1-7) is a random variable that depends on the observations Z and, through them, on (the realization of) x. Also, for a *given* Z, x is a random variable with a conditional pdf (2.2.2-1).
3. The MMSE estimation problem (2.4.1-6) is a particular case of Bayesian estimation where the expected value of a (positive definite) ***cost function*** $C(\hat{x} - x)$ is to be minimized. The MMSE cost function is a quadratic. The widespread use of the quadratic criterion is due primarily to the (relative) ease of obtaining the solution. (See also problem 2-3.)

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.4.2 Some LS Estimators

LS Estimator from a Single Measurement

For the problem of a single measurement of the unknown parameter x,

$$z = x + w \tag{2.4.2-1}$$

the least squares criterion leads to

$$\hat{x}^{\mathrm{LS}} = \arg\min_x [(z - x)^2] = z \tag{2.4.2-2}$$

which is the same result as $\hat{x}^{\mathrm{ML}}$ if w is zero-mean Gaussian. This is due to the fact that maximizing the likelihood function, which is a Gaussian pdf, is equivalent to minimizing the square in its exponent.

LS Estimator from Several Measurements

Assume now that k measurements are made

$$z(j) = x + w(j) \qquad j = 1, \ldots, k \tag{2.4.2-3}$$

where $w(j)$ are independent, identically distributed, normal, zero mean, and with common variance σ^2.

The likelihood function is, as in (2.4.1-5),

$$\Lambda_k(x) = ce^{-\frac{1}{2\sigma^2}\sum_{j=1}^{k}[z(j)-x]^2} \tag{2.4.2-4}$$

As before, the ML and LS estimates coincide and it can be easily shown that they are given by the following expression:

$$\hat{x}^{\mathrm{ML}}(k) = \hat{x}^{\mathrm{LS}}(k) = \frac{1}{k}\sum_{j=1}^{k} z(j) = \bar{z} \tag{2.4.2-5}$$

This estimate is known as the ***sample mean*** or ***sample average***, since it estimates the unknown mean x of the k random variables from (2.4.2-3).

2.4.3 MMSE vs. MAP Estimator in Gaussian Noise

In the single measurement example with a prior pdf on the parameter to be estimated, discussed in Subsection 2.3.2, the posterior pdf of x was obtained in (2.3.2-8) as

$$p(x|z) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{[x-\xi(z)]^2}{2\sigma_1^2}} \tag{2.4.3-1}$$

It is apparent by inspection that the mean of this Gaussian pdf is $\xi(z)$, which is also the ***mode*** (peak) of this pdf.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Thus

$$\hat{x}^{\text{MMSE}} = E[x|z] = \xi(z) = \hat{x}^{\text{MAP}} \tag{2.4.3-2}$$

i.e., the MMSE estimator (the conditional mean) *coincides* with the MAP estimator.

This is due to the fact that the *mean* and the *mode* of a Gaussian pdf, which is symmetric and ***unimodal***, coincide.

Note that, in view of (2.4.3-2), equation (2.4.3-1) can be also written as

$$p(x|z) = \mathcal{N}(x; \hat{x}^{\text{MMSE}}, \sigma_1^2) = \mathcal{N}(x; \hat{x}^{\text{MAP}}, \sigma_1^2) \tag{2.4.3-3}$$

2.5 UNBIASED ESTIMATORS

2.5.1 Definition

Non-Bayesian Case

For a nonrandom parameter, an estimator is said to be ***unbiased*** if

$$E[\hat{x}(k, Z^k)] = x_0 \tag{2.5.1-1}$$

where x_0 is the true value of the parameter. The expectation in (2.5.1-1) is over the estimate, which is a random variable since it is a function of the measurements (2.2.1-3), and is taken with respect to the conditional pdf $p(Z^k|x = x_0)$.

Bayesian Case

If x is a random variable with a prior pdf $p(x)$, then the unbiasedness property is written as

$$E[\hat{x}(k, Z^k)] = E[x] \tag{2.5.1-2}$$

where the expectation on the left-hand side above is with respect to the joint pdf $p(Z^k, x)$ and the one on the right-hand side is with respect to $p(x)$.

General Definition

The above unbiasedness requirements can be unified by requiring that the ***estimation error***

$$\tilde{x} \triangleq x - \hat{x} \tag{2.5.1-3}$$

be zero mean, that is,

$$\boxed{E[\tilde{x}] = 0} \tag{2.5.1-4}$$

Equation (2.5.1-4) covers both cases, with the expectation being taken over Z^k in the first case and over Z^k *and* x in the second case.

An estimator is unbiased if (2.5.1-4) holds for all k and is ***asymptotically unbiased*** if it holds in the limit as $k \to \infty$.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.5.2 Unbiasedness of an ML and a MAP Estimator

Consider the ML estimator (2.3.2-4) of the parameter x with true value x_0

$$\hat{x}^{\text{ML}} = z \tag{2.5.2-1}$$

from the single measurement

$$z = x + w \tag{2.5.2-2}$$

Its mean is

$$E[\hat{x}^{\text{ML}}] = E[z] = E[x_0 + w] = x_0 + E[w] = x_0 \tag{2.5.2-3}$$

since the mean of the Gaussian random variable w is zero.

For the MAP estimate (2.3.2-11) of x modeled as a Gaussian random variable with prior mean $\bar{x}$, prior variance σ^2, and independent of w,

$$\hat{x}^{\text{MAP}} = \xi(z) \triangleq \frac{\sigma^2}{\sigma_0^2 + \sigma^2}\bar{x} + \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}z \tag{2.5.2-4}$$

one has

$$\begin{aligned} E[\hat{x}^{\text{MAP}}] &= E[\xi(z)] = \frac{\sigma^2}{\sigma_0^2 + \sigma^2}\bar{x} + \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}E[z] \\ &= \frac{\sigma^2}{\sigma_0^2 + \sigma^2}\bar{x} + \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}[\bar{x} + E(w)] = \bar{x} = E[x] \end{aligned} \tag{2.5.2-5}$$

Thus, both of these estimates are unbiased.

2.5.3 Bias in the ML Estimation of Two Parameters

Consider the problem of estimating the unknown mean x of a set of k measurements as in (2.4.2-3), with the additional parameter to be estimated being the variance σ^2, now also assumed to be unknown.

The likelihood function for the unknown parameters x and σ is

$$\Lambda_k(x, \sigma) = p[z(1), \ldots, z(k)|x, \sigma] = \frac{1}{(2\pi)^{k/2}\sigma^k} e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{k}[z(j)-x]^2} \tag{2.5.3-1}$$

To maximize the above, one writes the ***likelihood equation*** by setting to zero the derivatives of Λ or, more conveniently, of $\ln \Lambda$, with respect to x and σ

$$\frac{\partial \ln \Lambda_k}{\partial x} = \frac{1}{\sigma^2}\sum_{j=1}^{k}[z(j) - x] = 0 \tag{2.5.3-2}$$

$$\frac{\partial \ln \Lambda_k}{\partial \sigma} = -\frac{k}{\sigma} + \frac{1}{\sigma^3}\sum_{j=1}^{k}[z(j) - x]^2 = 0 \tag{2.5.3-3}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The Solution of the Likelihood Equation

The first equation yields $\hat{x}^{\mathrm{ML}}$ as before in (2.4.2-5), that is, the sample mean — in this problem the estimate of x is not affected at all by the fact that σ is also unknown. Substituting this into the second equation yields

$$\frac{\partial \ln \Lambda_k}{\partial \sigma} = -\frac{k}{\sigma} + \frac{1}{\sigma^3} \sum_{j=1}^{k} [z(j) - \hat{x}^{\mathrm{ML}}]^2 = 0 \tag{2.5.3-4}$$

The resulting estimate, known as the ***sample variance*** based on k observations, is

$$[\hat{\sigma}^{\mathrm{ML}}(k)]^2 = \frac{1}{k} \sum_{j=1}^{k} [z(j) - \hat{x}^{\mathrm{ML}}]^2 = \frac{1}{k} \sum_{i=1}^{k} \left[z(j) - \frac{1}{k} \sum_{i=1}^{k} z(i) \right]^2 \tag{2.5.3-5}$$

The Means of the Sample Mean and Sample Variance

Denote the true values of the parameters by x_0 and σ_0. The expected value of the sample mean is

$$E[\hat{x}^{\mathrm{ML}}(k)] = E\left[\frac{1}{k} \sum_{j=1}^{k} z(j) \right] = x_0 \tag{2.5.3-6}$$

that is, the sample mean estimator (2.4.2-5) is unbiased.

The expected value of the sample variance (2.5.3-5) is

$$\begin{aligned} E\left\{ [\hat{\sigma}^{\mathrm{ML}}(k)]^2 \right\} &= E\left\{ \frac{1}{k} \sum_{j=1}^{k} \left[z(j) - \frac{1}{k} \sum_{i=1}^{k} z(i) \right]^2 \right\} \\ &= \frac{1}{k} \sum_{j=1}^{k} E\left\{ \left[w(j) - \frac{1}{k} \sum_{i=1}^{k} w(i) \right]^2 \right\} \\ &= \frac{1}{k^3} \sum_{j=1}^{k} E\left\{ \left[(k-1)w(j) - \sum_{\substack{i=1 \\ i \neq j}}^{k} w(i) \right]^2 \right\} \\ &= \frac{1}{k^2} [(k-1)^2 + k - 1] \sigma_0^2 \\ &= \frac{k-1}{k} \sigma_0^2 \end{aligned} \tag{2.5.3-7}$$

Thus the sample variance (2.5.3-5) is *biased*, even though it becomes unbiased as $k \to \infty$, i.e., it is *asymptotically unbiased*. In order to be unbiased, the denominator in (2.5.3-5) should be $k - 1$ rather than k:

$$[\hat{\sigma}(k)]^2 = \frac{1}{k-1} \sum_{j=1}^{k} \left[z(j) - \frac{1}{k} \sum_{i=1}^{k} z(i) \right]^2 \tag{2.5.3-8}$$

Expression (2.5.3-8) is the more common sample variance used. However, for reasonably large k this is not going to make a significant difference. (See also problem 2-4.)

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.6 THE VARIANCE AND MSE OF AN ESTIMATOR

2.6.1 Definitions of Estimator Variances

Non-Bayesian Case

For a non-Bayesian estimator, $\hat{x}(Z)$, (LS or ML) the ***variance of the estimator*** is

$$\text{var}[\hat{x}(Z)] \triangleq E\left[\{\hat{x}(Z) - E[\hat{x}(Z)]\}^2\right] \tag{2.6.1-1}$$

where the averaging is over the observation set Z.

If this estimator is *unbiased*, that is,

$$E[\hat{x}(Z)] = x_0 \tag{2.6.1-2}$$

where x_0 is the true value, then

$$\text{var}[\hat{x}(Z)] = E[[\hat{x}(Z) - x_0]^2] \tag{2.6.1-3}$$

If this estimator is *biased*, then (2.6.1-3) is its ***mean square error (MSE)***[3]

$$\text{MSE}[\hat{x}(Z)] = E[[\hat{x}(Z) - x_0]^2] \tag{2.6.1-4}$$

Bayesian Case

For a Bayesian estimator, the ***unconditional MSE*** is[4]

$$\text{MSE}[\hat{x}(Z)] \triangleq E[[\hat{x}(Z) - x]^2] \tag{2.6.1-5}$$

where the averaging is with respect to the joint pdf of the observations Z and the random parameter x. The above can be rewritten, using the smoothing property of expectations (see Subsection 1.4.12), as follows:

$$\text{MSE}[\hat{x}(Z)] = E\left[E\{[\hat{x}(Z) - x]^2|Z\}\right] = E[\text{MSE}[\hat{x}(Z)|Z]] \tag{2.6.1-6}$$

where the last expression inside the braces is the ***conditional MSE***, i.e., for a given realization (or value) of the observations Z.

For the MMSE estimator, the conditional MSE is

$$\begin{aligned} E[[\hat{x}^{\text{MMSE}}(Z) - x]^2|Z] &= E[[x - E(x|Z)]^2|Z] \\ &= \text{var}(x|Z) \end{aligned} \tag{2.6.1-7}$$

that is, the ***conditional variance*** of x *given* Z. Note that the expectations in (2.6.1-7) are with respect to $p(x|Z)$.

Averaging over Z yields

$$E[\text{var}(x|Z)] = E[[x - E(x|Z)]^2] \tag{2.6.1-8}$$

which is the unconditional MSE (2.6.1-5) of the estimate $\hat{x}^{\text{MMSE}}$. This is the "average squared error over all the possible observations."

[3]Mean square is a personality type from the official list compiled by psycho-statisticians.

[4]The variance, since it would be about the mean $E[x]$ (if unbiased), has no real meaning.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

General Definition

With the definition of the estimation error

$$\tilde{x} \triangleq x - \hat{x} \tag{2.6.1-9}$$

one can say in a unified manner that the expected value of the square of the estimation error is the estimator's variance or MSE:

$$E[\tilde{x}^2] = \begin{cases} \text{var}(\hat{x}) & \text{if } \hat{x} \text{ is unbiased and } x \text{ is nonrandom} \\ \text{MSE}(\hat{x}) & \text{in all cases} \end{cases} \tag{2.6.1-10}$$

where the expectations are to be taken according to the discussion above.

The square root of the variance (or MSE) of an estimator

$$\sigma_{\hat{x}} \triangleq \sqrt{\text{var}(\hat{x})} \tag{2.6.1-11}$$

is its ***standard error***, also called the ***standard deviation associated with the estimator*** or the ***standard deviation of the estimation error***.

The standard error provides a measure of the accuracy of the estimator: assuming the estimation error to be Gaussian, the difference between the estimate and the true value will be up to 2 standard errors with 95% probability.

2.6.2 Comparison of Variances of an ML and a MAP Estimator

The "qualities" of the ML and MAP estimators discussed in Subsection 2.3.2 (from a single observation), as measured by their variances, will be compared next.

For the MLE given by (2.3.2-4) one has

$$\text{var}(\hat{x}^{\text{ML}}) = E[(\hat{x}^{\text{ML}} - x_0)^2] = E[(z - x_0)^2] \triangleq \sigma^2 \tag{2.6.2-1}$$

For the MAP estimate given by (2.3.2-11), which has a Gaussian prior in this case, one has[5]

$$\begin{aligned} \text{var}(\hat{x}^{\text{MAP}}) &= E[(\hat{x}^{\text{MAP}} - x)^2] \\ &= E\left\{\left[\frac{\sigma^2}{\sigma_0^2 + \sigma^2}\bar{x} + \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}(x + w) - x\right]^2\right\} \\ &= E\left[\left[\frac{\sigma^2}{\sigma_0^2 + \sigma^2}(\bar{x} - x) + \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}w\right]^2\right] \\ &= \frac{\sigma_0^2\sigma^2}{\sigma_0^2 + \sigma^2} < \sigma^2 = \text{var}(\hat{x}^{\text{ML}}) \end{aligned} \tag{2.6.2-2}$$

[5] With abuse of notation because this is really the MSE; such abuses of notation are very common in the literature.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Thus it can be seen that the variance of the MAP estimator (given by the "parallel resistors formula") is *smaller* than that of the MLE — this is due to the availability of *prior information.*

Note that in (2.6.2-1) the averaging is only over z (or, equivalently, w) while in (2.6.2-2) the averaging is over w *and* x, which is assumed random. (See also problem 2-1.)

2.6.3 The Variances of the Sample Mean and Sample Variance

The ***variance of the sample mean*** (2.4.2-5) — the square of its ***standard error*** — is obtained as

$$E\left[[\hat{x}^{\mathrm{ML}}(k) - x_0]^2\right] = E\left\{\left[\frac{1}{k}\sum_{j=1}^{k}[z(j) - x_0]\right]^2\right\} = \frac{\sigma^2}{k} \tag{2.6.3-1}$$

which, as $k \to \infty$, converges to zero; that is, this estimator is consistent (the definition is given in the next subsection).

The ***variance of the sample variance*** (2.5.3-5) is computed next. For simplicity, it is assumed that the mean is zero and known. The estimator of the variance in this case is

$$(\hat{\sigma}^{\mathrm{ML}})^2 = \frac{1}{k}\sum_{j=1}^{k} z(j)^2 \tag{2.6.3-2}$$

and it can be easily shown that it is unbiased.

The variance of this estimator is, with the true value denoted by σ^2,

$$\begin{aligned} E\left[[(\hat{\sigma}^{\mathrm{ML}})^2 - \sigma^2]^2\right] &= E\left\{\left[\frac{1}{k}\sum_{j=1}^{k} z(j)^2 - \sigma^2\right]^2\right\} \\ &= \frac{1}{k^2}\sum_{j=1}^{k}\sum_{i=1}^{k} E[w(j)^2 w(i)^2] - 2\sigma^2\frac{1}{k}\sum_{j=1}^{k} E[w(j)^2] + \sigma^4 \\ &= \frac{1}{k^2}[k(k-1)\sigma^4 + k3\sigma^4] - \frac{2}{k}k\sigma^4 + \sigma^4 = \frac{2\sigma^4}{k} \end{aligned} \tag{2.6.3-3}$$

which also converges to zero as $k \to \infty$.

The following relationship has been used in (2.6.3-3)

$$E[w(i)^2 w(j)^2] = \begin{cases} \sigma^4 & \text{if } i \neq j \\ 3\sigma^4 & \text{if } i = j \end{cases} \tag{2.6.3-4}$$

The fourth moment of w is needed here and, assuming it to be Gaussian, use was made of (1.4.15-7).

The ***standard error of the sample variance*** from k samples is therefore, from (2.6.3-3), given by

$$\sigma_{(\hat{\sigma}^{\mathrm{ML}})^2} = \sigma^2\sqrt{2/k} \tag{2.6.3-5}$$

In the above the notation, σ_ξ has been used to denote the ***standard error of the estimate*** ξ.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Application — The Number of Samples Needed to Estimate a Variance with a Given Accuracy

Based on this result, the number of samples needed to obtain the sample variance *within 10% of the true value with probability of 95%* can be obtained as follows.

Assuming for convenience that the sample variance is normally distributed [6] about its mean (equal to the true variance) with standard error as above, one has the 95% confidence region

$$P\left\{|(\hat{\sigma}^{\mathrm{ML}})^2 - \sigma^2| \leq 1.96\sigma^2\sqrt{2/k}\right\} = 0.95 \tag{2.6.3-6}$$

The requirement of at most 10% error in the variance estimate, that is,

$$\frac{|(\hat{\sigma}^{\mathrm{ML}})^2 - \sigma^2|}{\sigma^2} = 0.1 \tag{2.6.3-7}$$

leads to setting

$$1.96\sqrt{2/k} = 0.1 \tag{2.6.3-8}$$

which yields

$$k \approx 800 \tag{2.6.3-9}$$

The resulting very large number of samples necessary for the required accuracy justifies the use of the CLT in (2.6.3-6). (See also problem 1-2.)

2.6.4 Estimation of the Probability of an Event

The ***estimation of the probability of an event*** can be done as follows. If in N independent identical experiments an event occurs N_0 times, then one can use the following estimate of the probability p of this event:

$$\hat{p} = \frac{N_0}{N} \tag{2.6.4-1}$$

It can be easily seen that the above is an unbiased estimate since

$$E\hat{p} = E\frac{N_0}{N} = \frac{Np}{N} = p \tag{2.6.4-2}$$

Now, since N_0 is a Bernoulli random variable (the sum of N i.i.d. binary random variables), its distribution is given exactly by the binomial distribution. However, one can gain more insight into this by assuming N large enough, in which case one can use, in view of the CLT, the Gaussian approximation of the pdf of $\hat{p}$. ICBES that

$$\mathrm{var}[\hat{p}] = \frac{p(1-p)}{N} \tag{2.6.4-3}$$

[6] In view of the CLT, see Subsection 1.4.23.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Thus, using the 95% (2σ) confidence region, one can say that

$$P\left\{|\hat{p}-p| \leq 2\sqrt{\frac{p(1-p)}{N}}\right\} = 0.95 \qquad (2.6.4\text{-}4)$$

Since we do not know p, one can use $\hat{p}$ instead of p in the expression of the variance above.

As an example, if $N = 1000$ and $\hat{p} = 0.1$, the standard deviation of the estimate is $\sqrt{0.1 \cdot 0.9/1000} \approx 0.01$. Therefore, the above confidence region becomes [0.08, 0.12].

Such techniques are commonly used in public opinion polls, but in a more conservative way: The worst case of $p = 0.5$ is assumed for the variance. For $N = 1000$ the standard deviation of the estimate is then $\sqrt{0.5 \cdot 0.5/1000} \approx 0.016$ with the confidence region being $[\hat{p} - 0.03, \hat{p} + 0.03]$. This is how to interpret the results from a poll, which state a ***margin of error*** of $\pm 3\%$.

2.7 CONSISTENCY AND EFFICIENCY OF ESTIMATORS

2.7.1 Consistency

An estimator of a *nonrandom parameter* is said to be a ***consistent estimator*** if the estimate (which is a random variable) converges to the true value in some stochastic sense.[7] Using the ***convergence in mean square*** criterion, then

$$\lim_{k\to\infty} E\left[[\hat{x}(k, Z^k) - x_0]^2\right] = 0 \qquad (2.7.1\text{-}1)$$

is the condition for ***consistency in the mean square sense***. The expectation is taken over Z^k, as in (2.5.1-1).

For a *random parameter*, convergence of its estimator in the mean square sense requires

$$\lim_{k\to\infty} E\left[[\hat{x}(k, Z^k) - x]^2\right] = 0 \qquad (2.7.1\text{-}2)$$

where the expectation is over Z^k *and* x, as in (2.5.1-2).

Similarly to the unbiasedness case, consistency can be expressed as the requirement that the estimation error converge to zero, that is,

$$\lim_{k\to\infty} \tilde{x}(k, Z^k) = 0 \qquad (2.7.1\text{-}3)$$

in some stochastic (e.g., mean square) sense.

[7] Another definition of consistency is "the last refuge of the unimaginative" (Oscar Wilde).

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Remark

The consistency defined above is an *asymptotic* property; that is, it is defined for the case when the sample size k tends to infinity and the object of estimation is a *fixed parameter*. Later, in the context of state estimation, where the object of estimation is an *evolving state*, there will be another definition of consistency as a *finite sample size* property.

2.7.2 The Cramer-Rao Lower Bound and the Fisher Information Matrix

According to the ***Cramer-Rao lower bound (CRLB)***, the mean square error corresponding to the estimator of a parameter *cannot be smaller* than a certain quantity related to the likelihood function.

Scalar Case

In the estimation of a scalar *nonrandom* parameter x with an *unbiased* estimator $\hat{x}(Z)$, the variance is bounded from below as follows:

$$\boxed{E\left[[\hat{x}(Z) - x_0]^2\right] \geq J^{-1}} \tag{2.7.2-1}$$

where

$$J \triangleq -E\left[\frac{\partial^2 \ln \Lambda(x)}{\partial x^2}\right]\Bigg|_{x=x_0} = E\left\{\left[\frac{\partial \ln \Lambda(x)}{\partial x}\right]^2\right\}\Bigg|_{x=x_0} \tag{2.7.2-2}$$

is the ***Fisher information***, $\Lambda(x) = p(Z|x)$ is the likelihood function (2.2.2-2) denoted for simplicity without subscript, and x_0 is the true value of the unknown constant x.

For a scalar *random* parameter x estimated by an unbiased estimator $\hat{x}(Z)$, the variance is bounded from below by a similar expression, namely,

$$E\left[[\hat{x}(Z) - x]^2\right] \geq J^{-1} \tag{2.7.2-3}$$

where

$$J \triangleq -E\left[\frac{\partial^2 \ln p(Z,x)}{\partial x^2}\right] = E\left\{\left[\frac{\partial \ln p(Z,x)}{\partial x}\right]^2\right\} \tag{2.7.2-4}$$

The expectations in (2.7.2-2) and (2.7.2-4) are taken as in (2.7.1-1) and (2.7.1-2), respectively.

Note that the Fisher information has two forms in (2.7.2-2) as well as in (2.7.2-4). The proof of (2.7.2-1) and the equivalence of the two forms in (2.7.2-2) — one with first partial derivatives and the other with second partials — is given later.

If an estimator's variance is equal to the CRLB, then such an estimator is called ***efficient***.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Multidimensional Case

For *nonrandom vector parameters*, the CRLB states that the covariance matrix of an unbiased estimator is bounded from below as follows:

$$\boxed{E[[\hat{x}(Z) - x_0][\hat{x}(Z) - x_0]'] \geq J^{-1}} \tag{2.7.2-5}$$

where the ***Fisher information matrix (FIM)*** is

$$\boxed{J \triangleq -E[\nabla_x \nabla_x' \ln \Lambda(x)]\,|_{x=x_0} = E[[\nabla_x \ln \Lambda(x)][\nabla_x \ln \Lambda(x)]']|_{x-x_0}} \tag{2.7.2-6}$$

and x_0 is the true value of the vector parameter x.

As in the scalar case, note the two forms of the FIM: one with the Hessian of the log-likelihood function and the other with the dyad of its gradient.

The matrix inequality in (2.7.2-5) is to be interpreted as follows:

$$A \geq B \qquad \Longleftrightarrow \qquad C \triangleq A - B \geq 0 \tag{2.7.2-7}$$

that is, the difference C of the two matrices is positive semidefinite.

A similar expression holds for the case of a multidimensional random parameter.

Remarks

The FIM can be seen as a quantification of the (maximum) ***existing information*** in the data about a parameter. Efficiency amounts to the ***extracted information*** being equal to the existing one, i.e., all the information has been extracted.

A necessary condition for an estimator to be consistent in the mean square sense is that there must be an increasing amount of information (in the sense of Fisher) about the parameter in the measurements — the Fisher information has to tend to infinity as $k \to \infty$. Then the CRLB converges to zero as $k \to \infty$ and thus the variance can also converge to zero.

Note

For estimators that are *biased*, there is a modified version of the CRLB (e.g., [Van Trees68]).

2.7.3 Proof of the Cramer-Rao Lower Bound

Let $\hat{x}(z)$ be an unbiased estimate of the nonrandom real-valued parameter x based on the observation (or set of observations) denoted now as z. The likelihood function of x is

$$\Lambda(x) = p(z|x) \tag{2.7.3-1}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

It will be assumed that the first and second derivatives of (2.7.3-1) with respect to x exist and are absolutely integrable.

From the unbiasedness condition on the estimate $\hat{x}(z)$, one has (the true value is denoted now also as x)

$$E[\hat{x}(z) - x] = \int_{-\infty}^{\infty} [\hat{x}(z) - x] p(z|x)\, dz = 0 \tag{2.7.3-2}$$

The derivative of the above with respect to x is

$$\begin{aligned} \frac{d}{dx} \int_{-\infty}^{\infty} [\hat{x}(z) - x] p(z|x)\, dz &= \int_{-\infty}^{\infty} \frac{\partial}{\partial x} \Big\{ [\hat{x}(z) - x] p(z|x) \Big\}\, dz \\ &= -\int_{-\infty}^{\infty} p(z|x)\, dz + \int_{-\infty}^{\infty} [\hat{x}(z) - x] \frac{\partial p(z|x)}{\partial x}\, dz \\ &= 0 \end{aligned} \tag{2.7.3-3}$$

Using the fact that the first integral in the last line above is equal to unity and the identity

$$\frac{\partial p(z|x)}{\partial x} = \frac{\partial \ln p(z|x)}{\partial x} p(z|x) \tag{2.7.3-4}$$

yields from (2.7.3-3)

$$\int_{-\infty}^{\infty} [\hat{x}(z) - x] \frac{\partial \ln p(z|x)}{\partial x} p(z|x)\, dz = 1 \tag{2.7.3-5}$$

Equation (2.7.3-5) can be rewritten as

$$\int_{-\infty}^{\infty} \Big\{ [\hat{x}(z) - x] \sqrt{p(z|x)} \Big\} \left\{ \frac{\partial \ln p(z|x)}{\partial x} \sqrt{p(z|x)} \right\} dz = 1 \tag{2.7.3-6}$$

The ***Schwarz inequality*** for real-valued functions, which is a generalized version of (1.3.4-2), is

$$|\langle f_1, f_2 \rangle| \triangleq \int_{-\infty}^{\infty} f_1(z) f_2(z)\, dz \leq \|f_1\|\, \|f_2\| \tag{2.7.3-7}$$

where

$$\|f_i\| \triangleq \{\langle f_i, f_i \rangle\}^{1/2} = \left\{ \int_{-\infty}^{\infty} f_i(z)^2\, dz \right\}^{1/2} \tag{2.7.3-8}$$

The equality in (2.7.3-7) holds if and only if

$$f_1(z) = c f_2(z) \qquad \forall z \tag{2.7.3-9}$$

Note that the left-hand side of (2.7.3-6) is an inner product of two functions as in (2.7.3-7). Using (2.7.3-7) to majorize the left-hand side of (2.7.3-6) yields

$$\left\{ \int_{-\infty}^{\infty} [\hat{x}(z) - x]^2 p(z|x)\, dz \right\}^{1/2} \left\{ \int_{-\infty}^{\infty} \left[\frac{\partial \ln p(z|x)}{\partial x} \right]^2 p(z|x)\, dz \right\}^{1/2} \geq 1 \tag{2.7.3-10}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

which can be rewritten as

$$E\{[\hat{x}(z) - x]^2\} \geq \left\{ E\left[\frac{\partial \ln p(z|x)}{\partial x} \right]^2 \right\}^{-1} \tag{2.7.3-11}$$

with equality holding if and only if

$$\frac{\partial \ln p(z|x)}{\partial x} = c(x)[\hat{x}(z) - x] \qquad \forall z \tag{2.7.3-12}$$

Equation (2.7.3-11) is equivalent to (2.7.2-1), which completes the proof of the CRLB for a nonrandom scalar parameter. In view of (2.7.3-2), which holds at the true value of the parameter, all the partial derivatives are to be evaluated at the true value of the parameter, which is indicated explicitly only in (2.7.2-2).

Equivalence of the Two Forms of the Fisher Information

To prove the equivalence of the two forms of the Fisher information in (2.7.2-2), consider the identity

$$\int_{-\infty}^{\infty} p(z|x)\, dz = 1 \tag{2.7.3-13}$$

Taking the derivative of the above with respect to x yields

$$\int_{-\infty}^{\infty} \frac{\partial p(z|x)}{\partial x}\, dz = 0 \tag{2.7.3-14}$$

Using identity (2.7.3-4), the above can be rewritten as

$$\int_{-\infty}^{\infty} \frac{\partial \ln p(z|x)}{\partial x} p(z|x)\, dz = 0 \tag{2.7.3-15}$$

Taking now the derivative of (2.7.3-15) with respect to x leads to

$$\int_{-\infty}^{\infty} \frac{\partial^2 \ln p(z|x)}{\partial x^2} p(z|x)\, dz + \int_{-\infty}^{\infty} \left[\frac{\partial \ln p(z|x)}{\partial x} \right]^2 p(z|x)\, dz = 0 \tag{2.7.3-16}$$

which proves the equivalence of the expression of the Fisher information with the second partial derivative of the log-likelihood function with the one that has the square of the first partial derivative, as in (2.7.2-2).

2.7.4 An Example of Efficient Estimator

Consider the likelihood function (2.4.2-4) for the estimation of the mean x from a set of k independent and identically distributed measurements with Gaussian noises.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The Fisher information in this case is the scalar quantity

$$J = -E\left[\frac{\partial^2 \ln \Lambda_k(x)}{\partial x^2}\right]\Bigg|_{x=x_0} = \frac{k}{\sigma^2} \tag{2.7.4-1}$$

Thus

$$E\left[[\hat{x}^{\mathrm{ML}}(k) - x_0]^2\right] \geq J^{-1} = \frac{\sigma^2}{k} \tag{2.7.4-2}$$

Comparing the above to (2.6.3-1), it is seen that the CRLB is met; that is, the ML estimator (which, in this case, is the sample mean) is efficient. This is because condition (2.7.3-12) is satisfied.

Since the variance (2.7.4-2) converges to zero as $k \to \infty$, this estimator is also consistent. (See also problem 2-2.)

Evaluation of the CRLB

In the simple case considered above, the expression of J is independent of x. In general this is not true and there is need to evaluate J at the *true value* of x.

If the true value of the parameter is not available, then the ***evaluation of the CRLB***, which amounts to a linearization, is done *at the estimate*. Caution has to be exercised in this case, since the unavoidable estimation errors can lead to a *possibly incorrect value* of the resulting Fisher information J, which, in general, is a matrix.

2.7.5 Large Sample Properties of the ML Estimator

The following are the ***large-sample properties of the ML estimator***:

1. It is ***asymptotically unbiased.***
2. It is ***asymptotically efficient.***

Thus, if there is "enough information" in the measurements, in which case the CRLB will tend to zero, the variance of the ML estimate will also converge to zero. Therefore, the ML estimate will converge to the true value — it will be consistent.

Another property of the ML estimator is the following:

3. It is ***asymptotically Gaussian.***

Combining all the above, *the ML estimate is asymptotically Gaussian with the mean equal to the true value of the parameter to be estimated and variance given by the CRLB.*

This can be summarized for a vector parameter as

$$\hat{x}^{\mathrm{ML}}(k) \sim \mathcal{N}(x, J^{-1}) \qquad \text{for large } k \tag{2.7.5-1}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where J is the Fisher information matrix.

Comparing the (non-Bayesian) MLE (2.7.5-1) with the (Bayesian) MMSE estimate, the conditional mean

$$\hat{x}^{\text{MMSE}} = E[x|z] \tag{2.7.5-2}$$

points out the contrast between these two philosophies:

1. In (2.7.5-1), given x, the estimate $\hat{x}^{\text{ML}}$ is a random variable, function of z;
2. In (2.7.5-2), given z, the true value x is a random variable.

2.8 SUMMARY

2.8.1 Summary of Estimators

Estimator of a parameter — a function of the measurements that yields a "best approximation" for the value of a parameter.

Estimate of a parameter — the value taken by the estimator for the given values (realizations) of the measurements.

Models for the parameter to be estimated:

1. *Unknown constant (nonrandom).*
2. *Random*: a (single) realization of a random variable according to a certain prior pdf.

Model 2 yields the ***Bayesian approach***, whereas model 1 leads to what is called the ***non-Bayesian approach***.

Likelihood function of a (nonrandom) parameter — pdf of the measurements conditioned on the parameter.

Bayes' formula — given a prior pdf of a (random) parameter, this formula yields its posterior pdf conditioned on the measurements.

ML estimate (of a nonrandom parameter) — the value of the parameter that maximizes its likelihood function.

MAP estimate (of a random parameter) — the value of the parameter that maximizes its posterior pdf.

The MAP estimate of a parameter with a *diffuse (noninformative)* prior pdf coincides with its MLE.

LS estimate (of a nonrandom parameter) — minimizes the sum of the squares of the errors between the measurements and the observed function of the parameter.

MMSE estimate (of a random parameter) — minimizes the expected value (mean) of the square of the parameter estimation error conditioned on the measurements. This estimate is the *conditional mean* of the parameter given the measurements.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

If in a given set of measurements the errors are additive, zero mean, *Gaussian*, and independent, then the *LS* estimate coincides with the *ML* estimate.

The *MAP* estimate of a *Gaussian* random variable coincides with its *MMSE* estimate (conditional mean).

2.8.2 Summary of Estimator Properties

Unbiased estimator — if the mean of the corresponding error is zero.

Variance/MSE of an estimator — the expected value of the square of the estimation error of an unbiased/biased estimator. The variance of the estimator of a parameter modeled as random (with some prior) is *smaller* than when it is modeled as an unknown constant.

Consistent estimator — if the corresponding error converges to zero in some stochastic sense (most common: in mean square).

CRLB — lower bound on the achievable variance in the estimation of a parameter. For an unbiased estimator, it is given by the inverse of the ***Fisher information matrix (FIM)***.

FIM — quantifies the existing total information about the parameter of interest in the observations.

Efficient estimator — if its variance meets the CRLB, that is, if all the existing information has been extracted.

On the Terminology

In (most of) the literature there is little or no distinction between the terms LS and MMSE estimation. The MMSE estimation is sometimes called LS, which is incorrect according to our definition, or ***least mean square (LMS)***, which is a valid alternate designation. Another term used is ***minimum variance (MV)***.

2.9 NOTES AND PROBLEMS

2.9.1 Bibliographical Notes

The basic concepts in estimation are discussed, for example, in [Van Trees68, Sage71, Melsa78]. The proof of the CRLB for vector-valued parameters can be found in [Van Trees68] and [Ljung87, p. 206].

Another model of uncertainty in parameter estimation is the "unknown but bounded" approach discussed in [Schweppe73].

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.9.2 Problems

2-1 Estimators for a discrete-valued parameter. A discrete-valued parameter with the prior pdf

$$p(x) = \sum_{i=1}^{2} p_i \delta(x - i)$$

is measured with the additive noise $w \sim \mathcal{N}(0, \sigma^2)$

$$z = x + w$$

1. Find the posterior pdf of the parameter.
2. Find its MAP estimate and the associated MSE conditioned on z.
3. Find its MMSE estimate and the associated variance.
4. Evaluate these estimates and MSE for

Case	p_1	σ	z
A	0.5	1	1.5
B	0.5	1	3
C	0.3	1	1.5
D	0.5	0.1	1.8

5. Comment on the meaningfulness of the two estimates in the above four cases.

2-2 ML Estimation with correlated noises. A parameter x is measured with correlated rather than independent additive Gaussian noises

$$z_k = x + w_k \qquad k = 1, \ldots, n$$

with

$$E[w_k] = 0 \qquad E[w_k w_j] = \begin{cases} 1 & k = j \\ \rho & |k - j| = 1 \\ 0 & |k - j| > 1 \end{cases}$$

For $n = 2$:

1. Write the likelihood function of the parameter x.
2. Find the MLE of x. What happens if $\rho = 1$? What happens if $\rho = -1$?
3. Find the CRLB for the estimation of x. Show the effect of $\rho > 0$ versus $\rho < 0$. Explain what happens at $\rho = -1$.
4. Is the MLE efficient? Can one have a perfect (zero-variance) estimate?

(The remaining items are more challenging.) For general n, let

$$z \triangleq [z_1 \ \cdots \ z_n]' \qquad \mathbf{1} \triangleq [1 \ \cdots \ 1]' \qquad w \triangleq [w_1 \ \cdots \ w_n]' \qquad P \triangleq E[ww']$$

5. Using the above notations, write the likelihood function of x.
6. Find the MLE of x.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2-3 Estimation criteria that lead to the conditional mean. Show that, in estimating a random vector x with the following criteria

1. $\min_{\hat{x}} E[(x-\hat{x})'A(x-\hat{x})|z], \quad \forall A > 0$ (positive definite)
2. $\min_{\hat{x}} \text{tr}[P]$ with $P \triangleq E[(x-\hat{x})(x-\hat{x})'|z]$
3. $\min_{\hat{x}} \text{tr}[AP]$ with A and P as above

all yield the same result $\hat{x} = E[x|z]$.

2-4 Estimate of the variance with the smallest MSE. Consider the problem of estimating the mean and the variance of a set of independent and identically distributed Gaussian random variables $z(j)$, $j = 1, \ldots, k$, as in Subsection 2.5.3, with the true mean x_0 and true variance σ_0^2.

1. Show that the value of n in

$$[\hat{\sigma}(k,n)]^2 = \frac{1}{n}\sum_{j=1}^{k}\left[z(j) - \frac{1}{k}\sum_{i=1}^{k} z(i)\right]^2$$

that minimizes the MSE of the above (defined according to (2.6.1-4) with respect to σ_0^2) is $n = k+1$.

2. Can a biased estimator have a smaller MSE than an unbiased one?

2-5 MAP estimate with two-sided exponential (Laplacian) prior pdf. Consider the same problem as in Subsection 2.3.2 but with a two-sided exponential prior

$$p(x) = \frac{a}{2}\, e^{-a|x|}$$

1. Write the posterior pdf of x.
2. Find $\hat{x}^{\text{MAP}}$.

2-6 Two-sided exponential prior made diffuse.

1. Specify the limiting process that will make the prior from problem 2-5 into a diffuse one.
2. Show that the resulting MAP estimate coincides with the MLE.

2-7 Minimum magnitude error estimate. Given $p(x,z)$, show that the Bayesian estimation that minimizes the expected value of the cost function

$$C(x-\hat{x}) \triangleq |x-\hat{x}|$$

yields $\hat{x} = x_m$, the ***median*** of x, defined as

$$\int_{-\infty}^{x_m} p(x|z)\, dx = \frac{1}{2}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2-8 MAP with Gaussian prior — vector version. Given $z = x + w$, where all the variables are n-vectors, with

$$w \sim \mathcal{N}(0, P) \qquad x \sim \mathcal{N}(\bar{x}, P_0)$$

and x independent of w. Find the MAP estimator of x in terms of z and the covariance of this estimator.

2-9 Conditional variance versus unconditional variance. Let

$$\bar{x} \triangleq E[x] \qquad \text{var}(x) \triangleq E[(x - \bar{x})^2]$$

$$\hat{x} \triangleq E[x|z] \qquad \text{var}(x|z) \triangleq E[(x - \hat{x})^2|z]$$

Prove that

$$\text{var}(x) \geq E[\text{var}(x|z)]$$

2-10 MMSE with exponential prior. Given the prior pdf of x as $p(x) = e^{-x}$, $x \geq 0$ and the observation $z = x + w$ where $w \sim \mathcal{N}(0, 1)$ and independent of x, find the following:

1. $p(z|x)$
2. $p(x|z)$
3. $E[x|z]$
4. $\text{var}[x|z]$

2-11 Altitude estimation from slant range. A sensor is located at $(0, 0)$. It is desired to estimate the height (altitude) y of a point (an aircraft) located at (d, y), where d (the horizontal range) is known, based on the "slant range" measurement $z = r + w$ where $r = h(y) = \sqrt{d^2 + y^2}$ and $w \sim \mathcal{N}(0, \sigma^2)$.

1. Write the likelihood function of y.
2. Find the CRLB for estimating y.
3. Evaluate the standard deviation of the estimate according to the CRLB for $d = 10^5$, $\sigma = 10^2$, and assumed true value $y = 10^3$. How useful would such an estimate be?
4. Find the expression of the MLE of y in terms of z and d.

2-12 Superefficiency? A scalar parameter is estimated in N Monte Carlo runs. The CRLB for this problem is $\sigma^2_{CRLB} = 10$. The sample variance obtained from $N = 100$ runs is $\hat{\sigma}^2$=7.811. Your best friend is concerned and tells you that you must have made a mistake somewhere. What is your answer to this? Give a quantitative justification for it — you can make any reasonable assumptions.

2-13 MLE from correlated measurements. Given the three estimates of the scalar x

$$z_i \triangleq \hat{x}_i = x + \tilde{x}_i \qquad i = 1, 2, 3$$

with the estimation errors $\tilde{x}_i$ jointly Gaussian, zero-mean, with

$$E[\tilde{x}_i \tilde{x}_j] = P_{ij} \qquad i, j = 1, 2, 3$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

find

1. The MLE $\hat{x}(z_1, z_2, z_3)$
2. The variance σ^2 of the above MLE

2-14 Measurement error variance. We have a measuring device for which we are to ascertain that its measurement error variance is less than the borderline acceptable value of 100. You can assume that the errors are zero-mean.

We carry out N independent trials and the result is that the estimate of the variance is $\hat{\sigma}^2 = 80$.

For what N are you willing to risk your job by saying that there is a 5% (or less) probability to get such an estimate while the true value is $\sigma^2 = 100$ (or higher)?

Hint: You can assume a Gaussian distribution for the error in the estimate of the variance and use a probability region about the borderline value.

2-15 Estimation with correlated measurements. Given the scalar random variables x_i, $i = 1, \ldots, N+1$, with $Ex_i = \mu$ and $\text{cov}[x_i, x_j] = \sigma^2 \rho^{|i-j|}$, where $|\rho < 1$.

1. Find the mean and variance of the following estimate of μ

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

You can assume $N \gg 1$.

2. Is this a consistent estimator? Justify the answer.

3. To how many i.i.d. random variables with the same first two moments are the above equivalent (i.e., they yield the same variance) for $\rho = 0.5$?

2-16 Public opinion polls.

1. Find the margin of error of a public opinion poll with $N = 625$ subjects.
2. How many subjects are needed for a margin of error of 1%?

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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)

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Chapter 11

ADAPTIVE ESTIMATION AND MANEUVERING TARGETS

11.1 INTRODUCTION

11.1.1 Adaptive Estimation — Outline

In the models considered previously, the only uncertainties consisted of additive white noises (process and measurement) with known statistical properties. In other words, the system model, consisting of the state transition matrix, the input gain (and input, if any), the measurement matrix, and noise covariances, were all assumed to be known. In the case of "colored" noise with known autocorrelation, it has been shown that it can be reduced to the standard situation in Chapter 8.

In many practical situations the above listed "parameters of the problem" are partially unknown and possibly time-varying. There are numerous techniques of ***system identification*** (e.g., [Ljung87]) that deal with the identification (estimation) of the *structural parameters* of a system.

The purpose of this chapter is to present state estimation techniques that can "adapt" themselves to certain types of uncertainties beyond those treated earlier — ***adaptive estimation algorithms***.

One type of uncertainty to be considered is the case of unknown inputs into the system, which typifies maneuvering targets. The other type will be a combination of system parameter uncertainties with unknown inputs where the system parameters (are assumed to) take values in a discrete set.

Maneuvering targets are characterized by an equation of the same form as (4.3.1-9), namely,

$$x(k+1) = F(k)x(k) + G(k)u(k) + v(k) \qquad \text{(11.1.1-1)}$$

but the input $u(k)$, which enters the system in addition to the process noise $v(k)$, is *unknown*.

For simplicity, linear models are considered — in the case of nonlinear models the same techniques that are discussed in the sequel can be used with linearization.

The approaches that can be used in such a situation fall into two broad categories:

1. The unknown input (the maneuver command) is modeled as a random process.
2. The unknown input is estimated in real time.

The random process type models can be classified according to the statistical properties of the process modeling the maneuver as

- White noise or
- Autocorrelated noise

Note that this ***unknown input as a random process*** approach amounts to treating the unknown input as (an additional) process noise, and, possibly, requiring augmentation of the system's state.

The use of fixed-level white noise to model maneuvers has been discussed in the context of kinematic models. However, maneuvers, by their nature, are of different magnitudes at different times. One way to accommodate this is by adjusting the level of the process noise. One option is a ***continuous noise level adjustment*** technique, discussed in Section 11.2. Alternatively, several discrete levels of noise can be assumed in the filter, with a ***noise level switching*** according to a certain rule, also discussed in Section 11.2.

The use of autocorrelated (colored) noise to model maneuvers has been presented in Section 8.2 — the approach is to "prewhiten" the noise and augment the state with the prewhitening subsystem. The noise level adjustment techniques can be used with the augmented system, which is driven by white noise.

All these methods relying on *modeling maneuvers as random processes* are approximations because maneuvers are, in general, not stochastic processes. Nevertheless, such approaches are simple and can be quite effective.

The second type of approach, ***input estimation***, is implemented assuming the input to be constant over a certain period of time. The estimation can be done based on the least squares criterion, and the result can be used in the following ways:

- The state estimate is corrected with the estimated input, or
- The state is augmented and the input becomes an extra state component that is reestimated sequentially within the augmented state.

The input estimation with state estimate correction technique is presented in Section 11.3. The technique of estimating the input and, when "statistically significant," augmenting the state with it (which leads to ***variable state dimension***), is the topic of Section 11.4. These two algorithms and the noise level switching technique of Section 11.2 are compared in Section 11.5.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The so-called ***multiple model (MM)*** algorithms are the topic of Section 11.6. These algorithms assume that the system behaves according to one of a finite number of models — it is in one of several ***modes*** (operating regimes). The models can differ in *noise levels* or their *structure* — different state dimensions and unknown inputs can be accommodated as well. Such systems are called ***hybrid systems*** — they have both discrete (structure/parameters) and continuous uncertainties (additive noises).

First the *static* MM algorithm — static from the point of view of the assumed evolution of the models, i.e., for fixed (nonswitching) models — is considered. Then the optimal *dynamic* algorithm — for switching models, according to a Markov chain — is presented. Since the optimal dynamic algorithm is not practical for implementation, two suboptimal approaches, one called ***generalized pseudo-Bayesian (GPB)*** and the other the ***interacting multiple model (IMM)***, are also presented.

The design of an IMM estimator for ***air traffic control (ATC)*** is discussed in detail. Guidelines are also developed for when an adaptive estimator is really needed, i.e., when a (single model based) Kalman filter is not adequate.

The use of the extended Kalman filter for state and system parameter estimation is briefly presented in Section 11.9.

11.1.2 Adaptive Estimation — Summary of Objectives

Show modeling of maneuvers as

- A random process (white or autocorrelated process noise)
  - continuously variable level;
  - several discrete levels with switching between them.
- A fixed input, estimated as an unknown constant, with
  - state estimate compensation;
  - variable state dimension.

Compare some of these algorithms.
Illustrate the optimal method of comparison of algorithms.
Multiple-model (hybrid system) approach:

Several models are used in parallel with a probabilistic weighting:

- Static — for fixed (nonswitching) models
- Dynamic — for switching models
  - Optimal
  - GPB1
  - GPB2
  - IMM

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Present the design of an IMM estimator for air traffic control.
Answer the question: When is an adaptive estimator really needed?
Show the use of the EKF for state and system parameter estimation.

11.2 ADJUSTABLE LEVEL PROCESS NOISE

11.2.1 Continuous Noise Level Adjustment

In this ***continuous noise level adjustment*** approach, the target is tracked with a filter in which a certain low level of process noise is assumed. The level of the noise is determined by its variance (or covariance matrix).

A maneuver manifests itself as a "large" innovation. A simple detection procedure for such an occurrence is based on the ***normalized innovation squared***

$$\epsilon_\nu(k) = \nu(k)'S(k)^{-1}\nu(k) \tag{11.2.1-1}$$

A threshold is set up such that based on the target model (for the nonmaneuvering situation)

$$P\{\epsilon_\nu(k) < \epsilon_{max}\} = 1 - \mu \tag{11.2.1-2}$$

with, say, tail probability $\mu = 0.01$.

If the threshold is exceeded, then the process noise variance $Q(k-1)$ is scaled up until ϵ_ν is reduced to the threshold ϵ_{max}.

Using a ***scaling factor*** $\phi(k)$ (i.e., a ***fudge factor***), the covariance of the innovations becomes

$$S(k) = H(k)[F(k-1)P(k-1|k-1)F(k-1)' + \phi(k)Q(k-1)]H(k)' + R(k) \tag{11.2.1-3}$$

In place of the single-time test statistic (11.2.1-1) one can also use a time average over a "sliding window." This is illustrated in the next subsection.

A similar technique can be used to lower the process noise level after the maneuver.

11.2.2 Process Noise with Several Discrete Levels

Another approach is to assume two or more levels of process noise and use a rule for ***noise level switching***.

Under normal conditions the filter is operating with the low level noise Q_1.

The normalized innovation squared

$$\epsilon_\nu(k) = \nu(k)'S(k)^{-1}\nu(k) \tag{11.2.2-1}$$

is monitored; and if it *exceeds* a certain threshold, the filter switches to a prechosen higher level of process noise, Q_2.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Under the linear-Gaussian assumptions, the pdf of ϵ_ν is chi-square distributed with n_z, the dimension of the measurement, degrees of freedom:

$$\epsilon_\nu \sim \chi^2_{n_z} \tag{11.2.2-2}$$

The switching threshold is chosen based on (11.2.2-2) such that the probability of being exceeded under normal conditions is small.

The decision statistic (11.2.2-1), which is based on a single sampling time, can be replaced by a ***moving average*** (or ***moving sum***) of the normalized innovations squared over a ***sliding window*** of s sampling times

$$\boxed{\epsilon^s_\nu(k) = \sum_{j=k-s+1}^{k} \epsilon_\nu(j)} \tag{11.2.2-3}$$

The above is chi-square distributed with sn_z degrees of freedom

$$\epsilon^s_\nu \sim \chi^2_{sn_z} \tag{11.2.2-4}$$

since (11.2.2-3) is the sum of s independent terms with distribution (11.2.2-2).

Alternatively, a ***fading memory average*** (also called ***exponentially discounted average***)

$$\boxed{\epsilon^\alpha_\nu(k) = \alpha\epsilon^\alpha_\nu(k-1) + \epsilon_\nu(k)} \tag{11.2.2-5}$$

where

$$0 < \alpha < 1 \tag{11.2.2-6}$$

and with initial condition $\epsilon^\alpha_\nu(0) = 0$, can be used.

The expected value of (11.2.2-5) in steady state is

$$E[\epsilon^\alpha_\nu(k)] = \frac{n_z}{1-\alpha} \tag{11.2.2-7}$$

The ***effective window length*** of (11.2.2-5) can be considered as the sum of the weights multiplying $\epsilon^\alpha_\nu(k)$, $k = 1, 2, \ldots$, i.e.,

$$s_\alpha = 1 + \alpha + \alpha^2 + \ldots = \frac{1}{1-\alpha} \tag{11.2.2-8}$$

For example, for $\alpha = 0.8$ one has $s_\alpha = 5$.

Based on (11.2.2-7), one can assume, as a first approximation (matching the first moment only), that $\epsilon^\alpha_\nu(k)$ is chi-square distributed with number of degrees of freedom given by (11.2.2-7).

Using the first *and* second moment-matching approximation described in Subsection 1.4.18, one obtains

$$\epsilon^\alpha_\nu \sim \frac{1}{1+\alpha}\chi^2_{n_\alpha} \tag{11.2.2-9}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where the number of degrees of freedom is

$$n_\alpha = n_z \frac{1+\alpha}{1-\alpha} \tag{11.2.2-10}$$

For example, for $\alpha = 0.8$ one obtains $n_\alpha = 18$. Using the first approximation for the same α with $n_z = 2$ leads to 10 degrees of freedom. In other words, the moment-matching approximation (with the first two moments) indicates a (relatively) narrower pdf. The "width" of a chi-square pdf can be taken as the ratio of its standard deviation to its mean. For χ^2_n, this is (see Subsection 1.4.17) $\sqrt{2n}/n$, which decreases as n increases.

Estimator Operation

Thus, the filter switches from the lower process noise covariance Q_1 to the higher Q_2 if the average (11.2.2-3) or (11.2.2-5) exceeds an ***upcrossing threshold***, determined based on (11.2.2-4) or (11.2.2-9), respectively, with a small tail probability.

After the filter starts running with the higher-level process noise the innovations are monitored again to see whether there is reason to switch back.

The change to the model with lower-level process noise is done when the normalized innovation, or a certain average of it, falls *below* another threshold — the ***downcrossing threshold***.

There is no exact way to choose these thresholds — even if the distributions were known exactly, the tail probabilities are still subjective. They can be chosen, to begin with, based on tail probabilities and then set following experimentation (Monte Carlo runs) and subjective evaluation of the results.

Remarks

This technique of noise level switching can be extended to more than two levels of process noise. It can also be used with white process noise or autocorrelated process noise; the latter has to be prewhitened, as discussed in Subsection 8.2.1.

In general, n_α, given in (11.2.2-10), is not an integer. In this case one has a gamma rather than chi-square distribution; for the sake of simplicity, however, one can use the chi-square tables with an interpolation.

11.2.3 Adjustable Level Process Noise — Summary

A maneuver can be detected by monitoring the *normalized innovation*; if it exceeds a threshold, then it can be assumed (in the absence of other factors) that the target has deviated from its previous "pattern."

A *fudge factor* can be used to scale up (continuously) the process noise (at the previous time) such that the modified prediction covariance is sufficiently

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

large for the normalized innovation to be below the set threshold; the increased innovation covariance should "cover" the observed deviation.

Alternatively, a number of levels of process noise can be assumed. In this approach, at any given time a *single filter* operates with the *"current" noise level.*

A simple rule of switching from the current noise level to the one above or below is followed.

The innovations are monitored via

- a moving average over a sliding window or
- a fading-memory average

of their normalized squared value.

This average is compared to a threshold based on the chi-square density for a small upper tail probability. If this *upcrossing threshold* is exceeded, then the filter switches to a higher level of process noise.

The downswitch is done using a *downcrossing threshold.*

11.3 INPUT ESTIMATION

11.3.1 The Model

Consider the system with state equation

$$x(k+1) = Fx(k) + Gu(k) + v(k) \tag{11.3.1-1}$$

where u is an ***unknown input*** modeling the target maneuvers and v is the process noise, zero mean white with known covariance Q.

The observations are

$$z(k) = Hx(k) + w(k) \tag{11.3.1-2}$$

with the observation noise w zero mean, white, with covariance R, and independent of the process noise.

The estimation of the state is done using the model without input (nonmaneuvering)

$$x(k+1) = Fx(k) + v(k) \tag{11.3.1-3}$$

Two Kalman filters are considered:

1. The *actual one* based on the nonmaneuvering model (11.3.1-3).
2. A *hypothetical one* based on the maneuvering model (11.3.1-1) with known u.

From the innovations of the ***nonmaneuvering filter*** based on (11.3.1-3), the input u is to be

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

- detected,
- estimated, and
- used to correct the state estimate.

This will be done using a sliding window of the latest s measurements, and during this window period the input will be assumed *constant*.

11.3.2 The Innovations as a Linear Measurement of the Unknown Input

Denote the present time by k and assume that the target starts maneuvering at time $k-s$; that is, the ***maneuver onset time*** is $k-s$. The unknown inputs during the interval $[k-s,\ldots,k]$ are $u(i)$, $i=k-s,\ldots,k-1$.

The state estimates from the *mismatched nonmaneuvering filter* based on (11.3.1-3) will be denoted by an asterisk. The recursion for these estimates is

$$\begin{aligned}\hat{x}^*(i+1|i) &= F[I-W(i)H]\hat{x}^*(i|i-1)+FW(i)z(i)\\ &\triangleq \Phi(i)\hat{x}^*(i|i-1)+FW(i)z(i)\\ &\qquad\qquad i=k-s,\ldots,k-1 \qquad (11.3.2\text{-}1)\end{aligned}$$

with the initial condition

$$\hat{x}^*(k-s|k-s-1)=\hat{x}(k-s|k-s-1) \qquad (11.3.2\text{-}2)$$

being the correct estimate (one-step prediction) before the maneuver started. This follows from the assumption that the maneuver onset time is $k-s$.

Similarly to (4.3.3-1), recursion (11.3.2-1) yields, in terms of (11.3.2-2),

$$\begin{aligned}\hat{x}^*(i+1|i) &= \Big[\prod_{j=0}^{i-k+s}\Phi(i-j)\Big]\hat{x}(k-s|k-s-1)\\ &+\sum_{j=k-s}^{i}\Big[\prod_{m=0}^{i-j-1}\Phi(i-m)\Big]FW(j)z(j) \quad i=k-s,\ldots,k-1 \qquad (11.3.2\text{-}3)\end{aligned}$$

If the inputs were known, the *hypothetical correct filter* based on (11.3.1-1) would yield estimates according to the recursion

$$\begin{aligned}\hat{x}(i+1|i) &= \Phi(i)\hat{x}(i|i-1)+FW(i)z(i)+Gu(i)\\ &= \Big[\prod_{j=0}^{i-k+s}\Phi(i-j)\Big]\hat{x}(k-s|k-s-1)\\ &\quad+\sum_{j=k-s}^{i}\Big[\prod_{m=0}^{i-j-1}\Phi(i-m)\Big][FW(j)z(j)+Gu(j)]\\ &\qquad\qquad i=k-s,\ldots,k-1 \qquad (11.3.2\text{-}4)\end{aligned}$$

which is the same as (11.3.2-3) except for the last term containing the inputs.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The innovations

$$\nu(i+1) = z(i+1) - H\hat{x}(i+1|i) \tag{11.3.2-5}$$

corresponding to the correct (but *hypothetical*) filter (11.3.2-4) are a zero-mean white sequence with covariance $S(i+1)$ given as in (5.2.3-9).

The innovations corresponding to the *nonmaneuvering filter* (11.3.2-3) are

$$\nu^*(i+1) = z(i+1) - H\hat{x}^*(i+1|i) \tag{11.3.2-6}$$

From (11.3.2-3) and (11.3.2-4) it follows that the innovations (11.3.2-6) of the nonmaneuvering filter — the *only real filter* — are the same as the "white noise sequence" (11.3.2-5) plus a "bias term" related to the inputs

$$\nu^*(i+1) = \nu(i+1) + H \sum_{j=k-s}^{i} \Big[\prod_{m=0}^{i-j-1} \Phi(i-m) \Big] G u(j) \tag{11.3.2-7}$$

Assuming the input to be constant over the interval $[k-s, \ldots, k-1]$, that is,

$$u(j) = u \qquad j = k-s, \ldots, k-1 \tag{11.3.2-8}$$

yields

$$\boxed{\nu^*(i+1) = \Psi(i+1)u + \nu(i+1) \qquad i = k-s, \ldots, k-1} \tag{11.3.2-9}$$

where

$$\Psi(i+1) \triangleq H \sum_{j-k-s}^{i} \Big[\prod_{m=0}^{i-j-1} \Phi(i-m) \Big] G \tag{11.3.2-10}$$

Equation (11.3.2-9) shows that

the *nonmaneuvering filter innovation* ν^*

is

a *linear measurement of the input* u

in the presence of the additive "white noise" ν given by (11.3.2-5) above.

11.3.3 Estimation of the Unknown Input

Based on (11.3.2-9), the input can be estimated via LS from

$$y = \Psi u + \epsilon \tag{11.3.3-1}$$

where

$$y = \begin{bmatrix} \nu^*(k) \\ \vdots \\ \nu^*(k-s+1) \end{bmatrix} \tag{11.3.3-2}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

is the stacked "measurement" vector,

$$\Psi = \begin{bmatrix} \Psi(k) \\ \vdots \\ \Psi(k-s+1) \end{bmatrix} \tag{11.3.3-3}$$

is the measurement matrix, and the "noise" ϵ, whose components are the innovations (11.3.2-5), is zero mean with block-diagonal covariance matrix

$$S = \text{diag}[S(i)] \tag{11.3.3-4}$$

The ***input estimate*** in batch form is

$$\boxed{\hat{u} = (\Psi' S^{-1} \Psi)^{-1} \Psi' S^{-1} y} \tag{11.3.3-5}$$

with the resulting covariance matrix

$$L = (\Psi' S^{-1} \Psi)^{-1} \tag{11.3.3-6}$$

An estimate is accepted, that is, a ***maneuver detection*** is declared if and only if (11.3.3-5) is *statistically significant* (see Subsection 1.5.2). The significance test for the estimate $\hat{u}$, which is an n_u-dimensional vector, is

$$\epsilon_{\hat{u}} = \hat{u}' L^{-1} \hat{u} \geq c \tag{11.3.3-7}$$

where c is a threshold chosen as follows. If the input is zero, then

$$\hat{u} \sim \mathcal{N}(0, L) \tag{11.3.3-8}$$

The statistic $\epsilon_{\hat{u}}$ from (11.3.3-7) is then chi-square distributed with n_u degrees of freedom, and c is such that the probability of false alarm is

$$P\{\epsilon_{\hat{u}} \geq c\} = \alpha \tag{11.3.3-9}$$

with, e.g., $\alpha \leq 0.01$.

11.3.4 Correction of the State Estimate

If a maneuver is detected, then the state estimate has to be corrected. This is done, as before, assuming that the maneuver onset time was $k - s$. The term reflecting the effect of the input in (11.3.2-4) is used to correct the predicted state as follows

$$\boxed{\hat{x}^u(k|k-1) = \hat{x}^*(k|k-1) + M(k)\hat{u}} \tag{11.3.4-1}$$

where

$$M(k) \triangleq \sum_{j=k-s}^{k-1} \Big[\prod_{m=0}^{k-j-2} \Phi(k-1-m) \Big] G \tag{11.3.4-2}$$

The covariance associated with the prediction (11.3.4-1) is

$$P^u(k|k-1) = P(k|k-1) + M(k) L M(k)' \tag{11.3.4-3}$$

where L is given in (11.3.3-6). Note that the covariance increases — the last term above is positive semidefinite or positive definite — to account for the uncertainty in the correction term used in (11.3.4-1).

The assumption about the maneuver onset time being $k - s$ *at every* k makes the algorithm very elegant, but not optimal.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Maneuver Onset Time Estimation

It is possible to estimate the onset time using a maximum likelihood criterion — this requires running in parallel a number of such algorithms. Each algorithm evaluates the likelihood function for its assumed maneuver onset time and the most likely effective window length s is chosen.

It is clear that this becomes quite expensive.

11.3.5 Input Estimation — Summary

The input estimation method is based on the following:

Given a linear system driven by a constant input, a filter based on the state model of the system *without input (input equal to zero)* estimates the state.

Then the innovations of this *mismatched filter* are

- a linear measurement of the input
- with additive zero-mean white noise with covariance equal to the filter's innovation covariance.

This allows the estimation of the input, assuming that

- it started at the beginning of the assumed sliding window, and
- it is approximately constant over this window in time.

The estimation is done via least squares.

The estimated input (e.g., acceleration) is tested for *statistical significance* (i.e., if it is "large enough" compared to its standard deviation).

If the estimated input is statistically significant, then the state estimate is corrected with the effect of the input from the time it was assumed to have started (at the beginning of the window).

The covariance of the state is adjusted (increased) because the correction was done with an estimated input, rather than the exact input, which is not available.

11.4 THE VARIABLE STATE DIMENSION APPROACH

11.4.1 The Approach

The target maneuver is considered here an inherent part of its dynamics, rather than noise. In the absence of maneuvers the filter operates using a "quiescent state model." Once a maneuver is detected, new state components are added — thus the name ***variable state dimension (VSD)*** approach.

The extent of the maneuver as detected is then used to yield an estimate for the extra state components, and corrections are made on the other state

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

components. The tracking is then done with the augmented state model until it will be reverted to the quiescent model by another decision.

The rationale for using a lower-order quiescent model and a higher-order maneuvering model is the following: This will allow good tracking performance in both situations rather than a compromise. For example, if the target does not have acceleration, using a third-order model increases the estimation errors for both position and velocity.

The two models used here, described in Section 6.3, are a (nearly) constant velocity model for the quiescent situation and a (nearly) constant acceleration model for the maneuvering situation.

The state vector, for a planar motion, is in the ***quiescent model***

$$x = \begin{bmatrix} \xi & \dot{\xi} & \eta & \dot{\eta} \end{bmatrix}' \tag{11.4.1-1}$$

In the ***maneuvering model***, the state is

$$x^m = \begin{bmatrix} \xi & \dot{\xi} & \eta & \dot{\eta} & \ddot{\xi} & \ddot{\eta} \end{bmatrix}' \tag{11.4.1-2}$$

The method is not limited to these models. For example, the maneuvering situation could be modeled using autocorrelated acceleration, as in Section 8.2.

11.4.2 The Maneuver Detection and Model Switching

The target tracking filter is initialized under the constant velocity model assumption as in Section 5.5, using a two-point differencing procedure.

The maneuver detection is done as follows:

A ***fading memory average*** of the innovations from the estimator based on the quiescent model is computed

$$\epsilon_\nu^\alpha(k) = \alpha \epsilon_\nu^\alpha(k-1) + \epsilon_\nu(k) \tag{11.4.2-1}$$

with

$$\epsilon_\nu(k) \triangleq \nu(k)' S(k)^{-1} \nu(k) \tag{11.4.2-2}$$

where $0 < \alpha < 1$, $\nu(k)$ is the innovation and $S(k)$ its covariance.

Since $\epsilon_\nu(k)$ is, under the Gaussian assumptions, chi-square distributed with n_z (dimension of the measurement) degrees of freedom, one has

$$\lim_{k \to \infty} E[\epsilon_\nu^\alpha(k)] = \frac{n_z}{1-\alpha} \tag{11.4.2-3}$$

As in Section 11.2, one can look at

$$s_\alpha = \frac{1}{1-\alpha} \tag{11.4.2-4}$$

as the ***effective window length*** over which the presence of a maneuver is tested.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The hypothesis that a maneuver is taking place is accepted if $\epsilon_\nu^\alpha(k)$ exceeds a certain threshold that can be determined based on the chi-square distribution as in Section 11.2. Then the estimator switches from the quiescent model to the maneuvering model.

The scheme for reverting to the quiescent model is as follows: The estimated accelerations are compared to their variances, and if they are not statistically significant the maneuver hypothesis is rejected.

Maneuver Termination vs. Maneuver Onset Detection

It is more difficult to detect ***maneuver termination*** than ***maneuver onset*** because a maneuvering model has a larger innovation covariance than a nonmaneuvering model. This is due to the fact that the latter has a larger state vector and assumes more motion uncertainty than the former.

Significance Test for the Acceleration Estimate

The test statistic for significance of the acceleration estimates is

$$\epsilon_{\hat{a}}(k) = \hat{a}(k|k)' P_a^m(k|k)^{-1} \hat{a}(k|k) \tag{11.4.2-5}$$

where $\hat{a}$ is the estimate of the acceleration components and P_a^m is the corresponding block from the covariance matrix of the maneuvering model.

When the ***moving average*** over a window of length p

$$\epsilon_{\hat{a}}^p(k) = \sum_{j=k-p+1}^{k} \epsilon_{\hat{a}}(j) \tag{11.4.2-6}$$

falls below a threshold, the acceleration is deemed *insignificant*. A fading memory average can also be used.

The situation where the acceleration drops suddenly to zero can lead to large innovations for the maneuvering model. This can be taken care of by allowing a switch to the lower-order model also when the maneuvering model's innovation exceeds, say, the 99% confidence region.

11.4.3 Initialization of the Augmented Model

When a maneuver is detected at time k, the filter assumes that the target had a constant acceleration starting at $k - s - 1$, where s is the effective window length. The state estimates for time $k - s$ are then modified as follows. It is assumed that only position measurements are available.

First, the estimates at $k - s$ for the acceleration are

$$\hat{x}_{4+i}^m(k-s|k-s) = \frac{2}{T^2}[z_i(k-s) - \hat{z}_i(k-s|k-s-1)] \qquad i = 1, 2 \tag{11.4.3-1}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The position components of the estimate at $k-s$ are taken as the corresponding measurements, that is,

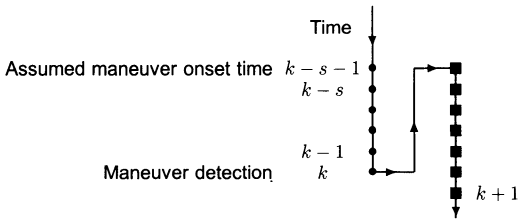
$$\hat{x}^m_{2i-1}(k-s|k-s) = z_i(k-s) \qquad i=1,2 \tag{11.4.3-2}$$

while the velocity components are corrected with the acceleration estimates as follows:

$$\hat{x}^m_{2i}(k-s|k-s) = \hat{x}_{2i}(k-s-1|k-s-1) + T\hat{x}^m_{4+i}(k-s|k-s) \qquad i=1,2 \tag{11.4.3-3}$$

The covariance matrix associated with the modified state estimate as given in (11.4.3-1) to (11.4.3-3) is $P^m(k-s|k-s)$, and its elements are presented in [Bar-Shalom82] together with their derivation. These equations specify in full the initialization of the filter based on the maneuvering model for time $k-s$.

Then, a recursive estimation algorithm (Kalman filter) based on this model reprocesses the measurements sequentially up to time k. Following this, the measurements are processed sequentially as they arrive. This is depicted in Fig. 11.4.3-1.



• Sequential estimation with quiescent model; ■ with maneuvering model

Figure 11.4.3-1: Switching from quiescent to maneuvering model.

11.4.4 VSD Approach — Summary

In the absence of maneuvers, the filter is based on a *quiescent* (low-order) model of the state.

If a maneuver is detected, the estimator switches to a higher dimension (augmented) *maneuvering* model that incorporates additional state components (acceleration).

The use of the lower-order model yields maximum estimation accuracy in the case where the target undergoes no accelerations — no information is "wasted" on estimating state components that are zero.

A maneuver is declared detected when a *fading memory average* of the normalized innovations exceeds a threshold. The fading memory has an *effective*

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

window length, and the onset of the maneuver is then taken as the beginning of this sliding window.

The augmented filter is initialized at the beginning of the maneuver detection window.

A maneuver is declared terminated when the extra state component's (acceleration) estimates become statistically insignificant.

11.5 A COMPARISON OF ADAPTIVE ESTIMATION METHODS FOR MANEUVERING TARGETS

11.5.1 The Problem

In this section the following methods of maneuver detection are illustrated and compared:

1. White process noise with two levels (Section 11.2)
2. Input estimation (Section 11.3)
3. Variable dimension filtering (Section 11.4)

The example considers a target whose position is sampled every $T = 10$ s. The target is moving in a plane with constant course and speed until $k = 40$ when it starts a slow $90°$ turn that is completed in 20 sampling periods. A second fast turn of $90°$ starts at $k = 61$ and is completed in 5 sampling times.

The initial condition of the target, with state

$$x = [\xi \;\; \dot{\xi} \;\; \eta \;\; \dot{\eta}]' \tag{11.5.1-1}$$

is, with position and velocity units m and m/s, respectively

$$x(0) = [2000 \;\; 0 \;\; 10000 \;\; -15]' \tag{11.5.1-2}$$

The slow turn is the result of acceleration inputs

$$u_\xi = u_\eta = 0.075\,\mathrm{m/s^2} \qquad 400\,\mathrm{s} \le t \le 600\,\mathrm{s} \tag{11.5.1-3}$$

and the fast turn has accelerations

$$u^\xi = u^\eta = -0.3\,\mathrm{m/s^2} \qquad 610\,\mathrm{s} \le t \le 660\,\mathrm{s} \tag{11.5.1-4}$$

Note that the changes in velocity are 0.75 and 3 m/s per sampling period for the slow and fast turn, respectively.

The measurements are position only according to the equation

$$z(k) = [1 \;\; 0 \;\; 1 \;\; 0]\, x(k) + w(k) \tag{11.5.1-5}$$

with $w(k)$ zero mean, white, independent of the process noise, and with variance

$$E[w(k)^2] = rI \tag{11.5.1-6}$$

where $r = 10^4\,\mathrm{m^2}$.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

11.5.2 The White Noise Model with Two Levels

The adaptive algorithm based on white process noise with switching between two levels for maneuver modeling used the following state representation per coordinate (same for each coordinate and decoupled between them)

$$x(k+1) = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} x(k) + \begin{bmatrix} T/2 \\ 1 \end{bmatrix} v(k) \tag{11.5.2-1}$$

with $v(k)$ zero mean white with variance

$$E[v(k)^2] = q \tag{11.5.2-2}$$

Note that here $v(k)$ represents the ***velocity increment over a sampling period.***

The two levels of process noise were q_0 for the quiescent period and q_1 after a maneuver was detected.

A maneuver was declared detected if the normalized innovation squared (11.2.2-1) exceeded a threshold ϵ_m. Then q_0 was replaced by q_1 in the filter. A maneuver was considered terminated when the normalized innovation squared, obtained from the filter with the higher q_1, fell below the same threshold ϵ_m — for simplicity, the upcrossing and downcrossing thresholds were taken the same.

The value for the quiescent process noise was $q_0 = 0$. Several values for the variance q_1 of the process noise modeling the maneuver and the threshold ϵ_m were tried. The best combination was determined based on the MSE in the position estimate of the target from 50 Monte Carlo runs. The resulting values were $q_1 = 10$ and $\epsilon_m = 3$.

The RMS position error (coordinate ξ) corresponding to these values is shown in Fig. 11.5.2-1. Note that the value $\sqrt{q_1} \approx 3$ matches approximately the *maximum change in velocity per sampling period* during the fast maneuver, indicated in Subsection 11.5.1.

11.5.3 The IE and VSD Methods

The same problem was simulated using the ***input estimation (IE)*** and ***variable state dimension (VSD)*** methods.

The IE method was run with a window of $s = 5$ samples to estimate the input. The state equation was the constant velocity model (6.3.2-1) for each coordinate and no process noise.

The VSD method was based on a quiescent model with state (11.4.1-1) and a maneuvering model with state (11.4.1-2). The quiescent model was, for each coordinate, the constant velocity state equation (6.3.2-1) with no process noise. The maneuvering model was a nearly constant acceleration state equation, given by (6.3.3-1) with process noise standard deviation σ_v in (6.3.3-4) taken equal to 5% of the estimated acceleration.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

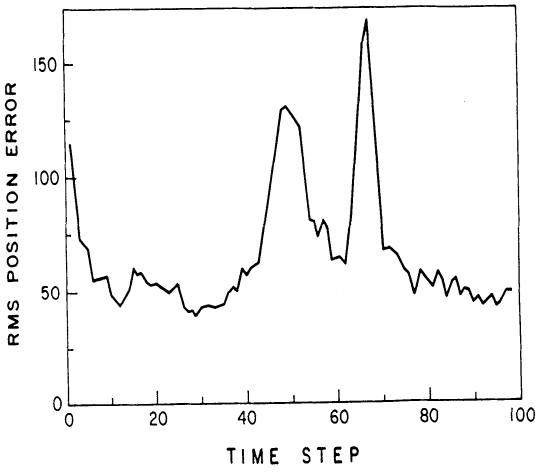


Figure 11.5.2-1: Position estimation error in coordinate ξ for the adaptive filter with two process noise levels (from 50 runs).

The following maneuver detection parameters were used in the VSD method:

1. The fading memory parameter from (11.4.2-1) was $\alpha = 0.8$, which corresponds to an effective window length (11.4.2-4) of $s = 5$.
2. The threshold for the fading memory average (11.4.2-1) was $\chi^2_{10}(95\%) = 18.3$. The maneuver detection was done together in the two coordinates based on both measurements (i.e., $n_z = 2$), which, with (11.4.2-3), yields that ϵ^{α}_{ν} is approximately chi-square distributed with 10 degrees of freedom.
3. The window for the calculation of the average normalized acceleration (11.4.2-6) for the significance test was $p = 2$.
4. The threshold for the acceleration significance test was $\chi^2_4(95\%) = 9.49$, based on a 4-degree-of-freedom chi-square random variable (the dimension of the acceleration, 2, multiplied by $p = 2$).

Figure 11.5.3-1a presents the position RMS errors (in coordinate ξ) for the IE and VSD filters based on a 50-run Monte Carlo average. These two filters were run with the same random numbers for further comparison.

The RMS errors are smaller in the VSD filter compared to the IE filter. Comparing with Fig. 11.5.2-1, it is seen that the simple approach of two-level white process noise modeling of the maneuver is quite effective — it is slightly worse than the VSD filter but better than the IE filter.

Figure 11.5.3-1b presents the velocity RMS errors for the IE and VSD filters based on the same 50 Monte Carlo runs. For velocity, the VSD filter is only slightly superior compared to the IE algorithm.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

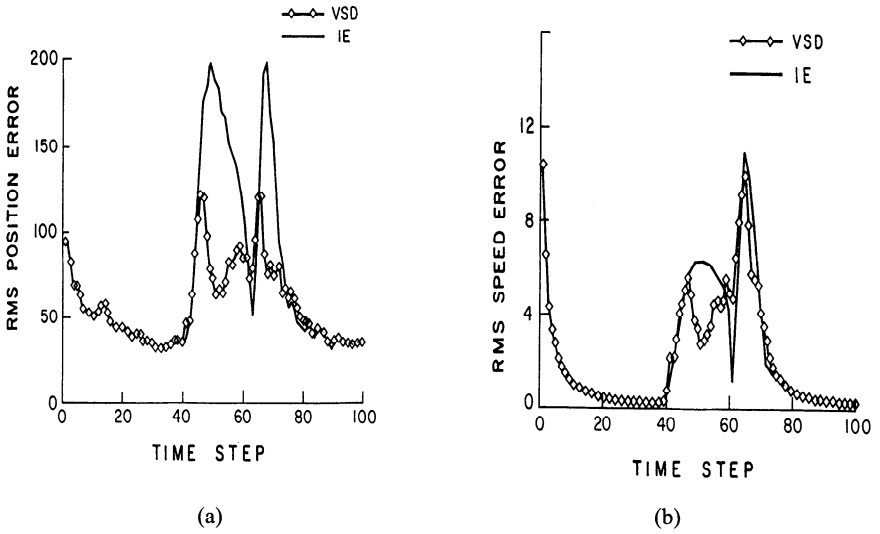


Figure 11.5.3-1: Estimation errors in coordinate ξ for the VSD and IE filters (from 50 runs).

Remark

It should be noted that none of these three adaptive algorithms considered here manage to keep the RMS position error in the estimate below the position measurement noise RMS value (which is, in one coordinate, 100 m) *during the entire maneuver period.* In other words, these algorithms have difficulty providing ***noise reduction*** during the critical time of a maneuver.

This is a general problem with adaptive estimation algorithms: The adaptation might not be rapid enough. In the present example, the errors reach their peak shortly after the start of the maneuver.

This issue will be considered again in a later example.

11.5.4 Statistical Test for Comparison of the IE and VSD Methods

Next, we shall examine more closely the results of the IE and VSD filters and carry out the systematic comparison of algorithms presented in Subsection 1.5.3. The question is whether, in view of the limited sample size, one can state that the VSD algorithm is superior to the IE algorithm.

The performance of interest will be the mean square error in the estimate of one state component, not indicated explicitly, for simplicity

$$\epsilon^{\text{VSD}}(k) \triangleq E[[\tilde{x}^{\text{VSD}}(k|k)]^2] \tag{11.5.4-1}$$

$$\epsilon^{\text{IE}}(k) \triangleq E[[\tilde{x}^{\text{IE}}(k|k)]^2] \tag{11.5.4-2}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The ***sample mean*** of the performance from N ***Monte Carlo runs*** is

$$\bar{\epsilon}^{\text{VSD}}(k) = \frac{1}{N}\sum_{i=1}^{N}[\tilde{x}^{\text{VSD}i}(k|k)]^2 \tag{11.5.4-3}$$

$$\bar{\epsilon}^{\text{IE}}(k) = \frac{1}{N}\sum_{i=1}^{N}[\tilde{x}^{\text{IE}i}(k|k)]^2 \tag{11.5.4-4}$$

where $\tilde{x}^{\text{VSD}i}(k|k)$ is the estimation error of the state component under consideration at time k in run i with algorithm VSD, and similarly for IE. These are the quantities shown in Figs. 11.5.3-1a and 11.5.3-1b.

We want to test if one can accept the hypothesis

$$H_1 : T^{\text{IV}}(k) \triangleq \epsilon^{\text{IE}}(k) - \epsilon^{\text{VSD}}(k) > 0 \tag{11.5.4-5}$$

that is, that algorithm VSD is superior to algorithm IE. Note that "superiority" is based on the *true mean square errors* defined in (11.5.4-1) and (11.5.4-2). A comparison of the *sample mean square errors* (11.5.4-3) and (11.5.4-4) corresponding to the two algorithms is not sufficient (actually is rather naive) since the inaccuracy in these sample means is not considered at all.

The correct statistical test is, as discussed in Subsection 1.5.3, based on the ***sample performance differences***

$$T^{\text{IV}i}(k) \triangleq [\tilde{x}^{\text{IE}i}(k|k)]^2 - [\tilde{x}^{\text{VSD}i}(k|k)]^2 \tag{11.5.4-6}$$

The test uses the sample mean $\bar{T}$ of the above differences and its standard error $\sigma_{\bar{T}}$ given in (1.5.3-9) and (1.5.3-10), respectively.

If the ratio $\bar{T}/\sigma_{\bar{T}}$ exceeds a threshold, then the *difference of the true means* in (11.5.4-5) is accepted as positive (H_1 is accepted) and the null hypothesis

$$H_0 : T^{\text{IV}} \leq 0 \tag{11.5.4-7}$$

is rejected because it has a "low level of significance." This is equivalent to saying that the estimated mean $\bar{T}$ is positive and statistically significant (i.e., H_1 can be accepted because it is "significant").

Hypothesis H_1 is usually accepted only if the significance level of H_0 is less than 5%, in which case the test is

$$\frac{\bar{T}}{\sigma_{\bar{T}}} > \mathcal{G}(95\%) = 1.65 \tag{11.5.4-8}$$

where $\mathcal{G}(1-\alpha)$ represents the point on the standard Gaussian distribution corresponding to upper tail probability of α.

The above statistical test was applied for comparing the performance over intervals $[k, l]$, rather than single points in time, by replacing (11.5.4-6) with

$$T^{\text{IV}i}(k,l) = \frac{1}{l-k+1}\sum_{m=k}^{l}\{[\tilde{x}^{\text{IE}i}(m|m)]^2 - [\tilde{x}^{\text{VSD}i}(m|m)]^2\} \tag{11.5.4-9}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 11.5.4-1 shows the test for the difference of the MSE between the two algorithms over several time intervals. Note that the slow maneuver took place during the time interval $[40, 60]$ and that the fast one took place during the interval $[61, 66]$. Statistically significant improvements of the VSD algorithm over the IE algorithm are observed in position over all the intervals considered and in velocity over two intervals. In the remaining two intervals, while there was improvement, it was not sufficient to reject H_0 at the 5% level. At the 8% level, however, H_0 would have been rejected in all cases.

Table 11.5.4-1: Test of means for comparison of VSD and IE algorithms from 50 runs.

	Velocity			*Position*		
Time interval	$\bar{T}$	$\sigma_{\bar{T}}$	*Test statistic*	$\bar{T}$	$\sigma_{\bar{T}}$	*Test statistic*
40-60	8.06	1.19	6.77	13,560	735	18.45
60-70	4.99	3.28	1.52	10,336	1,013	10.20
40-70	7.72	1.43	5.40	12,677	619	20.48
60-80	2.45	1.75	1.40	6,091	626	9.73

11.5.5 Comparison of Several Algorithms — Summary

The optimal statistical method of comparing two algorithms based on Monte Carlo runs has been illustrated in the evaluation of the variable state dimension and input estimation filters.

The VSD filter turned out to be superior (based on statistical significance analysis) and less demanding computationally than the IE filter.

The much simpler two-level white noise filter appears to have performance close to the VSD filter.

A multiple-level white or autocorrelated noise model seems to be the best compromise from the point of view of performance versus cost of implementation.

The computational requirements of the IE filter, the VSD filter, and the two-level process noise filter were, approximately, $8 : 2 : 1$.

Neither of these three filters considered here managed to keep the *peak position estimation error* (during the maneuver) below the *raw position measurement RMS error* — the measurement noise standard deviation. This is a major shortcoming of most adaptive estimation schemes.

As will be shown in the next section, the multiple model approach can accomplish this (almost[1]) while yielding good noise reduction during the constant velocity portions of the trajectory.

[1] In the spirit of "Our algorithms can almost do almost everything."

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

11.6 THE MULTIPLE MODEL APPROACH

11.6.1 Formulation of the Approach

In the ***multiple model (MM) approach*** it is assumed that the system obeys one of a finite number of models. Such systems are called ***hybrid***: They have both *continuous* (noise) uncertainties and *discrete* uncertainties — ***model*** or ***mode***, or ***operating regime*** uncertainties.

A *Bayesian framework* is used: Starting with prior probabilities of each model being correct (i.e., the system is in a particular mode), the corresponding posterior probabilities are obtained.

First the static case in which the model the system obeys is *fixed*, that is, no switching from one mode to another occurs during the estimation process (time-invariant mode) is considered. This will result in the ***static MM estimator***. While the model that is in effect stays fixed, each model has its own dynamics, so *the overall estimator is dynamic*.

The model, assumed to be in effect throughout the entire process, is one of r possible models (the system is in one of r modes)

$$M \in \{M_j\}_{j=1}^{r} \tag{11.6.1-1}$$

The prior probability that M_j is correct (the system is in mode j) is

$$P\{M_j|Z^0\} = \mu_j(0) \qquad j = 1, \ldots, r \tag{11.6.1-2}$$

where Z^0 is the prior information and

$$\sum_{j=1}^{r} \mu_j(0) = 1 \tag{11.6.1-3}$$

since the correct model is among the assumed r possible models.

It will be assumed that all models are linear-Gaussian. This approach can be used for nonlinear systems as well via linearization.

Subsequently, the dynamic situation of ***switching models*** or ***mode jumping*** is considered. In the latter case, the system undergoes transitions from one mode to another. The resulting estimator is a ***dynamic MM estimator***.

11.6.2 The Static Multiple Model Estimator

The ***static MM estimator*** — for fixed models — is obtained as follows.

Using Bayes' formula, the posterior probability of model j being correct, given the measurement data up to k, is given by the recursion

$$\begin{aligned}\mu_j(k) &\triangleq P\{M_j|Z^k\} = P\{M_j|z(k), Z^{k-1}\} = \frac{p[z(k)|Z^{k-1}, M_j]P\{M_j|Z^{k-1}\}}{p[z(k)|Z^{k-1}]} \\ &= \frac{p[z(k)|Z^{k-1}, M_j]P\{M_j|Z^{k-1}\}}{\sum_{i=1}^{r} p[z(k)|Z^{k-1}, M_i]P\{M_i|Z^{k-1}\}}\end{aligned} \tag{11.6.2-1}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

or

$$\boxed{\mu_j(k) = \frac{p[z(k)|Z^{k-1}, M_j]\mu_j(k-1)}{\sum_{i=1}^{r} p[z(k)|Z^{k-1}, M_i]\mu_i(k-1)} \qquad j = 1, \ldots, r} \tag{11.6.2-2}$$

starting with the given prior probabilities (11.6.1-2).

The first term on the right-hand side above is the ***likelihood function of mode*** j at time k, which, under the linear-Gaussian assumptions, is given by the expression (5.2.6-7)

$$\Lambda_j(k) \triangleq p[z(k)|Z^{k-1}, M_j] = p[\nu_j(k)] = \mathcal{N}[\nu_j(k); 0, S_j(k)] \tag{11.6.2-3}$$

where ν_j and S_j are the innovation and its covariance from the ***mode-matched filter*** corresponding to mode j. In a nonlinear and/or non-Gaussian situation the same Gaussian likelihood functions are used, even though they are clearly approximations.

Thus a Kalman filter matched to each mode is set up yielding ***mode-conditioned state estimates*** and the associated ***mode-conditioned covariances***. The probability of each mode being correct — the ***mode estimates*** — is obtained according to (11.6.2-2) based on its likelihood function (11.6.2-3) relative to the other filters' likelihood functions. In a nonlinear situation the filters are EKF instead of KF.

This modular estimator, which is a ***bank of filters***, is illustrated in Fig. 11.6.2-1.

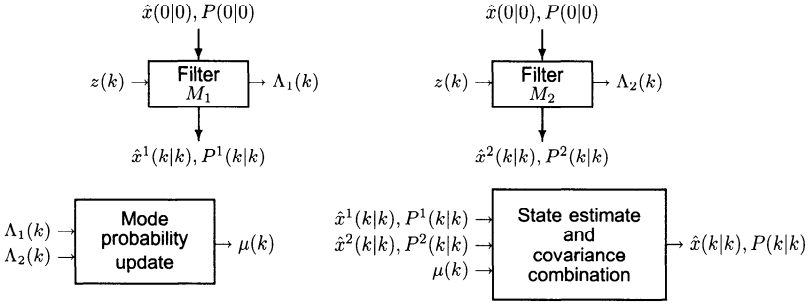


Figure 11.6.2-1: The static multiple model estimator for $r = 2$ fixed models.

The output of each mode-matched filter is the ***mode-conditioned state estimate*** $\hat{x}^j$, the associated covariance P^j and the ***mode likelihood function*** Λ_j.

After the filters are initialized, they run recursively *on their own estimates*. Their likelihood functions are used to update the mode probabilities. The latest mode probabilities are used to combine the mode-conditioned estimates and covariances.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Under the above assumptions the pdf of the state of the system is a Gaussian mixture with r terms

$$p[x(k)|Z^k] = \sum_{j=1}^{r} \mu_j(k)\mathcal{N}[x(k); \hat{x}^j(k|k), P^j(k|k)] \tag{11.6.2-4}$$

The combination of the mode-conditioned estimates is done therefore as follows

$$\boxed{\hat{x}(k|k) = \sum_{j=1}^{r} \mu_j(k)\hat{x}^j(k|k)} \tag{11.6.2-5}$$

and the covariance of the combined estimate is (see Subsection 1.4.16)

$$\boxed{P(k|k) = \sum_{j=1}^{r} \mu_j(k)\{P^j(k|k) + [\hat{x}^j(k|k) - \hat{x}(k|k)][\hat{x}^j(k|k) - \hat{x}(k|k)]'\}} \tag{11.6.2-6}$$

where the last term above is the ***spread of the means*** term.

The above is exact under the following assumptions:

1. The correct model is among the set of models considered,
2. The same model has been in effect from the initial time.

Assumption 1 can be considered a reasonable approximation; however, 2 is obviously not true if a maneuver has started at some time within the interval $[1, k]$, in which case a model change — ***mode jump*** — occurred.

Convergence of the Mode Estimates

If the mode set includes the correct one and no mode jump occurs, then the probability of the true mode will converge to unity, that is, this approach yields consistent estimates of the system parameters. Otherwise the probability of the model "nearest" to the correct one will converge to unity (this is discussed in detail in [Baram78]).

Ad Hoc Modifications for the Case of Switching Models

The following ad hoc modification can be made to the static MM estimator for estimating the state in the case of switching models: An artificial lower bound is imposed on the model probabilities (with a suitable renormalization of the remaining probabilities).

A shortcoming of the static MM estimator when used with switching models is that, in spite of the above ad hoc modification, the mismatched filters' errors can grow to unacceptable levels. Thus, reinitialization of the filters that are mismatched is, in general, needed. This is accomplished by using the estimate from filter corresponding to the best matched model in the other filters.

It should be pointed out that the above "fixes" are automatically (and rigorously) built into the dynamic MM estimation algorithms to be discussed next.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

11.6.3 The Dynamic Multiple Model Estimator

In this case the mode the system is in can undergo switching in time. The system is modeled by the equations

$$x(k) = F[M(k)]x(k-1) + v[k-1, M(k)] \quad (11.6.3\text{-}1)$$
$$z(k) = H[M(k)]x(k) + w[k, M(k)] \quad (11.6.3\text{-}2)$$

where $M(k)$ denotes the mode or model "at time k" — in effect *during the sampling period ending at* k. Such systems are also called ***jump-linear systems***. The mode jump process is assumed ***left-continuous*** (i.e., the impact of the new model starts at t_k^+).

The mode at time k is assumed to be among the possible r modes

$$M(k) \in \{M_j\}_{j=1}^{r} \quad (11.6.3\text{-}3)$$

The continuous-valued vector $x(k)$ and the discrete variable $M(k)$ are sometimes referred to as the ***base state*** and the ***modal state***, respectively.

For example,

$$F[M_j] = F_j \quad (11.6.3\text{-}4)$$
$$v(k-1, M_j) \sim \mathcal{N}(u_j, Q_j) \quad (11.6.3\text{-}5)$$

that is, the structure of the system and/or the statistics of the noises might be different from model to model. The mean u_j of the noise can model a maneuver as a deterministic input.

The lth ***mode history*** — or ***sequence of models*** — through time k is denoted as

$$M^{k,l} = \{M_{i_{1,l}}, \ldots, M_{i_{k,l}}\} \qquad l = 1, \ldots, r^k \quad (11.6.3\text{-}6)$$

where $i_{\kappa,l}$ is the model index at time κ from history l and

$$1 \leq i_{\kappa,l} \leq r \qquad \kappa = 1, \ldots, k \quad (11.6.3\text{-}7)$$

Note that the number of histories increases *exponentially with time*.

For example, with $r = 2$ one has at time $k = 2$ the following $r^k = 4$ possible sequences (histories) as shown below:

l	$i_{1,l}$	$i_{2,l}$
1	1	1
2	1	2
3	2	1
4	2	2

It will be assumed that the ***mode (model) switching*** — that is, the ***mode jump process*** — is a Markov process (Markov chain) with known mode transition probabilities

$$\boxed{p_{ij} \triangleq P\{M(k) = M_j | M(k-1) = M_i\}} \quad (11.6.3\text{-}8)$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

These ***mode transition probabilities*** will be assumed time-invariant and independent of the base state. In other words, this is a ***homogeneous Markov chain.***

The system (11.6.3-1), (11.6.3-2), and (11.6.3-8) is a generalized version of a ***hidden Markov model.***

The event that model j is in effect at time k is denoted as

$$M_j(k) \triangleq \{M(k) = M_j\} \tag{11.6.3-9}$$

The conditional probability of the lth history

$$\mu^{k,l} \triangleq P\{M^{k,l}|Z^k\} \tag{11.6.3-10}$$

will be evaluated next.

The lth sequence of models through time k can be written as

$$M^{k,l} = \{M^{k-1,s}, M_j(k)\} \tag{11.6.3-11}$$

where sequence s through $k-1$ is its *parent sequence* and M_j is its last element.

Then, in view of the Markov property,

$$P\{M_j(k)|M^{k-1,s}\} = P\{M_j(k)|M_i(k-1)\} \triangleq p_{ij} \tag{11.6.3-12}$$

where $i = s_{k-1}$, the index of the last model in the parent sequence s through $k-1$.

The conditional pdf of the state at k is obtained using the total probability theorem with respect to the mutually exclusive and exhaustive set of events (11.6.3-6), as a ***Gaussian mixture*** with an *exponentially increasing number of terms*

$$\boxed{p[x(k)|Z^k] = \sum_{l=1}^{r^k} p[x(k)|M^{k,l}, Z^k]P\{M^{k,l}|Z^k\}} \tag{11.6.3-13}$$

Since *to each mode sequence one has to match a filter*, it can be seen that an exponentially increasing number of filters are needed to estimate the (base) state, which makes the optimal approach impractical.

The probability of a mode history is obtained using Bayes' formula as

$$\begin{aligned}
\mu^{k,l} &= P\{M^{k,l}|Z^k\} \\
&= P\{M^{k,l}|z(k), Z^{k-1}\} \\
&= \frac{1}{c}\, p[z(k)|M^{k,l}, Z^{k-1}]P\{M^{k,l}|Z^{k-1}\} \\
&= \frac{1}{c}\, p[z(k)|M^{k,l}, Z^{k-1}]P\{M_j(k), M^{k-1,s}|Z^{k-1}\} \\
&= \frac{1}{c}\, p[z(k)|M^{k,l}, Z^{k-1}]P\{M_j(k)|M^{k-1,s}, Z^{k-1}\}\mu^{k-1,s} \\
&= \frac{1}{c}\, p[z(k)|M^{k,l}, Z^{k-1}]P\{M_j(k)|M^{k-1,s}\}\mu^{k-1,s}
\end{aligned} \tag{11.6.3-14}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where c is the normalization constant.

If the current mode depends only on the previous one (i.e., it is a Markov chain), then

$$\mu^{k,l} = \frac{1}{c}\, p[z(k)|M^{k,l}, Z^{k-1}]P\{M_j(k)|M_i(k-1)\}\mu^{k-1,s} \tag{11.6.3-15}$$

or

$$\boxed{\mu^{k,l} = \frac{1}{c}\, p[z(k)|M^{k,l}, Z^{k-1}]p_{ij}\mu^{k-1,s}} \tag{11.6.3-16}$$

where $i = s_{k-1}$ is the last model of the parent sequence s.

The above equation shows that *conditioning on the entire past history* is needed even if the random parameters are Markov.

Practical Algorithms

The only way to avoid the exponentially increasing number of histories, which have to be accounted for, is by going to suboptimal techniques.

A simple-minded suboptimal technique is to keep, say, the N histories with the largest probabilities, discard the rest, and renormalize the probabilities such that they sum up to unity.

The ***generalized pseudo-Bayesian (GPB)*** approaches combine histories of models that differ in "older" models. The first-order GPB, denoted as GPB1, considers only the possible models in the last sampling period. The second-order version, GPB2, considers all the possible models in the last two sampling periods. These algorithms require r and r^2 filters to operate in parallel, respectively.

Finally, the ***interacting multiple model (IMM)*** estimation algorithm will be presented. This algorithm is conceptually similar to GPB2, but requires only r filters to operate in parallel.

The Mode Transition Probabilities

The ***mode transition probabilities*** (11.6.3-8), indicated as assumed to be known, are actually ***estimator design parameters*** to be selected in the design process of the algorithm. This will be discussed in detail in Subsections 11.6.7 and 11.7.3.

Note

The GPB1 and IMM algorithms have approximately the same computational requirements as the static (fixed model) MM algorithm, but do not require ad hoc modifications as the latter, which is actually obsolete for switching models.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

11.6.4 The GPB1 Multiple Model Estimator for Switching Models

In the ***generalized pseudo-Bayesian estimator of first order (GPB1)***, at time k the state estimate is computed under each possible current model — a total of r possibilities (hypotheses) are considered. All histories that differ in "older" models are combined together.

The total probability theorem is thus used as follows:

$$\begin{aligned} p[x(k)|Z^k] &= \sum_{j=1}^{r} p[x(k)|M_j(k), Z^k] P\{M_j(k)|Z^k\} \\ &= \sum_{j=1}^{r} p[x(k)|M_j(k), z(k), Z^{k-1}] \mu_j(k) \\ &\approx \sum_{j=1}^{r} p[x(k)|M_j(k), z(k), \hat{x}(k-1|k-1), P(k-1|k-1)] \mu_j(k) \end{aligned} \tag{11.6.4-1}$$

Thus at time $k-1$ there is a single ***lumped estimate*** $\hat{x}(k-1|k-1)$ and the associated covariance that summarize (approximately) the past Z^{k-1}. From this, one carries out the prediction to time k and the update at time k under r hypotheses, namely,

$$\hat{x}^j(k|k) = \hat{x}[k|k; M_j(k), \hat{x}(k-1|k-1), P(k-1|k-1)] \qquad j = 1, \ldots, r \tag{11.6.4-2}$$

$$P^j(k|k) = P[k|k; M_j(k), P(k-1|k-1)] \qquad j = 1, \ldots, r \tag{11.6.4-3}$$

After the update, the estimates are combined with the weightings $\mu_j(k)$ (detailed later), resulting in the new combined estimate $\hat{x}(k|k)$. In other words, *the r hypotheses are merged into a single hypothesis at the end of each cycle.*

Finally, the mode (or model) probabilities are updated.

Figure 11.6.4-1 describes this estimator, which requires r filters in parallel.

The output of each model-matched filter is the ***mode-conditioned state estimate*** $\hat{x}^j$, the associated covariance P^j and the ***mode likelihood function*** Λ_j.

After the filters are initialized, they run recursively using *the previous combined estimate.* Their likelihood functions are used to update the mode probabilities. The latest mode probabilities are used to combine the model-conditional estimates and covariances.

The structure of this algorithm is

$$(N_e; N_f) = (1; r) \tag{11.6.4-4}$$

where N_e is the *number of estimates* at the start of the cycle of the algorithm and N_f is the *number of filters* in the algorithm.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

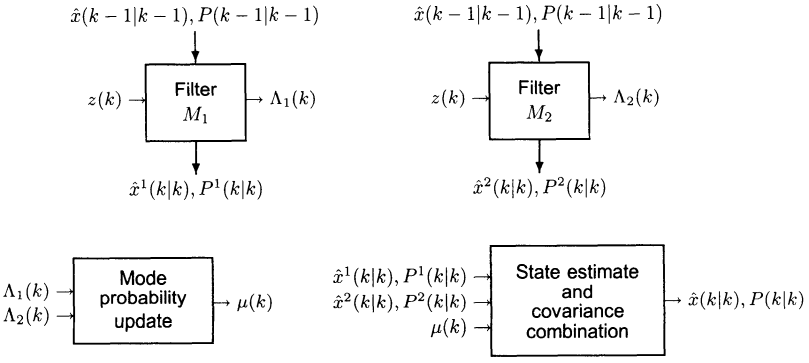


Figure 11.6.4-1: The GPB1 MM estimator for $r = 2$ switching models (one cycle).

The Algorithm

One cycle of the algorithm consists of the following:

1. ***Mode-matched filtering*** ($j = 1, \ldots, r$). Starting with $\hat{x}(k-1|k-1)$, one computes $\hat{x}^j(k|k)$ and the associated covariance $P^j(k|k)$ through a filter matched to $M_j(k)$. The likelihood functions

$$\Lambda_j(k) = p[z(k)|M_j(k), Z^{k-1}] \tag{11.6.4-5}$$

corresponding to these r filters are evaluated as

$$\boxed{\Lambda_j(k) = p[z(k)|M_j(k), \hat{x}(k-1|k-1), P(k-1|k-1)]} \tag{11.6.4-6}$$

2. ***Mode probability update*** ($j = 1, \ldots, r$). This is done as follows:

$$\begin{aligned}
\mu_j(k) &\triangleq P\{M_j(k)|Z^k\} \\
&= P\{M_j(k)|z(k), Z^{k-1}\} \\
&= \frac{1}{c}\, p[z(k)|M_j(k), Z^{k-1}] P\{M_j(k)|Z^{k-1}\} \\
&= \frac{1}{c}\, \Lambda_j(k) \sum_{i=1}^{r} P\{M_j(k)|M_i(k-1), Z^{k-1}\} \\
&\qquad \cdot P\{M_i(k-1)|Z^{k-1}\}
\end{aligned} \tag{11.6.4-7}$$

which yields with p_{ij} the known *mode transition probabilities*,

$$\boxed{\mu_j(k) = \frac{1}{c}\, \Lambda_j(k) \sum_{i=1}^{r} p_{ij}\mu_i(k-1)} \tag{11.6.4-8}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where c is the normalization constant

$$c = \sum_{j=1}^{r} \Lambda_j(k) \sum_{i=1}^{r} p_{ij}\mu_i(k-1) \tag{11.6.4-9}$$

3. ***State estimate and covariance combination.*** The latest combined state estimate and covariance are obtained according to the summation (11.6.4-1) as

$$\boxed{\hat{x}(k|k) = \sum_{j=1}^{r} \hat{x}^j(k|k)\mu_j(k)} \tag{11.6.4-10}$$

$$\boxed{P(k|k) = \sum_{j=1}^{r} \mu_j(k)\{P^j(k|k) + [\hat{x}^j(k|k) - \hat{x}(k|k)][\hat{x}^j(k|k) - \hat{x}(k|k)]'\}} \tag{11.6.4-11}$$

11.6.5 The GPB2 Multiple Model Estimator for Switching Models

In the ***generalized pseudo-Bayesian estimator of second order*** (or ***GPB2***), at time k the state estimate is computed under *each possible current and previous model* — a total of r^2 hypotheses (histories) are considered. All histories that differ only in "older" models are merged.

The total probability theorem is thus used as follows:

$$\begin{aligned} p[x(k)|Z^k] &= \sum_{j=1}^{r}\sum_{i=1}^{r} p[x(k)|M_j(k), M_i(k-1), Z^k]P\{M_i(k-1)|M_j(k), Z^k\} \\ &\quad \cdot P\{M_j(k)|Z^k\} \\ &= \sum_{j=1}^{r}\sum_{i=1}^{r} p[x(k)|M_j(k), z(k), M_i(k-1), Z^{k-1}]\mu_{i|j}(k-1|k)\mu_j(k) \\ &\approx \sum_{j=1}^{r}\sum_{i=1}^{r} p[x(k)|M_j(k), z(k), \hat{x}^i(k-1|k-1), P^i(k-1|k-1)] \\ &\quad \cdot \mu_{i|j}(k-1|k)\mu_j(k) \end{aligned} \tag{11.6.5-1}$$

that is, the past $\{M_i(k-1), Z^{k-1}\}$ is approximated by the ***mode-conditioned estimate*** $\hat{x}^i(k-1|k-1)$ and associated covariance.

Thus at time $k-1$ there are r estimates and covariances, each to be predicted to time k and updated at time k under r hypotheses, namely,

$$\hat{x}^{ij}(k|k) \triangleq \hat{x}[k|k; M_j(k), \hat{x}^i(k-1|k-1), P^i(k-1|k-1)] \qquad i,j = 1,\ldots,r \tag{11.6.5-2}$$

$$P^{ij}(k|k) \triangleq P[k|k; M_j(k), P^i(k-1|k-1)] \qquad i,j = 1,\ldots,r \tag{11.6.5-3}$$

After the update, the estimates corresponding to the same latest model hypothesis are combined with the weightings $\mu_{i|j}(k-1|k)$, detailed later, resulting

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

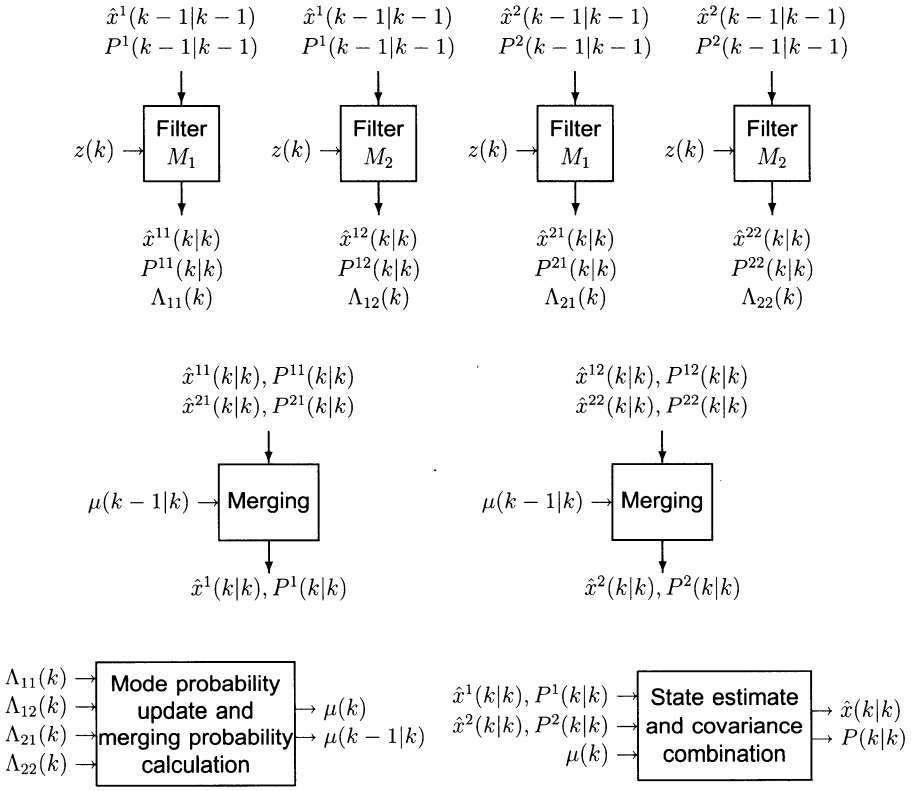


Figure 11.6.5-1: The GPB2 MM estimator for $r = 2$ models (one cycle).

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

in r estimates $\hat{x}^j(k|k)$. In other words, the r^2 hypotheses are *merged* into r *at the end of each estimation cycle.*

Figure 11.6.5-1 describes this algorithm, which requires r^2 parallel filters.

The structure of the GPB2 algorithm is

$$(N_e; N_f) = (r; r^2) \tag{11.6.5-4}$$

where N_c is the *number of estimates* at the start of the cycle of the algorithm and N_f is the *number of filters* in the algorithm.

The Algorithm

One cycle of the algorithm consists of the following:

1. ***Mode-matched filtering*** $(i, j = 1, \ldots, r)$. Starting with $\hat{x}^i(k-1|k-1)$, one computes $\hat{x}^{ij}(k|k)$ and the associated covariance $P^{ij}(k|k)$ through a filter matched to $M_j(k)$. The likelihood functions corresponding to these r^2 filters

$$\Lambda_{ij}(k) = p[z(k)|M_j(k), M_i(k-1), Z^{k-1}] \tag{11.6.5-5}$$

are evaluated as

$$\boxed{\Lambda_{ij}(k) = p[z(k)|M_j(k), \hat{x}^i(k-1|k-1), P^i(k-1|k-1)] \qquad i, j = 1, \ldots, r} \tag{11.6.5-6}$$

2. ***Calculation of the merging probabilities*** $(i, j = 1, \ldots, r)$. The probability that mode i was in effect at $k-1$ if mode j is in effect at k is, conditioned on Z^k,

$$\begin{aligned}
\mu_{i|j}(k-1|k) &\triangleq P\{M_i(k-1)|M_j(k), Z^k\} \\
&= P\{M_i(k-1)|z(k), M_j(k), Z^{k-1}\} \\
&= \frac{1}{c_j} P[z(k), M_j(k)|M_i(k-1), Z^{k-1}] P\{M_i(k-1)|Z^{k-1}\} \\
&= \frac{1}{c_j} p[z(k)|M_j(k), M_i(k-1), Z^{k-1}] \\
&\quad \cdot P\{M_j(k)|M_i(k-1), Z^{k-1}\} P\{M_i(k-1)|Z^{k-1}\}
\end{aligned} \tag{11.6.5-7}$$

where $P[\cdot]$ denotes a mixed pdf-probability. Thus the ***merging probabilities*** are

$$\boxed{\mu_{i|j}(k-1|k) = \frac{1}{c_j} \Lambda_{ij}(k) p_{ij} \mu_i(k-1) \qquad i, j = 1, \ldots, r} \tag{11.6.5-8}$$

where

$$c_j = \sum_{i=1}^{r} \Lambda_{ij}(k) p_{ij} \mu_i(k-1) \tag{11.6.5-9}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The ***mode transition probabilities*** p_{ij} are assumed to be known — their selection is part of the algorithm design process.

3. ***Merging*** $(j = 1, \ldots, r)$. The state estimate corresponding to $M_j(k)$ is obtained by combining the estimates (11.6.5-2) according to the inner summation in (11.6.5-1) as follows

$$\boxed{\hat{x}^j(k|k) = \sum_{i=1}^{r} \hat{x}^{ij}(k|k)\mu_{i|j}(k-1|k) \qquad j = 1, \ldots, r} \tag{11.6.5-10}$$

The covariance corresponding to the above is

$$\boxed{\begin{aligned} P^j(k|k) = \textstyle\sum_{i=1}^{r} \mu_{i|j}(k-1|k)\{&P^{ij}(k|k) \\ &+[\hat{x}^{ij}(k|k) - \hat{x}^j(k|k)][\hat{x}^{ij}(k|k) - \hat{x}^j(k|k)]'\} \end{aligned}} \tag{11.6.5-11}$$

4. ***Mode probability updating*** $(j = 1, \ldots, r)$. This is done as follows

$$\begin{aligned} \mu_j(k) &\triangleq P\{M_j(k)|z(k), Z^{k-1}\} \\ &= \frac{1}{c} P[z(k), M_j(k)|Z^{k-1}] \\ &= \frac{1}{c} \sum_{i=1}^{r} P[z(k), M_j(k)|M_i(k-1), Z^{k-1}]P\{M_i(k-1)|Z^{k-1}\} \\ &= \frac{1}{c} \sum_{i=1}^{r} p(z(k)|M_j(k), M_i(k-1), Z^{k-1}) \\ &\quad \cdot P\{M_j(k)|M_i(k-1), Z^{k-1}\}\mu_i(k-1) \end{aligned} \tag{11.6.5-12}$$

Thus the updated ***mode probabilities*** are

$$\boxed{\mu_j(k) = \frac{1}{c} \sum_{i=1}^{r} \Lambda_{ij}(k) p_{ij} \mu_i(k-1) = \frac{c_j}{c} \qquad j = 1, \ldots, r} \tag{11.6.5-13}$$

where c_j is the expression from (11.6.5-9) and c is the normalization constant

$$c = \sum_{j=1}^{r} c_j \tag{11.6.5-14}$$

5. ***State estimate and covariance combination.*** The latest state estimate and covariance *for output only* are

$$\boxed{\hat{x}(k|k) = \sum_{j=1}^{r} \hat{x}^j(k|k)\mu_j(k)} \tag{11.6.5-15}$$

$$\boxed{P(k|k) = \sum_{j=1}^{r} \mu_j(k)\{P^j(k|k) + [\hat{x}^j(k|k) - \hat{x}(k|k)][\hat{x}^j(k|k) - \hat{x}(k|k)]'\}} \tag{11.6.5-16}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

11.6.6 The Interacting Multiple Model Estimator

In the ***interacting multiple model (IMM) estimator***, at time k the state estimate is computed under *each possible current model* using r filters, with each filter using a different combination of the previous model-conditioned estimates — *mixed initial condition*.

The total probability theorem is used as follows to yield r filters running in parallel:

$$\begin{aligned} p[x(k)|Z^k] &= \sum_{j=1}^{r} p[x(k)|M_j(k), Z^k] P\{M_j(k)|Z^k\} \\ &= \sum_{j=1}^{r} p[x(k)|M_j(k), z(k), Z^{k-1}] \mu_j(k) \end{aligned} \tag{11.6.6-1}$$

The model-conditioned posterior pdf of the state, given by

$$p[x(k)|M_j(k), z(k), Z^{k-1}] = \frac{p[z(k)|M_j(k), x(k)]}{p[z(k)|M_j(k), Z^{k-1}]} p[x(k)|M_j(k), Z^{k-1}] \tag{11.6.6-2}$$

reflects one cycle of the state estimation filter matched to model $M_j(k)$ starting with the prior, which is the last term above.

The total probability theorem is now applied to the last term above (the prior), yielding

$$\begin{aligned} p[x(k)|M_j(k), Z^{k-1}] &= \sum_{i=1}^{r} p[x(k)|M_j(k), M_i(k-1), Z^{k-1}] \\ &\quad \cdot P\{M_i(k-1)|M_j(k), Z^{k-1}\} \\ &\approx \sum_{i=1}^{r} p[x(k)|M_j(k), M_i(k-1), \{\hat{x}^l(k-1|k-1), P^l(k-1|k-1)\}_{l=1}^{r}] \\ &\quad \cdot \mu_{i|j}(k-1|k-1) \\ &= \sum_{i=1}^{r} p[x(k)|M_j(k), M_i(k-1), \hat{x}^i(k-1|k-1), P^i(k-1|k-1)] \\ &\quad \cdot \mu_{i|j}(k-1|k-1) \end{aligned} \tag{11.6.6-3}$$

The second line above reflects the approximation that the past through $k-1$ is summarized by r model-conditioned estimates and covariances. The last line of (11.6.6-3) is a mixture with weightings, denoted as $\mu_{i|j}(k-1|k-1)$, different for each current model $M_j(k)$. This mixture is assumed to be a mixture of Gaussian pdfs (a Gaussian sum) and then approximated via moment matching

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

by a single Gaussian (details given later):

$$\begin{aligned} p[x(k)|M_j(k), Z^{k-1}] &= \sum_{i=1}^{r} \mathcal{N}\left[x(k); E[x(k)|M_j(k), \hat{x}^i(k-1|k-1)], \text{cov}[\cdot]\right] \\ &\quad \cdot \mu_{i|j}(k-1|k-1) \\ &\approx \mathcal{N}\Big[x(k); \sum_{i=1}^{r} E\left[x(k)|M_j(k), \hat{x}^i(k-1|k-1)\right] \mu_{i|j}(k-1|k-1), \text{cov}[\cdot]\Big] \\ &= \mathcal{N}\Big[x(k); E[x(k)|M_j(k), \sum_{i=1}^{r} \hat{x}^i(k-1|k-1)\mu_{i|j}(k-1|k-1)], \text{cov}[\cdot]\Big] \end{aligned} \tag{11.6.6-4}$$

The last line above follows from the linearity of the Kalman filter and amounts to the following:

> The input to the filter matched to model j is obtained from an ***interaction*** of the r filters, which consists of the ***mixing*** of the estimates $\hat{x}^i(k-1|k-1)$ with the weightings (probabilities) $\mu_{i|j}(k-1|k-1)$, called the ***mixing probabilities***.

The above is equivalent to hypothesis merging *at the beginning* of each estimation cycle [Blom88]. More specifically, the r hypotheses, instead of "fanning out" into r^2 hypotheses (as in the GPB2 — see Fig. 11.6.5-1), are "mixed" into a new set of r hypotheses as shown in Fig. 11.6.6-1. This is the key feature that yields r hypotheses with r filters, rather than with r^2 filters as in the GPB2 algorithm.

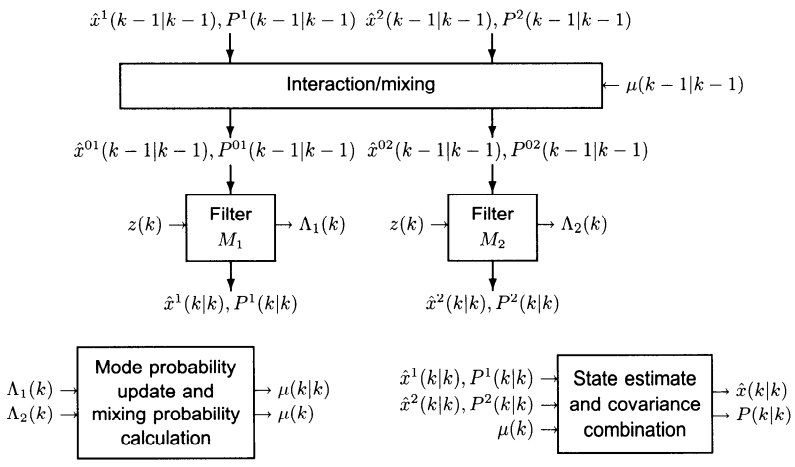


Figure 11.6.6-1: The IMM estimator (one cycle).

Figure 11.6.6-1 describes this algorithm, which consists of r interacting filters operating in parallel. The mixing is done at the input of the filters with

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the probabilities, detailed later in (11.6.6-6), conditioned on Z^{k-1}. In contrast to this, the GPB2 algorithm has r^2 filters and a somewhat similar mixing is done, but at their outputs, with the probabilities (11.6.5-7), conditioned on Z^k.

The structure of the IMM algorithm is

$$(N_e; N_f) = (r; r) \tag{11.6.6-5}$$

where N_e is the *number of estimates* at the start of the cycle of the algorithm and N_f is the *number of filters* in the algorithm.

The Algorithm

One cycle of the algorithm consists of the following:

1. ***Calculation of the mixing probabilities*** $(i, j = 1, \ldots, r)$. The probability that mode M_i was in effect at $k-1$ given that M_j is in effect at k conditioned on Z^{k-1} is

$$\begin{aligned} \mu_{i|j}(k-1|k-1) &\triangleq P\{M_i(k-1)|M_j(k), Z^{k-1}\} \\ &= \frac{1}{\bar{c}_j} P\{M_j(k)|M_i(k-1), Z^{k-1}\} P\{M_i(k-1)|Z^{k-1}\} \end{aligned} \tag{11.6.6-6}$$

The above are the ***mixing probabilities***, which can be written as

$$\boxed{\mu_{i|j}(k-1|k-1) = \frac{1}{\bar{c}_j} p_{ij}\mu_i(k-1) \qquad i, j = 1, \ldots, r} \tag{11.6.6-7}$$

where the normalizing constants are

$$\bar{c}_j = \sum_{i=1}^{r} p_{ij}\mu_i(k-1) \qquad j = 1, \ldots, r \tag{11.6.6-8}$$

Note the difference between (11.6.6-6), where the conditioning is Z^{k-1}, and (11.6.5-7), where the conditioning is Z^k. This is what makes it possible to carry out the mixing at the *beginning* of the cycle, rather than the standard merging at the *end* of the cycle.

2. ***Mixing*** $(j = 1, \ldots, r)$. Starting with $\hat{x}^i(k-1|k-1)$, one computes the mixed initial condition for the filter matched to $M_j(k)$ according to (11.6.6-4) as

$$\boxed{\hat{x}^{0j}(k-1|k-1) = \sum_{i=1}^{r} \hat{x}^i(k-1|k-1)\mu_{i|j}(k-1|k-1) \qquad j = 1, \ldots, r} \tag{11.6.6-9}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The covariance corresponding to the above is

$$\boxed{\begin{aligned} P^{0j}(k-1|k-1) &= \sum_{i=1}^{r} \mu_{i|j}(k-1|k-1)\Big\{P^{i}(k-1|k-1) \\ &\quad + [\hat{x}^{i}(k-1|k-1) - \hat{x}^{0j}(k-1|k-1)] \\ &\quad \cdot [\hat{x}^{i}(k-1|k-1) - \hat{x}^{0j}(k-1|k-1)]'\Big\} \\ &\quad j = 1, \ldots, r \end{aligned}} \tag{11.6.6-10}$$

3. ***Mode-matched filtering*** ($j = 1, \ldots, r$). The estimate (11.6.6-9) and covariance (11.6.6-10) are used as input to the filter matched to $M_j(k)$, which uses $z(k)$ to yield $\hat{x}^{j}(k|k)$ and $P^{j}(k|k)$.

The likelihood functions corresponding to the r filters

$$\Lambda_j(k) = p[z(k)|M_j(k), Z^{k-1}] \tag{11.6.6-11}$$

are computed using the mixed initial condition (11.6.6-9) and the associated covariance (11.6.6-10) as

$$\Lambda_j(k) = p[z(k)|M_j(k), \hat{x}^{0j}(k-1|k-1), P^{0j}(k-1|k-1)] \qquad j = 1, \ldots, r \tag{11.6.6-12}$$

that is,

$$\boxed{\Lambda_j(k) = \mathcal{N}[z(k); \hat{z}^{j}[k|k-1; \hat{x}^{0j}(k-1|k-1)], S^{j}[k; P^{0j}(k-1|k-1)]] \qquad j = 1, \ldots, r} \tag{11.6.6-13}$$

4. ***Mode probability update*** ($j = 1, \ldots, r$). This is done as follows:

$$\begin{aligned} \mu_j(k) &\triangleq P\{M_j(k)|Z^k\} \\ &= \frac{1}{c}\, p[z(k)|M_j(k), Z^{k-1}] P\{M_j(k)|Z^{k-1}\} \\ &= \frac{1}{c}\, \Lambda_j(k) \sum_{i=1}^{r} P\{M_j(k)|M_i(k-1), Z^{k-1}\} P\{M_i(k-1)|Z^{k-1}\} \\ &= \frac{1}{c}\, \Lambda_j(k) \sum_{i=1}^{r} p_{ij}\mu_i(k-1) \qquad j = 1, \ldots, r \end{aligned} \tag{11.6.6-14}$$

or

$$\boxed{\mu_j(k) = \frac{1}{c}\, \Lambda_j(k)\bar{c}_j \qquad j = 1, \ldots, r} \tag{11.6.6-15}$$

where $\bar{c}_j$ is the expression from (11.6.6-8) and

$$c = \sum_{j=1}^{r} \Lambda_j(k)\bar{c}_j \tag{11.6.6-16}$$

is the normalization constant for (11.6.6-15).

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

5. ***Estimate and covariance combination.*** Combination of the model-conditioned estimates and covariances is done according to the mixture equations

$$\boxed{\hat{x}(k|k) = \sum_{j=1}^{r} \hat{x}^j(k|k)\mu_j(k)} \tag{11.6.6-17}$$

$$\boxed{P(k|k) = \sum_{j=1}^{r} \mu_j(k)\{P^j(k|k) + [\hat{x}^j(k|k) - \hat{x}(k|k)][\hat{x}^j(k|k) - \hat{x}(k|k)]'\}} \tag{11.6.6-18}$$

This combination is *only* for output purposes — it is not part of the algorithm recursions.

Note

One possible generalization of the IMM estimator is the "second-order IMM" with an extra period depth. While the derivations are rather lengthy, it has been reported that this algorithm is *identical* to the GPB2 [Barret90].

11.6.7 An Example with the IMM Estimator

The use of the IMM estimator is illustrated on the example simulated in Section 11.5 where several of the earlier techniques were compared. The results presented in the sequel, which were obtained with DynaEst^{TM}, deal with the turn of 90° over 20 sampling periods.

A two-model IMM, designated as IMM2, was first used. This algorithm consisted of

1. A constant velocity model (second-order, with no process noise); and
2. A Wiener process acceleration model (third-order model) with process noise (acceleration increment over a sampling period) $q = 10^{-3} = (0.0316\,\mathrm{m/s^2})^2$.

Note that the acceleration in this case ($0.075\,\mathrm{m/s^2}$) corresponds to about $2.4\sqrt{q}$.

The Markov chain transition matrix between these models was taken as

$$[p_{ij}] = \begin{bmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{bmatrix} \tag{11.6.7-1}$$

The final results were not very sensitive to these values (e.g., p_{11} can be between 0.8 and 0.98). The lower (higher) value will yield less (more) peak error during maneuver but higher (lower) RMS error during the quiescent period; that is, it has a higher (lower) bandwidth.

A three-model IMM, designated as IMM3, was also used. This consisted of the above two models plus another one:

3. A constant acceleration (third-order) model without process noise.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

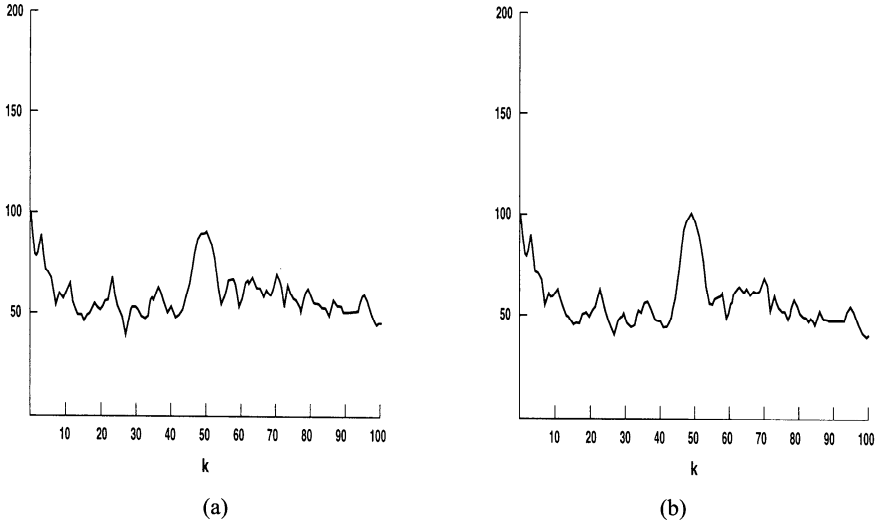


Figure 11.6.7-1: Position RMS error. (a) IMM2, (b) IMM3.

The Markov chain transition matrix was

$$[p_{ij}] = \begin{bmatrix} 0.95 & 0.05 & 0 \\ 0.33 & 0.34 & 0.33 \\ 0 & 0.05 & 0.95 \end{bmatrix} \tag{11.6.7-2}$$

Figure 11.6.7-1 shows the position (coordinate ξ) RMS error from 50 Monte Carlo runs for the IMM2 and IMM3, with the estimator design parameters as indicated above. Comparing with Fig. 11.5.3-1a, it can be seen that the peak errors are approximately equal to the measurement noise standard deviation (which is 100 m), that is, substantially smaller than with the IE or VSD. During the nonaccelerating period the errors are somewhat larger, but still there is an *RMS* ***noise reduction factor*** of 2; this corresponds to a *variance noise reduction factor* of 4.

Similar results for the velocity (in coordinate ξ) can be observed by comparing Figs. 11.6.7-2 with Fig. 11.5.3-1b.

Overall, the IMM2 and the IMM3 perform similarly in the position estimation; the IMM3 has a somewhat better velocity estimation capability.

Figure 11.6.7-3 shows the evolution in time of the maneuvering mode probabilities in IMM2 and IMM3. The figure illustrates the ***soft switching*** that takes place when a maneuver is initiated: It takes a few samples in this case to "detect" the maneuver. The "detection" of a maneuver manifests itself here as a sharp increase in the probability of the maneuvering mode — mode 2 in IMM2 and mode 3 in IMM3.

Meta 1018
p. 189

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

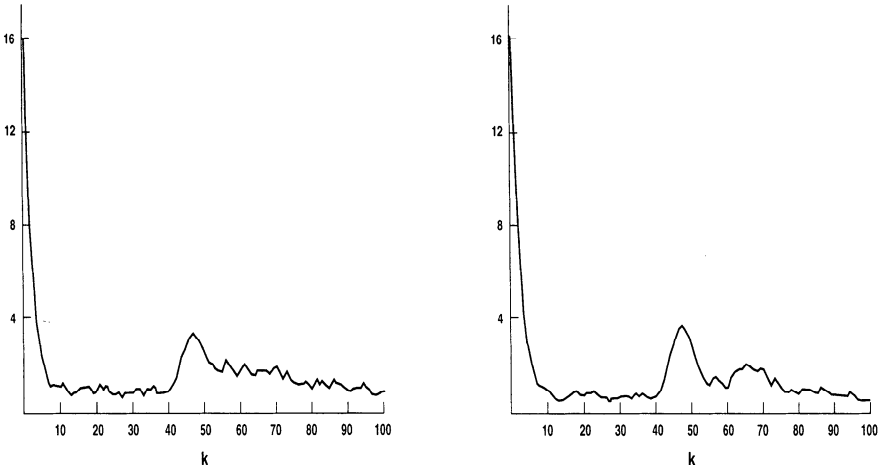


Figure 11.6.7-2: Velocity RMS error. (a) IMM2, (b) IMM3.

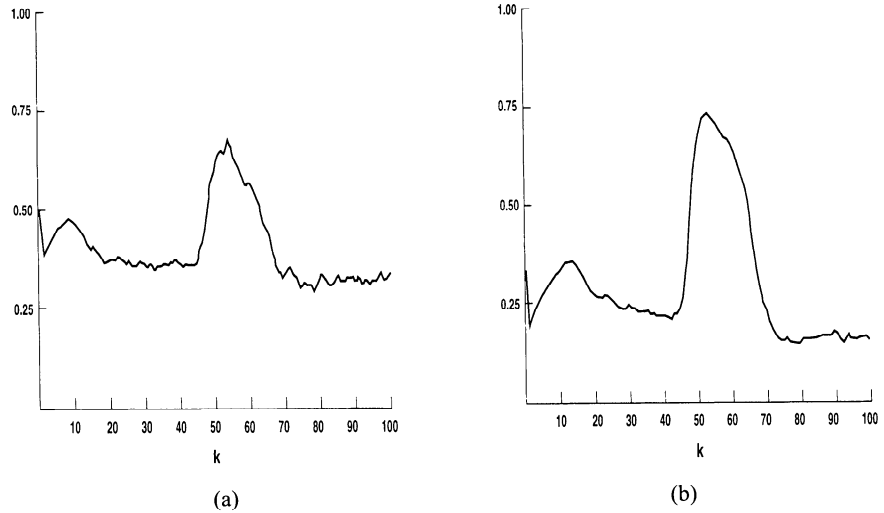


Figure 11.6.7-3: Maneuvering mode probability. (a) IMM2, (b) IMM3.

Meta 1018
p. 190

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Remark

These results, based on [Bar-Shalom89], are not only superior to the IE technique of Section 11.3 (illustrated in Subsection 11.5.3), but also superior to the augmented version of the IE algorithm that estimates the maneuver onset time [Bogler87].

This shows that the IMM, at the cost of about three Kalman filters, is preferable to the augmented IE technique, which has a much larger implementation complexity — of 30 to 100 Kalman filters.

11.6.8 Use of DynaEst^{TM} to Design an IMM Estimator

In Subsection 5.3.3 the use of the companion software DynaEst^{TM} for specifying a simulation scenario and designing a Kalman filter for estimation was demonstrated. In this section, the design of an IMM cstimator using DynaEst is illustrated. One of the IMM estimator configurations presented in Subsection 11.6.7, IMM2, which consisted of two filter modules, is used for this. Before proceeding to design the IMM estimator, DynaEst has to be used to define the simulation scenario in Section 11.5 as shown in Figs. 5.3.3-3–5.3.3-11.

Following the design of the simulation scenario, the IMM estimator design steps are as follows.

Selection of a Tracker The selection of a Kalman filter as the estimator was shown in Fig. 5.3.3-13. Now, the IMM estimator is selected as the tracker as shown in Fig. 11.6.8-1.

Structure Selection of an IMM Estimator First, the number of filter modules in the IMM estimator has to be specified. In addition, the nature of the mode transition probability matrix $[p_{ij}]$ has to be selected. DynaEst offers a choice between using a fixed mode transition probability matrix and using an adaptive one whose elements are modified at each sampling time based on the revisit interval and the ***mean sojourn time*** of each mode.

The simulation scenario defined in Section 11.5 uses a constant sampling interval $T = 10\,\text{s}$. Hence, using a fixed mode transition probability matrix, as shown in Fig. 11.6.8-2, is appropriate.

The Mode Transition Probability Matrix The fixed mode transition probability matrix of the IMM2 estimator, given in (11.6.7-1), is specified in DynaEst as shown in Fig. 11.6.8-3.

Selection of an IMM Filter Module to Design Since the IMM estimator consists of a number of filter modules, possibly with different state-space models, the parameters for them have to be specified individually for each module. The first filter module is selected for design as shown in Fig. 11.6.8-4.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

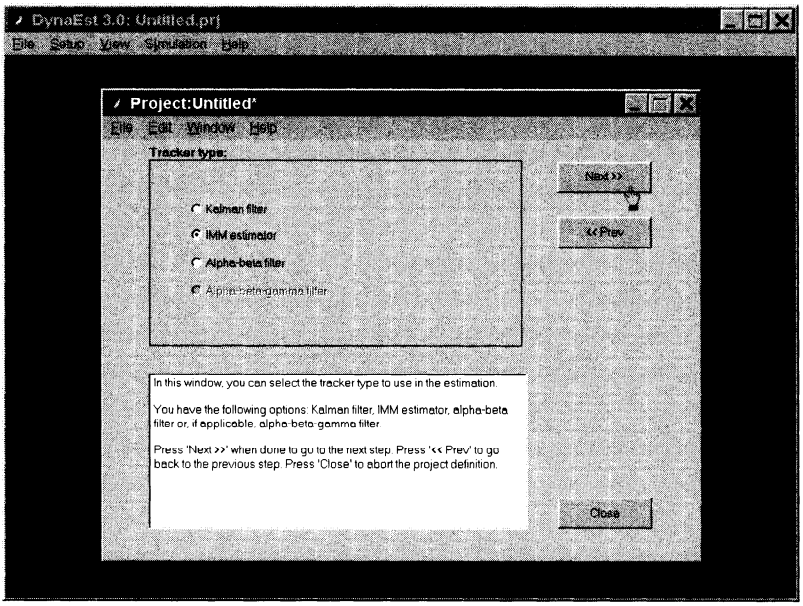


Figure 11.6.8-1: Selection of an IMM estimator as the tracker.

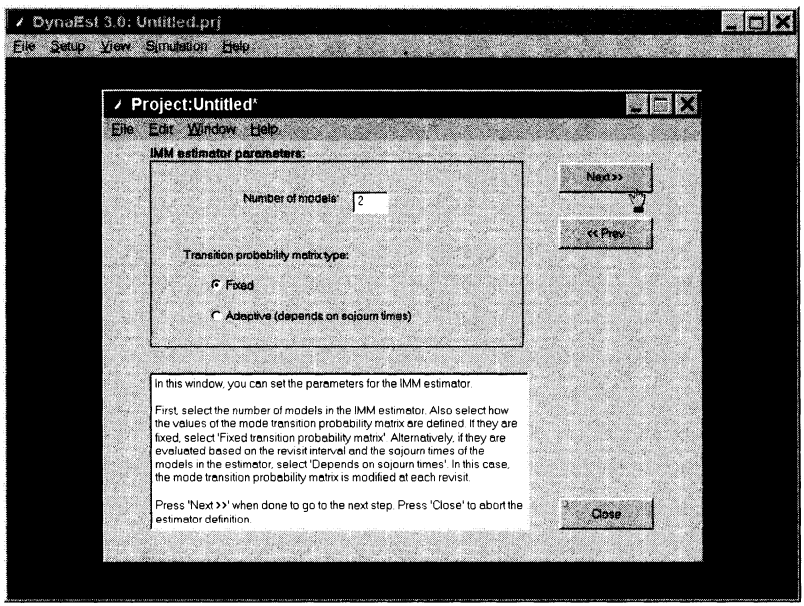


Figure 11.6.8-2: Structure selection of an IMM estimator.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

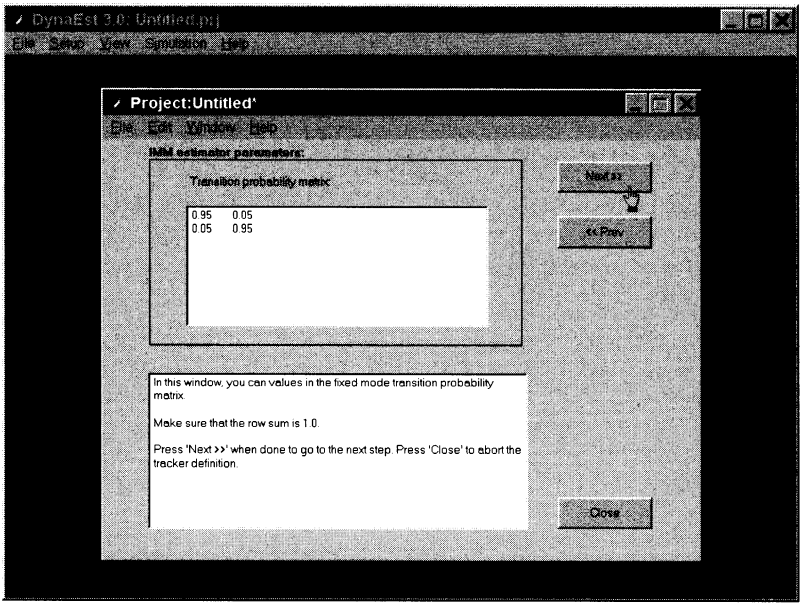


Figure 11.6.8-3: Specification of the mode transition probability matrix.

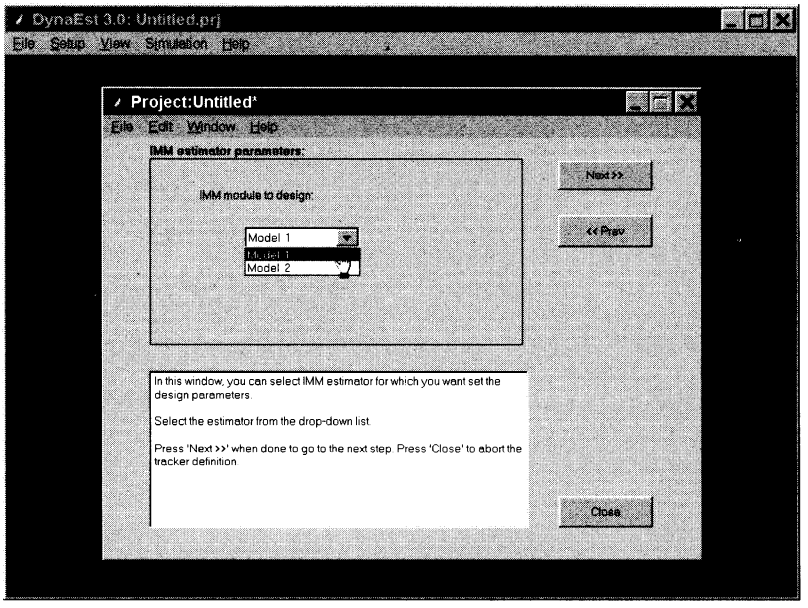


Figure 11.6.8-4: Selection of an IMM filter module to design.

Meta 1018
p. 193

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

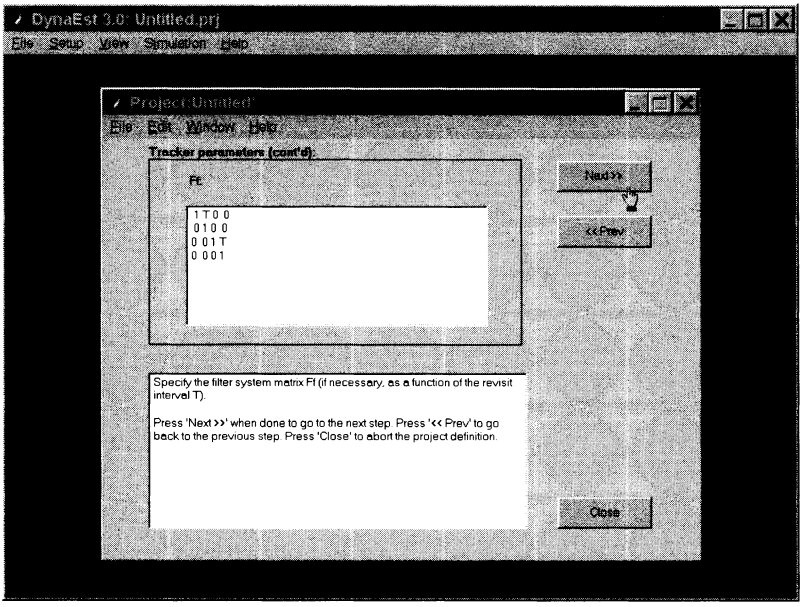


Figure 11.6.8-5: Design of a filter module.

Design of a Filter Module Selecting the parameters in a module is the crucial step in the IMM estimator design. This step involves selecting the assumed motion model and setting the corresponding process noise parameters in the module. DynaEst offers a sequence of windows similar to those in Figs. 5.3.3-6–5.3.3-11 for setting the motion model, noise parameters, initial estimates, and the corresponding covariance matrix. The first window in the IMM filter module design process is shown in Fig. 11.6.8-5.

Selection of the Next Filter Module to Design After defining the target motion model, process noise parameters, initial estimate, and the corresponding covariance matrix for the first IMM filter module, the next step is to select another filter module for designing as shown in Fig. 11.6.8-6. After selecting this filter module, its parameters are defined as described earlier.

Viewing Estimation Results After the IMM estimator is designed, DynaEst executes the specified number of Monte Carlo runs and offers choices to plot the estimation results as shown in Fig. 11.6.8-7. Note that the mode probability plot option is enabled only when the IMM estimator is used for estimation.

The design of the other IMM estimator, IMM3, can be carried out in a similar

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

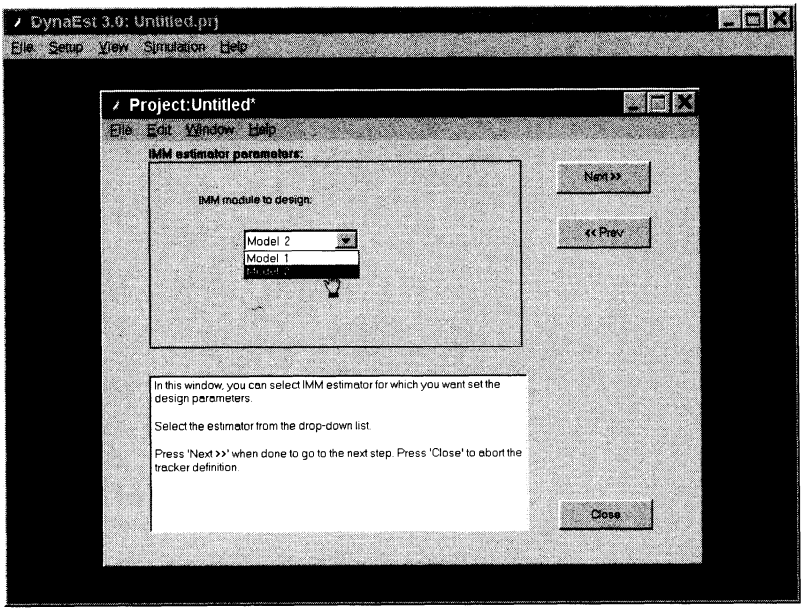


Figure 11.6.8-6: Selection of the next filter module to design.

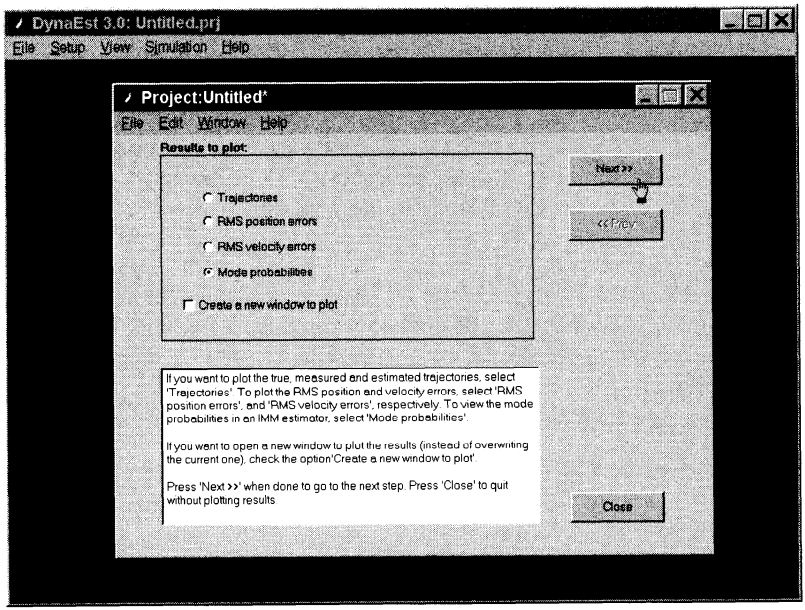


Figure 11.6.8-7: Viewing estimation results.

Meta 1018
p. 195

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

manner, but with the additional constant acceleration (third-order) model. This is left as an exercise to the reader.

11.6.9 The Multiple Model Approach — Summary

The *multiple model* or *hybrid system* approach assumes the system to be in one of a finite number of modes (i.e., that it is described by one out of a finite number of models).

Each model is characterized by its parameters — the models can differ, for instance, in the level of the process noise (its variance), a deterministic input, and/or any other parameter (different dimension state vectors are also possible).

For the *fixed model* case the estimation algorithm consists of the following:

- For each model a filter "matched" to its parameters is yielding *model-conditioned estimates and covariances.*
- A *mode probability calculator* — a *Bayesian model comparator* — updates the probability of each mode using
  - the likelihood function (innovations) of each filter
  - the prior probability of each model
- An *Estimate combiner* computes the overall estimate and the associated covariance as the weighted sum of the model-conditioned estimates and the corresponding covariance — via the (Gaussian) mixture equations.

The typical ad hoc modification of the fixed model (static) approach to handle switching models is to impose a lower bound on the probability of each model.

For systems that undergo changes in their mode during their operation — *mode jumping (model switching)* — one can obtain the *optimal multiple model estimator* which, however, consists of an exponentially increasing number of filters.

This is because the optimal approach requires conditioning on each *mode history*, and their number is increasing exponentially. Thus, suboptimal algorithms are necessary for the (realistic) mode transition situation.

The *first-order generalized pseudo-Bayesian* (GPB1) MM approach computes the state estimate accounting for each possible current model.

The *second-order generalized pseudo-Bayesian* (GPB2) MM approach computes the state estimate accounting for

- Each possible current model
- Each possible model at the previous time

The *interacting multiple model* (IMM) approach computes the state estimate that accounts for *each possible current model* using a suitable mixing of the previous model-conditioned estimates depending on the current model.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

These algorithms are ***decision-free*** — no maneuver detection decision is needed — the algorithms undergo a ***soft switching*** according to the latest updated mode probabilities.

Table 11.6.9-1 presents a comparison of the complexities (in terms of the functions to be performed) of the static algorithm and the three algorithms presented for switching models.

Table 11.6.9-1: Comparison of complexities of the MM algorithms.

	Static	*GPB1*	*GPB2*	*IMM*
Number of filters	r	r	r^2	r
Number of combinations of r estimates and covariances	1	1	$r+1$	$r+1$
Number of probability calculations	r	r	r^2+r	r^2+r

As can be seen from the above, the static algorithm has the same requirements as the GPB1. The IMM has only slightly higher requirements than the GPB1, but clearly significantly lower than GPB2.

In view of this, the modifications of the static algorithm for the switching situation are considered obsolete.

In view of the fact that the IMM performs significantly better than GPB1 and almost as well as GPB2 [Blom88], the IMM is considered to be the best compromise between complexity and performance.

Furthermore, the IMM has been shown to be able to keep the position estimation error not worse than the raw measurement error during the critical maneuver periods (onset and termination) and provide significant improvement (noise reduction) at other times.

11.7 DESIGN OF AN IMM ESTIMATOR FOR ATC TRACKING

11.7.1 ATC Motion Models

In ***air traffic control (ATC)***, civilian aircraft have two basic modes of flight:

- ***Uniform motion (UM)*** — the straight and level flight with a constant speed and course
- ***Maneuver*** — turning or climbing/descending

The horizontal and vertical motion models can be, typically, assumed to be decoupled [Wang99]. The design and evaluation of several estimators for the horizontal motion will be discussed next. The flight modes in the *horizontal plane* can be modeled by

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

- A nearly constant velocity model for the uniform motion, implemented as a WNA (white noise acceleration, or second-order kinematic, discussed in Section 6.5.3) model with low-level process noise
- A maneuvering model, which can be implemented as
  - a WNA model with significant process noise, commensurate with the expected maneuvers, or
  - a *nearly "coordinated turn"* model

The nearly constant velocity model is given by

$$x(k+1) = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix} x(k) + \begin{bmatrix} \frac{1}{2}T^2 & 0 \\ T & 0 \\ 0 & \frac{1}{2}T^2 \\ 0 & T \end{bmatrix} v(k) \qquad (11.7.1\text{-}1)$$

where T is the sampling interval; x is the state of the aircraft, defined as

$$x = [\xi \;\; \dot{\xi} \;\; \eta \;\; \dot{\eta}]' \qquad (11.7.1\text{-}2)$$

with ξ and η denoting the Cartesian coordinates of the horizontal plane; and v is a zero-mean Gaussian white noise used to model ("cover") small accelerations, the turbulence, wind change, and so on, with an appropriate covariance Q, which is a design parameter.

The turn of a civilian aircraft usually follows a pattern known as ***coordinated turn (CT)*** (Subsection 4.2.2) — characterized by constant turn rate and constant speed. Although the actual turns are not exactly "coordinated" since the ground speed is the airspeed plus the wind speed, it can be suitably described by the "coordinated turn" model plus a fairly small noise representing the modeling error, resulting in the nearly coordinated turn model. The ***CT model*** is necessarily a nonlinear one if the turn rate is not a known constant. Augmenting the state vector (11.7.1-2) by one more component — the turn rate Ω, that is,

$$x = [\xi \;\; \dot{\xi} \;\; \eta \;\; \dot{\eta} \;\; \Omega]' \qquad (11.7.1\text{-}3)$$

the ***nearly coordinated turn model*** is then given by

$$x(k+1) = \begin{bmatrix} 1 & \frac{\sin\Omega(k)T}{\Omega(k)} & 0 & -\frac{1-\cos\Omega(k)T}{\Omega(k)} & 0 \\ 0 & \cos\Omega(k)T & 0 & -\sin\Omega(k)T & 0 \\ 0 & \frac{1-\cos\Omega(k)T}{\Omega(k)} & 1 & \frac{\sin\Omega(k)T}{\Omega(k)} & 0 \\ 0 & \sin\Omega(k)T & 0 & \cos\Omega(k)T & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} x(k)$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$+\begin{bmatrix} \frac{1}{2}T^2 & 0 & 0 \\ T & 0 & 0 \\ 0 & \frac{1}{2}T^2 & 0 \\ 0 & T & 0 \\ 0 & 0 & T \end{bmatrix} v(k) \qquad (11.7.1\text{-}4)$$

Note that the process noise v in (11.7.1-1) has different dimension from the one in (11.7.1-4).

Assuming only position measurements are available, this yields the following observation equation

$$z(k) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} x(k) + w(k) \qquad (11.7.1\text{-}5)$$

where w is the measurement noise. Although the original radar measurements are in polar coordinates, they can be converted to Cartesian coordinates, as discussed in Subsection 10.4.3.

11.7.2 The EKF for the Coordinated Turn Model

Since the model in (11.7.1-4) is nonlinear, the estimation of the state (11.7.1-3) will be done via the EKF.

The equations for state prediction and the corresponding covariance, the only ones affected by the nonlinearity of (11.7.1-4), are given below. The dynamic equation (11.7.1-4) can be rewritten compactly as

$$x(k+1) = f[k, x(k)] + \Gamma_{\mathrm{CT}}(k) v(k) \qquad (11.7.2\text{-}1)$$

To obtain the predicted state $\hat{x}(k+1|k)$, the nonlinear function in (11.7.2-1) is expanded in Taylor series around the latest estimate $\hat{x}(k|k)$ with terms up to first order to yield the (first-order) EKF [2]. The vector Taylor series expansion of (11.7.2-1) up to first order is

$$\begin{aligned} x(k+1) &= f[k, \hat{x}(k|k)] + f_x(k)[x(k) - \hat{x}(k|k)] \\ &\quad + \mathrm{HOT} + \Gamma_{\mathrm{CT}}(k) v(k) \end{aligned} \qquad (11.7.2\text{-}2)$$

where HOT represents the higher-order terms and

$$f_x(k) = [\nabla_x f(k, x)']' \,|_{x=\hat{x}(k|k)}$$

[2] The second-order EKF was found not to provide any benefit in this problem.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$= \begin{bmatrix} 1 & \dfrac{\sin\hat{\Omega}(k)T}{\hat{\Omega}(k)} & 0 & -\dfrac{1-\cos\hat{\Omega}(k)T}{\hat{\Omega}(k)} & f_{\Omega,1}(k) \\ 0 & \cos\hat{\Omega}(k)T & 0 & -\sin\hat{\Omega}(k)T & f_{\Omega,2}(k) \\ 0 & \dfrac{1-\cos\hat{\Omega}(k)T}{\hat{\Omega}(k)} & 1 & \dfrac{\sin\hat{\Omega}(k)T}{\hat{\Omega}(k)} & f_{\Omega,3}(k) \\ 0 & \sin\hat{\Omega}(k)T & 0 & \cos\hat{\Omega}(k)T & f_{\Omega,4}(k) \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{11.7.2-3}$$

is the Jacobian of the vector f evaluated at the latest estimate of the state. The partial derivatives with respect to Ω are given by

$$\begin{bmatrix} f_{\Omega,1}(k) \\ f_{\Omega,2}(k) \\ f_{\Omega,3}(k) \\ f_{\Omega,4}(k) \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{\left(\cos\hat{\Omega}(k)T\right)T\dot{\hat{\xi}}(k)}{\hat{\Omega}(k)} - \dfrac{\left(\sin\hat{\Omega}(k)T\right)\dot{\hat{\xi}}(k)}{\hat{\Omega}(k)^2} - \dfrac{\left(\sin\hat{\Omega}(k)T\right)T\hat{\eta}(k)}{\hat{\Omega}(k)} - \dfrac{\left(-1+\cos\hat{\Omega}(k)T\right)\hat{\eta}(k)}{\hat{\Omega}(k)^2} \\ -\left(\sin\hat{\Omega}(k)T\right)T\dot{\hat{\xi}}(k) - \left(\cos\hat{\Omega}(k)T\right)T\hat{\eta}(k) \\ \dfrac{\left(\sin\hat{\Omega}(k)T\right)T\dot{\hat{\xi}}(k)}{\hat{\Omega}(k)} - \dfrac{\left(1-\cos\hat{\Omega}(k)T\right)\dot{\hat{\xi}}(k)}{\hat{\Omega}(k)^2} + \dfrac{\left(\cos\hat{\Omega}(k)T\right)T\hat{\eta}(k)}{\hat{\Omega}(k)} - \dfrac{\left(\sin\hat{\Omega}(k)T\right)\hat{\eta}(k)}{\hat{\Omega}(k)^2} \\ \left(\cos\hat{\Omega}(k)T\right)T\dot{\hat{\xi}}(k) - \left(\sin\hat{\Omega}(k)T\right)T\hat{\eta}(k) \end{bmatrix} \tag{11.7.2-4}$$

Based on the above expansion, the state prediction and state prediction covariance in the EKF are

$$\hat{x}(k+1|k) = f[k, \hat{x}(k|k)] \tag{11.7.2-5}$$

$$P(k+1|k) = f_x(k)P(k|k)f_x(k)' + \Gamma_{\mathrm{CT}}(k)Q(k)\Gamma_{\mathrm{CT}}(k)' \tag{11.7.2-6}$$

where Q is the covariance of the process noise in (11.7.2-1).

The initial estimate of the turn rate is zero since one does know whether it will be a left turn ($\Omega > 0$) or[3] a right turn ($\Omega < 0$). Since, from this initial estimate, the algorithm is capable of estimating the turn rate including its sign, there is no need for left- and right-turn models, as in some other IMM versions. Note that at the initialization step and when the estimated turn rate is zero (or when it is sufficiently close to zero), the limiting form of $f_x(k)$, given in

[3]This sign convention follows the trigonometric convention (the navigation convention has the opposite sign for the turn rates.)

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(11.7.2-7), should be used, namely,

$$f_x(k)|_{\hat{\Omega}(k)=0} = \begin{bmatrix} 1 & T & 0 & 0 & -\frac{1}{2}T^2\hat{\dot{\eta}}(k) \\ 0 & 1 & 0 & 0 & -T\hat{\dot{\eta}}(k) \\ 0 & 0 & 1 & T & \frac{1}{2}T^2\hat{\dot{\xi}}(k) \\ 0 & 0 & 0 & 1 & T\hat{\dot{\xi}}(k) \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{11.7.2-7}$$

Mixing of States of Different Dimension in the IMM

Since the dimension of the state (11.7.1-3) is 5, while that of (11.7.1-2) is 4, the mixing in the IMM estimator is done by augmenting the latter with a component that is zero. This amounts to having a zero turn rate in the constant velocity motion, which can be also seen to reduce (11.7.1-4) to (11.7.1-1). Similarly, the covariance associated with (11.7.1-2) is to be augmented by a column and a row of zeros.

11.7.3 Selection of Models and Parameters

To obtain the best possible results, the IMM algorithm has to be properly designed to meet the following requirements of the ATC tracking simultaneously:

- Reduce as much as possible the estimation errors during the uniform motion.
- Maintain the peak estimation error during the maneuver lower than that of the unfiltered raw measurements.
- Provide correct and timely indication of the flight mode, especially rapid detection of the maneuver.

These requirements are fulfilled by means of

- Design of aircraft motion models for all modes of flight
- Selection of the model parameters, such as the noise levels
- Determination of the parameters of the underlying Markov chain, that is, the transition probabilities

Model selection should consider both the quality and complexity of the model. Typically, the models used in the IMM configuration for ATC tracking will include one (a nearly constant velocity model) for the uniform motion and one or more for the maneuver — the nearly coordinated turn model of (11.7.1-4) is a typical one.

Selection of noise levels for each model is an important part of the estimator design. Although the uniform motion is better modeled in (11.7.1-1) with a small process noise v to model the air turbulence, winds aloft changes, and

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

so forth, it may be legitimate to use a somewhat larger process noise for the uniform motion to cover slow turns as well as small linear accelerations so as to ease the burden of modeling a broad range of maneuvers. The right choice of the noise level of the nearly coordinated turn model depends on what turn rate range is expected and how many models are to be used for the maneuvers.

The performance of the IMM algorithm is not very sensitive to the *choice of the transition probabilities*. However, this choice provides to a certain degree the trade-off between the peak estimation errors at the onset of the maneuver and the maximum reduction of the estimation errors during the uniform motion. The guideline for a proper choice is to match roughly the transition probabilities with the system's actual ***mean sojourn time*** in each mode. See problem 11-6 for the expression of the mean sojourn time in units of the sampling interval.

11.7.4 The ATC Scenario

As a generic ATC tracking problem, the following scenario is considered. The radar, stationed at [0 m, 0 m], provides direct position only measurements (after the polar-to-Cartesian conversion) with RMS errors of 100 m in each of the two Cartesian coordinates.[4] The intervals between the samples are $T = 5$ s. In the scenario under consideration, starting from [25,000 m, 10,000 m] at time $t = 0$ s, the aircraft flies westward for 125 s at 120 m/s, before executing a $1^\circ/\text{s}$ coordinated turn (which amounts to an acceleration of $0.2g$ at this speed), for 90 s. Then it flies southward for another 125 s, followed by a $3^\circ/\text{s}$ turn (an acceleration of $0.6g$ at this speed) for 30 s. After the turn, it continues to fly westward at constant velocity. The target trajectory is shown in Fig. 11.7.4-1. This scenario leads to a maneuvering index (see Chapter 6) that is quite high (almost 1.5) and thus very little noise reduction can be achieved by a single model based state estimator, which will have to be, by necessity, conservative, i.e., designed for the maximum acceleration.

The following estimator configurations were considered:

- KF: A Kalman filter using a second order linear kinematic model (WNA) with process noise of standard deviation 1 m/s^2 modeling the maneuver. This is a compromise between the maximum acceleration (6 m/s^2) and the constant velocity trajectory portions.
- IMM-L: An IMM estimator with two second-order linear kinematic models (WNA) with two noise levels. The one with the lower noise level with standard deviation 0.1 m/s^2 is used to model the uniform motion and the other one with standard deviation 2 m/s^2 for the maneuvers. The mode

[4]Its location under these assumptions is not relevant. In practice, the measurements are in polar coordinates and it is best to convert them into Cartesian (see Section 10.4.3). In the polar measurement case, the location of the sensor matters because the angular measurement errors are multiplied by the range from the sensor to the target.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

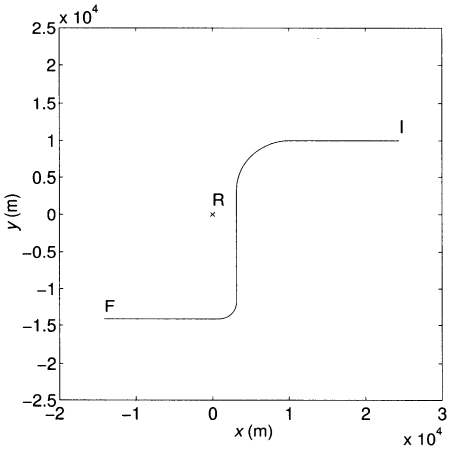


Figure 11.7.4-1: The target trajectory (I — initial point, F — final point, R — radar location).

transition probability matrix π_{L} was

$$\pi_{\mathrm{L}} = \begin{bmatrix} 0.95 & 0.05 \\ 0.10 & 0.90 \end{bmatrix} \tag{11.7.4-1}$$

- IMM-CT: An IMM estimator with one second-order kinematic model (a nearly constant velocity model with process noise standard deviation 0.1 $\mathrm{m/s^2}$) for the uniform motion and a (nearly) coordinated turn model with the turn rate (as part of the state) being estimated using an EKF. The process noise standard deviations used in the coordinated turn model were 0.5 $\mathrm{m/s^2}$ and $0.2°/\mathrm{s^2}$ for the linear and turn portions of the state, respectively.
  The mode transition probability matrix π_{CT} was

$$\pi_{\mathrm{CT}} = \begin{bmatrix} 0.95 & 0.05 \\ 0.10 & 0.90 \end{bmatrix} \tag{11.7.4-2}$$

The initial estimates were based on one-point initializations; that is, the initial position estimate was at the first measured position and the initial velocity/turn rate was zero. The initial velocity/turn rate variances were based on their assumed maximum values. The mode set, the transition mode probability matrices, process noise standard deviations and initial estimates/variances are estimator design parameters, which should be selected based on the expected scenarios and the designer's experience.

11.7.5 Results and Discussion

The performances of the various estimators are shown in Table 11.7.5-1, where all the position estimation errors are for the ξ and η coordinates *combined*. The

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

maneuver detection delay is defined to be the latency, measured in sampling periods, from the maneuver onset time to the time that the probability of the uniform motion mode falls below 0.5. The "UM probability error" stands for the (steady-state) probability of the maneuver modes during the uniform motion. In other words, this is the estimator-calculated probability that the UM mode is not in effect while, in truth, it is.

Table 11.7.5-1: Estimation errors in the estimators (raw position measurement error 141m).

	Estimator		
	KF	IMM-L	IMM-CT
Peak position error (m)	143	138	109
UM position error (m)	123	83	71
Peak speed error (m/s)	22.5	9.3	4.8
UM speed error (m/s)	3.2	1.9	1.3
UM course error (°)	3.8	2.7	1.8
1°/s Maneuver detection delay (scans)	—	4	1
3°/s Maneuver detection delay (scans)	—	3	1
UM probability error (%)	—	5.0	3.5

As presented in Table 11.7.5-1, the IMM-CT estimator yields a peak RMS position error during the mode of flight change that is almost 25% below the raw measurement error and 50% RMS error [5] reduction during the uniform motion. The speed estimation error during the uniform motion is 1% of the aircraft speed. The detection of the maneuver is quick: in two scans (i.e., one scan delay) for both turns in the target trajectory.

Since the coordinate-combined raw measurement error (1σ) is $100\sqrt{2} = 141$ m, the angle deviation from the straight line in one scan (sampling period) during the uniform motion due to 1σ measurement error can be as large as

$$\tan^{-1} \frac{141 \text{ m}}{120 \text{ m/s} \cdot 5 \text{ s}} = 13.2^\circ \qquad (11.7.5\text{-}1)$$

This is about the same as the expected course change caused by a $3^\circ/\text{s}$ turn in one scan, which is $(3^\circ/\text{s}) \cdot (5\,\text{s})) = 15^\circ$. In view of this, the UM probability error of only 3.5% seems quite low. The rapid detection of the maneuver, together with this small probability error, verifies the good reliability of this design in terms of providing the correct and timely information of the flight mode.

If the sampling period is shorter, one can achieve more reduction of the errors in the state estimate compared to the unfiltered radar measurements during the maneuver, and substantially more noise reduction can be achieved during the uniform motion.

[5] Square root of the sample average (mean) of the squared error from the Monte Carlo simulations.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Figure 11.7.5-1 presents the RMS position errors (ξ and η coordinates *combined*) for the IMM-L and the IMM-CT estimators as well as a baseline Kalman filter (tuned specially for this case). The initial large estimation errors are due to the fact that the initial probability of each model was set to be equal to account for the worst case of ignorance.

The following observations about the relative performances of the above estimators can be made:

IMM-CT has the peak RMS position error of about 110 m, superior to the single measurement position RMS error, which is $100\sqrt{2} = 141$ m. Moreover, when the aircraft is not maneuvering, the RMS error is around 70 m — that is, a reduction of about 50% in the RMS error (70% in MSE). The Kalman filter has a peak RMS position error of 143 m. During the uniform motion, however, its RMS error stays at about 120 m. This estimation error of the specially tuned Kalman filter for the uniform motion can be reduced only slightly at a cost of a much higher peak error. This is the typical behavior of a single-model based Kalman filter. Between the IMM-CT estimator and the IMM-L, the former yields faster maneuver detection and lower peak estimation errors.

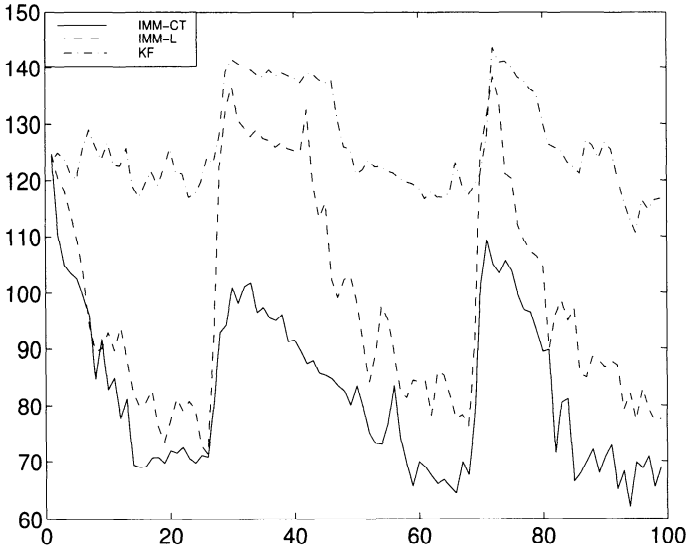


Figure 11.7.5-1: RMS position estimation errors.

Figure 11.7.5-2 shows the evolution of the average mode 2 (turning mode) probabilities in the IMM-L and IMM-CT estimators. The figure clearly indicates the rapid "detection" capability of IMM-CT — in two samples for both $1^\circ/s$ and $3^\circ/s$ turns. The less sophisticated IMM-L detects the turns in three to four samples.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

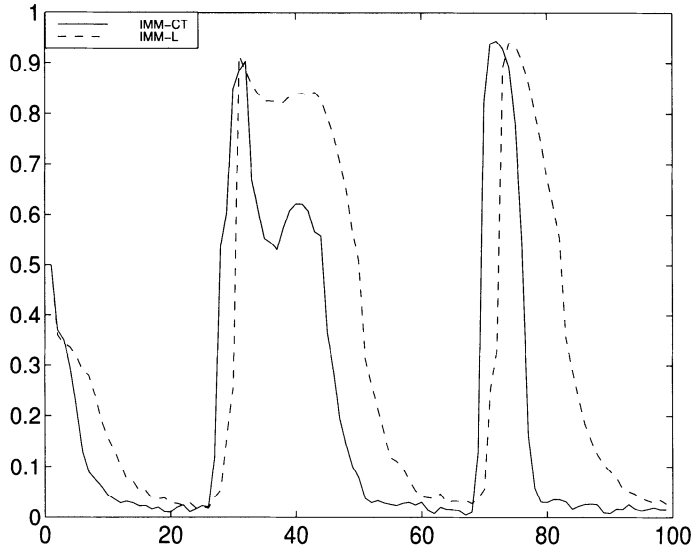


Figure 11.7.5-2: Mode 2 probabilities in IMM-L and IMM-CT estimators.

Sensitivity of the IMM to the Design Parameters

When the IMM uses several linear models with different process noise variances, a large ratio of these variances between models is needed; otherwise the models cannot be "distinguished" by the mode probability evaluator of the IMM (if they are the same, the IMM becomes a disguised KF).

Values of the maneuver onset probability between 0.05 and 0.1 yield similar results. Overall, the sensitivity of the performance to such changes in the transition probabilities are moderate.

Variable sampling interval, which is common in real systems, can be accounted for easily — see [Bar-Shalom95].

Some Remarks

The main findings from this example are as follows:

- The KF is clearly inferior to any (reasonably designed) IMM estimator.
- The accuracy of the turn rate estimate is not very important as far as the quality of the position, speed, and course estimates are concerned. What is important is the correct and timely detection of the maneuver and the fast response of the filter to this detection.
- The IMM-CT estimator is the best choice for tracking maneuvering with its capability to track the linear as well as turn motion of the target.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

When the aircraft is in uniform motion, the estimates given by the filter(s) based on the nearly coordinated turn model(s) have little effect on the overall (combined) estimate of the IMM algorithm (and on the interaction estimates for the filter based on the linear model at the next cycle) because their estimates have small weights. Consequently, while the aircraft is indeed in uniform motion, the filter matched to the uniform motion model is dominant and yields accurate estimates.

Once the aircraft starts to maneuver, the filter matched to the coordinated turn model takes over rapidly, leading to a significant reduction in the peak estimation errors. The IMM-CT estimator performs better than the KF and the IMM-L estimators during maneuvers.

A noise level switching version of the KF (with one model at a time: running a filter with a low variance process noise during the constant velocity portions of the trajectory and, following a maneuver detection, switching to a filter with a high level process noise) is not practical due to the very short duration of the maneuver. Also, in real systems where the origin of the measurements is uncertain, the "switching" technique is totally inadequate [Bar-Shalom95].

In view of the nature of maneuvers, which are not zero-mean white random processes, as the standard assumptions of the KF require, one cannot expect any of the filters to be perfectly consistent. The judgment as to which estimator design is best is a somewhat subjective one, and it should take into account the following:

- The performance (RMS errors) in both position and velocity (or speed and course): the maximum error and the average during the uniform motion
- Our desire for filter consistency
- The complexity of the implementation

Based on the above consideration, the IMM-CT is the best choice.

11.8 WHEN IS AN IMM ESTIMATOR NEEDED?

In Section 11.7.4 the design of an IMM estimator for a generic ATC problem was presented and it was shown that the IMM estimator performed significantly better than a Kalman filter. Specifically, the use of an IMM estimator with two filter modules, namely, a white-noise acceleration model and a coordinated turn model, resulted in about 50% reduction in RMS position errors vs. the Kalman filter. This improvement is due to the adaptive bandwidth (or the dynamic range) capability of the IMM estimator, that is, its ability to handle uniform motion (in which case it acts as a narrowband lowpass filter), as well as high-frequency maneuver motion (in which case it acts as wideband lowpass filter).

The above results raise an important question: When is an adaptive estimator — in particular, the IMM estimator — necessary? Intuitively, the higher the uncertainty — that is, the higher the maneuverability of the target — the more

Meta 1018
p. 207

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

a versatile tracker like the IMM estimator is needed. In the sequel we present an answer to the above question using simulation-based estimation results.

It is shown below that the choice between a (single model) Kalman filter and a multiple model tracker like the IMM estimator depends on the maneuvering index λ of the target, defined in (6.5.3-14). That is, it is shown that above a certain maneuvering index threshold, the performance of the IMM estimator, in terms of estimation errors, is significantly better than that of a Kalman filter, and therefore the former's use is recommended above that particular maneuvering index.

11.8.1 Kalman Filter vs. IMM Estimator

The test scenario, which is similar to the one used in Section 11.7.4, is constructed as follows:

1. The sensor measures the two-dimensional Cartesian positions with RMS errors of $\sigma_w = 100$ m in each coordinate.
2. The revisit interval of the sensor is fixed at $T = 5$ s.
3. Initially, the target moves at nearly a constant velocity for 60 s with process noise standard deviation $\sigma_v = 0.2$ m/s^2, which yields a very small maneuvering index of $\lambda = 0.05$ in each coordinate.
4. Following this, the target executes a series of three maneuvers, each of which
   - lasts for 60 s
   - has a given (fixed) maneuvering index λ (with the corresponding white process noise)
   - is interleaved with nonmaneuvering periods lasting for 60s with $\lambda = 0.05$.

The level of maneuvering is varied from $\lambda = 0.1$ to 2.5, in increments of 0.1. That is, the process noise standard deviation ranges from $\sigma_v = 0.4$ m/s^2 to 10 m/s^2 for the above values of σ_w and T. The performance of the trackers based on Kalman filter and the IMM estimator are evaluated from 100 Monte Carlo runs for each of these maneuvering index values.

For the Kalman filter, a second order (WNA) linear kinematic model with σ_v corresponding to 0.8 of the maximum process noise standard deviation (during the maneuvering intervals) was used. This choice was made in order to retain as much as possible the ability to track maneuvering and nonmaneuvering intervals. Other values of σ_v yielded similar results.

The IMM estimator consisted of two WNA models designed as follows:

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

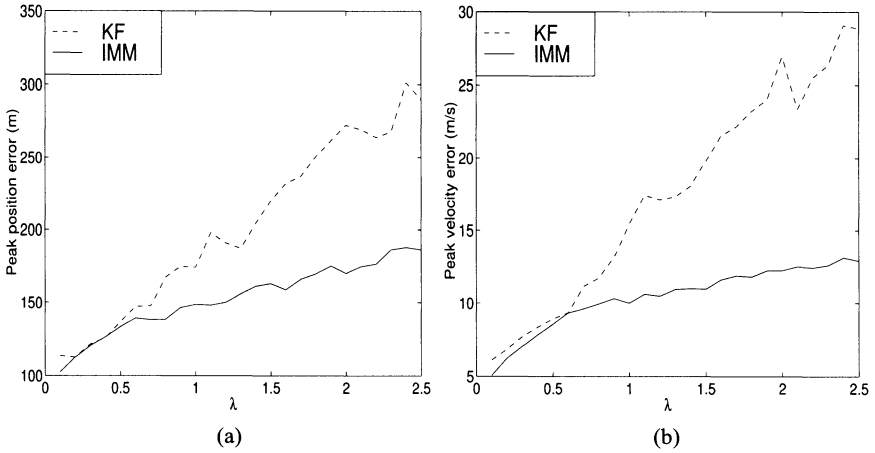


Figure 11.8.1-1: Peak RMS estimation errors. (a) Position errors, (b) velocity errors.

1. One model with low process noise with $\sigma_v = 0.2\,\text{m/s}^2$, the process noise standard deviation during the nonmaneuvering intervals.
2. One model with high process noise corresponding to the maneuver process noise standard deviation.
3. The mode transition probability matrix π was

$$\pi = \begin{bmatrix} 0.90 & 0.10 \\ 0.10 & 0.90 \end{bmatrix} \tag{11.8.1-1}$$

which corresponds to a sojourn time of about 60 s in each mode at $T = 5$ s.

Figure 11.8.1-1 shows the peak position and velocity errors obtained with the two estimators. It can be seen that both estimators result in nearly the same peak estimation errors up to $\lambda = 0.5$ while beyond that value of λ the IMM estimator yields significantly lower estimation errors.

Figure 11.8.1-2 shows the overall RMS position and velocity errors. Here also, one can notice that the IMM estimator results in significant error reduction over the Kalman filter for $\lambda > 0.5$.

Figure 11.8.1-3 shows the position and velocity RMS errors during the non-maneuvering intervals. It can be seen that the errors obtained with the IMM estimator remain almost constant, whereas those obtained with the Kalman filter increase with λ; the latter is unable to provide noise reduction during non-maneuvering intervals because, in order to reduce the overall error, its parameters are tuned to the maneuvering intervals.

This is a direct result of KF's lack of adaptive bandwidth capability. In contrast, the IMM estimator is able to maintain the noise reduction during non-maneuvering intervals even with higher levels of target maneuverability — by

Meta 1018
p. 209

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

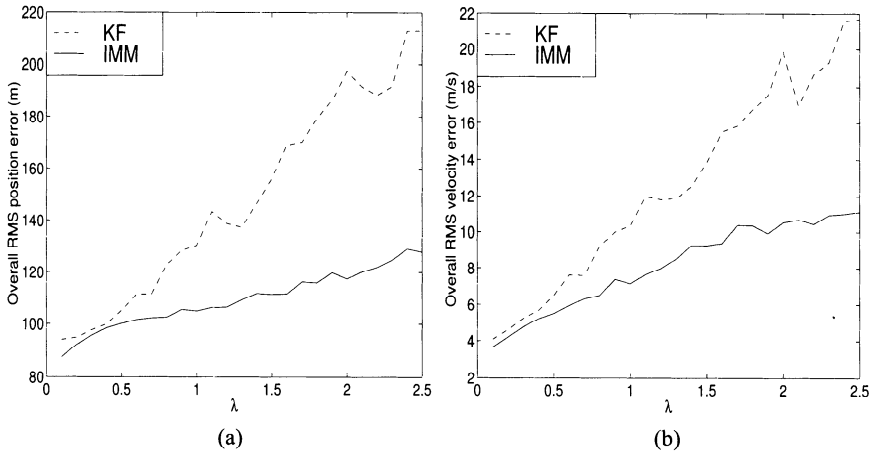


Figure 11.8.1-2: Overall RMS estimation errors. (a) Position errors, (b) velocity errors.

adapting the weights given to the two filter modules (mode probabilities), it gives consistently better results.

The above observation is confirmed further in Fig. 11.8.1-4, where the position and velocity RMS errors during the maneuvering intervals are shown. Notice that even with the Kalman filter's noise levels adapted to the increasing maneuverability, it is unable to match the IMM estimator's performance levels. Also notice that the breaking point for the Kalman filter is, once again, around $\lambda = 0.5$, up to which both estimators perform nearly the same.

These results indicate that below about $\lambda = 0.5$, the Kalman filter and the IMM estimator performs equally well. Other values of measurements error standard deviations and sampling intervals yielded similar results. Above this (approximate) threshold, the more sophisticated IMM estimator yields consistently better results. The ***adaptive bandwidth*** or ***dynamic range capability*** of the IMM estimator becomes a significant factor in the estimation process when the maneuverability of the target is sufficiently high. The adaptation via mode probability update helps the IMM estimator keep the estimation errors consistently low, both during maneuvers as well as during the benign motion intervals.

In summary, based on the above observations it is recommended that the ***maneuvering index threshold*** of $\lambda = 0.5$ be used as a guideline in making a choice between the Kalman filter and the IMM estimator.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

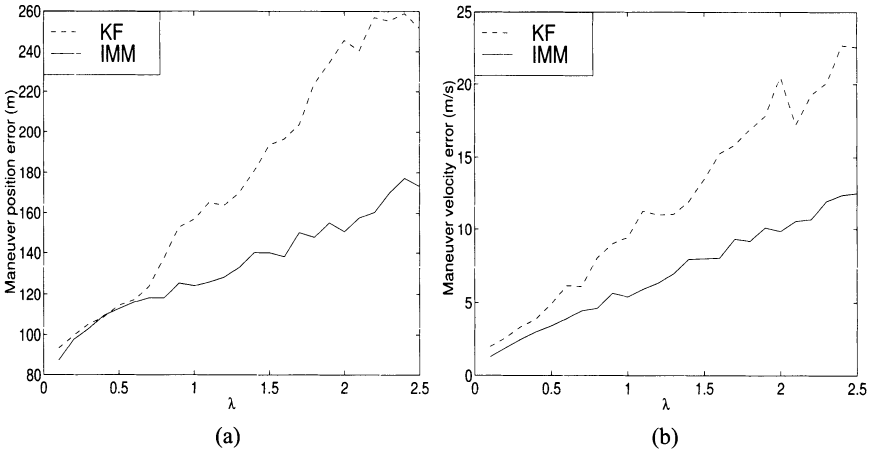


Figure 11.8.1-3: Uniform motion RMS estimation errors. (a) Position errors, (b) velocity errors.

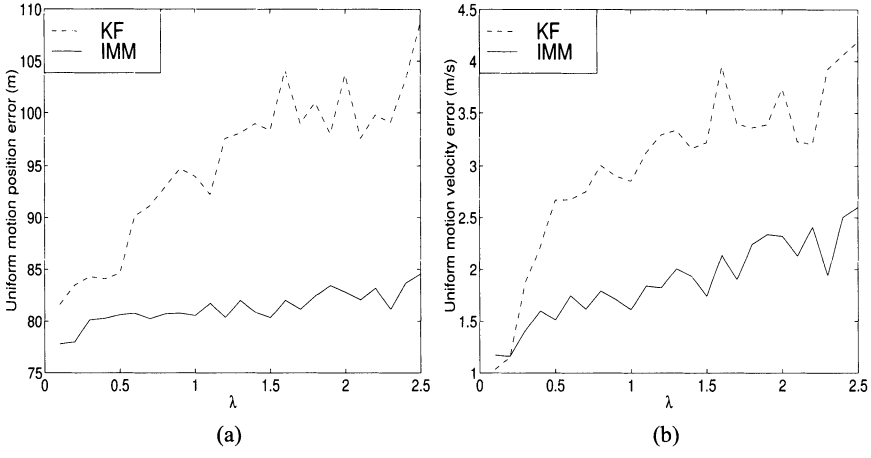


Figure 11.8.1-4: Maneuver RMS estimation errors. (a) Position errors, (b) velocity errors.

Meta 1018
p. 211

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

11.9 USE OF EKF FOR SIMULTANEOUS STATE AND PARAMETER ESTIMATION

11.9.1 Augmentation of the State

The extended Kalman filter can be used for suboptimal state estimation in nonlinear dynamic systems.

The situation of linear systems with unknown parameters that are continuous-valued can be put in the framework of nonlinear state estimation by augmenting the base state. The ***base state*** is the state of the system with the parameters assumed known.

Denoting the unknown parameters as a vector θ, the ***augmented state*** will be the ***stacked vector*** consisting of the base state x and θ

$$y(k) \triangleq \begin{bmatrix} x(k) \\ \theta \end{bmatrix} \tag{11.9.1-1}$$

The linear dynamic equation of x, with the known input u and process noise v,

$$x(k+1) = F(\theta)x(k) + G(\theta)u(k) + v(k) \tag{11.9.1-2}$$

and the "dynamic equation" of the parameter vector (assumed time invariant)

$$\theta(k+1) = \theta(k) \tag{11.9.1-3}$$

can be rewritten as a nonlinear dynamic equation for the augmented state

$$y(k+1) = f[y(k), u(k)] + v(k) \tag{11.9.1-4}$$

The resulting nonlinear equation (11.9.1-4) is then used for an EKF to estimate the entire augmented state.

The same technique can be used if the dynamic equation of the base state is nonlinear.

The model represented by (11.9.1-3) for the parameter dynamics assumes it to be constant. Therefore, the covariance yielded by the EKF will, asymptotically, tend to zero, and consequently the filter gain for these components will tend to zero. This is because there is no process noise entering into these components of the augmented state — the ***controllability condition*** 2 from Subsection 5.2.5, pertaining to the Riccati equation for the state estimation covariance, is not satisfied.

Since the EKF is not an optimal estimation algorithm, it will in general not yield consistent estimates for the parameters — the estimates will not converge to the true values. Thus the situation where the parameter variances tend to zero is undesirable because it will lead in practice to estimation errors much larger than the filter-calculated variances.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

This can be remedied (to some extent) by assuming an ***artificial process noise*** (or ***pseudo-noise***) entering into the parameter equation. This amounts to replacing (11.9.1-3) by

$$\theta(k+1) = \theta(k) + v_\theta(k) \tag{11.9.1-5}$$

where the parameter process noise is assumed zero mean and white. Thus the parameter is modeled as a (discrete time) Wiener process.

Any nonzero variance of this process noise will prevent the filter-calculated variances of the parameter estimates from converging to zero. Furthermore, this also gives the filter the ability to estimate ***slowly varying parameters***.

The choice of the variance of the artificial process noise for the parameters — the ***tuning of the filter*** — can be done as follows:

1. Choose the standard deviation of the process noise as a few percent of the (estimated/guessed) value of the parameter.
2. Simulate the system and the estimator with random initial estimates (for the base state as well as the parameters) and monitor the normalized estimation errors.
3. Adjust the noise variances until, for the problem of interest, the filter is consistent — it yields estimation errors commensurate with the calculated augmented state covariance matrix. The criteria to be used are those from Section 5.4 — the estimation bias and the normalized estimation error squared (NEES).

The example of the coordinated turn motion with an unknown rate from (11.7.1-2) to (11.7.1-4) falls into this category.

11.9.2 An Example of Use of the EKF for Parameter Estimation

Consider the scalar system, that is, its base state x is a scalar, given by

$$x(k+1) = a(k)x(k) + b(k)u(k) + v_1(k) \tag{11.9.2-1}$$

where $v_1(k)$ is the base state process noise and the two unknown parameters are $a(k)$ and $b(k)$, possibly time-varying.

The observations are

$$z(k) = x(k) + w(k) \tag{11.9.2-2}$$

Following the procedure of the previous subsection, the augmented state is

$$y(k) \triangleq \begin{bmatrix} y_1(k) \\ y_2(k) \\ y_3(k) \end{bmatrix} \triangleq \begin{bmatrix} x(k) \\ a(k) \\ b(k) \end{bmatrix} \tag{11.9.2-3}$$

With this the nonlinear dynamic equation corresponding to (11.9.2-1) can be written as

$$y_1(k+1) = f^1[y(k), u(k)] + v_1(k) \triangleq y_1(k)y_2(k) + y_3(k)u(k) + v_1(k) \tag{11.9.2-4}$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and the "dynamic equation" of the parameters is

$$y_i(k+1) = f^i[y(k), u(k)] + v_i(k) \triangleq y_i(k) + v_i(k) \qquad i = 2, 3 \quad (11.9.2\text{-}5)$$

The augmented state equation is then

$$y(k+1) = f[y(k), u(k)] + v(k) \quad (11.9.2\text{-}6)$$

with the augmented process noise

$$v(k) \triangleq \begin{bmatrix} v_1(k) \\ v_2(k) \\ v_3(k) \end{bmatrix} \quad (11.9.2\text{-}7)$$

assumed zero mean and with covariance

$$Q = \text{diag}(q_1, q_2, q_3) \quad (11.9.2\text{-}8)$$

With the linear measurements (11.9.2-2), the only place in the EKF where linearizations are to be carried out is the base state prediction.

The second-order EKF will use the following augmented state prediction equations. From (11.9.2-4) the base state prediction can be obtained directly as

$$\hat{y}_1(k+1|k) = \hat{y}_2(k|k)\hat{y}_1(k|k) + \hat{y}_3(k|k)u(k) + P_{21}(k|k) \quad (11.9.2\text{-}9)$$

since

$$E[a(k)x(k)|Z^k] = E[y_2(k)y_1(k)|Z^k] = \hat{y}_2(k|k)\hat{y}_1(k|k) + \text{cov}[y_2(k), y_1(k)|Z^k] \quad (11.9.2\text{-}10)$$

Note that the last term in (11.9.2-9) is the only second-order term.

The predicted values of the remaining two components of the augmented state, which are the system's unknown parameters, follow from (11.9.2-5) as

$$\hat{y}_i(k+1|k) = \hat{y}_i(k|k) \qquad i = 2, 3 \quad (11.9.2\text{-}11)$$

Equations (11.9.2-9) and (11.9.2-11), in general, have to be obtained using the series expansion of the EKF, which requires evaluation of the Jacobian of the vector f and the Hessians of its components. In the present problem, the Jacobian is

$$F(k) = \begin{bmatrix} \hat{y}_2(k|k) & \hat{y}_1(k|k) & u(k) \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (11.9.2\text{-}12)$$

and the Hessians are

$$f_{yy}^1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad f_{yy}^2 = 0 \qquad f_{yy}^3 = 0 \quad (11.9.2\text{-}13)$$

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Then, with the components of f given in (11.9.2-4) and (11.9.2-5), it can be easily shown (see problem 11-3) that using (10.3.2-4) for the augmented state, that is,

$$\hat{y}(k+1|k) = f[\hat{y}(k|k), u(k)] + \frac{1}{2}\sum_{i=1}^{n_x} e_i \text{tr}[f_{yy}^i P(k|k)] \qquad (11.9.2\text{-}14)$$

yields (11.9.2-9) and (11.9.2-11).

The prediction covariance of the base state can be obtained, using (10.3.2-6) as

$$\begin{aligned} P_{11}(k+1|k) \;=\; & \hat{y}_2(k|k)^2 P_{11}(k|k) + 2\hat{y}_2(k|k)\hat{y}_1(k|k)P_{21}(k|k) \\ & + 2\hat{y}_2(k|k)u(k)P_{13}(k|k) + \hat{y}_1(k|k)^2 P_{22}(k|k) + 2\hat{y}_1(k|k)u(k)P_{23}(k|k) \\ & + u(k)^2 P_{33}(k|k) + P_{21}(k|k)^2 + P_{22}(k|k)P_{11}(k|k) + q_1 \end{aligned} \qquad (11.9.2\text{-}15)$$

11.9.3 EKF for Parameter Estimation — Summary

The EKF can be used to estimate simultaneously

- the base state and
- the unknown parameters

of a system.

This is accomplished by stacking them into an augmented state and carrying out the series expansions of the EKF for this augmented state.

Since the EKF is a suboptimal technique, significant care has to be exercised to avoid filter inconsistencies. The filter has to be tuned with artificial process noise so that its estimation errors are commensurate with the calculated variances.

11.10 NOTES, PROBLEMS, AND TERM PROJECT

11.10.1 Bibliographical Notes

The "adjustment" of the process noise covariance, presented in Section 11.2, has become part of the Kalman filtering folklore. It can be found in [Jazwinski69, Jazwinski70]. Among its applications, the one presented in [Chang77] deals with state and parameter estimation for maneuvering reentry vehicles. In [Tenney77b, Tenney77a] similar ideas are used for passive tracking of maneuvering targets including reinitialization of the filter upon detection of the maneuver. In [Castella80] a continuous update of the process noise covariance is presented based on a fading memory average of the residuals. An adaptation of the filter gain based on the residuals' deviation from orthogonality is presented in [Hampton73]. In [Spingarn72] a least squares approach with exponential discount of older measurements is used for maneuvering targets. Estimation of the noise covariances has been discussed in, for example, [Li94] and [Myers76].

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The switching from one model to another with different noise parameters is discussed in [McAulay73] in a manner somewhat similar to Subsection 11.2.2.

The input estimation method of Section 11.3 is based on [Chan79]. Estimation of the input and the measurement noise covariance has been discussed in [Moghaddamjoo86]. The generalized likelihood ratio technique [Willsky76b, Korn82] deals simultaneously with the estimation of the input and its onset time. Such an approach was taken by [Bogler87], who augmented the IE technique of [Chan79] to include maneuver onset time estimation; however, the resulting algorithm was very costly and performed less well than the IMM [Bar-Shalom89]. In [Demirbas87] an approach consisting of hypothesis testing via dynamic programming is proposed.

The variable dimension approach to tracking maneuvering targets discussed in Section 11.4 is from [Bar-Shalom82], which also presented the comparison of algorithms discussed in Section 11.5. An extensive comparison of various maneuver detection approaches is described in [Woolfson85].

A weighted state estimate, along the lines of Section 11.6, is presented in [Thorp73]. The multiple model (MM) approach was originally presented in [Magill65]. The Markov switching of models is discussed in [Moose73, Moose75, Moose79, Moose80, Gholson77]. The generalized pseudo-Bayesian (GPB) algorithms are proposed by [Ackerson70, Jaffer71b, Jaffer71a]. The GPB2 MM approach is based on [Chang78]. A survey of the MM techniques and their connection with failure detection is given in [Tugnait82]. Recent work in failure detection using the static MM estimator includes [Hanlon98]. The use of the IMM for failure detection, which is substantially more effective than the static MM estimator, was presented in [Zhang98]. Related work in the area of failure detection is [Willsky76a, Caglayan80, Kerr89a]; the connection between failure detection, multiobject tracking and general modeling uncertainties is discussed in [Pattipati83]. Detection-estimation algorithms that approximate the optimal algorithm of Subsection 11.6.3 are discussed in [Hadidi78, Tugnait79]. The Interacting MM algorithm is from [Blom84, Blom88]. The application of the IMM to air traffic control can be found in [Bar-Shalom98b] (Chapters 1 and 2). Section 11.7 is based on [Wang99]. Reference [Li93] discusses how one can predict the performance of an IMM algorithm without Monte Carlo simulations and illustrates this technique on examples, including a simplified ATC problem. The generalization of the IMM to second-order Markov systems was presented in [Blom85]. Smoothing with an IMM was discussed in [Helmick95]. A survey on the IMM and its numerous applications can be found in [Mazor98]. The book [Sworder99] presents a mathematical treatise on hybrid systems. A survey of MM methods can be found in [Li96].

In [Li95] a theoretical framework for evaluating the performance of hybrid estimation algorithms (with continuous and discrete uncertainties) is presented.

Adaptive sampling techniques for a maneuvering target are presented in [Blackman81]. This topic is discussed in the larger context of radar resource allocation in [Blair98, Kirubarajan98]. The use of target orientation measurements for tracking maneuvering targets is explored in [Kendrick81, Lefas84, Andrisani86, Sworder89].

11.10.2 Problems

11-1 CRLB for a two-model parameter estimation problem. Consider an estimation problem where the noise behaves according to one of two models. Using the binary random

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

variable α, the observation can be written as

$$z = x + \alpha w_1 + (1 - \alpha) w_2$$

where x is the (unknown constant) parameter to be estimated,

$$P\{\alpha = 1\} = p_1 \qquad P\{\alpha = 0\} = p_2 = 1 - p_1$$

with $w_i \sim \mathcal{N}(0, \sigma_i^2)$ independent of each other and of α. In other words, the measurement has with probability p_1 accuracy σ_1 and with probability p_2 accuracy σ_2.

1. Write the likelihood function of x, $\Lambda(x) = p(z|x)$.
2. Find $\hat{x}^{\mathrm{ML}}$.
3. Evaluate the corresponding MSE.
4. Evaluate the CRLB for this estimation problem assuming $p_1 = 0.5$, $\sigma_1 = 1$, $\sigma_2 = 100$. (*Hint*: Use the fact that $\sigma_1 \ll \sigma_2$ to approximate the integral that requires otherwise numerical evaluation.)
5. Compare the results of items 3 and 4 for the above numbers. Comment on the usefulness of the bound in this problem.

11-2 Two-model parameter estimation problem with a prior. Given the random parameter $x \sim \mathcal{N}(x_0, \sigma_0^2)$ to be estimated based on the same measurement as in problem 11-1.

1. Find the MMSE estimate of x given z.
2. Find the conditional variance of x given z.

Assume x, α and w_i are all independent.

11-3 EKF prediction equations for a system with unknown parameters.

1. Derive the predicted augmented state from (11.9.2-14).
2. Prove (11.9.2-15) and derive the remaining terms of this covariance.

11-4 EKF simulation for state and parameter estimation. Consider the system (11.9.2-1) with $x(0) = 0$, $a(k) = 0.5$, $b(k) = 0$, process noise zero mean with variance $q_1 = 0.09$. The observations are given by (11.9.2-2) with measurement noise zero mean with variance $r = 0.09$. The covariance matrix of the initial augmented state estimate is $P(0|0) = \mathrm{diag}(0.09, 0.04)$.

1. Implement an EKF for this problem (to estimate x, a, and b) without an artificial process noise. Initialize with random initial estimates according to $P(0|0)$. List the normalized estimation error squared (NEES) for each component of the augmented state for 100 steps.
2. Perform 100 Monte Carlo runs for the above. List the averages of the normalized estimation error for bias check and the NEES for each component for 100 steps. Indicate the corresponding 95% confidence regions.
3. If the filter needs tuning, do it and present the results. Compare also the absolute RMS estimation errors with those from (2).

11-5 Fading memory average. Prove (11.2.2-7).

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

11-6 Sojourn time in a state for a Markov chain. Prove

$$E[\tau_i] = \frac{1}{1 - p_{ii}}$$

where τ_i is the expected sojourn time (in units of the sampling interval) of a Markov chain in state i (mode i) and p_{ii} is the transition probability from state i to state i.

11-7 Hybrid system with i.i.d. modal state sequence.

1. Find the Markov chain transition matrix that, regardless of the initial condition, yields (in one step) the fixed probability vector $[p_1 \; p_2]'$.

2. Given that the mean sojourn time (MST) in state 1 is τ_1 periods, find, for the above case, τ_2, the MST in state 2.

11-8 Direct mode observation in a hybrid system. Consider the prior pdf of x as

$$p(x) = \sum_{i=1}^{n} \mu_i \delta(x - x_i)$$

and the observation

$$z = x + w \qquad w \sim \mathcal{N}(0, 1)$$

Find the closed-form expressions of the posterior weights $\hat{\mu}_i$ conditioned on z.

11-9 Fault detection. Consider a direct discrete-time WNA motion model in one coordinate with $T = 1$. Let the measurement be

$$z = \begin{bmatrix} h_{11} & 0 \\ 0 & h_{22} \end{bmatrix} x + w$$

with all noises zero-mean white with unity variance.

The system in the normal mode has $h_{11} = 1$ and $h_{22} = 1$. The failure modes are:

FM1. $h_{11} = 0$
FM2. $h_{22} = 0$
FM3. $h_{11} = 0$ and $h_{22} = 0$.

The system starts in the normal mode, then goes into FM1 at $k = 10$, recovers at $k = 20$, goes into FM2 at $k = 30$, recovers at $k = 40$, goes into FM3 at $k = 50$, and recovers at $k = 60$.

Let the initial condition be

Scenario A. $x = [0 \; 0]'$
Scenario B. $x = [0 \; 10]'$

both with identity matrix initial covariance.

1. Design an IMM estimator for this system. Specify the design parameters and implement it with DynaEst^{TM}.

2. Based on 100 runs with the same failure sequence (but independent noises and random initial state estimates), find the average number of samples it takes to detect each failure mode and recovery for the two scenarios. Define what you mean by failure mode detection. Explain the difference in performance between the two scenarios.

Meta 1018
p. 218

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

11-10 Maneuvering index threshold vs. debiasing threshold. Show that the maneuvering index threshold $\lambda = 0.5$ from Section 11.8 is (almost) equivalent to the debiasing threshold (10.4.3-17).

11.10.3 Term Project — IMM Estimator for Air Traffic Control

Design IMM estimators for the following air traffic control problem:

The Ground Truth is a target moving with a constant speed of 250 m/s with initial state in Cartesian coordinates (with position in m)

$$x = [\xi \;\; \dot{\xi} \;\; \eta \;\; \dot{\eta}]' = [0 \;\; 0 \;\; 0 \;\; 250]'$$

The sampling period is $T = 10$ s. At $k = 10$ ($t = 100$ s) it starts a left turn of 2°/s for 30 s, then continues straight until $k = 20$, at which time it turns right with 1°/s for 45 s, then left with 1°/ s for 90 s, then right with 1°/s for 45 s, then continues straight until $k = 50$. (The turns are not known to the filter!)

Measurements are made starting from $k = 0$ on the position of this target in polar coordinates (range r and azimuth θ) by a radar located at $[\xi_0, \eta_0] = [-10^4, 0]$, with

$$r = \sqrt{(\xi - \xi_0)^2 + (\eta - \eta_0)^2} \qquad \theta = \tan^{-1}\left(\frac{\eta - \eta_0}{\xi - \xi_0}\right)$$

with additive white Gaussian noise with covariance $R = \text{diag}[2500\,\text{m}^2, (1°)^2]$.

Design the following two IMM estimators:

(A) IMM-L consisting of two (linear) WNA models (Section 6.3.2), with process noise (assumed zero-mean white Gaussian) with variances $\sigma^2_{v_m}$, $m = 1, 2$. These, together with the Markov chain transition matrix between the models, are the IMM estimator design parameters.

(B) IMM-CT consisting of one WNA model and a coordinated turn model with the turn rate as an additional state component. The CT model has an additional design parameter, the turn rate process noise variance.

The IMM estimator is initialized from the measurements at $k = 0$ and $k = 1$ and starts running from $k = 2$. Each model has initial probability 0.5 and the same initial estimate.

1. Calculate the true state according to the above specifications for $k = 0, \ldots, 50$ (the truth evolves without noise).

2. Generate the noisy measurements along the trajectory.

3. Implement the IMM estimators for this problem. Indicate the rationale for the choice of the estimator design parameters.

4. Calculate the following average results for $N = 100$ runs for the two designs (try to have the estimated position RMS errors not to exceed the single measurement position RMS error during the maneuver — a few percent excess is OK — while having as much as possible

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

"noise reduction" in the straight portions of the trajectory):

$$
\begin{aligned}
\text{NORXE} &\triangleq \text{AVE}\{\tilde{x}_1(k|k)/\sqrt{P_{11}(k|k)}\} \\
\text{FPOS} &\triangleq \sqrt{\text{AVE}\{P_{11}(k|k)+P_{33}(k|k)\}} \\
\text{FVEL} &\triangleq \sqrt{\text{AVE}\{P_{22}(k|k)+P_{44}(k|k)\}} \\
\text{RMSPOS} &\triangleq \text{RMS position error (}\textit{Note}\text{: both coordinates combined)} \\
\text{RMSVEL} &\triangleq \text{RMS velocity error (}\textit{Note}\text{: both coordinates combined)} \\
\text{RMSSPD} &\triangleq \text{RMS speed error (}\textit{Note}\text{: speed} \triangleq \text{magnitude of velocity vector)} \\
\text{RMSCRS} &\triangleq \text{RMS course error (}\textit{Note}\text{: course} \triangleq \text{direction of velocity vector)} \\
\text{NEES} &\triangleq \text{AVE}\{\tilde{x}(k|k)'P(k|k)^{-1}\tilde{x}(k|k)\} \\
\text{MOD2PR} &\triangleq \text{AVE}\{\text{model 2 probability}\}
\end{aligned}
$$

where

$$
\text{RMS}(y) \triangleq \sqrt{\frac{1}{N}\sum_{j=1}^{N}(y^j)^2} \qquad \text{AVE}(y) \triangleq \frac{1}{N}\sum_{j=1}^{N} y^j
$$

and y^j is the outcome of y in run j. Provide the expressions you used for calculating RMSPOS, RMSVEL, RMSSPD and RMSCRS. Indicate which is your final ("best") design.

5. Indicate the distributions of NORXE and NEES and their 95% probability regions. Justify the final choice of the design parameters. Summarize the effect of the Markov chain transition matrix on the final results based on the designs illustrated in item 4.

6. Plot for both designs RMSPOS and FPOS; RMSVEL and FVEL; NORXE with its probability region; RMSSPD; RMSCRS; NEES with its probability region; and MODPR. Comment on the estimator bias and consistency.

7. Repeat item 4 for the best single-model filter (a standard KF) of your choice (a single design). Indicate how the design parameter was chosen and comment on its performance as compared to the best IMM.

DELIVERABLES:

A concise report is due.

Each student will have to make a 10- to 15-min. presentation in class. It is suggested to prepare 8–10 viewgraphs for this presentation, describing the designs and performance plots. The most important plots are:

- A single run with the true and estimated trajectories (best IMM). Plot the true trajectory as a solid line, marking the position (+) at the sampling time; the measurement (□) connected with a line to the corresponding true position; and the updated position (○), also connected with a line to the corresponding true position at each sampling time. Provide magnified plots of the turn portions.

and the following comparison plots:

- RMSPOS for the two IMMs (RMSPOSIMM-L, RMSPOSIMM-CT) and the KF (RMSPOSKF), together, for comparison; also, indicate the (baseline) raw unfiltered position error (RMSPOSRAW).
- RMSSPD for the two IMMs and the KF (together, for comparison).
- RMSCRS for the two IMMs and the KF (together, for comparison).

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

- NEES for the two IMMs and the KF (together, for comparison).
- MOD2PR for the two IMMs (together, for comparison).

Develop an incremental animated display for the tracker with the following specifications:

- Plot the true position (black dot) every 0.1 s at a rate n times faster than real time, with the display speedup $n \in [1, 10]$. This high density sequence of points should appear as a continuous line.
- Each position measurement should appear as a box "□" connected with a red line to the true position at the corresponding time.
- The display of the measurement should be delayed from its true time by a constant Δ, which is independent of the display speedup. Use $\Delta = 0.1$ s to start with, then adjust to make the display more appealing if necessary.
- Each updated position should appear as bullet "•", the starting point of a black arrow of length and direction given by the estimated velocity vector multiplied by the sampling interval T. The tip of this arrow will be then the predicted measurement for the next time under a linear motion model.
- The starting point of the arrow should be connected with a blue line to the true position at the corresponding time.
- The display of the measurement should be delayed from its true time by a constant 2Δ, which is independent of the display speedup. Use $\Delta = 0.1$ s to start with, then adjust to make the display more appealing if necessary.
- You can use your creativity to provide sound effects (e.g., at each measurement time) to keep the audience entertained.

The display should grow incrementally, rather than being restarted at each sampling time.

Meta 1018
p. 221

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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)

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Bibliography

[Ackerson70] G. A. Ackerson and K. S. Fu. On State Estimation in Switching Environments. *IEEE Trans. Automatic Control*, AC-15(1):10–17, Jan. 1970.

[Aidala83] V. J. Aidala and S. E. Hammel. Utilization of Modified Polar Coordinates for Bearings-Only Tracking. *IEEE Trans. Automatic Control*, AC-28:283–293, March 1983.

[Anderson79] B. O. D. Anderson and J.B. Moore. *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, NJ, 1979.

[Andrisani86] D. Andrisani, F. P. Kuhl, and D. Gleason. A Nonlinear Tracker Using Attitude Measurements. *IEEE Trans. Aerospace and Electronic Systems*, AES-22:533–539, Sept. 1986.

[Athans68] M. Athans, R. P. Wishner, and A. Bertolini. Suboptimal State Estimation for Continuous-Time Nonlinear Systems from Discrete Noisy Measurements. *IEEE Trans. Automatic Control*, AC-13:504–514, Oct. 1968.

[Baheti90] R. S. Baheti, A. J. Laub, and D. M. Wiberg (Eds.). Bierman Memorial Issue on Factorized Estimation Applications. *IEEE Trans. Automatic Control*, AC-35:1282–1319, Dec. 1990.

[Balakrishnan84] A. V. Balakrishnan. *Kalman Filtering Theory*. University Series in Modern Engineering, Springer, Berlin, Germany, 1984.

[Bar-Shalom82] Y. Bar-Shalom and K. Birmiwal. Variable Dimension Filter for Maneuvering Target Tracking. *IEEE Trans. Aerospace and Electronic Systems*, AES-18(5):621–629, Sept. 1982.

[Bar-Shalom83] Y. Bar-Shalom and K. Birmiwal. Consistency and Robustness of PDAF for Target Tracking in Cluttered Environments. *Automatica*, 19:431–437, July 1983.

[Bar-Shalom88] Y. Bar-Shalom and T. E. Fortmann. *Tracking and Data Association*. Academic Press, New York, 1988.

[Bar-Shalom89] Y. Bar-Shalom, K. C. Chang, and H. A. P. Blom. Tracking a Maneuvering Target Using Input Estimation vs. the Interacting Multiple Model Algorithm. *IEEE Trans. Aerospace and Electronic Systems*, AES-25:296–300, March 1989.

[Bar-Shalom95] Y. Bar-Shalom and X. R. Li. *Multitarget-Multisensor Tracking: Principles and Techniques*. YBS Publishing, 1995.

[Bar-Shalom98a] Y. Bar-Shalom (Ed.). *Multitarget-Multisensor Tracking: Advanced Applications*. Artech House, Norwood, MA, 1990. Reprinted by YBS Publishing, 1998.

[Bar-Shalom98b] Y. Bar-Shalom (Ed.). *Multitarget-Multisensor Tracking: Applications and Advances*, Vol. II. Artech House, Norwood, MA, 1992. Reprinted by YBS Publishing, 1998.

[Baram78] Y. Baram and N. R. Sandell. Consistent Estimation on Finite Parameter Sets with Application to Linear Systems Identification. *IEEE Trans. Automatic Control*, AC-23(3):451–454, June 1978.

[Barniv85] Y. Barniv. Dynamic Programing Solution to Detecting Dim Moving Targets. *IEEE Trans. Aerospace and Electronic Systems*, AES-21:144–156, Jan. 1985.

[Barret90] I. Barret. *Synthese d'algorithmes de Poursuite Multi-Radars d'avions Civils Manoeuvrants*. PhD thesis, Ecole Nationale Superieure de l'aeronautique et de l'espace, Toulouse, France, 1990.

Meta 1018
p. 222

[Bellman61] R. Bellman. *Adaptive Control Processes: A Guided Tour*. Princeton University Press, Princeton, NJ, 1961.

[Benedict62] T. R. Benedict and G. W. Bordner. Synthesis of an Optimal Set of Radar Track-While-Scan Smoothing Equations. *IRE Trans. Automatic Control*, AC-7:27–32, July 1962.

[Bertsekas92] D. P. Bertsekas. *Linear Network Optimization: Algorithms and Codes*. The MIT Press, Cambridge, MA, 1992.

[Bethel87] R. E. Bethel and R. G. Rahikka. An Optimum First-Order Time Delay Tracker. *IEEE Trans. Aerospace and Electronic Systems*, AES-23:718–725, Nov. 1987.

[Bierman77] G. J. Bierman. *Factorization Methods for Discrete Sequential Estimation*. Academic Press, New York, 1977.

[Blackman81] S. S. Blackman, T. J. Broida, and M. F. Cartier. Applications of a Phased Array Antenna in a Multiple Maneuvering Target Environment. In *Proc. 20th IEEE Conf. Decision and Control*, San Diego, CA, Dec. 1981.

[Blackman86] S. S. Blackman. *Multiple Target Tracking with Radar Applications*. Artech House, Norwood, MA, 1986.

[Blackman99] S. S. Blackman and R. Popoli. *Design and Analysis of Modern Tracking Systems*. Artech House, Norwood, MA, 1999.

[Blair98] W. D. Blair, G. A. Watson, T. Kirubarajan, and Y. Bar-Shalom. Benchmark for Radar Resource Allocation and Tracking Targets in the Presence of ECM. *IEEE Trans. Aerospace and Electronic Systems*, AES-34, Oct. 1998.

[Blom84] H. A. P. Blom. An Efficient Filter for Abruptly Changing Systems. In *Proc. 23rd IEEE Conf. Decision and Control*, Las Vegas, NV, Dec. 1984.

[Blom85] H. A. P. Blom. An Efficient Decision-Making-Free Filter for Processes with Abrupt Changes. In *Proc. 7th IFAC Symp. Identif. and System Parameter Estimation*, York, U.K., July 1985.

[Blom88] H. A. P. Blom and Y. Bar-Shalom. The Interacting Multiple Model Algorithm for Systems with Markovian Switching Coefficients. *IEEE Trans. Automatic Control*, AC-33(8):780–783, Aug. 1988.

[Bogler87] P. L. Bogler. Tracking a Maneuvering Target Using Input Estimation. *IEEE Trans. Aerospace and Electronic Systems*, AES-23(3):298–310, May 1987.

[Braasch90] M. S. Braasch. A Signal Model for GPS. *Navigation: Journal of The Institute of Navigation*, 37(4), Winter 1990.

[Brookner98] E. Brookner. *Tracking and Kalman Filtering Made Easy*. John Wiley & Sons, New York, 1998.

[Brown72] R. G. Brown. Integrated Navigation Systems and Kalman Filtering: A Perspective. *Navigation: Journal of The Institute of Navigation*, 19(4), Winter 1972.

[Brown97] R. G. Brown and P. Y. C. Hwang. *Introduction to Random Signals and Applied Kalman Filtering*. John Wiley & Sons, New York, 1997.

[Broxmeyer64] C. Broxmeyer. *Inertial Navigation Systems*. McGraw-Hill, New York, 1964.

[Bryson75] A. Bryson and Y. C. Ho. *Applied Optimal Control*. Hemisphere, Cambridge, MA, 1975.

[Caglayan80] A. K. Caglayan. Simultaneous Failure Detection and Estimation in Linear Systems. In *Proc. 19th IEEE Conf. Decision and Control*, Albuquerque, NM, Dec. 1980.

[Carlson90] N. A. Carlson. Federated Square Root Filter for Decentralized Parallel Processes. *IEEE Trans. AES*, 26(3), May 1990.

[Carvalho97] H. Carvalho, P. Del Moral, A. Monin, and G. Salut. Optimal Nonlinear Filtering in GPS/INS Integration. *IEEE Trans. AES*, 33(3), July 1997.

[Castella80] F. R. Castella. An Adaptive Two-Dimensional Kalman Tracking Filter. *IEEE Trans. Aerospace and Electronic Systems*, AES-16:822–829, Nov. 1980.

Meta 1018
p. 223

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[Castella81] F. R. Castella. Tracking Accuracies with Position and Rate Measurements. *IEEE Trans. Aerospace and Electronic Systems*, AES-17:433–437, May 1981.

[Chan79] Y. T. Chan, A. G. C. Hu, and J. B. Plant. A Kalman Filter Based Tracking Scheme with Input Estimation. *IEEE Trans. Aerospace and Electronic Systems*, AES-15(2):237–244, March 1979.

[Chang77] C. B. Chang, R. H. Whiting, and M. Athans. On the State and Parameter Estimation for Maneuvering Reentry Vehicles. *IEEE Trans. Automatic Control*, AC-22(2):99–105, Feb. 1977.

[Chang78] C. B. Chang and M. Athans. State Estimation for Discrete Systems with Switching Parameters. *IEEE Trans. Aerospace and Electronic Systems*, AES-14(5):418–425, May 1978.

[Chang84] C. B. Chang, K. P. Dunn, and L. C. Youens. A Tracking Algorithm for Dense Target Environments. In *Proc. 1984 American Control Conf.*, San Diego, CA, June 1984.

[Chen84] C. T. Chen. *Linear System Theory and Design*. Holt, New York, 1984.

[Cortina91] E. Cortina, D. Otero, and C. E. D'Attelis. Maneuvering Target Tracking Using EKF. *IEEE Trans. Aerospace and Electronic Systems*, AES-27:155–158, Jan. 1991.

[Daum88] F. E. Daum. New Exact Nonlinear Filters. In J. C. Spall (Ed.), *Bayesian Analysis of Time Series and Dynamic Models*. Dekker, New York, 1988.

[Daum94] F. E. Daum. New exact nonlinear filters: Theory and applications. In *Proc. SPIE Conf.*, Vol. 2235, pages 636–649, Orlando, FL, April 1994.

[Daum95] F. E. Daum. Beyond Kalman Filters: Practical Design of Nonlinear Filters. In *SPIE Proceedings*, Vol. 2561, July 1995.

[Demirbas87] K. Demirbas. Maneuvering Target Tracking with Hypothesis Testing. *IEEE Trans. Aerospace and Electronic Systems*, AES-23:757–766, Nov. 1987.

[Doucet00] A. Doucet, S. Godsill, and C. Andrieu. On Sequential Simulation-Based Methods for Bayesian Filtering. *Statistics and Computing*, 10(3):197–208, April 2000.

[Doucet01] A. Doucet, J. F. G. de Freitas, and N. J. Gordon. *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, New York, 2001.

[Dougherty95] J. J. Dougherty, H. El-Shaerief, D. J. Simon, and G. A. Whitmer. GPS Modeling for Designing Aerospace Vehicle Navigation Systems. *IEEE Trans. AES*, 31(2), April 1995.

[Drummond93] O. E. Drummond and G. Frenkel. Glossary of Tracking Terms of SDI Panels on Tracking. In *Proc. SPIE Conf. on Signal and Data Processing of Small Targets*, Vol. 1954, April 1993.

[Ekstrand83] B. Ekstrand. Analytical Steady State Solution for a Kalman Tracking Filter. *IEEE Trans. Aerospace and Electronic Systems*, AES-19:815–819, Nov. 1983.

[Ekstrand85] B. Ekstrand. Analytical Steatdy-State Solution for a Continuous Time Kalman Filter. *IEEE Trans. Aerospace and Electronic Systems*, AES-21:746–750, Nov. 1985.

[Enge99] P. Enge and P. Misra (Eds.). Global Positioning System, Special Issue. *Proceedings of IEEE*, 87(1), 1999.

[Farina85] A. Farina and F. A. Studer. *Radar Data Processing,* Vol. I: *Introduction and Tracking,* Vol. II: *Advanced Topics and Applications*. Research Studies Press, Letchworth, Hertfordshire, England, 1985.

[Fitzgerald74] R. J. Fitzgerald. Effects of Range-Doppler Coupling on Chirp Radar Tracking Accuracy. *IEEE Trans. Aerospace and Electronic Systems*, AES-10:528–532, July 1974.

[Fitzgerald80] R. J. Fitzgerald. Simple Tracking Filters: Filtering and Smoothing Performance. *IEEE Trans. Aerospace and Electronic Systems*, AES-16:860–864, Nov. 1980.

[Fitzgerald81] R. J. Fitzgerald. Simple Tracking Filters: Closed Form Solutions. *IEEE Trans. Aerospace and Electronic Systems*, AES-17:781–785, Nov. 1981.

[Fletcher87] R. Fletcher. *Practical Methods of Optimization*. John Wiley & Sons, Chichester, England, 1987.

Meta 1018
p. 224

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[Fogel88] E. Fogel and M. Gavish. Nth-order dynamics target observability from angle measurements. *IEEE Trans. Aerospace and Electronic Systems*, AES-24(3):305–308, May 1988.

[Fortmann77] T. E. Fortmann and K. L. Hitz. *An Introduction to Linear Control Systems*. Dekker, New York, 1977.

[Fraser69] D. C. Fraser and J. E. Potter. The Optimum Smoother as a Combination of Two Optimum Linear Filters. *IEEE Trans. Automatic Control*, AC-14(4):387–390, August 1969.

[Friedland69] B. Friedland. Treatment of Bias in Recursive Filtering. *IEEE Trans. Automatic Control*, AC-14:359–367, Aug. 1969.

[Friedland73] B. Friedland. Optimum Steady State Position and Velocity Estimation Using Noisy Sampled Position Data. *IEEE Trans. Aerospace and Electronic Systems*, AES-9:906–911, Nov. 1973.

[Galkowski91] P. J. Galkowski and M. A. Islam. An Alternative Derivation of the Modified Gain Function of Song and Speyer. *IEEE Trans. Automatic Control*, 36:1323–1326, Nov. 1991.

[Gao93] Y. Gao, E. J. Krakiwsky, and M. A. Abousalem. Comparison and Analysis of Centralized, Decentralized, and Federated Filters. *Navigation: Journal of The Institute of Navigation*, 40(1), Spring 1993.

[Gelb74] A. Gelb. *Applied Optimal Estimation*. MIT Press, Cambridge, MA, 1974.

[Gholson77] N. H. Gholson and R. L. Moose. Maneuvering Target Tracking Using Adaptive State Estimation. *IEEE Trans. Aerospace and Electronic Systems*, AES-13:310–317, May 1977.

[Gordon93] N. J. Gordon, D. J. Salmond, and A. F. M. Smith. A Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation. *IEE Proceedings on Radar and Signal Processing*, 140:107–113, July 1993.

[Gray93] J. Gray and W. Murray. A Derivation of an Analytic Expression for the Alpha-Beta-Gamma Filter. *IEEE Trans. Aerospace and Electronic Systems*, AES-29:1064 1065, July 1993.

[Grewal93] M. S. Grewal and A. P. Andrews. *Kalman Filtering: Theory and Practice*. Prentice-Hall, Englewood Cliffs, NJ, 1993.

[Hadidi78] M. T. Hadidi and S. C. Schwartz. Sequential Detection with Markov Interrupted Observations. In *Proc. 16th Allerton Conf. on Communication, Control and Computing*, Univ. of Illinois, Oct. 1978.

[Ham83] F. M. Ham and R. G. Brown. Observability, Eigenvalues and Kalman Filtering. *IEEE Trans. Aerospace and Electronic Systems*, AES-34(3):269–273, Mar. 1983.

[Hampton73] R. L. T. Hampton and J. R. Cooke. Unsupervised Tracking of Maneuvering Vehicles. *IEEE Trans. Aerospace and Electronic Systems*, AES-9:197–207, March 1973.

[Hanlon98] P. D. Hanlon and P. S. Maybeck. Interrelationship of Single-Filter and Multiple-Model Adaptive Algorithms. *IEEE Trans. Aerospace and Electronic Systems*, AES-34:934–946, July 1998.

[Helmick95] R. E. Helmick, W. D. Blair, and S. A. Hoffman. Fixed-Interval Smoothing for Markovian Switching Systems. *IEEE Trans. Information Theory*, 41:1845–1855, Nov. 1995.

[Itzkowitz89] H. R. Itzkowitz and R. S. Baheti. Demonstration of Square Root Kalman Filter on WARP Parallel Computer. In *Proc. 1989 American Control Conf.*, Pittsburgh, PA, June 1989.

[Jaarsma91] D. Jaarsma and R. A. Davis. Joint Estimation of Time Delay and Doppler Ratio from Broadband Using Discrete Estimation Techniques. In *Proc. 1991 International Conf. Acoustics, Speech, and Signal Processing*, Toronto, Canada, May 1991.

[Jaffer71a] A. G. Jaffer and S. C. Gupta. Optimal Sequential Estimation of Discrete Processes with Markov Interrupted Observations. *IEEE Trans. Automatic Control*, AC-16:417–475, Oct. 1971.

[Jaffer71b] A. G. Jaffer and S. C. Gupta. Recursive Bayesian Estimation with Uncertain Observation. *IEEE Trans. Information Theory*, IT-17:614–616, Sept. 1971.

[Jazwinski69] A. H. Jazwinski. Adaptive Filtering. *Automatica*, 5(4):475–485, July 1969.

Meta 1018
p. 225

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[Jazwinski70] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New York, 1970.

[Johnston72] J. Johnston. *Econometric Methods*. McGraw-Hill, New York, 1972.

[Kailath81] T. Kailath. *Lectures on Wiener and Kalman Filtering*. Springer-Verlag, New York, 1981.

[Kalata84] P. R. Kalata. The Tracking Index: A Generalized Parameter for Alpha-Beta-Gamma Target Trackers. *IEEE Trans. Aerospace and Electronic Systems*, AES-20:174–182, March 1984.

[Kalman60] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Trans. ASME, J. Basic Engineering*, 82:34–45, March 1960.

[Kalman61] R. E. Kalman and R. Bucy. New Results in Linear Filtering and Prediction Theory. *Trans. ASME, J. Basic Engineering*, 83:95–108, March 1961.

[Kayton90] M. Kayton (Ed.). *Navigation: Land, Sea, Air, and Space*. IEEE Press, New York, 1990.

[Kendrick81] J. D. Kendrick, P. S. Maybeck, and J. G. Reid. Estimation of Aircraft Target Motion Using Orientation Measurements. *IEEE Trans. Aerospace and Electronic Systems*, 17:254–260, March 1981.

[Kerr89a] T. H. Kerr. Duality Between Failure Detection and Radar/Optical Maneuver Detection. *IEEE Trans. Aerospace and Electronic Systems*, AES-25:520–528, July 1989.

[Kerr89b] T. H. Kerr. Status of CR-like Lower Bounds for Nonlinear Filtering. *IEEE Trans. Aerospace and Electronic Systems*, AES-25:590–601, September 1989.

[Kirubarajan96] T. Kirubarajan and Y. Bar-Shalom. Target Motion Analysis in Clutter for Passive Sonar Using Amplitude Information. *IEEE. Trans. Aerospace and Electronic Systems*, AES-32:1367–1384, Oct. 1996.

[Kirubarajan98] T. Kirubarajan, Y. Bar-Shalom, W. D. Blair, and G. A. Watson. IMMPDA Solution to Benchmark for Radar Resource Allocation and Tracking Targets in the Presence of ECM. *IEEE. Trans. Aerospace and Electronic Systems*, AES-34:1115–1134, Oct. 1998.

[Kirubarajan00] T. Kirubarajan and Y. Bar-Shalom. Kalman Filter vs. IMM Estimator: When Do We Need the Latter? In *Proc. SPIE Conf. Signal Processing, Sensor Fusion and Target Recogntion (#4048)*, Orlando, FL, April 2000.

[Kleinman89] D. L. Kleinman. Private communication, 1989.

[Korn82] J. Korn, S. W. Gully, and A. S. Willsky. Application of the Generalized Likelihood Ratio Algorithm to Maneuver Detection and Estimation. In *Proc. 1982 American Control Conf.*, Arlington, VA, June 1982.

[Kramer87] S. C. Kramer and H. W. Sorenson. Bayesian Parameter Estimation. In *Proc. 1987 American Control Conf.*, Minneapolis, MN, June 1987.

[Kramer88] S. C. Kramer and H. W. Sorenson. Recursive Bayesian Estimation Using Piece-wise Constant Approximations. *Automatica*, 24(6):789–801, Nov. 1988.

[Larson66] R. E. Larson and J. Peschon. A Dynamic Programming Approach to Trajectory Estimation. *IEEE Trans. Automatic Control*, AC-11:537–540, July 1966.

[Lefas84] C. C. Lefas. Using Roll-Angle Measurement to Track Aircraft Maneuvers. *IEEE Trans. Aerospace and Electronic Systems*, AES-20:672–681, Nov. 1984.

[Lerro93] D. Lerro and Y. Bar-Shalom. Tracking with Debiased Consistent Converted Measurements vs. EKF. *IEEE Trans. Aerospace and Electronic Systems*, AES-29(3):1015–1022, July 1993.

[Lewis86] F. L. Lewis. *Optimal Estimation*. John Wiley & Sons, New York, 1986.

[Li93] X. R. Li and Y. Bar-Shalom. Performance Prediction of the Interacting Multiple Model Algorithm. *IEEE Trans. Aerospace and Electronic Systems*, AES-29(3):755–771, July 1993.

[Li94] X. R. Li and Y. Bar-Shalom. A Recursive Multiple Model Approach to Noise Identification. *IEEE Trans. Aerospace and Electronic Systems*, AES-30(3):671–684, July 1994.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[Li95] X. R. Li and Y. Bar-Shalom. Performance Prediction of Hybrid Algorithms. In C. T. Leondes (Ed.), *Advances in Aerospace Systems Dynamics and Control Systems*, Vol. 72. Academic Press, New York, 1995.

[Li96] X. R. Li. Hybrid Estimation Techniques. In C. T. Leondes (Ed.), *Advances in Aerospace Systems Dynamics and Control Systems*, Vol. 76. Academic Press, New York, 1996.

[Li97] X. R. Li. Canonical Transformation for Tracking with Kinematic Models. *IEEE. Trans. Aerospace and Electronic Systems*, AES-33:1212–1224, October 1997.

[Liggins97] M. Liggins, C. Y. Chong, I. Kadar, M. G. Alford, V. Vannicola, and S. Thomopoulos. Distributed Fusion Architectures and Algorithms for Target Tracking. *Proceedings of IEEE*, 85(1), 1997.

[Lin01] X. Lin, T. Kirubarajan, Y. Bar-Shalom, and X. R. Li. Enhanced Accuracy GPS Navigation Using the Interacting Multiple Model Estimator. In *Proc. 2000 IEEE Aerospace Conf.*, Big Sky, MT, 2001.

[Ljung87] L. Ljung. *System Identification*. Prentice-Hall, Englewood Cliffs, NJ, 1987.

[Magill65] D. T. Magill. Optimal Adaptive Estimation of Sampled Stochastic Processes. *IEEE Trans. Automatic Control*, AC-10:434–439, 1965.

[Maybeck79] P. S. Maybeck. *Stochastic Models, Estimation and Control,* Vol. I. Academic Press, New York, 1979.

[Maybeck82] P. S. Maybeck. *Stochastic Models, Estimation and Control,* Vols. II and III. Academic Press, New York, 1982.

[Mazor98] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J. Dayan. Interacting Multiple Model Methods in Target Tracking: A Survey. *IEEE. Trans. Aerospace and Electronic Systems*, AES-34:103–123, Jan. 1998.

[McAulay73] R. J. McAulay and E. J. Denlinger. A Decision-Directed Adaptive Tracker. *IEEE Trans. Aerospace and Electronic Systems*, AES-9(2):229–236, March 1973.

[Meditch69] J. Meditch. *Stochastic Linear Estimation and Control*. McGraw-Hill, New York, 1969.

[Mehra71] R. K. Mehra. A Comparison of Several Nonlinear Filters for Reentry Vehicle Tracking. *IEEE Trans. Automatic Control*, AC-16:307–319, Aug. 1971.

[Melsa78] J. L. Melsa and D. L. Cohn. *Decision and Estimation Theory*. McGraw-Hill, New York, 1978.

[Mo98] L. Mo, X. Song, Y. Zhou, Z. Sun, and Y. Bar-Shalom. Unbiased Converted Measurements in Tracking. *IEEE. Trans. Aerospace and Electronic Systems*, AES-34:1023–1027, July 1998.

[Moghaddamjoo86] A. Moghaddamjoo and R. L. Kirlin. Robust Adaptive Kalman Filtering with Unknown Inputs. In *Proc. 1986 American Control Conf.*, Seattle, WA, June 1986.

[Moose73] R. L. Moose and P. L. Wang. An Adaptive Estimator with Learning for a Plant Containing Semi-Markov Switching Parameters. *IEEE Trans. Systems, Man and Cybernetics*, SMC-3:277–281, May 1973.

[Moose75] R. L. Moose. An Adaptive State Estimator Solution to the Maneuvering Target Problem. *IEEE Trans. Automatic Control*, AC-20(3):359–362, June 1975.

[Moose79] R. L. Moose, H. F. VanLandingham, and D. H. McCabe. Modeling and Estimation of Tracking Maneuvering Targets. *IEEE Trans. Aerospace and Electronic Systems*, AES-15(3):448–456, May 1979.

[Moose80] R. L. Moose, H. F. VanLandingham, and D. H. McCabe. Application of Adaptive State Estimation Theory. In *Proc. 19th IEEE Conf. Decision and Control*, Albuquerque, NM, Dec. 1980.

[Mutambara98] A. G. O. Mutambara. *Decentralized Estimation and Control for Multisensor Systems*. CRC Press, Boca Raton, LA, 1998.

Meta 1018
p. 227

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[Myers76] K. A. Myers and B. D. Tapley. Adaptive Sequential Estimation with Unknown Noise Statistics. *IEEE Trans. Automatic Control*, AC-21:520–523, Aug. 1976.

[Nardone81] S. C. Nardone and V. J. Aidala. Observability Criteria for Bearings-Only Target Motion Analysis. *IEEE Trans. Aerospace and Electronic Systems*, AES-17:162–166, March 1981.

[Navarro77] A. M. Navarro. General Properties of Alpha-Beta and Gamma Tracking Filters, 1977. Report PHL 1977-02, National Defense Research Organization, Physics Lab., The Hague, The Netherlands, Jan. 1977 (US Dept. of Commerce, National Technical Information Servies, Report 77-24347).

[Ng83] L. C. Ng and R. A. LaTourette. Equivalent Bandwidth of a General Class of Polynominal Smoothers. *J. Acoustical Society of America*, 74:814–816, Sept. 1983.

[O'Halloran88] D. R. O'Halloran and R. K. Baheti. Fast Mapping of a Kalman Filter on WARP, GE-CRD, Schenectady, NY, 1988.

[Pachter87] M. Pachter. Comments on "Analytical Steady-State Solution for a Continuous Time Kalman Filter". *IEEE Trans. Aerospace and Electronic Systems*, AES-23:596–597, July 1987.

[Papoulis84] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York, 1984.

[Parkinson96] B. W. Parkinson and J. J. Spilker (Eds.). *Global Positioning System: Theory and Applications*. AIAA, 1996.

[Pattipati83] K. R. Pattipati and N. R. Sandell. A Unified View of State Estimation in Switching Environments. In *Proc. 1983 American Control Conf.*, pages 458–465, 1983.

[Pitman62] G. R. Pitman (Ed.). *Inertial Guidance*. Wiley, New York, 1962.

[Poor88] H. V. Poor. *An Introduction to Signal Detection and Estimation*. Springer-Verlag, New York, 1988.

[Press92] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, New York, 1992.

[Raiffa72] H. Raiffa and R. Schlaifer. *Applied Statistical Decision Theory*. M.I.T. Press, Cambridge, MA, 1972.

[Rankin94] J. Rankin. GPS and Differential GPS: An Error Model for Sensor Simulation. In *Proceedings of 1994 IEEE Position, Location, and Navigation Symposium*, Las Vegas, NV, April 1994.

[Rauch65] H. E. Rauch, F. Tung, and C. T. Striebel. Maximum Likelihood Estimation of Linear Dynamic Systems. *AIAA Journal*, 3:1445–1450, Aug. 1965.

[Rhodes71] I. B. Rhodes. A Tutorial Introduction to Estimation and Filtering. *IEEE Trans. Automatic Control*, AC-16(6):608–704, Dec. 1971.

[Rogers88] S. R. Rogers. Steady State Performance of Decoupled Kalman Filter. In *Proc. NAECON*, Dayton, OH, May 1988.

[Sage71] A. Sage and J. Melsa. *Estimation Theory with Applications to Communications and Control*. McGraw-Hill, New York, 1971.

[Schmidt93] G. C. Schmidt. Designing Nonlinear Filters Based on Daum's Theory. *Journal of Guidance, Control, and Dynamics*, 16(2):371–376, March-April 1993.

[Schweppe73] F. Schweppe. *Uncertain Dynamic Systems*. Prentice-Hall, Englewood Cliffs, NJ, 1973.

[Singer70] R. A. Singer. Estimating Optimal Tracking Filter Performance for Manned Maneuvering Targets. *IEEE Trans. Aerospace and Electronic Systems*, AES-6:473–483, July 1970.

[Sittler64] R. W. Sittler. An Optimal Data Association Problem in Surveillance Theory. *IEEE Trans. Mil. Electron.*, MIL-8:125–139, April 1964.

[Sklansky57] J. Sklansky. Optimizing the Dynamic Parameters of a Track-While-Scan System. *RCA Review*, 18:163–185, June 1957.

Meta 1018
p. 228

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[Song85] T. L. Song and J. L. Speyer. A Stochastic Analysis of a Modified Gain Extended Kalman Filter with Applications to Estimation with Bearings Only Measurements. *IEEE Trans. Automatic Control*, AC-30:940–979, July 1985.

[Sorenson85] H. W. Sorenson (Ed.). *Kalman Filtering: Theory and Application*. IEEE Press, New York, 1985.

[Sorenson88] H. W. Sorenson. Recursive Estimation for Nonlinear Dynamic Systems. In J. C. Spall (Ed.), *Bayesian Analysis of Time Series and Dynamic Models*, chapter 6. Marcel Dekker, Inc., New York, 1988.

[Spingarn72] K. Spingarn and H. L. Weidemann. Linear Regression Filtering and Prediction for Tracking Maneuvering Aircraft Targets. *IEEE Trans. Aerospace and Electronic Systems*, AES-8:800–810, Nov. 1972.

[Stallard87] D. V. Stallard. An Angle-Only Tracking Filter in Modified Spherical Coordinates. In *Proc. 1987 AIAA Guid., Navig. and Control Conf.*, pages 542–550, 1987.

[Striebel65] C. Striebel. Sufficient Statistics in the Optimum Control of Stochastic Systems. *J. Math. Anal. and Appl.*, 12:576–592, Dec. 1965.

[Swerling59] P. Swerling. First Order Error Propagation in a Stagewise Smoothing Procedure for Satellite Observations. *J. Astronautical Sciences*, 12:46–52, Autumn 1959.

[Sworder89] D. D. Sworder and R. G. Hutchins. Image-Enchanced Tracking. *IEEE Trans. Aerospace and Electronic Systems*, AES-25:701–710, Sept. 1989.

[Sworder99] D. D. Sworder and J. E. Boyd. *Estimation Problems in Hybrid Systems*. Cambridge University Press, New York, 1999.

[Tenney77a] R. R. Tenney, T. B. Ballard, and L. E. Miller. A Dual Passive Tracking Filter for Maneuvering Targets. In *Proc. ONR Passive Target Tracking Conf.*, Naval Postgrad. School, Monterey, CA, May 1977.

[Tenney77b] R. R. Tenney, R. S. Hebbert, and N. R. Sandell. A Tracking Filter for Maneuvering Sources. *IEEE Trans. Automatic Control*, AC-22:246–251, April 1977.

[Thorp73] J. S. Thorp. Optimal Tracking of Maneuvering Targets. *IEEE Trans. Aerospace and Electronic Systems*, AES-9:512–519, July 1973.

[Tichavsky98] P. Tichavsky, C. H. Muravchik, and A. Nehorai. Posterior Cramer-Rao Bounds for Discrete-Time Nonlinear Filtering. *IEEE Trans. Signal Processing*, SP-46:1386–1396, May 1998.

[Tugnait79] J. K. Tugnait and A. H. Haddad. A Detection-Estimation Scheme for State Estimation in Switching Environments. *Automatica*, 15(4):477–481, July 1979.

[Tugnait82] J. K. Tugnait. Detection and Estimation for Abruptly Changing Systems. *Automatica*, 18(5):607–615, Sept. 1982.

[Van Trees68] H. L. Van Trees. *Detection, Estimation, and Modulation Theory, Part I*. John Wiley & Sons, New York, 1968.

[Wang99] H. Wang, T. Kirubarajan, and Y. Bar-Shalom. Precision Large Scale Air Traffic Surveillance Using an IMM Estimator with Assignment. *IEEE. Trans. Aerospace and Electronic Systems*, AES-35(1):255–266, Jan. 1999.

[Weiss80] H. Weiss and J. B. Moore. Improved Extended Kalman Filter Design for Passive Tracking. *IEEE Trans. Automatic Control*, AC-25:807–811, Aug. 1980.

[Willsky76a] A. S. Willsky. A Survey of Design Methods for Failure Detection in Dynamic Systems. *Automatica*, 12(6):601–611, Nov. 1976.

[Willsky76b] A. S. Willsky and H. L. Jones. A Generalized Likelihood Ratio Approach to the Detection and Estimation of Jumps in Linear Systems. *IEEE Trans. Automatic Control*, AC-21:108–112, Feb. 1976.

[Wishner70] R. P. Wishner, R. E. Larson, and M. Athans. Status of Radar Tracking Algorithms. In *Proc. Symp. Nonlinear Estimation*, San Diego, CA, Sept. 1970.

Meta 1018
p. 229

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[Woolfson85] M. S. Woolfson. An Evaluation of Manoeuvre Detector Algorithms. *GEC J. of Research (Chelmsford, England)*, 3(3):181–190, 1985.

[Yeddanapudi97] M. Yeddanapudi, Y. Bar-Shalom, and K. R. Pattipati. IMM Estimation for Multitarget-Multisensor Air Traffic Surveilance. *IEEE Proceedings*, 85(1):80–94, Jan 1997.

[Zadeh63] L. A. Zadeh and C. A. Desoer. *Linear System Theory*. McGraw-Hill, New York, 1963.

[Zhang98] Y. M. Zhang and X. R. Li. Detection and diagnosis of sensor and actuator failures using IMM estimator. *IEEE Trans. Aerospace and Electronic Systems*, AES-34:1293–1313, Oct. 1998.

Downloaded from https://onlinelibrary.wiley.com/doi/ by University Of Texas Libraries, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Meta 1018
p. 230