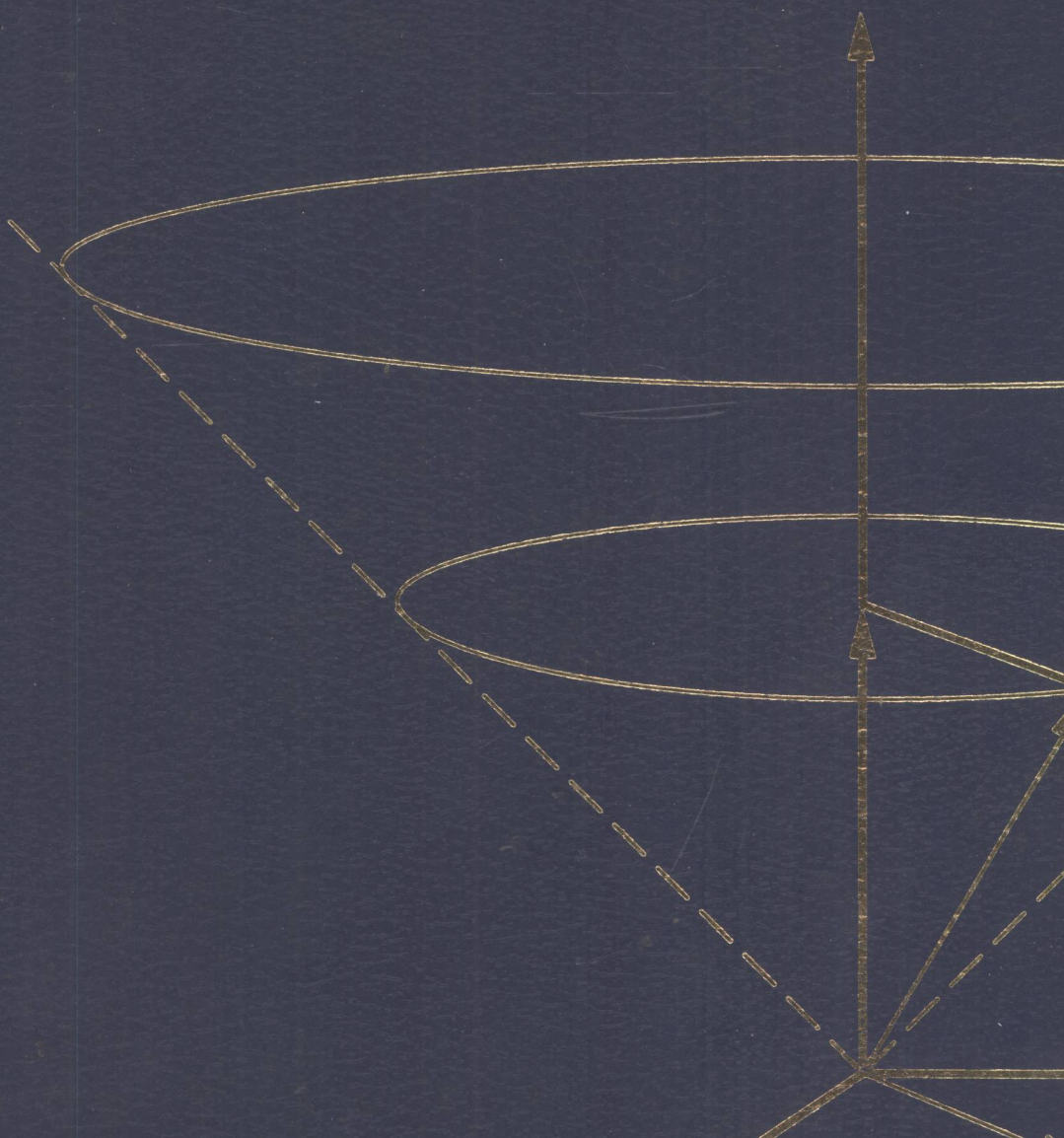


Statistical Signal Processing

Detection, Estimation, and Time Series Analysis

Louis L. Scharf



TN 971.33
S311

9260327

Statistical Signal Processing

Detection, Estimation, and Time Series Analysis

Louis L. Scharf

University of Colorado at Boulder

with Cédric Demeure collaborating on Chapters 10 and 11



E9260327



ADDISON-WESLEY PUBLISHING COMPANY

Reading, Massachusetts • Menlo Park, California • New York
Don Mills, Ontario • Wokingham, England • Amsterdam
Bonn • Sydney • Singapore • Tokyo • Madrid • San Juan

**This book is in the Addison-Wesley Series in Electrical and Computer Engineering:
Digital Signal Processing**

Consulting Editor: Richard Roberts

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and Addison-Wesley was aware of a trademark claim, the designations have been printed in initial caps or all caps.

The programs and applications presented in this book have been included for their instructional value. They have been tested with care, but are not guaranteed for any particular purpose. The publisher does not offer any warranties or representations, nor does it accept any liabilities with respect to the programs or applications.

Library of Congress Cataloging-in-Publication Data

Scharf, Louis L.

Statistical signal processing : detection, estimation, and time
series analysis / Louis L. Scharf.

p. cm.

Includes bibliographical references and index.

ISBN 0-201-19038-9

1. Signal processing—Statistical methods. I. Title.

TK5102.5.S3528 1990

621.382'2—dc20

90-42747

CIP

Reprinted with corrections July, 1991

Copyright ©1991 by Addison-Wesley Publishing Company, Inc.

All rights reserved. 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, or otherwise, without the prior written permission of the publisher. Printed in the United States of America.

2 3 4 5 6 7-MA-939291

To Carol, Greg, and Heidi

Preface



The field of statistical signal processing embraces the many mathematical procedures that engineers and statisticians use to draw inference from imperfect or incomplete measurements. The major domains of the field are detection, estimation, and time series analysis.

Abstractly, statistical signal processing is a theory for using experimental measurements to transform a prior model for a signal into a posterior model that may be used to make informed decisions. The quality of the decision is measured by a loss function that depends on ground truth and the decision taken. It is the intricate interplay between prior models, measurement schemes, loss functions, and decision rules that gives statistical signal processing its great variety.

ORGANIZATION AND USE

This book is my personal statement about the fundamental ideas in statistical signal processing. The book breaks down along four distinct topical lines: mathematical and statistical preliminaries; detection theory; estimation theory; and time series analysis. There is enough material to support a two-semester course in statistical signal processing, but the book may be used for separate one-semester courses in detection theory, estimation theory, or time series analysis. In a detection theory course, Chapters 1 through 5 may be covered in their entirety. In an estimation theory course, Chapters 1 through 3 and 6 through 8 may be covered. In a time series course, Chapters 1 through 3 and 9 through 11 are appropriate. Chapter 9 on least squares is a swing chapter that may be treated as a topic in estimation theory or time series analysis.

A GUIDED TOUR OF THE BOOK

Mathematical and Statistical Preliminaries

I begin in Chapter 2 with a fairly comprehensive review of linear algebra, matrix theory, and multivariate normal theory. Linear algebra, and the geometrical pictures that

bring life to it, forms the basis for our prior structural information about a signal. Matrix theory provides an algebra for manipulating and composing linear transformations, and multivariate normal theory provides the statistical methodology for computing the distribution of linear and quadratic forms in normal random vectors. When teaching from this chapter, I pick and choose from the topics, making sure to develop the ideas of linear independence, subspaces and their spans, orthogonal subspaces, QR factorizations, projections, the singular value decomposition, the multivariate normal distribution, and the distribution of quadratic forms in projection operators. In Chapter 3, I develop the main results in the theory of sufficient statistics and show the fundamental role they play in the computation of minimum variance unbiased estimators.

Detection Theory

Chapters 4 and 5 are dedicated to detection theory. In Chapter 4 I treat the many faces of the Neyman-Pearson theory of hypothesis testing. I cover the rudiments of decision theory, discuss the roles of sufficiency and invariance in hypothesis testing, and develop the theory of uniformly most powerful tests. I then apply the principles of sufficiency and invariance to signal detection when the signal model, or the noise model, is incompletely known. This produces matched filters, CFAR matched filters, matched subspace filters, and CFAR matched subspace filters. The final two sections of Chapter 4 treat reduced-rank detectors and linear discriminant functions for detecting Gaussian signals in Gaussian noise. Chapter 4 is long, so some instructors may wish to omit the linear discriminant functions and, perhaps, the sections on matched subspace filtering, although the latter is very important in the modern study of detection theory. In Chapter 5, I treat the Bayesian theory of hypothesis testing, wherein a prior distribution is assigned to the hypotheses under test. Minimax tests are constructed as Bayes tests against least favorable priors. The study of M -orthogonal symboling produces insights into channel capacity, and the study of composite matched filters produces insights into associative memories.

Estimation Theory

Chapters 6 through 9 are devoted to estimation theory. I begin in Chapter 6 with the maximum likelihood theory of parameter estimation, where I discuss the roles that sufficiency and invariance play in the maximum likelihood theory and discuss the Cramer-Rao bound for the variance of unbiased estimators. Nuisance parameters are explored in depth. In the last several sections of the chapter, I apply maximum likelihood theory to the identification of subspaces, to the identification of ARMA parameters, and to the identification of structured covariance matrices. Chapter 6 is long, so some instructors may want to give nuisance parameters a once-over-lightly, and select just a few of the applications. In Chapter 7 parameters are endowed with prior distributions, and the Bayes theory is developed for turning prior distributions into posterior distributions. The Bayes theory produces the Gauss-Markov

theorem for multivariate normal parameters and measurements. I interpret the Gauss-Markov theorem by showing how it transforms a channel model for measurements into an inverse channel model, or estimator-plus-noise model. The Gauss-Markov theorem is applied to sequential Bayes estimators, the Kalman filter, and the Wiener filter. In Chapter 8 we explore in more detail minimum mean-squared error estimators and the role that the conditional mean estimator plays. I derive conditional mean estimators to solve a number of problems in signal processing: low-rank Wiener filters, linear predictors, Kalman filters, low-rank approximations of random vectors, scalar and block quantizers, and reduced-rank block quantizers. The last application produces rate-distortion formulas of the Shannon variety. In Chapter 9 we develop the theory of least squares for estimating parameters in the linear statistical model and stress the singular value decomposition for the insight that it brings to least squares problems. We then study the performance of least squares when errors are normally distributed. This study produces order selection rules for reducing the rank of least squares estimators. The middle sections of the chapter are devoted to special topics such as sequential, weighted, constrained, and total least squares. The last several sections are devoted to applications: inverse problems, mode identification, parameter estimation in ARMA systems, linear prediction, and the identification of structured covariance matrices. Chapter 9 is long, so some instructors may want to select just a few of the applications.

Time Series Analysis

In Chapters 10 and 11 we cover linear prediction and modal analysis. In our treatment of linear prediction we begin with the classical stationary theory of Wold and Kolmogorov and establish the connection between linear prediction and maximum entropy extension of correlation sequences. We then develop the nonstationary theory of fitting order-increasing whiteners to finite covariance matrices, paying special attention to the Levinson and Schur recursions for computing the reflection coefficients that keep the recursions going. We study the least squares theory of linear prediction and derive the lattice recursions for computing reflection coefficients. Linear prediction in ARMA time series produces the MSK algorithms for fast Kalman filtering. In the last two sections of the chapter we apply linear prediction to the computation of likelihood and the design of a differential PCM system. In Chapter 11 we draw a distinction between linear prediction and modal analysis. We study Prony's method, exact least squares, the total least squares of Golub and Van Loan, the principal components of Tufts and Kumaresan, and MUSIC of Bienvenu and Schmidt as the most prominent techniques for estimating modes. We outline pencil methods and then present Kumaresan's procedure for estimating modes from frequency domain data.

ACKNOWLEDGMENTS

Since the early 70s I have had the opportunity to talk with Henry Cox and Ben Friedlander about many problems in signal processing. Both have generously volunteered interesting problems to explore, shared their own elegant solutions, and offered insight-

ful interpretations of others' solutions. In 1980 Claude Gueguen spent six months at Colorado State University, where he taught his course on Parametric Signal Processing. In this course, and in our collaborative research, I gained my first appreciation for the power of linear algebraic models as structural models for signals. I want to thank Claude for sharing with me his ideas and his deep understanding of system theory and signal processing. Since 1985 I have resided in the office adjacent to Tom Mullis. This has provided me with the opportunity to follow Tom's courses in linear algebra and spectrum analysis, to collaborate with him on several pieces of writing, and to attend his research seminars. I thank him for sharing his gift for clear thinking with me and for getting me out of a few tight spots in the final stages of writing this manuscript.

I first conceived this book in 1984, while teaching a graduate course in signal processing at the University of Rhode Island. During my stay at the University of Rhode Island I profited immensely from my association with Don Tufts. His insights profoundly influenced my own thinking, and I would like to acknowledge my intellectual debt to him. Ramdas Kumaresan improved my understanding of modal analysis, and Steve Kay directed me to a deeper understanding of signal detection in the linear statistical model. Many other people at the University of Rhode Island made my stay there a happy time of intellectual growth and professional development. I would like to acknowledge their collegiality.

Dick Roberts, late Professor of Electrical and Computer Engineering at the University of Colorado, reviewed my writing of several sections of the book. He encouraged me to complete it and recommended that I use his editor, Tom Robbins at Addison-Wesley. I followed Dick's advice, and as a consequence I have had the pleasure of working with Tom for the past 18 months, as I put the final touches on the manuscript. During this period, Lynn Kirlin used the manuscript at the University of Victoria and offered many improvements to the presentation.

For the past 20 years or so I have been able to maintain a single-minded interest in statistical signal processing because four program directors at the Office of Naval Research have supported our work. I thank Drs. Bruce McDonald, Doug Depriest, Ed Wegman, and Neil Gerr for their consistent management of ONR's programs in signal processing and mathematical statistics.

I thank my many students for their contributions to my understanding of signal processing. Cédric Demeure helped me write Chapters 10 and 11, so much so that they are as much his as mine. Finally, let me express my sincere thanks to Julie Fredlund, secretary to the group in digital signal processing at the University of Colorado, for her masterful and patient preparation of a demanding manuscript.

Louis Scharf
Boulder, CO

Contents

CHAPTER 1

Introduction 1

- 1.1 Statistical Signal Processing and Related Topics 1
- 1.2 The Structure of Statistical Reasoning 3
- 1.3 A Detection Problem 4
- 1.4 An Estimation Problem 7
- 1.5 A Time Series Problem 9
- 1.6 Notation and Terminology 11
 - Probability Distributions 11
 - Linear Models 12
- References and Comments 18
- Problems 19

CHAPTER 2

Rudiments of Linear Algebra and Multivariate Normal Theory 23

- 2.1 Vector Spaces 24
 - Euclidean Space 24
 - Hilbert Space 25
 - Matrices 25
- 2.2 Linear Independence 25
 - Gram Determinant 26*
 - Sequences of Gram Determinants 27
 - Cholesky Factors of the Gram Matrix 29

2.3	<i>QR</i> Factors	31
	Gram-Schmidt Procedure	32
	Householder Transformation	33
	Givens Transformation	36
2.4	Linear Subspaces	37
	Basis	38
	Direct Subspaces	38
	Unicity	38
	Dimension, Rank, and Nullity	39
	Linear Equations	40
	Decomposition of R^n	41
2.5	Hermitian Matrices	42
	The Eigenvalues of a Hermitian Matrix Are Real	42
	The Eigenvectors of a Hermitian Matrix Are Orthogonal	42
	Hermitian Matrices Are Diagonalizable	42
2.6	Singular Value Decomposition	43
	Range and Null Space	45
	Low Rank Approximation	45
	Resolution (or Decomposition) of Identity	46
2.7	Projections, Rotations, and Pseudoinverses	46
	Projections	47
	Rotations	48
	Pseudoinverse	49
	Orthogonal Representations	49
2.8	Quadratic Forms	51
2.9	Matrix Inversion Formulas	52
2.10	The Multivariate Normal Distribution	55
	Characteristic Function	56
	The Bivariate Normal Distribution	57
	Linear Transformations	59
	Analysis and Synthesis	60
	Diagonalizing Transformations	61
2.11	Quadratic Forms in MVN Random Variables	62
	Quadratic Forms Using Projection Matrices	64
	Asymptotics	66
	References and Comments	66
	Problems	67

CHAPTER 3

Sufficiency and MVUB Estimators 77

3.1	Sufficiency	78
	Discrete Random Variables	79
	Continuous Random Variables	82
	Nonsingular Transformations and Sufficiency	85

3.2 Minimal and Complete Sufficient Statistics	85
Minimality	87
Completeness	88
Unbiasedness	88
Completeness Ensures Minimality	88
Summary	89
3.3 Sufficiency and Completeness in the Exponential Family	89
Sufficiency	89
Completeness	90
3.4 Sufficiency in the Linear Statistical Model	91
Recursive Computation of the Sufficient Statistic in the Linear Statistical Model	91
Partitioned Matrix Inverse	92
Sufficient Statistic	92
White Noise	93
3.5 Sufficiency in the Components of Variance Model	93
3.6 Minimum Variance Unbiased (MVUB) Estimators	94
Interpretation	96
References and Comments	97
Problems	97

CHAPTER 4

Neyman-Pearson Detectors 103

4.1 Classifying Tests	104
4.2 The Testing of Binary Hypotheses	105
Size or Probability of False Alarm	105
Power or Detection Probability	106
Receiver Operating Characteristics (ROC)	107
4.3 The Neyman-Pearson Lemma	107
Choosing the Threshold	108
Interpretation	109
Geometrical Properties of the ROC Curve	109
North-by-Northwest: Birdsall's Insight	110
4.4 The Multivariate Normal Model	111
Uncommon Means and Common Covariance	111
Linear Statistical Model	115
Common Means and Uncommon Covariances	116
Uncommon Means and Uncommon Variances	117
4.5 Binary Communication	117
4.6 Sufficiency	121
4.7 The Testing of Composite Binary Hypotheses	123
Uniformly Most Powerful Test	124
4.8 Invariance	127
Invariance of an Hypothesis Testing Problem	128

Invariant Tests and Maximal Invariant Statistics	132
Uniformly Most Powerful Invariant Test	135
Reduction by Sufficiency and Invariance	135
4.9 Matched Filters (Normal)	136
4.10 CFAR Matched Filters (t)	140
4.11 Matched Subspace Filters (χ^2)	145
4.12 CFAR Matched Subspace Filters (F)	148
A Comparative Summary and a Partial Ordering of Performance	149
4.13 Signal Design	153
Signal Design for Detection	153
Constrained Signal Design	154
4.14 Detectors for Gaussian Random Signals	157
Likelihood Ratios and Quadratic Detectors	157
Orthogonal Decomposition	158
Distribution of Log Likelihood	158
Rank Reduction	160
4.15 Linear Discriminant Functions	162
Linear Discrimination	163
An Extremization Problem	163
Maximizing Divergence	164
References and Comments	166
Problems	167
Appendix: The t , χ^2 , and F Distributions	174
Central χ^2	174
Central t	175
Central F	175
Noncentral χ^2	176
Noncentral t	177
Noncentral F	178
Size and Power	178

CHAPTER 5

Bayes Detectors 179

5.1 Bayes Risk for Hypothesis Testing	180
Loss	181
Risk	181
Bayes Risk	183
5.2 Bayes Tests of Simple Hypotheses	183
5.3 Minimax Tests	185
Risk Set	185
Bayes Tests	187
Minimax and Maximin Tests	188
Computing Minimax Tests	189
Least Favorable Prior	190

5.4	Bayes Tests of Multiple Hypotheses	191
5.5	M -Orthogonal Signals	193
5.6	Composite Matched Filters and Associative Memories	198
	Application to Associative Memories	201
	Summary	202
5.7	Likelihood Ratios, Posterior Probabilities, and Odds	202
	Bayes Tests	203
5.8	Balanced Tests	204
	References and Comments	205
	Problems	206

CHAPTER 6

Maximum Likelihood Estimators 209

6.1	Maximum Likelihood Principle	210
	Random Parameters	213
6.2	Sufficiency	216
6.3	Invariance	217
6.4	The Fisher Matrix and the Cramer-Rao Bound	221
	Cramer-Rao Bound	222
	Concentration Ellipses	225
	Efficiency	226
	Cramer-Rao Bound for Functions of Parameters	229
	Numerical Maximum Likelihood and the Stochastic Fisher Matrix	230
6.5	Nuisance Parameters	231
6.6	Entropy, Likelihood, and Nonlinear Least Squares	233
	Entropy	233
	Likelihood	234
	Nonlinear Least Squares	234
	Comments	235
6.7	The Multivariate Normal Model	235
6.8	The Linear Statistical Model	238
6.9	Mode Identification in the Linear Statistical Model	239
	Maximum Likelihood Equations	240
	The Fisher Information Matrix	241
6.10	Maximum Likelihood Identification of ARMA Parameters	242
	Maximum Likelihood Equations	245
	The Projector $\mathbf{P}(\mathbf{a})$	245
	Interpretations	246
	KiSS	247
	The Fisher Information Matrix	248
	Mode Identification	250
6.11	Maximum Likelihood Identification of a Signal Subspace	252

6.12 Estimation of Parameters in Sinusoidal Models 256

Joint Density 256

Likelihood 256

Summary and Algorithm 258

6.13 Structured Covariance Matrices 260

Linear Structure 261

Low-Rank Orthogonal Correlation Matrix 261

References and Comments 266

Problems 269

Appendix: Vector and Matrix Gradients 274

Vector Gradients 274

Matrix Gradients 275

CHAPTER 7**Bayes Estimators 277**

7.1 Bayes Risk for Parameter Estimation 277

Loss 278

Risk 278

Bayes Risk 279

Bayes Risk Estimator 280

Minimax Estimator 280

7.2 The Risk Set 281

Convexity 281

Bayes Rules 282

Minimax Rules 284

7.3 Computing Bayes Risk Estimators 285

Continuous Case 285

Quadratic Loss and the Conditional Mean 286

Uniform Loss and the Maximum of the A Posteriori Density 287

7.4 Bayes Sufficiency and Conjugate Priors 290

7.5 The Multivariate Normal Model 292

Conditional Distribution of \mathbf{y} , Given \mathbf{x} 292Conditional Distribution of \mathbf{x} , Given \mathbf{y} 296

Filtering Diagrams 298

Interpretations 300

Innovations 301

7.6 The Linear Statistical Model 303

7.7 Sequential Bayes 305

7.8 The Kalman Filter 307

7.9 The Wiener Filter 312

References and Comments 314

Problems 314

CHAPTER 8

Minimum Mean-Squared Error Estimators 323

- 8.1 Conditional Expectation and Orthogonality 324
- 8.2 Minimum Mean-Squared Error Estimators 326
- 8.3 Linear Minimum Mean-Squared Error (LMMSE) Estimators 327
 - Wiener-Hopf Equation 327
 - Summary and Interpretations 328
- 8.4 Low-Rank Wiener Filter 330
- 8.5 Linear Prediction 331
- 8.6 The Kalman Filter 333
 - Prediction 333
 - Estimation 335
 - Covariance Recursions 336
 - The Kalman Recursions 337
- 8.7 Low-Rank Approximation of Random Vectors 337
 - Interpretation 338
 - Order Selection 339
- 8.8 Optimum Scalar Quantizers 339
 - Scalar Quantizers 340
 - Designing the Optimum Quantizer 342
- 8.9 Optimum Block Quantizers 346
- 8.10 Rank Reduction and Rate Distortion 349
- References and Comments 351
- Problems 352

CHAPTER 9

Least Squares 359

- 9.1 The Linear Model 360
 - Interpretations 360
 - The Normal Error Model 364
- 9.2 Least Squares Solutions 365
 - Projections 365
 - Signal and Orthogonal Subspaces 366
 - Orthogonality 368
- 9.3 Structure of Subspaces in Least Squares Problems 371
- 9.4 Singular Value Decomposition 372
 - Synthesis Representation 373
 - Analysis Representations 373
- 9.5 Solving Least Squares Problems 374
 - Cholesky-Factoring the Gram Matrix 374
 - QR-Factoring the Model Matrix 375
 - Singular Value Decomposition 376

9.6	Performance of Least Squares	377
	Posterior Model	377
	Performance	378
9.7	Goodness of Fit	378
	Statistician's Pythagorean Theorem	379
9.8	Improvement in SNR	379
9.9	Rank Reduction	380
9.10	Order Selection in the Linear Statistical Model	381
9.11	Sequential Least Squares	384
9.12	Weighted Least Squares	386
9.13	Constrained Least Squares	387
	Interpretations	388
	Condition Adjustment	388
9.14	Quadratic Minimization with Linear Constraints	389
9.15	Total Least Squares	392
9.16	Inverse Problems and Underdetermined Least Squares	393
	Characterizing the Class of Solutions	394
	Minimum-Norm Solution	395
	Reducing Rank	395
	Bayes	397
	Maximum Entropy	398
	Image Formation	398
	Newton-Raphson	400
9.17	Mode Identification in the Linear Statistical Model	401
9.18	Identification of Autoregressive Moving Average Signals and Systems	402
9.19	Linear Prediction and Prony's Method	405
	Modified Least Squares	405
	Linear Prediction	406
	Prony's Method	407
9.20	Least Squares Estimation of Structured Correlation Matrices	409
	Linear Structure	409
	Toeplitz Matrix	410
	Low-Rank Matrix	411
	Orthonormal Case	413
	More on the Orthonormal Case	413
	References and Comments	415
	Problems	415

CHAPTER 10

Linear Prediction 423

10.1	Stationary Time Series	424
	Wold Representation	426
	Kolmogorov Representation	427
	Filtering Interpretations	428

10.2	Stationary Prediction Theory	429
	Prediction Error Variance	430
	Prediction Error Variance and Poles and Zeros	431
	Spectral Flatness	433
	Filtering Interpretations	434
10.3	Maximum Entropy and Linear Prediction	436
10.4	Nonstationary Linear Prediction	438
	Synthesis	440
	Nonstationary Innovations Representation	441
	Analysis	441
	Nonstationary Predictor Representation	442
10.5	Fast Algorithms of the Levinson Type	442
	Interpretation	446
	Backward Form	447
	Filtering Interpretations	447
	AR Synthesis Lattice	448
	ARMA Lattice	450
10.6	Fast Algorithms of the Schur Type	451
10.7	Least Squares Theory of Linear Prediction	452
	QR Factors and Sliding Windows	456
	Summary and Interpretations	459
10.8	Lattice Algorithms	460
	Initialization	462
	Recursions for k_i	462
	Solving for σ_i^2	462
	Algebraic Interpretations	463
	Lattice Interpretations	463
10.9	Prediction in Autoregressive Moving Average Time Series	464
	Stationary State-Space Representations	464
	Markovian State-Space Model	466
	Nonstationary (or Innovations) State-Space Representations	468
10.10	Fast Algorithms of the Morf-Sidhu-Kailath Type	470
10.11	Linear Prediction and Likelihood Formulas	472
10.12	Differential PCM	473
	References and Comments	475
	Problems	480

CHAPTER 11

Modal Analysis 483

11.1	Signal Model	484
	Modal Decomposition	485
	ARMA Impulse Response	486
	Linear Prediction	488