

Statistical Image Processing and Multidimensional Modeling

Information Science and Statistics

Series Editors:

M. Jordan

Robert Nowak

Bernhard Schölkopf

For other titles published in this series, go to
www.springer.com/series/3816

Paul Fieguth

Statistical Image Processing and Multidimensional Modeling

 Springer

Paul Fieguth
Department of Systems Design Engineering
Faculty of Engineering
University of Waterloo
Waterloo Ontario N2L-3G1
Canada
pfieguth@uwaterloo.ca

ISSN 1613-9011
ISBN 978-1-4419-7293-4 e-ISBN 978-1-4419-7294-1
DOI 10.1007/978-1-4419-7294-1
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2010938436

© Springer Science+Business Media, LLC 2011

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

As a young professor in 1997 I taught my graduate course in *Stochastic Image Processing* for the first time. Looking back on my rough notes from that time, the course must have been a near impenetrable disaster for the graduate students enrolled, with a long list of errors, confusions, and bad notation.

With every repetition the course improved, with significant changes to notation, content, and flow. However, at the same time that a cohesive, large-scale form of the course took shape, the absence of any textbook covering this material became increasingly apparent. There are countless texts on the subjects of image processing, Kalman filtering, and signal processing, however precious little for random fields or spatial statistics. The few texts that *do* cover Gibbs models or Markov random fields tend to be highly mathematical research monographs, not well suited as a textbook for a graduate course.

More than just a graduate course textbook, this text was developed with the goal of being a useful reference for graduate students working in the areas of image processing, spatial statistics, and random fields. In particular, there are many concepts which are known and documented in the research literature, which are useful for students to understand, but which do not appear in many textbooks. This perception is driven by my own experience as a PhD student, which would have been considerably simplified if I had had a text accessible to me addressing some of the following gaps:

- FFT-based estimation (Section 8.3)
- A nice, simple, clear description of multigrid (Section 9.2.5)
- The inference of dynamic models from cross-statistics (Chapter 10)
- A clear distinction and relationship between squared and unsquared kernels (Chapter 5)

- A graphical summary relating Gibbs and Markov models (Figure 6.11)

To facilitate the use of this textbook and the methods described within it, I am making available online (see page XV) much of the code which I developed for this text. This code, some colour figures, and (hopefully few) errata can be found from this book's home page:

<http://ocho.uwaterloo.ca/book>

This text has benefited from the work, support, and ideas of a great many people. I owe a debt of gratitude to the countless researchers upon whose work this book is built, and who are listed in the bibliography. Please accept my apologies for any omissions.

The contents of this book are closely aligned with my research interests over the past ten years. Consequently the work of a number of my former graduate students appears in some form in this book, and I would like to recognize the contributions of Simon Alexander, Wesley Campaigne, Gabriel Carballo, Michale Jamieson, Fu Jin, Fakhry Khellah, Ying Liu, and Azadeh Mohebi.

I would like to thank my Springer editor, John Kimmel, who was an enthusiastic supporter of this text, and highly tolerant of my slow pace in writing. Thanks also to copy editor Valerie Greco for her careful examination of grammar and punctuation (and where any remaining errors are mine, not hers). I would like to thank the anonymous reviewers, who read the text thoroughly and who provided exceptionally helpful constructive criticism. I would also like to thank the *non*-anonymous reviewers, friends and students who gave the text another look: Werner Fieguth, Betty Pries, Akshaya Mishra, Alexander Wong, Li Liu, and Gerald Mwangi.

I would like to thank Christoph Garbe and Michael Winckler, the two people who coordinated my stay at the University of Heidelberg, where this text was completed. My thanks to the Deutscher Akademischer Austausch Dienst, the Heidelberg Graduate School, and to the Heidelberg Collaboratory for Image Processing for supporting my visit.

Many thanks and appreciation to Betty for encouraging this project, and to the kids in Appendix C for just being who they are.

Contents

Preface	V
Table of Contents	VII
List of Examples	XIII
List of Code Samples	XV
Nomenclature	XVII
1 Introduction	1
Part I Inverse Problems and Estimation	11
2 Inverse Problems	13
2.1 Data Fusion	16
2.2 Posedness	19
2.3 Conditioning	23
2.4 Regularization and Prior Models	29
2.4.1 Deterministic Regularization	34
2.4.2 Bayesian Regularization	37
2.5 Statistical Operations	40
2.5.1 Canonical Problems	40
2.5.2 Prior Sampling	42
2.5.3 Estimation	44
2.5.4 Posterior Sampling	49
2.5.5 Parameter Estimation	50
Application 2: Ocean Acoustic Tomography	50
Summary	52
For Further Study	53
Sample Problems	53
3 Static Estimation and Sampling	57
3.1 Non-Bayesian Estimation	58

3.2	Bayesian Estimation	64
3.2.1	Bayesian Static Problem	67
3.2.2	Bayesian Estimation and Prior Means	68
3.2.3	Approximate Bayesian Estimators	70
3.2.4	Bayesian / NonBayesian Duality	73
3.3	Static Sampling	74
3.4	Data Fusion	76
	Application 3: Atmospheric Temperature Inversion [282]	78
	Summary	80
	For Further Study	81
	Sample Problems	81
4	Dynamic Estimation and Sampling	85
4.1	The Dynamic Problem	86
4.1.1	First-Order Gauss–Markov Processes	88
4.1.2	Static — Dynamic Duality	89
4.2	Kalman Filter Derivation	93
4.3	Kalman Filter Variations	100
4.3.1	Kalman Filter Algorithms	102
4.3.2	Steady-State Kalman Filtering	107
4.3.3	Kalman Filter Smoother	109
4.3.4	Nonlinear Kalman Filtering	114
4.4	Dynamic Sampling	118
4.5	Dynamic Estimation for Discrete-State Systems	119
4.5.1	Markov Chains	119
4.5.2	The Viterbi Algorithm	120
4.5.3	Comparison to Kalman Filter	122
	Application 4: Temporal Interpolation of Ocean Temperature [191]	125
	Summary	125
	For Further Study	127
	Sample Problems	127
Part II Modelling of Random Fields		131
5	Multidimensional Modelling	133
5.1	Challenges	134
5.2	Coupling and Dimensionality Reduction	135
5.3	Sparse Storage and Computation	139
5.3.1	Sparse Matrices	139
5.3.2	Matrix Kernels	141
5.3.3	Computation	143
5.4	Modelling	148
5.5	Deterministic Models	149
5.5.1	Boundary Effects	153

5.5.2	Discontinuity Features	155
5.5.3	Prior-Mean Constraints	156
5.6	Statistical Models	158
5.6.1	Analytical Forms	160
5.6.2	Analytical Forms and Nonstationary Fields	164
5.6.3	Recursive / Dynamic Models	166
5.6.4	Banded Inverse-Covariances	167
5.7	Model Determination	169
5.8	Choice of Representation	172
	Application 5: Synthetic Aperture Radar Interferometry [53]	173
	For Further Study	175
	Sample Problems	175
6	Markov Random Fields	179
6.1	One-Dimensional Markovianity	180
6.1.1	Markov Chains	181
6.1.2	Gauss–Markov Processes	181
6.2	Multidimensional Markovianity	182
6.3	Gauss–Markov Random Fields	185
6.4	Causal Gauss–Markov Random Fields	189
6.5	Gibbs Random Fields	192
6.6	Model Determination	199
6.6.1	Autoregressive Model Learning	199
6.6.2	Noncausal Markov Model Learning	201
6.7	Choices of Representation	207
	Application 6: Texture Classification	208
	Summary	211
	For Further Study	212
	Sample Problems	212
7	Hidden Markov Models	215
7.1	Hidden Markov Models	216
7.1.1	Image Denoising	216
7.1.2	Image Segmentation	219
7.1.3	Texture Segmentation	220
7.1.4	Edge Detection	221
7.2	Classes of Joint Markov Models	222
7.3	Conditional Random Fields	225
7.4	Discrete-State Models	227
7.4.1	Local Gibbs Models	228
7.4.2	Nonlocal Statistical-Target Models	229
7.4.3	Local Joint Models	230
7.5	Model Determination	231
	Application 7: Image Segmentation	233
	For Further Study	237

Sample Problems	237
8 Changes of Basis	241
8.1 Change of Basis	243
8.2 Reduction of Basis	247
8.2.1 Principal Components	248
8.2.2 Multidimensional Basis Reduction	253
8.2.3 Local Processing	259
8.3 FFT Methods	262
8.3.1 FFT Diagonalization	263
8.3.2 FFT and Spatial Models	265
8.3.3 FFT Sampling and Estimation	266
8.4 Hierarchical Bases and Preconditioners	269
8.4.1 Interpolated Hierarchical Bases	272
8.4.2 Wavelet Hierarchical Bases	273
8.4.3 Wavelets and Statistics	275
8.5 Basis Changes and Markov Random Fields	278
8.6 Basis Changes and Discrete-State Fields	281
Application 8: Global Data Assimilation [111]	285
Summary	287
For Further Study	288
Sample Problems	288
Part III Methods and Algorithms	291
9 Linear Systems Estimation	293
9.1 Direct Solution	295
9.1.1 Gaussian Elimination	295
9.1.2 Cholesky Decomposition	295
9.1.3 Nested Dissection	296
9.2 Iterative Solution	298
9.2.1 Gauss–Jacobi / Gauss–Seidel	299
9.2.2 Successive Overrelaxation (SOR)	303
9.2.3 Conjugate Gradient and Krylov Methods	306
9.2.4 Iterative Preconditioning	310
9.2.5 Multigrid	313
Application 9: Surface Reconstruction	320
For Further Study	321
Sample Problems	322
10 Kalman Filtering and Domain Decomposition	325
10.1 Marching Methods	327
10.2 Efficient, Large-State Kalman Filters	330
10.2.1 Large-State Kalman Smoother	331

10.2.2	Steady-State KF	334
10.2.3	Strip KF	334
10.2.4	Reduced-Update KF	336
10.2.5	Sparse KF	337
10.2.6	Reduced-Order KF	338
10.3	Multiscale	339
	Application 10: Video Denoising [178]	347
	Summary	350
	For Further Study	351
	Sample Problems	351
11	Sampling and Monte Carlo Methods	355
11.1	Dynamic Sampling	357
11.2	Static Sampling	358
11.2.1	FFT	361
11.2.2	Marching	362
11.2.3	Multiscale Sampling	363
11.3	MCMC	365
11.3.1	Stochastic Sampling	366
11.3.2	Continuous-State Sampling	369
11.3.3	Large-Scale Discrete-State Sampling	370
11.4	Nonparametric Sampling	374
	Application 11: Multi-Instrument Fusion of Porous Media [236]	377
	For Further Study	379
	Sample Problems	379
Part IV	Appendices	381
A	Algebra	383
A.1	Linear Algebra	383
A.2	Matrix Operations	385
A.3	Matrix Positivity	388
A.4	Matrix Positivity of Covariances	389
A.5	Matrix Types	391
A.6	Matrix / Vector Derivatives	391
A.7	Matrix Transformations	395
A.7.1	Eigendecompositions	396
A.7.2	Singular Value Decomposition	400
A.7.3	Cholesky, Gauss, LU, Gram–Schmidt, QR, Schur	401
A.8	Matrix Square Roots	407
A.9	Pseudoinverses	409
B	Statistics	411
B.1	Random Variables, Random Vectors, and Random Fields	411

XII CONTENTS

B.1.1	Random Variables	411
B.1.2	Joint Statistics	412
B.1.3	Random Vectors	414
B.1.4	Random Fields	415
B.2	Transformation of Random Vectors	416
B.3	Multivariate Gaussian Distribution	417
B.4	Covariance Matrices	419
C	Image Processing	423
C.1	Convolution	424
C.2	Image Transforms	428
C.3	Image Operations	430
	Reference Summary	433
	References	437
	Index	451

List of Examples

Application 2	Ocean Acoustic Tomography	50
Application 3	Atmospheric Temperature Inversion [282]	78
Application 4	Temporal Interpolation of Ocean Temperature [191]	122
Application 5	Synthetic Aperture Radar Interferometry [53]	173
Application 6	Texture Classification	208
Application 7	Image Segmentation	232
Application 8	Global Data Assimilation [111]	285
Application 9	Surface Reconstruction	318
Application 10	Video Denoising [178]	345
Application 11	Multi-Instrument Fusion of Porous Media [236]	377
Example 2.1	Root Finding is an Inverse Problem	16
Example 2.2	Interpolation and Posedness	20
Example 2.3	Regression and Posedness	23
Example 2.4	Measurement Models and Conditioning	25
Example 2.5	Matrix Conditioning	27
Example 2.6	Interpolation and Regularization	32
Example 2.7	Interpolation and Smoothness Models	36
Example 2.8	Interpolation and Cross Validation	38
Example 2.9	State Estimation and Sampling	46
Example 3.1	Linear Regression is Least Squares	62
Example 3.2	Simple Scalar Estimation	69
Example 3.3	Prior Mean Removal	71
Example 3.4	Static Estimation and Sampling	75
Example 4.1	Dynamic Estimation and Sampling	92
Example 4.2	Kalman Filtering and Interpolation	98
Example 4.3	Recursive Smoothing and Interpolation	112
Example 5.1	Multidimensional Interpolation	159
Example 5.2	Model Inference	170
Example 6.1	Example Markov Kernels	190
Example 6.2	Robust Estimation	198
Example 6.3	Markov Model Inference and Model Order	204
Example 6.4	Model Inference and Conditioning	206
Example 7.1	Multiple Hidden Fields	223

Example 8.1	Principal Components for Remote Sensing	252
Example 8.2	Overlapped Local Processing	261
Example 8.3	FFT and Spatial Statistics	268
Example 8.4	Hierarchical Bases	276
Example 9.1	Iterative Solutions to Interpolation	302
Example 9.2	Eigendistributions of Iterative Interpolators	305
Example 10.1	Interpolation by Marching	332
Example 11.1	Annealing and Energy Barriers	372
Algorithm 1	The Kalman Filter	97
Algorithm 2	Simplified Expectation Maximization	232
Algorithm 3	FFT Estimation and Sampling	267
Algorithm 4	The Cholesky Decomposition	296
Algorithm 5	Gauss–Seidel	300
Algorithm 6	Conjugate Gradient	309
Algorithm 7	Multigrid	315
Algorithm 8	MATLAB Multigrid implementation I	318
Algorithm 9	MATLAB Multigrid implementation II	319
Algorithm 10	Single Gibbs Sample	366
Algorithm 11	Single Metropolis Sample	367
Algorithm 12	Basic, Single-Scale Annealing	369
Algorithm 13	Multilevel Annealing, Initializing from Coarser Scale	373
Algorithm 14	Multilevel Annealing, Constrained by Coarser Scale	374

List of Code Samples

To make the methods and algorithms described in this book as useful and accessible as possible, a variety of MATLAB scripts, written as part of the development of this textbook, are being made available.

The intent of the scripts is certainly not to be directly applicable to large multidimensional problems. Instead, the code is provided to give the reader some simple examples which actually work, which illustrate the concepts documented in the text, and which hopefully more rapidly lead the reader to be able to do implementations of his or her own.

The code can be found from the text website at

<http://ocho.uwaterloo.ca/book>

Example 2.4:	cond_num.m	25
Example 2.5:	cond_num.m	28
Example 2.7:	interp_reg.m	36
Example 2.8:	cross_val.m	38
Figure 5.7:	bands.m	145
Figure 5.8:	kernels.m	152
Figure 5.15:	corr_len.m	158
Example 5.1:	twod_interp.m	159
Figure 5.16:	posdef.m	161
Figure 5.18:	bands.m	168
Figure 6.2:	mrf.m	182
Example 6.1:	mrf.m	190
Example 6.2:	robust.m	198
Example 6.3:	mrf.m	204
Example 6.3:	mrf.m	207
Figure 8.2:	iterative.m	243
Figure 8.7:	basisreduc.m	255
Figure 8.8:	basisreduc.m	258
Example 8.2:	ovlp_demo.m	261
Example 8.3:	fft_mrf.m	268
Example 8.4:	precond_hier.m	276

Figure 8.17:	mrf.m	280
Example 9.2:	iterative.m	302
Example 9.2:	iterative.m	305
Figure 9.10:	multigrid.m	316
Example 10.1:	marching.m	332
Figure 11.3:	twod_sample.m	359
Figure 11.3:	twod_sample.m	361
Figure 11.4:	fft_discrete.m	362
Figure A.1:	posdef.m	390
Figure B.3:	pdf.m	421

Nomenclature

The following tables of nomenclature are designed to assist the reader in understanding the mathematical language used throughout this text. In the author's opinion this is of considerable value particularly for readers who seek to use the book as a reference and need to be able to understand individual equations or sections without reading an entire chapter for context. Four sets of definitions follow:

1. Basic syntax
2. Mathematical functions
3. Definitions of commonly-used variables
4. Notation for spatial models

Page references are given to provide a few examples of use and some context to the notation, but are in no way intended to be exhaustive.

We limit ourselves here to just defining the notation. For an explanation of algebraic concepts (matrix transpose, eigendecomposition, inverse, etc.) the reader is referred to Appendix A. For an explanation of related statistical concepts (expectation, covariance, etc.), see Appendix B. A brief overview of image processing can be found in Appendix C. Most of the spatial models are explained in Chapters 5 and 6.

Syntax	Definition	Page References
a	scalar, random variable	16 411
\underline{a}	column vector, random vector	13 414
a_i	i th element of vector \underline{a}	16 414
\underline{a}_i	i th vector in a sequence	15 294
a_{ij}	i, j th element of matrix A	299 385
A	matrix	13 383
A^T	matrix transpose	31 385
A^H	matrix Hermitian (complex transpose)	392
A^{-1}	matrix inverse	20 386
$ A $	matrix determinant	63 386 418
\mathcal{A}	kernel corresponding to stationary matrix A	142 143 151
\mathcal{A}^{-1}	kernel corresponding to A^{-1} , but $\mathcal{A}^{-1} \neq (\mathcal{A})^{-1}$	146
\mathcal{A}^T	kernel corresponding to A^T , but $\mathcal{A}^T \neq (\mathcal{A})^T$	146 152 166
\mathbb{R}^n	real vector of length n	19 253 384
$\mathbb{R}^{k \times n}$	real $k \times n$ array	143 383
$[A]_:$, $[A]_{n \times 1}$	reordering of matrix to column vector	133 141 166
$[\underline{a}]_{n \times m}$	reordering of column vector to $n \times m$ matrix	133 146
$[\underline{a}]_{\underline{n}}$	reordering to $n_1 \times n_2 \times \dots$ multidimensional array	265
$\hat{a}, \hat{\underline{a}}, \hat{A}$	estimate of a, \underline{a}, A	22 58
$\hat{\underline{a}}$	estimation error in $\hat{\underline{a}}$	64 108
$\bar{a}, \bar{\underline{a}}, \bar{A}$	transformation of a, \underline{a}, A	41 241
$\check{\underline{a}}$	given sample data of \underline{a}	201 412
\bar{P}	estimation error covariance	68 72 294
$\Pr(Q)$	probability of some event Q	119 181 411
$p(x)$	probability density of x	42 65 411
$p(x y)$	conditional probability density	45 179 413
$\{\dots\}$	a set	15 415
$ S $	number of elements in set S	15 121 203
$\ \underline{x}\ , \ A\ $	vector norm for \underline{x} , matrix norm for A	22 59
$\ \underline{x}\ _P = \underline{x}^T P \underline{x}$	vector squared-norm for \underline{x} with respect to covariance P	31 63 73
\textcircled{a}	convolution kernel origin	145 152 268
\sim	is distributed as ...	49 355
$\underline{x} \sim P$	\underline{x} has covariance P ; the mean is zero or not of interest	28 141 241
$\underline{x} \sim (\underline{\mu}, P)$	\underline{x} has mean $\underline{\mu}$ and covariance P , distribution unknown	37 69 74
$\underline{x} \sim \mathcal{N}(\underline{\mu}, P)$	\underline{x} is Gaussian with mean $\underline{\mu}$ and covariance P	28 63 417

Tab. Notation.1. Basic vector, matrix, and statistical syntax

Function	Definition	Page References
$\text{sign}(a)$	the sign of scalar a , $\text{sign}(a) = \begin{cases} -1 & a < 0 \\ 0 & a = 0 \\ 1 & a > 0 \end{cases}$	144 400
$a \bmod b$	division modulus (remainder)	155 262 392
$\min(\cdot)$	the minimum value in a set	198 385 385
$\min_{x \in X}(\cdot)$	the minimum value of a function over range X	401
$\arg_x \min(\cdot)$	the value of x which minimizes the function	30 63 409
$\text{Ra}(A)$	the range space of A	19 53 385
$\text{Nu}(A)$	the null space of A	19 53 384
$\text{rank}(A)$	the rank of A	53 385
$\text{dim}(\cdot)$	the dimension of a space	384
$\text{tr}(A)$	the trace, the sum of the diagonal elements of A	247 387 395
$\det(A)$	the determinant of A	24 386 387
$\kappa(A)$	matrix condition number of A	26 104 246
$\text{diag}(A)$	a vector containing the diagonal elements of A	54 140 266
$\text{Diag}(\underline{x})$	a diagonal matrix with \underline{x} along the diagonal	165 264 294
I	the identity matrix	37 277 357
FFT_d	the d -dimensional fast Fourier transform	267 361 428
FFT_d^{-1}	the d -dimensional inverse fast Fourier transform	265 361
WT	the wavelet transform	273 347 428
WT^{-1}	the inverse wavelet transform	275 290
\odot	element-by-element matrix multiplication	146 165 266
\oslash	element-by-element matrix division	146 254 265
$*$	convolution	142 146 424
\circledast	circular convolution	427 428
$\text{var}(\cdot)$	variance	164 389
$\text{cov}(\cdot)$	covariance	42 67 87
$E[\cdot]$	expectation	42 64 412
$E_a[\cdot]$	expectation over variable a , if otherwise ambiguous	232 363
\equiv	is equivalent to, identical to	31 61
\triangleq	is defined as	15 58
$< > \leq \geq$	inequalities, in positive-definite sense for matrices	81 100 388

Tab. Notation.2. Mathematical functions and operations (see Appendix A)

Symbol	Definition	Page	References
b	linear system target	245	293 403
b	estimator bias		65 66
c	random field clique		193 368
d	spatial dimensionality		262 267
e	error	54	58 298
\underline{e}_i	the i th unit vector: all zeros with a one in the i th position		384
f	forward problem	13	15 30
g	Markov random field model coefficient		186 202
i, j	general indices		17 37
k, n, q	matrix and vector dimensions	19	140 148
m	measurement	13	40 58
p	probability density	42	65 411
r	linear system residual	298	306 314
s, t	time		42 86
v	measurement noise	13	40 58
v	eigenvector	249	304 396
w	dynamic process noise	42	86 325
x, y	spatial location or indices	35	150 221
z	system state	13	40 58
A	linear system, normal equations	245	293 403
A	dynamic model predictor	86	143 325
B	dynamic model stochastic weight	42	86 325
C	measurement model	13	40 58
E	expectation	42	64 412
F	Fourier transform		257 263
F	change of basis (forwards)	241	247 314
G	Markov random field model		186 201
H	energy function	192	222 371
I	identity matrix	37	277 357
J	optimization criterion		198 249
K	estimator gain	97	108 330
L	constraints matrix	150	157 293
M	measurements field (multidimensional)	170	215 267
N	image or patch size	134	214 329
P	state covariance	40	143 160
Q	squared system constraints		152 152
R	measurement noise covariance	40	63 87
S	change of basis (backwards)	246	247 314
T	annealing temperature		192 368
U, V	orthogonal matrices	245	250 400
W	wavelet transform		277 348
Z	random field (multidimensional)	133	141 327

Tab. Notation.3. Symbol definitions

Symbol	Definition	Page	References
α, β, γ	constants	41	98 164
β	Gibbs inverse temperature	192	228 355
δ	Dirac delta	87	197 200
δ	small offset or perturbation	26	160 184
ϵ	small amount		21 305
κ	matrix condition number	26	104 246
θ	model parameters		50 170
λ	regularization parameter	31	35 63
λ	eigenvalue	304	396 420
μ	mean	37	67 412
ν	estimator innovations	95	96 110
ρ	correlation coefficient		390 413
ρ	spectral radius		300 398
σ	standard deviation	69	102 412
σ	singular value	27	256 400
τ	time offset or period		44 111
ξ	correlation length		28 124
ζ	threshold	198	300 432
Γ	covariance square root	104	166 401
Λ, Σ	covariance	66	93 389
Ψ	state space or alphabet	119	122 181
Ω	problem space (multidimensional lattice)	184	193 415
Ξ	region subset operator		167 342
\mathcal{B}	matrix banding structure	144	146 167
\mathcal{C}	clique set		193 194
\mathcal{N}	neighbourhood	184	189 226
$\mathcal{N}(\mu, P)$	Gaussian distribution with mean $\underline{\mu}$, covariance P	28	63 417
$\mathcal{O}(\cdot)$	complexity order		143 326
\mathbb{R}	real		374
\mathbb{R}^n	real vector of length n	19	253 384
$\mathbb{R}^{k \times n}$	real $k \times n$ array		143 383
\mathbb{Z}	Gibbs partition function	192	355
0, 1	scalar constant zero, one		24 59
$\underline{0}, \underline{1}$	vector constants of all zeros, all ones	19	45 72
$\mathbf{0}, \mathbf{1}$	matrix constants of all zeros, all ones		66 76

Tab. Notation.3. Symbol definitions (cont'd)

Nonstationary Model (Dense)	Stationary Model (Kernel)	Model Type	Page References
A, B	\mathcal{A}, \mathcal{B}	Square root model (dynamic)	86 143 325
Γ	$\mathbf{\Gamma}$	Square root model (static)	104 166 401
P	\mathcal{P}	Squared model (covariance)	40 143 160
V	\mathcal{V}	Square root inverse model (Gibbs field)	193
G	\mathcal{G}	Squared inverse model (Markov field)	186 201
L	\mathcal{L}	Square root deterministic model (constraints)	150 157 293
Q	\mathcal{Q}	Squared deterministic model	152 152
C, R	\mathcal{C}, \mathcal{R}	Measurement model	13 40 58

Tab. Notation.4. Spatial model definitions

Introduction

Images are all around us! Inexpensive digital cameras, video cameras, computer webcams, satellite imagery, and images off the Internet give us access to spatial imagery of all sorts. The vast majority of these images will be of scenes at human scales — pictures of animals / houses / people / faces and so on — relatively complex images which are not well described statistically or mathematically. Many algorithms have been developed to process / denoise / compress / segment such images, described in innumerable textbooks on image processing [36, 54, 143, 174, 210], and briefly reviewed in Appendix C.

Somewhat less common, but of great research interest, are images which do allow some sort of mathematical characterization, and to which standard image-processing algorithms may not apply. In most cases we do not necessarily have *images* here, per se, but rather spatial datasets, with one or more measurements taken over a two- or higher-dimensional space.

There are many important problems falling into this latter group of scientific images, and where this text seeks to make a contribution. Examples abound throughout remote sensing (satellite data mapping, data assimilation, sea-ice / climate-change studies, land use), medical imaging (denoising, organ segmentation, anomaly detection), computer vision (textures, image classification, segmentation), and other 2D / 3D problems (groundwater, biological imaging, porous media, etc.).

Although a great deal of research has been applied to scientific images, in most cases the resulting methods are not well documented in common textbooks, such that many experienced researchers will be unfamiliar with the use of the FFT method (Section 8.3) or of posterior sampling (Chapter 11), for example.

The goal, then, of this text is to address methods for solving multidimensional inverse problems. In particular, the text seeks to avoid the pitfall of being entirely mathematical / theoretical at one extreme, or primarily applied / algorithmic on the other, by deliberately developing the basic theory (Part I), the mathematical mod-

elling (Part II), and the algorithmic / numerical methods (Part III) of solving a given problem.

Inverse Problems

So, to begin, why would we want to solve an inverse problem?

There are a great many spatial phenomena that a person might want to study ...

- The salinity of the ocean surface as a function of position;
- The temperature of the atmosphere as a function of position;
- The height of the grass growing in your back yard, as a function of location;
- The proportions of oil and water in an oil reservoir.

In each of these situations, you aren't just handed a map of the spatial process you wish to study, rather you have to *infer* such a map from given measurements. These measurements might be a simple function of the spatial process (such as measuring the height of the grass using a ruler) or might be complicated nonlinear functions (such as microwave spectra for inferring temperature).

The process by which measurements are generated from the spatial process is normally relatively straightforward, and is referred to as a *forward problem*. More difficult, then, is the *inverse problem*, discussed in detail in Chapter 2, which represents the mathematical inverting of the *forward problem*, allowing you to infer the process of interest from the measurements. A simple illustration is shown in Figure 1.1.

Large Multidimensional Problems

So why is it that we wish to study large multidimensional problems?

The solution to linear inverse problems (see Chapter 3) is easily formulated analytically, and even a nonlinear inverse problem can be reformulated as an optimization problem and solved. The challenge, then, is not the solving of inverse problems *in principle*, but rather actually solving them *in practice*.

For example, the solution to a linear inverse problem involves a matrix inversion. As the problem is made larger and larger, eventually the matrix becomes computationally or numerically impossible to invert. However, this is not just an abstract limit — even a modest two-dimensional problem at a resolution of 1000×1000 pixels contains one million unknowns, which would require the inversion of a one-million by one-million matrix: completely unfeasible.

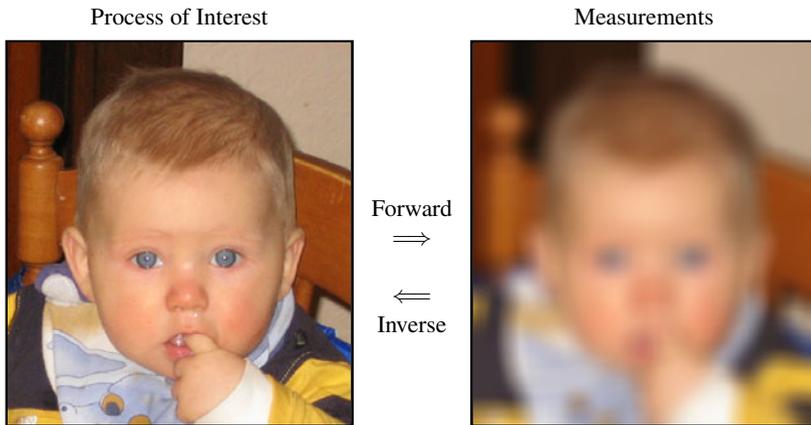


Fig. 1.1. An inverse problem: You want a nice clear photo of a face, however your camera yields blurry measurements. To solve this inverse problem requires us to mathematically invert the forward process of blurring.

Therefore even rather modestly sized two- and higher-dimensional problems become impossible to solve using straightforward techniques, yet these problems are very common. Problems having one million or more unknowns are littered throughout the fields of remote sensing, oceanography, medical imaging, and seismology, to name a few.

To be clear, a problem is considered to be multidimensional if it is a function of two or more independent variables. These variables could be spatial (as in a two-dimensional image or a three-dimensional volume), spatio-temporal (such as a video, a sequence of two-dimensional images over time), or a function of other variables under our control.

Multidimensional Methods versus Image Processing

What is it that the great diversity of algorithms in the image processing literature cannot solve?

The majority of images which are examined and processed in image processing are “real” images, pictures and scenes at human scales, where the images are not well described mathematically. Therefore the focus of image processing is on making relatively few explicit, mathematical assumptions about the image, and instead focusing on the development of algorithms that perform image-related tasks (such as compression, segmentation, edge detection, etc.).



Fig. 1.2. Which of these might be best characterized mathematically? Many natural phenomena, when viewed at an appropriate scale, have a behaviour which is sufficiently varied or irregular that it can be modelled via relatively simple equations, as opposed to a human face, which would need a rather complex model to be represented accurately.

In contrast, of great research interest are images taken at microscopic scales (cells in a Petri dish, the crystal structure of stone or metal) or at macroscopic scales (the temperature distribution of the ocean or of the atmosphere, satellite imagery of the earth) which do, in general, allow some sort of mathematical characterization, as explored in Figure 1.2. That is, the focus of this text is on the assumption or inference of rather *explicit* mathematical models of the unknown process.

Next, in order to be able to say something about a problem, we need measurements of it. These measurements normally suffer from one of three issues, any one of which would preclude the use of standard image-processing techniques:

1. For measurements produced by a scientific instrument, acquiring a measurement normally requires time and/or money, therefore the number of measurements is constrained. Frequently this implies that the multidimensional problem of interest is only sparsely sampled, as illustrated in Figure 1.3.

There exist many standard methods to interpolate gaps in a sequence of data, however standard interpolation knows nothing about the underlying phenomenon being studied. That is, surely a grass-like texture should be interpolated differently from a map of ocean-surface temperature.

2. Most measurements are not exact, but suffer from some degree of noise. Ideally we would like to remove this noise, to infer a more precise version of the underlying multidimensional phenomenon.

There exist many algorithms for noise reduction in images, however these are necessarily heuristic, because they are designed to work on photographic images, which might contain images of faces / cars / trees and the like. Given a scientific dataset, surely we would wish to undertake denoising in a more systematic (ideally optimal) manner, somehow dependent on the behaviour of the underlying phenomenon.

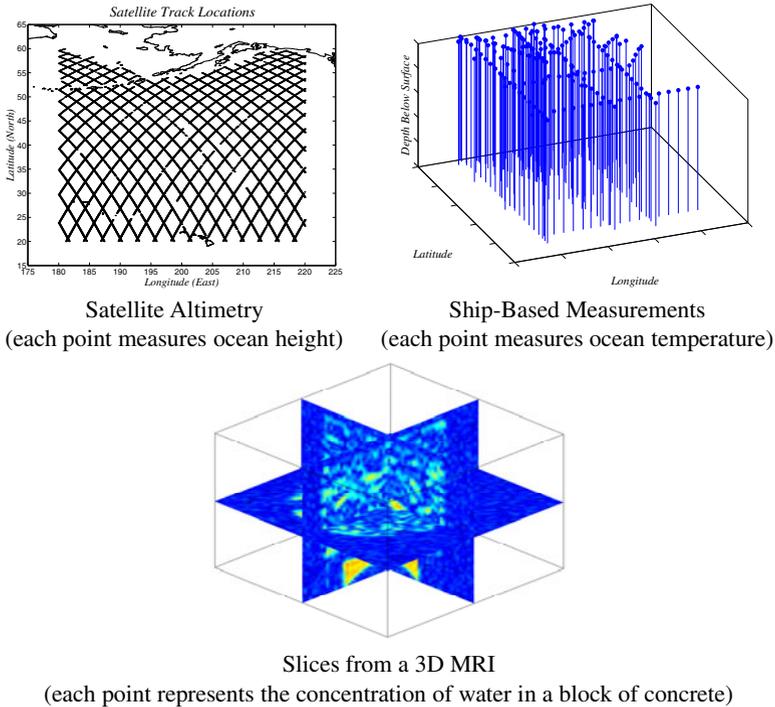


Fig. 1.3. Multidimensional measurements: Three examples of two- or three-dimensional measurements which could not be processed by conventional means of image processing. The altimetric measurements are sparse, following the orbital path of a satellite; the ship-based measurements are irregular and highly sparse, based on the paths that a ship followed in towing an instrument array; the MRI measurements are dense, but at poor resolution and with substantial noise.

3. In many cases of scientific imaging, the raw measurement produced by an instrument is *not* a direct measurement of the multidimensional field, but rather some function of it. For example, in Application 3 we wish to study atmospheric temperature based on radiometric measurements of microwave intensities: the air temperature and microwave intensity are indeed related, but are very different quantities.

Standard methods in image processing normally assume that the measurements (possibly noisy, possibly blurred) form an image. However, having measurements being some complicated function of the field of interest (an inverse problem) is more subtle and requires a careful formulation.

Statistics and Random Fields

What is it that makes a problem statistical, and why do we choose to focus on statistical methods?

An interest in *spatial statistics* goes considerably beyond the modelling of phenomena which are inherently *random*. In particular, multidimensional random fields offer the following advantages:

1. Even if an underlying process is not random, in most cases measurements of the process are corrupted by noise, and therefore a statistical representation may be appropriate.
2. Many processes exhibit a degree of irregularity or complexity that would be extremely difficult to model deterministically. Two examples are shown in Figure 1.4; although there are physics which govern the behaviour of both of these examples (e.g., the Navier–Stokes differential equation for water flow) the models are typically highly complex, containing a great number of unknown parameters, and are computationally difficult to simulate.

A random-fields approach, on the other hand, would implicitly approximate these complex models on the basis of observed statistics.

A random field¹ X is nothing but a large collection of random variables arranged on some set of points (possibly a two- or three-dimensional grid, perhaps on a sphere, or perhaps irregularly distributed in a high-dimensional space). The random field is characterized by the statistical interrelationships between its random variables.

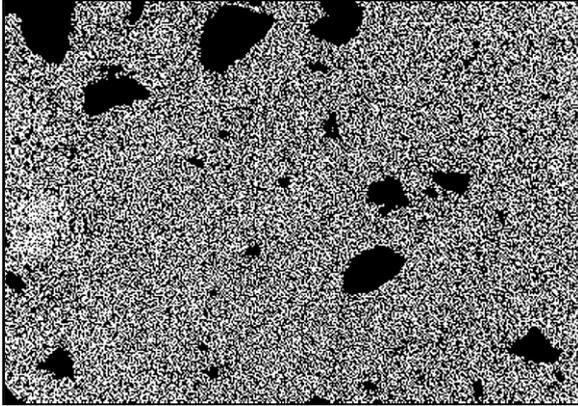
The main problem associated with a statistical formulation is the computational complexity of the resulting solution. However, as we shall see, there exists a comprehensive set of methods and algorithms for the manipulation and efficient solving of problems involving random fields. The development of this theory and of associated algorithms is the fundamental goal of this text.

Specifically, the key problem explored in this text is representational and computational efficiency in the solving of large problems. The question of efficiency is easily motivated: even a very modestly sized 256×256 image has 65 536 elements, and the glass beads image in Figure 1.4 contains in excess of 100 million elements! It comes as no surprise that a great part of the research into random fields involves the discovery or definition of *implicit* statistical forms which lead to effective or faithful representations of the true statistics, while admitting computationally efficient algorithms.

Broadly speaking there are four typical problems associated with random fields [112]:

¹ Random variables, random vectors, and random fields are reviewed in Appendix B.1.

A Porous Medium of Packed Glass Beads



(Microscopic Data from M. Ioannidis, Dept. Chemical Engineering, University of Waterloo)

Global Ocean Surface Temperature

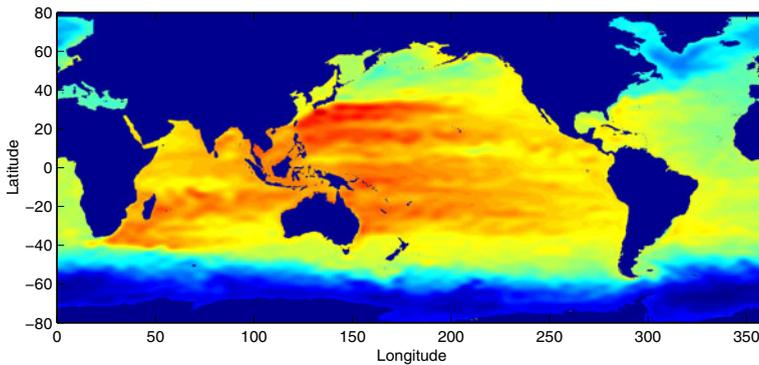


Fig. 1.4. Two examples of phenomena which may be modelled via random fields: packed glass beads (top), and the ocean surface temperature (bottom). Alternatives to random fields do exist to model these phenomena, such as ballistic methods for the glass beads, and coupled differential equations for the ocean, however such approaches would be greatly more complex than approximating the observed phenomena on the basis of inferred spatial statistics.

1. Representation: how is the random field represented and parametrized?
2. Synthesis: how can we generate “typical” realizations of the random field?
3. Parameter estimation: given a parametrized statistical model and sample image, how can we estimate the unknown parameters in the model?

4. Random fields estimation: given noisy observations of the random field, how can the unknown random field be estimated?

All four of these issues are of interest to us, and are developed throughout the text.

For each of these there are separate questions of formulation,

How do I write down the equations that need to be solved?

as opposed to those of solution,

How do I actually find a solution to these equations?

Part I of this text focuses mostly on the former question, establishing the mathematical fundamentals that are needed to express a solution, *in principle*. This gives us a solution which we might call

1. **Brute Force:** The direct implementation of the solution equations, irrespective of computational storage, complexity, and numerical robustness issues.

Parts II and III then examine the latter question, seeking practical, elegant, or indirect solutions to the problems of interest. However, *practical* should not be interpreted to mean that the material is only of dry interest to the specialist sitting at a computer, about to develop a computer program. Many of the most fundamental ideas expressed in this text are particularly in Part II, where deep insights into the nature of spatial random fields are explored.

A few kinds of efficient solutions, alternatives to the direct implementations from Part I, are summarized as follows:

2. **Dimensionality Reduction:** Transforming a problem into one or more lower-dimensional problems.
3. **Change of Basis:** A mathematical transformation of the problem which simplifies its computational or numerical complexity.
4. **Approximate Solution:** An approximation to the exact analytical solution.
5. **Approximated Problem:** Rather than solving the given problem, identifying a similar problem which can be solved exactly.
6. **Special Cases:** Circumstances in which the statistics or symmetry of the problem gives rise to special, efficient solutions.

These six points give a broad sense of what this text is about.