

Statistics for Engineering and Information Science

Series Editors

M. Jordan, S.L. Lauritzen, J.F. Lawless, V. Nair

Statistics for Engineering and Information Science

Akaike and Kitagawa: The Practice of Time Series Analysis.

Cowell, Dawid, Lauritzen, and Spiegelhalter: Probabilistic Networks and Expert Systems.

Doucet, de Freitas, and Gordon: Sequential Monte Carlo Methods in Practice.

Fine: Feedforward Neural Network Methodology.

Hawkins and Olwell: Cumulative Sum Charts and Charting for Quality Improvement.

Jensen: Bayesian Networks and Decision Graphs.

Marchette: Computer Intrusion Detection and Network Monitoring:
A Statistical Viewpoint.

Vapnik: The Nature of Statistical Learning Theory, Second Edition.

Arnaud Doucet
Nando de Freitas
Neil Gordon
Editors

Sequential Monte Carlo Methods in Practice

Foreword by Adrian Smith

With 168 Illustrations

 Springer

Arnaud Doucet
Department of Electrical and
Electronic Engineering
The University of Melbourne
Victoria 3010
Australia
doucet@ee.mu.oz.au

Nando de Freitas
Computer Science Division
387 Soda Hall
University of California
Berkeley, CA 94720-1776
USA
jfgf@cs.berkeley.edu

Neil Gordon
Pattern and Information Processing
Defence Evaluation and Research
Agency
St. Andrews Road
Malvern, Worcs, WR14 3PS
UK
N. Gordon@signal.dera.gov.uk

Series Editors

Michael Jordan
Department of Computer Science
University of California, Berkeley
Berkeley, CA 94720
USA

Steffen L. Lauritzen
Department of Mathematical Sciences
Aalborg University
DK-9220 Aalborg
Denmark

Jerald F. Lawless
Department of Statistics
University of Waterloo
Waterloo, Ontario N2L 3G1
Canada

Vijay Nair
Department of Statistics
University of Michigan
Ann Arbor, MI 48109
USA

Library of Congress Cataloging-in-Publication Data

Doucet, Arnaud.

Sequential Monte Carlo methods in practice / Arnaud Doucet, Nando de Freitas, Neil Gordon.

p. cm. — (Statistics for engineering and information science)

Includes bibliographical references and index.

ISBN 978-1-4419-2887-0

ISBN 978-1-4757-3437-9 (eBook)

DOI 10.1007/978-1-4757-3437-9

1. Monte Carlo method. I. de Freitas, Nando. II. Gordon, Neil (Neil James), 1967–
III. Title. IV. Series.

QA298 .D68 2001

519.2'82—dc21

00-047093

Printed on acid-free paper.

© 2001 Springer Science+Business Media New York

Originally published by Springer Science+Business Media Inc. in 2001

Softcover reprint of the hardcover 1st edition 2001

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), 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 of general descriptive names, trade names, trademarks, etc., in this publication, even if the former are not especially identified, is not to be taken as a sign that such names, as understood by the Trade Marks and Merchandise Marks Act, may accordingly be used freely by anyone.

Foreword

It is a great personal pleasure to have the opportunity to contribute the foreword of this important volume.

Problems arising from new data arriving sequentially in time and requiring on-line decision-making responses are ubiquitous in modern communications and control systems, economic and financial data analysis, computer vision and many other fields in which real-time estimation and prediction are required.

As a beginning postgraduate student in Statistics, I remember being excited and charmed by the apparent simplicity and universality of Bayes' theorem as the key logical and computational mechanism for sequential learning. I also recall my growing sense of disappointment as I read Aoki's (1967) volume *Optimisation of Stochastic Systems* and found myself increasingly wondering how any of this wonderful updating machinery could be implemented to solve anything but the very simplest problems. This realisation was deeply frustrating. The elegance of the Kalman Filter provided some solace, but at the cost of pretending to live in a linear Gaussian world. Once into the nonlinear, non-Gaussian domain, we were without a universally effective approach and driven into a series of ingenious approximations, some based on flexible mixtures of tractable distributions to approximate and propagate uncertainty, or on local linearisations of nonlinear systems. In particular, Alspach and Sorenson's (1972) Gaussian sum approximations were developed into a systematic filtering approach in Anderson and Moore (1979); and the volume edited by Gelb (1974) provides an early overview of the use of the extended Kalman Filter.

In fact, the computational obstacles to turning the Bayesian handle in the sequential applications context were just as much present in complex, non-sequential applications and considerable efforts were expended in the 1970s and 1980s to provide workable computational strategies in general Bayesian Statistics.

Towards the end of the 1980s, the realisation emerged that the only universal salvation to hand was that of simulation. It involved combining increasingly sophisticated mathematical insight into, and control over, Monte Carlo techniques with the extraordinary increases we have witnessed

in computer power. The end result was a powerful new toolkit for Bayesian computation.

Not surprisingly, those primarily concerned with sequential learning moved in swiftly to adapt and refine this simulation toolkit to the requirements of on-line estimation and prediction problems.

This volume provides a comprehensive overview of what has been achieved and documents the extraordinary progress that has been made in the past decade. A few days ago, I returned to Aoki's 1967 volume. This time as I turned the pages, there was no sense of disappointment. We now really can compute the things we want to!

Adrian Smith
Queen Mary and Westfield College
University of London
November 2000

Acknowledgments

During the course of editing this book, we were fortunate to be assisted by the contributors and several individuals. To address the challenge of obtaining a high quality publication each chapter was carefully reviewed by many authors and several external reviewers. In particular, we would like to thank David Fleet, David Forsyth, Simon Maskell and Antonietta Mira for their kind help. The advice and assistance of John Kimmel, Margaret Mitchell, Jenny Wolkowicki and Fred Bartlett from Springer-Verlag has eased the editing process and is much appreciated.

Special thanks to our loved ones for their support and patience during the editing process.

Arnaud Doucet

Electrical & Electronic Engineering Department
The University of Melbourne,
Victoria 3010, Australia
`doucet@ee.mu.oz.au`

Nando de Freitas

Computer Science Division,
387 Soda Hall,
University of California,
Berkeley, CA 94720-1776, USA
`jfgf@cs.berkeley.edu`

Neil Gordon

Pattern and Information Processing,
Defence Evaluation and Research Agency,
St Andrews Road,
Malvern, Worcs, WR14 3PS, UK.
`N.Gordon@signal.dera.gov.uk`

Contents

Foreword	v
Acknowledgments	vii
Contributors	xxi
I Introduction	1
1 An Introduction to Sequential Monte Carlo Methods	3
Arnaud Doucet, Nando de Freitas, and Neil Gordon	
1.1 Motivation	3
1.2 Problem statement	5
1.3 Monte Carlo methods	6
1.3.1 Perfect Monte Carlo sampling	7
1.3.2 Importance sampling	8
1.3.3 The Bootstrap filter	10
1.4 Discussion	13
II Theoretical Issues	15
2 Particle Filters - A Theoretical Perspective	17
Dan Crisan	
2.1 Introduction	17
2.2 Notation and terminology	17
2.2.1 Markov chains and transition kernels	18
2.2.2 The filtering problem	19
2.2.3 Convergence of measure-valued random variables	20
2.3 Convergence theorems	21
2.3.1 The fixed observation case	21
2.3.2 The random observation case	24
2.4 Examples of particle filters	25
2.4.1 Description of the particle filters	25

2.4.2	Branching mechanisms	28
2.4.3	Convergence of the algorithm	31
2.5	Discussion	33
2.6	Appendix	35
2.6.1	Conditional probabilities and conditional expectations	35
2.6.2	The recurrence formula for the conditional distribution of the signal	38
3	Interacting Particle Filtering With Discrete Observations	43
	Pierre Del Moral and Jean Jacod	
3.1	Introduction	43
3.2	Nonlinear filtering: general facts	46
3.3	An interacting particle system under Case A	48
3.3.1	Subcase A1	48
3.3.2	Subcase A2	55
3.4	An interacting particle system under Case B	60
3.4.1	Subcase B1	60
3.4.2	Subcase B2	67
3.5	Discretely observed stochastic differential equations	71
3.5.1	Case A	72
3.5.2	Case B	73
III	Strategies for Improving Sequential Monte Carlo Methods	77
4	Sequential Monte Carlo Methods for Optimal Filtering	79
	Christophe Andrieu, Arnaud Doucet, and Elena Punskeya	
4.1	Introduction	79
4.2	Bayesian filtering and sequential estimation	79
4.2.1	Dynamic modelling and Bayesian filtering	79
4.2.2	Alternative dynamic models	80
4.3	Sequential Monte Carlo Methods	82
4.3.1	Methodology	82
4.3.2	A generic algorithm	85
4.3.3	Convergence results	86
4.4	Application to digital communications	88
4.4.1	Model specification and estimation objectives	89
4.4.2	SMC applied to demodulation	91
4.4.3	Simulations	93

5	Deterministic and Stochastic Particle Filters in State-Space Models	97
	Erik Bølviken and Geir Storvik	
5.1	Introduction	97
5.2	General issues	98
5.2.1	Model and exact filter	98
5.2.2	Particle filters	99
5.2.3	Gaussian quadrature	100
5.2.4	Quadrature filters	101
5.2.5	Numerical error	102
5.2.6	A small illustrative example	104
5.3	Case studies from ecology	104
5.3.1	Problem area and models	104
5.3.2	Quadrature filters in practice	107
5.3.3	Numerical experiments	110
5.4	Concluding remarks	112
5.5	Appendix: Derivation of numerical errors	114
6	RESAMPLE–MOVE Filtering with Cross-Model Jumps	117
	Carlo Berzuini and Walter Gilks	
6.1	Introduction	117
6.2	Problem statement	118
6.3	The RESAMPLE–MOVE algorithm	119
6.4	Comments	124
6.5	Central limit theorem	125
6.6	Dealing with model uncertainty	126
6.7	Illustrative application	129
6.7.1	Applying RESAMPLE–MOVE	131
6.7.2	Simulation experiment	134
6.7.3	Uncertainty about the type of target	135
6.8	Conclusions	138
7	Improvement Strategies for Monte Carlo Particle Filters	139
	Simon Godsill and Tim Clapp	
7.1	Introduction	139
7.2	General sequential importance sampling	140
7.3	Markov chain moves	143
7.3.1	The use of bridging densities with MCMC moves	144
7.4	Simulation example: TVAR model in noise	145
7.4.1	Particle filter algorithms for TVAR models	146
7.4.2	Bootstrap (SIR) filter	148
7.4.3	Auxiliary particle filter (APF)	149
7.4.4	MCMC resampling	150
7.4.5	Simulation results	152
7.5	Summary	157

7.6	Acknowledgements	158
8	Approximating and Maximising the Likelihood for a General State-Space Model	159
	Markus Hürzeler and Hans R. Künsch	
8.1	Introduction	159
8.2	Bayesian methods	159
8.3	Pointwise Monte Carlo approximation of the likelihood . .	161
8.3.1	Examples	161
8.4	Approximation of the likelihood function based on filter samples	164
8.5	Approximations based on smoother samples	166
8.5.1	Approximation of the likelihood function	167
8.5.2	Stochastic EM-algorithm	167
8.6	Comparison of the methods	168
8.6.1	AR(1) process	168
8.6.2	Nonlinear example, 3 parameters	171
8.6.3	Nonlinear model, 5 parameters	173
8.6.4	Discussion	173
8.7	Recursive estimation	173
9	Monte Carlo Smoothing and Self-Organising State-Space Model	177
	Genshiro Kitagawa and Seisho Sato	
9.1	Introduction	177
9.2	General state-space model and state estimation	178
9.2.1	The model and the state estimation problem . . .	178
9.2.2	Non-Gaussian filter and smoother	179
9.3	Monte Carlo filter and smoother	180
9.3.1	Approximation of non-Gaussian distributions . . .	180
9.3.2	Monte Carlo filtering	181
9.3.3	Derivation of the Monte Carlo filter	182
9.3.4	Monte Carlo smoothing	183
9.3.5	Non-Gaussian smoothing for the stochastic volatility model	186
9.3.6	Nonlinear Smoothing	188
9.4	Self-organising state-space models	189
9.4.1	Likelihood of the model and parameter estimation .	189
9.4.2	Self-organising state-space model	191
9.5	Examples	192
9.5.1	Self-organising smoothing for the stochastic volatility model	192
9.5.2	Time series with trend and stochastic volatility . .	194
9.6	Conclusion	195

10 Combined Parameter and State Estimation in Simulation-Based Filtering	197
Jane Liu and Mike West	
10.1 Introduction and historical perspective	197
10.2 General framework	199
10.2.1 Dynamic model and analysis perspective	199
10.2.2 Filtering for states	200
10.2.3 Filtering for states and parameters	202
10.3 The treatment of model parameters	202
10.3.1 Artificial evolution of parameters	202
10.3.2 Kernel smoothing of parameters	203
10.3.3 Reinterpreting artificial parameter evolutions . . .	204
10.4 A general algorithm	206
10.5 Factor stochastic volatility modelling	208
10.6 Discussion and future directions	217
11 A Theoretical Framework for Sequential Importance Sampling with Resampling	225
Jun S. Liu, Rong Chen, and Tanya Logvinenko	
11.1 Introduction	225
11.2 Sequential importance sampling principle	227
11.2.1 Properly weighted sample	227
11.2.2 Sequential build-up	228
11.3 Operations for enhancing SIS	229
11.3.1 Reweighting, resampling and reallocation	230
11.3.2 Rejection control and partial rejection control . . .	231
11.3.3 Marginalisation	234
11.4 Monte Carlo filter for state-space models	234
11.4.1 The general state-space model	235
11.4.2 Conditional dynamic linear model and the mixture Kalman filter	236
11.5 Some examples	237
11.5.1 A simple illustration	237
11.5.2 Target tracking with MKF	239
11.6 Discussion	241
11.7 Acknowledgements	242
12 Improving Regularised Particle Filters	247
Christian Musso, Nadia Oudjane, and Francois Le Gland	
12.1 Introduction	247
12.2 Particle filters	249
12.2.1 The (classical) interacting particle filter (IPF) . . .	250
12.2.2 Regularised particle filters (RPF)	251
12.3 Progressive correction	255
12.3.1 Focus on the correction step	256

12.3.2	Principle of progressive correction	257
12.3.3	Adaptive choice of the decomposition	258
12.4	The local rejection regularised particle filter (L2RPF) . .	260
12.4.1	Description of the filter	260
12.4.2	Computing the coefficient $c_t^{(i)}(\alpha_t)$	263
12.5	Applications to tracking problems	264
12.5.1	Range and bearing	265
12.5.2	Bearings-only	266
12.5.3	Multiple model particle filter (MMPF)	269
13	Auxiliary Variable Based Particle Filters	273
	Michael K. Pitt and Neil Shephard	
13.1	Introduction	273
13.2	Particle filters	274
13.2.1	The definition of particle filters	274
13.2.2	Sampling the empirical prediction density	274
13.2.3	Weaknesses of particle filters	276
13.3	Auxiliary variable	277
13.3.1	The basics	277
13.3.2	A generic SIR based auxiliary proposal	278
13.3.3	Examples of adaption	283
13.4	Fixed-lag filtering	288
13.5	Reduced random sampling	289
13.5.1	Basic ideas	289
13.5.2	Simple outlier example	290
13.6	Conclusion	292
13.7	Acknowledgements	293
14	Improved Particle Filters and Smoothing	295
	Photis Stavropoulos and D.M. Titterton	
14.1	Introduction	295
14.2	The methods	296
14.2.1	The smooth bootstrap	296
14.2.2	Adaptive importance sampling	300
14.2.3	The kernel sampler of Hürzeler and Künsch	302
14.2.4	Partially smooth bootstrap	303
14.2.5	Roughening and sample augmentation	305
14.2.6	Application of the methods in particle filtering and smoothing	306
14.3	Application of smooth bootstrap procedures to a simple control problem	308
14.3.1	Description of the problem	308
14.3.2	An approach to the continuous-time version of the problem	309
14.3.3	An adaptation of Titterton's method	310

14.3.4	Probabilistic criterion 1	310
14.3.5	Probabilistic criterion 2: working directly with the cost	311
14.3.6	Unknown variances	311
14.3.7	Resampling implementation	312
14.3.8	Simulation results	314
14.3.9	Further work on this problem	317

IV Applications 319

15 Posterior Cramér-Rao Bounds for Sequential Estimation 321

Niclas Bergman

15.1	Introduction	321
15.2	Review of the posterior Cramér-Rao bound	322
15.3	Bounds for sequential estimation	323
15.3.1	Estimation model	324
15.3.2	Posterior Cramér-Rao bound	325
15.3.3	Relative Monte Carlo evaluation	327
15.4	Example – terrain navigation	329
15.5	Conclusions	338

16 Statistical Models of Visual Shape and Motion 339

Andrew Blake, Michael Isard, and John MacCormick

16.1	Introduction	339
16.2	Statistical modelling of shape	341
16.3	Statistical modelling of image observations	343
16.4	Sampling methods	345
16.5	Modelling dynamics	346
16.6	Learning dynamics	349
16.7	Particle filtering	352
16.8	Dynamics with discrete states	354
16.9	Conclusions	355

17 Sequential Monte Carlo Methods for Neural Networks 359

N de Freitas, C Andrieu, P Højén-Sørensen, M Niranjan, and
A Gee

17.1	Introduction	359
17.2	Model specification	360
17.2.1	MLP models for regression and classification . . .	360
17.2.2	Variable dimension RBF models	362
17.3	Estimation objectives	365
17.4	General SMC algorithm	366
17.4.1	Importance sampling step	367
17.4.2	Selection step	368

17.4.3	MCMC Step	369
17.4.4	Exact step	371
17.5	On-line classification	371
17.5.1	Simple classification example	372
17.5.2	An application to fault detection in marine diesel engines	373
17.6	An application to financial time series	375
17.7	Conclusions	379
18	Sequential Estimation of Signals under Model Uncertainty	381
	Petar M. Djurić	
18.1	Introduction	381
18.2	The problem of parameter estimation under uncertainty	383
18.3	Sequential updating of the solution	384
18.4	Sequential algorithm for computing the solution	389
18.4.1	A Sequential-Importance-Resampling scheme	390
18.4.2	Sequential sampling scheme based on mixtures	395
18.5	Example	397
18.6	Conclusions	400
18.7	Acknowledgment	400
19	Particle Filters for Mobile Robot Localization	401
	Dieter Fox, Sebastian Thrun, Wolfram Burgard, and Frank Dellaert	
19.1	Introduction	401
19.2	Monte Carlo localization	403
19.2.1	Bayes filtering	403
19.2.2	Models of robot motion and perception	404
19.2.3	Implementation as particle filters	405
19.2.4	Robot results	408
19.2.5	Comparison to grid-based localization	410
19.3	MCL with mixture proposal distributions	414
19.3.1	The need for better sampling	414
19.3.2	An alternative proposal distribution	416
19.3.3	The mixture proposal distribution	419
19.3.4	Robot results	420
19.4	Multi-robot MCL	423
19.4.1	Basic considerations	423
19.4.2	Robot results	425
19.5	Conclusion	426
20	Self-Organizing Time Series Model	429
	Tomoyuki Higuchi	
20.1	Introduction	429

20.1.1	Generalised state-space model	429
20.1.2	Monte Carlo filter	430
20.2	Self-organizing time series model	432
20.2.1	Genetic algorithm filter	432
20.2.2	Self-organizing state-space model	434
20.3	Resampling scheme for filtering	435
20.3.1	Selection scheme	435
20.3.2	Comparison of performance: simulation study	436
20.4	Application	438
20.4.1	Time-varying frequency wave in small count data	438
20.4.2	Self-organizing state-space model for time-varying frequency wave	439
20.4.3	Results	440
20.5	Conclusions	444
21	Sampling in Factored Dynamic Systems	445
	Daphne Koller and Uri Lerner	
21.1	Introduction	445
21.2	Structured probabilistic models	448
21.2.1	Bayesian networks	448
21.2.2	Hybrid networks	449
21.2.3	Dynamic Bayesian networks	451
21.3	Particle filtering for DBNs	454
21.4	Experimental results	457
21.5	Conclusions	464
22	In-Situ Ellipsometry Solutions Using Sequential Monte Carlo	465
	Alan D. Marrs	
22.1	Introduction	465
22.2	Application background	465
22.3	State-space model	467
22.3.1	Ellipsometry measurement model	468
22.3.2	System evolution model	471
22.3.3	Particle filter	472
22.4	Results	474
22.5	Conclusion	475
22.6	Acknowledgments	477
23	Manoeuvring Target Tracking Using a Multiple-Model Bootstrap Filter	479
	Shaun McGinnity and George W. Irwin	
23.1	Introduction	479
23.2	Optimal multiple-model solution	481
23.3	The IMM algorithm	483

23.4	Multiple model bootstrap filter	484
23.4.1	Example	486
23.5	Target tracking examples	488
23.5.1	Target scenarios	488
23.5.2	Linear, Gaussian tests	488
23.5.3	Polar simulation results	492
23.6	Conclusions	495
23.7	Acknowledgments	496
24	Rao-Blackwellised Particle Filtering for Dynamic Bayesian Networks	499
	Kevin Murphy and Stuart Russell	
24.1	Introduction	499
24.2	RBPF in general	500
24.2.1	How do we choose which nodes to sample?	503
24.3	The RBPF algorithm in detail	506
24.4	Application: concurrent localisation and map learning for a mobile robot	508
24.4.1	Results on a one-dimensional grid	511
24.4.2	Results on a two-dimensional grid	514
24.5	Conclusions and future work	515
25	Particles and Mixtures for Tracking and Guidance	517
	David Salmond and Neil Gordon	
25.1	Introduction	517
25.1.1	Guidance as a stochastic control problem	518
25.1.2	Information state	519
25.1.3	Dynamic programming and the dual effect	520
25.1.4	Separability and certainty equivalence	521
25.1.5	Sub-optimal strategies	522
25.2	Derivation of control laws from particles	523
25.2.1	Certainty equivalence control	523
25.2.2	A scheme based on open-loop feedback control . .	524
25.3	Guidance in the presence of intermittent spurious objects and clutter	525
25.3.1	Introduction	525
25.3.2	Problem formulation	525
25.3.3	Simulation example	526
25.3.4	Guidance results	528
26	Monte Carlo Techniques for Automated Target Recognition	533
	Anuj Srivastava, Aaron D. Lanterman, Ulf Grenander, Marc Loizeaux, and Michael I. Miller	
26.1	Introduction	533

	Contents	xix
26.1.1	The Bayesian posterior	535
26.1.2	Inference engines	536
26.2	Jump-diffusion sampling	539
26.2.1	Diffusion Processes	540
26.2.2	Jump processes	541
26.2.3	Jump-diffusion algorithm	544
26.3	Sensor models	545
26.4	Experiments	547
26.5	Acknowledgments	552
	Bibliography	553
	Index	577

Contributors

- Christophe Andrieu** Department of Mathematics,
University of Bristol,
Bristol, BS8 1TW, England
C.Andrieu@bristol.ac.uk
- Niclas Bergman** Data Fusion,
SaabTech Systems,
SE-175 88 Jarfalla, Sweden
ncbe@celsiustech.se
- Carlo Berzuini** Dipartimento di Informatica e Sistemistica,
Via Ferrata 1, University of Pavia,
27100 Pavia, Italy
carlo@laplace.unipv.it
- Andrew Blake** Microsoft Research Center,
1 Guildhall Street
Cambridge, CB2 3NH, England
ablake@microsoft.com
- Erik Bølviken** Department of Mathematics,
University of Oslo,
N-0316 Oslo, Norway
erikb@math.uio.no
- Wolfram Burgard** Institut für Informatik,
Albert-Ludwigs-Universität Freiburg,
D-79110 Freiburg, Germany
burgard@informatik.uni-freiburg.de
- Rong Chen** Department of Statistics,
Texas A&M University,
College Station TX 77845, USA
chen@stat.tamu.edu

- Tim Clapp** Astrium UK,
Gunnels Wood Road,
Stevenage, SG1 2AS, England
- Dan Crisan** Department of Mathematics,
Huxley Building,
Imperial College, 180 Queens Gate,
London, SW7 2BZ, England
dcrisan@ic.ac.uk
- Nando de Freitas** Computer Science Division
University of California, Berkeley
Berkeley, CA 94720-1776, USA
jfgf@cs.berkeley.edu
- Pierre Del Moral** LSP, Bât. 1R1,
Université Paul Sabatier,
118 route de Narbonne,
31062 Toulouse, France
delmoral@cict.fr
- Frank Dellaert** School of Computer Science,
Carnegie Mellon University,
Pittsburgh, PA 15213-3891, USA
dellaert@cs.cmu.edu
- Petar Djurić** Department of Electrical and
Computer Engineering,
State University of New York,
Stony Brook, NY 11794, USA
djuric@ece.sunysb.edu
- Arnaud Doucet** Department of Electrical & Electronic Engineering,
The University of Melbourne,
Victoria 3010, Australia
doucet@ee.mu.oz.au
- Dieter Fox** Department of Computer Science and Engineering,
University of Washington,
Seattle, WA 98195-2350, USA
fox@cs.washington.edu
- Andrew Gee** Speech, Vision and Robotics Group,
Department of Engineering, Trumpington Street,
Cambridge University,
Cambridge, CB2 1PZ, England
ahg@eng.cam.ac.uk

- Walter Gilks** Medical Research Council,
Biostatistics Unit,
Robinson Way, Cambridge, CB2 2SR, England
wally.gilks@mrc-bsu.cam.ac.uk
- Simon Godsill** Signal Processing Group,
Department of Engineering, Trumpington Street,
Cambridge University,
Cambridge, CB2 1PZ, England
sjg@eng.cam.ac.uk
- Neil Gordon** Pattern and Information Processing,
Defence Evaluation and Research Agency,
St. Andrews Road, Malvern, WR14 3PS, England
N.Gordon@signal.dera.gov.uk
- Ulf Grenander** Division of Applied Mathematics,
Brown University,
Providence, RI 02912, USA
ulf@brownvm.brown.edu
- Tomoyuki Higuchi** The Institute of Statistical Mathematics,
4-6-7 Minami-Azabu, Minato-ku,
Tokyo 106-8569, Japan
higuchi@ism.ac.jp
- Pedro Højen-Sørensen** Section for Digital Signal Processing,
Department of Mathematical Modelling,
Technical University of Denmark,
DK-2800 Lyngby, Denmark
phs@eivind.imm.dtu.dk
- Markus Hürzeler** UBS AG,
Nüscherstrasse 22,
CH-8098, Zurich, Switzerland
markus.huerzeler@ubs.com
- George Irwin** School of Electrical and Electronic Engineering,
The Queen's University of Belfast,
Belfast, BT9 5AH, Northern Ireland
g.irwin@ee.qub.ac.uk
- Michael Isard** Computer Systems Research Center,
Palo Alto, CA 94301, USA
misard@pa.dec.com

- Jean Jacod** Laboratoire de Probabilités,
Université de Paris 6, 4 place Jussieu,
75252 Paris, France
jj@cer.jussieu.fr
- Genshiro Kitagawa** The Institute of Statistical Mathematics,
4-6-7 Minami-Azabu, Minato-ku,
Tokyo 106-8569, Japan
kitagawa@ism.ac.jp
- Daphne Koller** Computer Science Department
Stanford University
Stanford, CA 94305-9010, USA
koller@cs.stanford.edu
- Hans R. Künsch** Seminar für Statistik,
ETH Zentrum,
CH-8092 Zürich, Switzerland
kuensch@stat.math.ethz.ch
- Aaron Lanterman** Coordinated Science Laboratory,
University of Illinois,
Urbana, IL 61801, USA
lanterma@ifp.uiuc.edu
- François Le Gland** IRISA-INRIA, Campus de Beaulieu,
Campus de Beaulieu
35042 Rennes, France
legland@irisa.fr
- Uri Lerner** Computer Science Department,
Stanford University,
Stanford, CA 94305-9010, USA
uri@cs.stanford.edu
- Jane Liu** CDC Investment Management Corporation,
1251 Avenue of the Americas, 16th floor,
New York, NY 10020, USA
- Jun Liu** Department of Statistics,
Harvard University,
1 Oxford Street,
Cambridge, MA 02138, USA
jliu@stat.harvard.edu

- Tanya Logvinenko** Department of Statistics,
Stanford University,
Stanford, CA 94305, USA
tanyalog@stat.stanford.edu
- Marc Loizeaux** Department of Statistics,
Florida State University,
Tallahassee, FL 32306-4330, USA
- John MacCormick** Systems Research Center,
Compaq Computer Corporation,
Palo Alto, CA 94301, USA
jmac@pa.dec.com
- Alan Marrs** Pattern and Information Processing,
Defence Evaluation and Research Agency,
St. Andrews Road, Malvern, WR14 3PS, England
A.Marrs@signal.dera.gov.uk
- Shaun McGinnity** Openwave,
Charles House,
103-111 Donegall Street,
Belfast, BT1 2FJ, Northern Ireland
shaun.mcginnity@openwave.com
- Michael Miller** Center for Imaging Science,
Johns Hopkins University,
Baltimore, MD 21218, USA
mim@cis.jhu.edu
- Kevin Murphy** Computer Science Division,
University of California, Berkeley
Berkeley, CA 94720-1776, USA
murphyk@cs.berkeley.edu
- Christian Musso** ONERA-DTIM-MCT,
92322 Chatillon, France
christian.musso@onera.fr
- Mahesan Niranjan** Department of Computer Science,
University of Sheffield,
Sheffield, S1 4DP, England
M.Niranjan@dcs.shef.ac.uk

- Nadia Oudjane** Universite de Paris 13
Department of Mathematics,
Avenue Jean-Baptiste Clement,
93430 Villetaneuse, France
- Michael Pitt** Department of Economics,
University of Warwick,
Coventry, CV4 7AL, UK
M.K.Pitt@warwick.ac.uk
- Elena Punskeya** Signal Processing Group,
Department of Engineering, Trumpington Street,
Cambridge University,
Cambridge, CB2 1PZ, England
op205@eng.cam.ac.uk
- Stuart Russell** Computer Science Division
University of California, Berkeley
Berkeley, CA 94720-1776, USA
russell@cs.berkeley.edu
- David Salmond** Guidance and Associated Processing,
Defence Evaluation and Research Agency,
Ively Road, Farnborough, GU14 0LX, England
djsalmond@dera.gov.uk
- Seisho Sato** The Institute of Statistical Mathematics,
4-6-7 Minami-Azabu, Minato-ku,
Tokyo 106-8569, Japan
sato@ism.ac.jp
- Neil Shephard** Nuffield College,
Oxford University,
Oxford, OX1 1NF, England
neil.shephard@nuffield.ox.ac.uk
- Adrian Smith** Queen Mary and Westfield College,
Mile End Road, London, E1 4NS, England
principal@qmw.ac.uk
- Anuj Srivastava** Department of Statistics,
Florida State University,
Tallahassee, FL 32306-4330, USA
anuj@stat.fsu.edu

- Photis Stavropoulos** Department of Statistics,
University of Glasgow,
Glasgow G12 8QQ, Scotland
- Geir Storvik** Department of Mathematics,
University of Oslo,
N-0316 Oslo, Norway
geirs@math.uio.no
- Sebastian Thrun** School of Computer Science,
Carnegie Mellon University,
Pittsburgh, PA 15213-3891, USA
thrun@cs.cmu.edu
- Mike Titterington** Department of Statistics,
University of Glasgow,
Glasgow G12 8QQ, Scotland
m.titterington@stats.gla.ac.uk
- Mike West** Institute of Statistics & Decision Sciences,
Duke University,
Durham, NC 27708-0251, USA
mw@stat.duke.edu

Part III

Strategies for Improving Sequential Monte Carlo Methods

Deterministic and Stochastic Particle Filters in State-Space Models

Erik Bølviken
Geir Storvik

5.1 Introduction

Optimal or Bayesian filtering in state-space models is a question of computing series of linked numerical integrals where output from one is input to the other (Bucy and Senne 1971). Particle filtering can be regarded as comprising techniques for solving these integrals by replacing the complicated posterior densities involved by *discrete* approximations, based on *particles* (Kitagawa 1996). There is evidence that the numerical errors as the process is iterated often stabilise, or at least do not accumulate sharply (see section 5.2.5). Filters of this type can be constructed in many ways. Most of the contributions to this volume employ Monte Carlo designs (see also (Doucet 1998) and the references therein). Particles are then random drawings of state vectors under the current posterior. This amounts to Monte Carlo evaluations of integrals. Numerically inaccurate, but often practical and easy to implement, general methods to run the sampling have been developed. Alternatively, particles can be laid out through a deterministic plan, using more sophisticated and more accurate numerical integration techniques. Such an approach has been discussed in (Kitagawa 1987), (Pole and West 1988) and (Pole and West 1990), but recently most work has been based on Monte Carlo methods. To some extent, Monte Carlo and deterministic particle filters are complementary approaches, and one may also wonder whether they may be usefully combined (see (Monahan and Genz 1997) for such a combination in a non-dynamic setting). Emphasis in this paper is on deterministic filtering. A general framework can be found in the above-mentioned references and in (West and Harrison 1997)[Section 13.5]. We shall present a common perspective in the next section, where our contribution will be on design issues.

Deterministic filtering could be the method of choice when the state process has low dimension and high numerical accuracy is wanted. Our applications, which are a group of statistical problems from population biology, are of that type. The biological signals in some of the examples in Section 5.3 are extremely weak, yet governed by well-defined stochastic models derived from basic theory on fluctuations of animal populations. The scientific objective is in understanding the underlying mechanisms. This means model identification, and in a frequency context, computation and maximization of likelihood functions. Filtering is then a mean, to compute such quantities, rather than an end in itself, but the computational problems remain much the same as in engineering, where the interest is usually in the filtering estimates. Even though our statistical applications are off-line, computer speed is still highly relevant because the computations are often repeated many times to estimate uncertainty through bootstrapping. One difference from engineering, however, may be the attitude towards Monte Carlo error. In ad-hoc situations in which a filtered estimate is used for, say, guidance or control, added uncertainty that is a small fraction of the total may not matter much. In a context of basic science, as in Section 5.3, this could be different, and Monte Carlo error in likelihood evaluations can be awkward if the function is to be maximised. Various types of computing likelihoods will be compared in Section 5.3.

5.2 General issues

5.2.1 *Model and exact filter*

We are dealing with a discrete time vectorial process $\{\mathbf{x}_t\}$ in $\mathcal{R}^{n_x}$, observed indirectly through another process $\{\mathbf{y}_t\}$ in $\mathcal{R}^{n_y}$. The framework is a Markov model for $\{\mathbf{x}_t\}$ based on (possibly time-varying) transition densities $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ and conditionally independent observations $\{\mathbf{y}_t\}$ given $\{\mathbf{x}_t\}$. The latter usually means that the conditional density of the vector $\mathbf{y}_{1:t} \triangleq (\mathbf{y}_1, \dots, \mathbf{y}_t)$ given the corresponding $\mathbf{x}_{1:t} \triangleq (\mathbf{x}_1, \dots, \mathbf{x}_t)$ factors into $p(\mathbf{y}_{1:t}|\mathbf{x}_{1:t}) = \prod_{s=1}^t p(\mathbf{y}_s|\mathbf{x}_s)$, where $p(\mathbf{y}_s|\mathbf{x}_s)$ is the density of $\mathbf{y}_s$ given $\mathbf{x}_s$. The symbol p will throughout be used to designate density functions of various kinds, for example in addition to those above, $p(\mathbf{y}_{1:t})$ for the density function of $\mathbf{y}_{1:t}$. These functions may depend on t themselves, say $p_t(\mathbf{y}_{1:t})$ or $p_s(\mathbf{y}_s|\mathbf{x}_s)$ rather than $p(\mathbf{y}_{1:t})$ or $p(\mathbf{y}_s|\mathbf{x}_s)$, but it is not necessary to include this in the notation.

The exact filter for $\{\mathbf{x}_t\}$ can be written as a set of recursive integration equations (Bucy and Senne 1971). Start with $p(\mathbf{x}_0) \triangleq p(\mathbf{x}_0|\mathbf{y}_0)$ as prior for

$\mathbf{x}_0$ and calculate recursively

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})d\mathbf{x}_{t-1} \quad (5.2.1)$$

$$C_t = \int p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})d\mathbf{x}_t \quad (5.2.2)$$

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = C_t^{-1}p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1}), \quad (5.2.3)$$

for $t = 1, \dots, T$. The normalisation constants $\{C_t\}$ then produce the log-likelihood function of observations $\mathbf{y}_{1:T}$ through

$$\log\{p(\mathbf{y}_{1:T})\} = \sum_{t=1}^T \log(C_t). \quad (5.2.4)$$

The relationships (5.2.1-5.2.4) are well known and are consequences of $\{\mathbf{x}_t\}$ being a Markov process and $\{\mathbf{y}_t\}$ being conditionally independent given $\{\mathbf{x}_t\}$. The proof is an elementary application of Bayes' formula.

5.2.2 Particle filters

Computation of the posterior densities $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ and the log likelihood function $p(\mathbf{y}_{1:T})$ is an exercise in high-dimensional numerical integration. General purpose integration methods are bound to be inefficient, and it seems better to design iterative schemes imitating the exact one (5.2.1-5.2.3). Particle filters proceed in this manner through *discrete* approximations to the exact posterior distributions. Let $\hat{p}(\mathbf{x}_t|\mathbf{y}_{1:t})$ be some discrete analogue to the exact density $p(\mathbf{x}_t|\mathbf{y}_{1:t})$. The points $\mathbf{x}_t^{(i)}$ on which $\hat{p}(\mathbf{x}_t)$ assigns positive probabilities are known as the *particles*. Their number N_t may vary. Suppose a reasonable approximation $\hat{p}_{t-1}(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})$ is available at time $t-1$. When inserted for the exact density $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})$ on the right in (5.2.1), we obtain

$$\hat{p}(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \sum_{j=1}^{N_{t-1}} p(\mathbf{x}_t|\mathbf{x}_{t-1}^{(j)})\hat{p}(\mathbf{x}_{t-1}^{(j)}|\mathbf{y}_{1:t-1}). \quad (5.2.5)$$

The main point of the design is to ensure a good approximation to the exact predictive density $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$. When (5.2.5) replaces its exact counterpart in (5.2.3), we immediately have

$$\tilde{p}(\mathbf{x}_t|\mathbf{y}_{1:t}) = \tilde{C}_t^{-1}p(\mathbf{y}_t|\mathbf{x}_t)\hat{p}(\mathbf{x}_t|\mathbf{y}_{1:t-1}), \quad (\tilde{C}_t \text{ a constant}), \quad (5.2.6)$$

as an approximate update density. To complete the recursion, (5.2.6) must be replaced by a particle approximation $\hat{p}(\mathbf{x}_t|\mathbf{y}_{1:t})$. Filters proposed in the literature vary as to how this step should be carried out. Most of the papers in this volume use Monte Carlo sampling. If the particles $\mathbf{x}_t^{(i)}$ are drawn randomly from (5.2.6), the probabilities become $\hat{p}(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t}) = N_t^{-1}$. We may enhance numerical accuracy by importance sampling (Doucet 1998),

the probabilities are then the importance weights. In recent years, many tricks have been invented to obtain computationally fast sampling.

Stochastic generation of particles corresponds to Monte Carlo evaluations of integrals. An alternative is to use other, more accurate integration methods. This raises, as we shall see in Section 5.3, various issues regarding design, but to understand this we shall first review elements of the theory of numerical integration.

5.2.3 Gaussian quadrature

Consider the case of a *scalar* process x_t , denoted x in this sequel. The integrands in (5.2.1) and (5.2.2) are of the form $h(x)p(x)$ with h some function and p a density. Evaluations of integrals are often efficiently carried out by Gaussian quadrature. The simplest among these rules is the Gauss-Legendre method. The standard form is

$$\int_{-1}^1 h(x)p(x)dx = \sum_{i=1}^m \gamma^{(i)} h(\xi^{(i)}) p(\xi^{(i)}) + \mathcal{E}_0, \quad (5.2.7)$$

where the abscissas $\xi^{(1)}, \dots, \xi^{(m)}$ and the positive weights $\gamma^{(1)}, \dots, \gamma^{(m)}$ are tabulated in numerical literature; see also (Press, Teukolsky, Vetterling and Flannery 1992) for a simple computer program. For a general interval (5.2.7) changes into

$$\int_A^B h(x)p(x)dx = \sum_{i=1}^m w^{(i)} h(x^{(i)}) p(x^{(i)}) + \mathcal{E}, \quad (5.2.8)$$

where

$$x^{(i)} = \frac{1}{2}(A + B + (B - A)\xi^{(i)}), \quad w^{(i)} = \frac{1}{2}(B - A)\gamma^{(i)}. \quad (5.2.9)$$

The error term $\mathcal{E}$ is

$$\mathcal{E} = \frac{1}{2}(B - A)^{2m+1} \frac{(m!)^4}{\{(2m)!\}^3(2m+1)} (ph)^{(2m)}(\omega), \quad (5.2.10)$$

where $(ph)^{(2m)}(\omega)$ is the derivative of order $2m$ of ph at some $\omega \in (A, B)$. The error term thus vanishes if the integrand is a polynomial of degree $< 2m$, and the accuracy is almost startling otherwise. Assuming $B - A = 10$, $\mathcal{E}$ being of order 10^{-10} times the derivative for $m = 10$ and of 10^{-32} for $m = 20$ for $B - A = 10$ (but higher order derivatives often grow!).

The integral approximation (5.2.8-5.2.9) has a probabilistic interpretation. Suppose the probability mass of p is negligible outside the interval (A, B) . Define

$$\hat{p}(x^i) = w^{(i)} p(x^{(i)}) \quad (5.2.11)$$

and insert $h \equiv 1$ in (5.2.8). Then

$$\sum_{i=1}^m \hat{p}(x^{(i)}) \approx \int_A^B p(x) dx \approx 1,$$

so that $\hat{p}(x^{(1)}), \dots, \hat{p}(x^{(m)})$ is almost a proper probability vector. Normalisation is usually not worth the bother. Let $\hat{p}$ be the distribution assigning these probabilities to $x^{(i)}$, $i = 1, \dots, m$. Then (5.2.8) expresses that $\hat{p}$ is an approximation to p in the sense that the expectation of $h(x)$ for any *smooth* function h is almost equal for the two distributions. Note that $\hat{p}(x^{(i)})$ and $p(x^{(i)})$ deviate.

Other quadrature rules produce similar particle approximations through an idea similar to importance sampling. Choose some density ψ and note that $\int h p = \int h R \psi$ where $R = p/\psi$. Quadrature with ψ as a weight function yields

$$\int h(x) p(x) dx = \int h(x) R(x) \psi(x) dx = \sum_{i=1}^m \gamma^{(i)} h(\xi^{(i)}) R(\xi^{(i)}) + \mathcal{E}, \quad (5.2.12)$$

where abscissas $\xi^{(i)}$ and weights $\gamma^{(i)}$, which depend on ψ , differ from those in (5.2.7). The error term now vanishes if hR is polynomial of degree less than $2m$. The particle approximation (similar to (5.2.11)) becomes $x^{(i)} = \xi^{(i)}$ and $\hat{p}(x^{(i)}) = \gamma^{(i)} p(x^{(i)}) / \psi(x^{(i)})$.

5.2.4 Quadrature filters

Quadrature filters are constructed by replacing the density $\tilde{p}(\mathbf{x}_t | \mathbf{y}_{1:t-1})$ in (5.2.6) by a particle approximation based on the quadrature formulas in Section 5.2.3. This leads to the following recursive scheme, where

$$\hat{p}(\mathbf{x}_t^{(i)} | \mathbf{y}_{1:t-1}) = \sum_{j=1}^{N_{t-1}} p(\mathbf{x}_t^{(i)} | \mathbf{x}_{t-1}^{(j)}) \hat{p}(\mathbf{x}_{t-1}^{(j)} | \mathbf{y}_{1:t-1}) \quad (5.2.13)$$

$$\hat{C}_t = \sum_{i=1}^{N_t} p(\mathbf{y}_t | \mathbf{x}_t^{(i)}) w_t^{(i)} \hat{p}(\mathbf{x}_t^{(i)} | \mathbf{y}_{1:t-1}) \quad (5.2.14)$$

$$\hat{p}(\mathbf{x}_t^{(i)} | \mathbf{y}_{1:t}) = \hat{C}_t^{-1} p(\mathbf{y}_t | \mathbf{x}_t^{(i)}) w_t^{(i)} \hat{p}(\mathbf{x}_t^{(i)} | \mathbf{y}_{1:t-1}), \quad (5.2.15)$$

for $i = 1, \dots, N_t$. Implementing the scheme (5.2.13-5.2.15) when $n_x > 1$ raises a number of problems. The mathematical theory of Gaussian quadrature described above is inherently one-dimensional and must be applied sequentially, one state variable at a time. The weights $w^{(i)}$ in (5.2.14) and (5.2.15) will then be *products* of weights from each of the n_x variables; see Section 5.3.2 for details. This also suggests that the computational requirements will grow rapidly with n_x . The number of operations in quadrature filtering is proportional to N_t^2 for each step of the recursion.

When $n_x = 1$, as in the example, N_t equaled the order m of the quadrature rule. But if n_x abscissas are used for each variable of the state vector, then $N_t = m^{n_x}$, easily a huge number. A particle representation of n_x *independent* variables actually needs such a large N_t to be accurate, but if there are correlations it may be possible reduce it.

5.2.5 Numerical error

It is possible to gain some insight into how numerical error propagates by elementary methods. Let

$$\delta_t = \log\{\hat{p}(\mathbf{y}_{1:t})\} - \log\{p(\mathbf{y}_{1:t})\} = \sum_{s=1}^t \log \hat{C}_s - \sum_{s=1}^t \log C_s \quad (5.2.16)$$

be the accumulated error in the log-likelihood function, and define

$$\varepsilon(\mathbf{x}_t | \mathbf{y}_{1:t-1}) = \exp(\delta_{t-1}) \hat{p}(\mathbf{x}_t | \mathbf{y}_{1:t-1}) - p(\mathbf{x}_t | \mathbf{y}_{1:t-1}) \quad (5.2.17)$$

as an indicator of numerical error present in the distribution $\hat{p}(\mathbf{x}_t | \mathbf{y}_{1:t-1})$. The factor $\exp(\delta_{t-1})$ is to avoid normalisations in the recursion (5.2.18) below.

By elementary manipulations it is easily proved (see appendix) that

$$\varepsilon(\mathbf{x}_{t+1} | \mathbf{y}_{1:t}) = C_t^{-1} \int p(\mathbf{x}_{t+1} | \mathbf{x}_t) p(\mathbf{y}_t | \mathbf{x}_t) \varepsilon(\mathbf{x}_t | \mathbf{y}_{1:t-1}) d\mathbf{x}_t + \eta(\mathbf{x}_{t+1}), \quad (5.2.18)$$

where

$$\eta(\mathbf{x}_{t+1}) = \exp(\delta_t) \left[\sum_i p(\mathbf{x}_{t+1} | \mathbf{x}_t^{(i)}) \hat{p}(\mathbf{x}_t^{(i)} | \mathbf{y}_t) - \int p(\mathbf{x}_{t+1} | \mathbf{x}_t) \hat{p}(\mathbf{x}_t | \mathbf{y}_{1:t}) d\mathbf{x}_t \right]. \quad (5.2.19)$$

If we insert (5.2.3) into (5.2.1) it emerges that the sequence of posterior densities $\{p(\mathbf{x}_t | \mathbf{y}_{t-1})\}$ themselves satisfy exactly the same recursion (5.2.18) but for the term $\eta(\mathbf{x}_{t+1})$, which signifies the contribution of the numerical error at time t . Suppose now that r_t is the smallest real number for which

$$|\varepsilon(\mathbf{x}_t | \mathbf{y}_{1:t-1})| \leq r_t p(\mathbf{x}_t | \mathbf{y}_{1:t-1}) \quad (5.2.20)$$

so that r_t controls the *relative error* in $\hat{p}(\mathbf{x}_t | \mathbf{y}_{1:t-1})$. Then, by inserting (5.2.20) in the integrand (5.2.18) it follows that

$$r_{t+1} \leq r_t + \tilde{\eta}(\mathbf{x}_{t+1}), \quad (5.2.21)$$

where

$$\tilde{\eta}(\mathbf{x}_{t+1}) = \eta(\mathbf{x}_{t+1}) / p(\mathbf{x}_{t+1} | \mathbf{y}_{1:t}) \quad (5.2.22)$$

is the relative numerical error introduced at time t .

Relative error in the approximations diminishes through the next step of the scheme. New error is brought, as well, but the old relative error does not blow up as the recursion is progressed.

It is possible for Gaussian quadrature filters to give a rigorous error statement. Suppose we have been able to find approximations so that

$$\left| \sum_{j=1}^{N_{t-1}} w_t^{(j)} p(\mathbf{x}_t^{(i)} | \mathbf{x}_{t-1}^{(j)}) p(\mathbf{x}_{t-1}^{(j)} | \mathbf{y}_{1:t-1}) - p(\mathbf{x}_t^{(i)} | \mathbf{y}_{t-1}) \right| \leq \varepsilon p(\mathbf{x}_t^{(i)} | \mathbf{y}_{1:t-1}) \quad (5.2.23)$$

for all i, t . Notice that this condition is for $p(\mathbf{x}_t^{(j)} | \mathbf{y}_{1:t})$ known and for a finite set of $\mathbf{x}_{t+1}^{(i)}$ values, which is always possible to obtain by setting N_t large enough. It is then proved (in the Appendix) that

$$r_{t+1} \leq r_t(1 + \varepsilon)^t + \varepsilon, \quad (5.2.24)$$

which implies that

$$r_t \leq (1 + \varepsilon)^t - 1. \quad (5.2.25)$$

By defining

$$\varepsilon(\mathbf{y}_t | \mathbf{y}_{1:t-1}) = \exp(\delta_{t-1}) \hat{p}(\mathbf{y}_t | \mathbf{y}_{1:t-1}) - p(\mathbf{y}_t | \mathbf{y}_{1:t-1}) \quad (5.2.26)$$

and replacing condition (5.2.23) with

$$\left| \sum_{i=1}^{N_t} w_t^{(i)} p(\mathbf{y}_t | \mathbf{x}_t^{(i)}) p(\mathbf{x}_t^{(i)} | \mathbf{y}_{1:t-1}) - p(\mathbf{y}_t | \mathbf{y}_{t-1}) \right| \leq \varepsilon p(\mathbf{x}_t^{(i)} | \mathbf{y}_{1:t-1}) \quad (5.2.27)$$

for all t , it is by similar reasoning possible to show that

$$|\varepsilon(\mathbf{y}_t | \mathbf{y}_{1:t-1})| \leq [r_t(1 + \varepsilon) + \varepsilon] p(\mathbf{y}_t | \mathbf{y}_{1:t-1}). \quad (5.2.28)$$

However,

$$\varepsilon(\mathbf{y}_t | \mathbf{y}_{1:t-1}) = \exp(\delta_{t-1}) \hat{C}_t - C_t = [\exp(\delta_t) - 1] p(\mathbf{y}_t | \mathbf{y}_{1:t-1}),$$

which results in

$$|\exp(\delta_t) - 1| \leq r_t(1 + \varepsilon) + \varepsilon \leq (1 + \varepsilon)^{t+1} - 1, \quad (5.2.29)$$

and consequently that

$$\delta_t \leq (t + 1) \log(1 + \varepsilon), \quad (5.2.30)$$

meaning the error is approximately linear in t . This shows that, although the error will increase with time, it will not accumulate sharply. Notice that, in general, a better limit for the error in the likelihood evaluation is not possible. This can be seen by considering a situation in which all observations are independent, such that it is straightforward to show that the error involved in the computation of the log-likelihood is the sum of errors at each time point.

5.2.6 A small illustrative example

In order to compare the quadrature filter with the Monte Carlo filter, we shall consider the one-dimensional nonlinear reference model (Doucet 1998)

$$x_t = \frac{1}{2}x_{t-1} + 25\frac{x_{t-1}}{1 + x_{t-1}^2} + 8\cos(1.2t) + e_t \quad (5.2.31)$$

$$z_t = \frac{x_t^2}{20} + v_t, \quad (5.2.32)$$

where $x_0 \sim \mathcal{N}(0, 5)$, e_t and v_t are mutually independent white Gaussian noises, $e_t \sim \mathcal{N}(0, \sigma_e^2)$ and $v_t \sim \mathcal{N}(0, \sigma_v^2)$ with $\sigma_e^2 = 10$ and $\sigma_v^2 = 1$. Data were simulated according to this model for $t = 1, \dots, T = 100$. The quadrature filter was implemented as described in Section 5.2.4, while rejection sampling with the predictive distribution as a proposal was used for the Monte Carlo filter. Although more effective Monte Carlo filters could be used, our main concern here is to examine the variability in the result, making the choice of Monte Carlo filter less important.

The filters were first run with a large number of particles ($N_t = N = 500$ for the quadrature filter and $N_t = N = 2000$ for the Monte Carlo filter). The results from the two filters agreed in this case, giving us the true posterior as a reference. In Figure 5.1, the estimates of the predictive density $p(x_{t+1}|y_{1:t})$ for $t = 100$ are displayed. For both filters, $N = 50$ were used. The true density is given by the solid curve. The Monte Carlo filter (long dashed curve) is able to find the two modes of the density, but is somewhat inaccurate on the weights on these modes. The quadrature filter (dotted) curve, on the other hand, is clearly able to yield the true density on almost all parts.

5.3 Case studies from ecology

5.3.1 Problem area and models

Population biology is a field in which optimal filtering is likely to yield substantial benefits, and we shall in this section demonstrate what can be achieved with such a technique. It is generally recognised that, in the stationary regime, animal populations oscillate due to feedback links between predator and prey and interactions with environment. A reasonable mathematical framework, see (Royama 1992), (Stenseth, Bjørnstad and Falck 1996), imposes nonlinear autoregressive models of low order on logarithms of abundance. The parameters of the models, influenced by vegetation and climatic conditions, might vary between sites. There is a dearth of data. The usual data reflect the number of animals caught in traps annually or seasonally, sometimes gathered by means of scientific experiment. Other data is arrived at deductively, e.g., from the number of skins delivered to

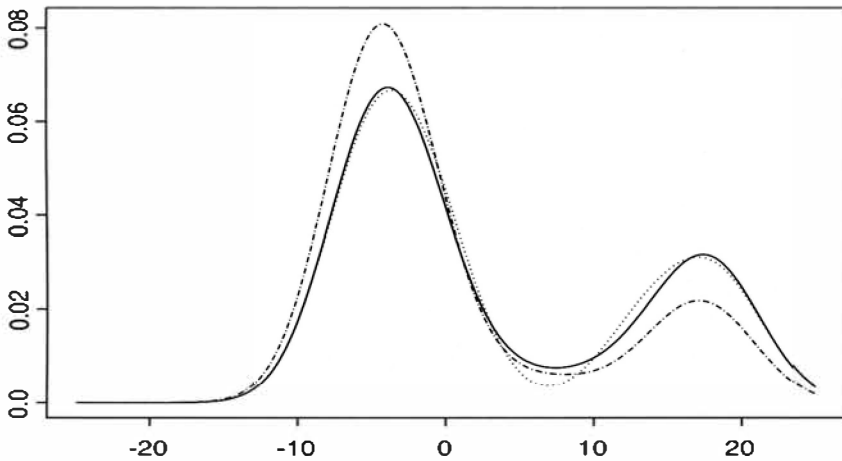


Figure 5.1. True and estimated predictive density $p(x_{101}|y_{1:100})$ for the reference model. The true density is given as a solid curve, the Monte Carlo filter estimate is given as a long dashed curve, and the quadrature estimate is given by dotted curve.

a company. The famous lynx series data, used in countless textbooks on statistical time series analysis (for example (Tong 1990)), are of the latter type.

The traditional approach in ecology (Stenseth, Bjørnstad and Saitoh 1996, Stenseth, Falck, Bjørnstad and Krebs 1997, Stenseth, Falck, Chain, Bjørnstad, Donoghue and Tong 1998)) is to identify the data with the animal population, but it is really more correct to regard data and population as different entities and employ a state model. This generates a more realistic picture of statistical uncertainties, deals with zero observations and opens for utilisation sources of data that have traditionally *not* been used scientifically; see below. Populations are, on first approximation, described by linear autoregressive models on log-scale, e.g.,

$$x_t = a_1 x_{t-1} + \dots + a_r x_{t-q} + e_t, \quad (5.3.1)$$

where $\{e_t\}$ is independent noise with zero mean and constant variance. The mean of x_t is not estimable under the measurement models introduced below, so has been subtracted out. Thus, in (5.3.1) x_t is the difference from the long-term average. Interest is directed towards the period (i.e. the location of the maximum of the spectral density), the autoregressive coefficients (which have ecological interpretations) and the order q of the series (which relates to ecological hypotheses).

The linear model (5.3.1) is useful, but fails to deal with the hypothesis that the built-up phase of animal abundance tends to be faster for a population in decline. There is for the lynx empirical evidence supporting such a hypothesis; see (Stenseth et al. 1997). One possibility, utilised by these authors, is to impose a TAR (truncated autoregressive) model of the type outlined in (Tong 1990). The single autoregressive relationship is then split into two separate ones, i.e.

$$x_t = \begin{cases} a_1 x_{t-1} + \dots a_r x_{t-q} + e_t, & \text{if } x_{t-d} \leq \theta; \\ b_1 x_{t-1} + \dots b_r x_{t-q} + e_t, & \text{if } x_{t-d} > \theta. \end{cases} \quad (5.3.2)$$

This split expresses that the population changes into a different phase (i.e. decline) d time units after the threshold θ has been passed. Mathematical properties of these models are summarised in (Tong 1990).

We shall consider three different measurement regimes $\{y_t\}$. The most common ones are *counts* (based on trappings). The natural model for y_t given x_t is then

$$p(y_t|x_t) = \frac{\lambda_t^{y_t}}{y_t!} \exp(-\lambda_t), \quad \log(\lambda_t) = \alpha + \beta x_t. \quad (5.3.3)$$

Notice that the parameter β influences the distribution of $\{y_t\}$ through $\beta\sigma_e$, where σ_e is the standard deviation of the series $\{e_t\}$ in (5.3.1) and (5.3.2). It is therefore not possible to estimate β and σ_e jointly. The same defect occurs in the other observation models (5.3.4) and (5.3.5) below.

Other types of data are categorical, based on judgments by experienced field observers. These sources of information have never been utilised scientifically, but would be a welcome additional source of information in population biology. One type of assessment is binary, y_t being 0 or 1, $y_t = 0$ signifying a population below average and $y_t = 1$ above. Such judgments are with errors. There can be no precise definition as to the precise meaning of these categories. A plausible model is

$$p(y_t|x_t) = p_t^{y_t} (1 - p_t)^{1-y_t}, \quad p_t = \frac{\exp(\alpha + \beta x_t)}{1 + \exp(\alpha + \beta x_t)}. \quad (5.3.4)$$

Here β is close to zero for bad observers and very large for good ones. The other parameter α captures *bias* in observer evaluation (upwards if $\alpha > 0$ and downwards if $\alpha < 0$). Other observations may be in terms of *differences* $x_t - x_{t-1}$ (*ratios* on the original scale). We shall in the next section discuss an example where $y_t \in \{-1, 0, 1\}$, where $y_t = -1$ means decline from one time point to the next, $y_t = 0$ signifies (at best) moderate changes and $y_t = 1$ growth. A possible model is now

$$p(y_t|x_t) = \begin{cases} c \exp(-\alpha_n - \beta(x_t - x_{t-1})) & \text{if } y_t = -1 \\ c & \text{if } y_t = 0 \\ c \exp(-\alpha_p + \beta(x_t - x_{t-1})) & \text{if } y_t = 1, \end{cases} \quad (5.3.5)$$

where c is a constant forcing the three probabilities to sum to one. Interpretations of the three parameters α_n , α_p and β are similar to those in (5.3.4).

5.3.2 Quadrature filters in practice

Constructing quadrature filters is a question of selecting particles $\mathbf{x}_t^{(i)}$ and their associated weights $w_t^{(i)}$ in the scheme (5.2.13-5.2.15) above. In the present instance, several issues are involved in this selection. How should joint particle grids in several variables be defined? How is the discontinuity in the TAR model (5.3.2) handled? The process $\{x_t\}$ describing the population is stationary and varies within a well-defined, and in practice limited, region. Should that process be utilised?

It is most convenient to answer these questions when the autoregressive processes (5.3.1) and (5.3.2) have order $q = 2$. The state vector at t is then (x_t, x_{t-1}) with state representation

$$\begin{pmatrix} x_t \\ x_{t-1} \end{pmatrix} = \begin{pmatrix} a_1 & a_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ x_{t-2} \end{pmatrix} + \begin{pmatrix} e_t \\ 0 \end{pmatrix}. \quad (5.3.6)$$

Notice the zero noise in the second component. The relation (5.2.1) of the exact filter becomes

$$p(x_t, x_{t-1} | y_{1:t-1}) = \int_{-\infty}^{\infty} p(x_t | x_{t-1}, x_{t-2}) p(x_{t-1}, x_{t-2} | y_{1:t-1}) dx_{t-2}, \quad (5.3.7)$$

which is a uni-dimensional integral despite the state vector's having two variables, it would have been univariate even for general q . Autoregressive processes are special because only one (of q) relations of the state representation has non-zero noise. This is responsible for the univariate integral (5.3.7) and is actually desirable. A general quadrature filter requires $O(N_t^2)$ operations for each step t of the recursion, but the constructions below are much faster.

However, the degenerate noise in (5.3.6) means that it is less clear how the joint grid for (x_t, x_{t-1}) should be constructed. At first glance, it may seem natural to use the version on the left of Figure 5.2, which utilises the correlation structure between x_t and x_{t-1} . Skew grids, as in Figure 5.2 left, follow by applying univariate particle approximations to the two factors sequentially. First, quadrature abscissas $x_{t-1}^{(j)}$ (needed for later integration) are laid out for x_{t-1} and the second approximation is conditioned on the first. The problem with this design is that the number of particles will grow explosively. Suppose there are N_{t-1} different particles for x_{t-1} . Each of them is unchanged as it moves ahead and each gives birth to m different particles for x_t , making $N_t = mN_{t-1}$. One way to keep this under control is to change the grid with fewer particles before progressing, but this

approach requires the posterior densities to be interpolated numerically, adding to the error. Further it is not cheap computationally. Interpolation from skew grids, such as that in Figure 5.2 left, is also a cumbersome process, especially in higher dimensions.

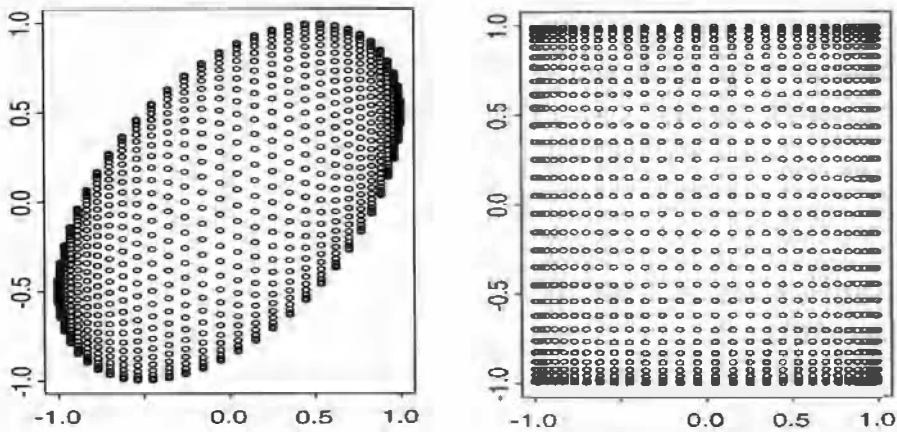


Figure 5.2. Particles for (x_t, x_{t-1}) as Gauss-Legendre abscissas in each dimension. On the left: particles constructed by conditioning x_t on x_{t-1} , on the right: particles constructed independently in each direction.

We have tested versions such as these, and they certainly work, but for stationary autoregressive processes we tend to prefer the simpler solution in Figure 5.2 right where the particles have been laid out as quadrature abscissas for the *marginal* densities of the two variables. This is not so parsimonious as approximations to the joint distribution, but now the number of particles can be kept constant as the algorithm progresses without interpolation. In detail, these methods work as described below when based on Gauss-Legendre quadrature.

First, a particle approximation to the density $p(x_0, x_{-1})$ is needed for initialisation. In the linear, Gaussian case the corresponding distribution has zero means and an easily computable covariance matrix; see (Priestley 1981). If σ_x is the standard deviation, we may apply (5.2.9) to each variable with $A = -B_0\sigma_x$ and $B = B_0\sigma_x$ for, say $B_0 = 5$. The square with edges $(\pm B_0\sigma_x, \pm B_0\sigma_x)$ then contains virtually the whole probability mass of the initial density $p(x_0, x_{-1})$, and the bivariate particle approximation becomes:

Initialisation

$$\begin{aligned}
x_0^{(i)} &= x_{-1}^{(i)} = B_0 \sigma_x \xi^{(i)} \\
\hat{p}(x_0^{(i)}, x_{-1}^{(j)}) &= p(x_0^{(i)}, x_{-1}^{(j)}) B_0^2 \gamma^{(i)} \gamma^{(j)}.
\end{aligned} \tag{5.3.8}$$

Closed-form expressions for the joint density $p(x_0, x_{-1})$ are rarely available apart from in the Gaussian, linear case. A technique that works well (and which was used for the TAR model) is that of running the main recursion below a suitable number of steps *without data* from a start such as in (5.3.8). Since $\{x_t\}$ by definition is stationary, the recursion will settle at a steady state. For the TAR model, where simple necessary and sufficient conditions for stationarity are not available (see (Tong 1990)), this is actually a way of checking for stationarity.

Next comes the main recursive step. By adapting the general scheme (5.2.13-5.2.15) by means of Gauss-Legendre integration, we obtain the following method (with $N_t = N$ for all t).

Algorithm 5.3.1 (Main recursion). Choose A_t and B_t (see text). Put

$$x_t^i = \frac{1}{2}(A_t + B_t + (B_t - A_t)\xi^{(i)}), \tag{5.3.9}$$

$$w_t^i = \frac{1}{2}(B_t - A_t)\gamma^{(i)}. \tag{5.3.10}$$

Update the probabilities by

$$\hat{p}(x_t^{(i)}, x_{t-1}^{(j)} | y_{1:t-1}) = \sum_{k=1}^N p(x_t^{(i)} | x_{t-1}^{(j)}, x_{t-2}^{(k)}) \hat{p}(x_{t-1}^{(j)}, x_{t-2}^{(k)}), \tag{5.3.11}$$

$$\hat{q}(x_t^{(i)}, x_{t-1}^{(j)} | y_{1:t}) = p(y_t | x_t^{(i)}, x_{t-1}^{(j)}) \hat{p}(x_t^{(i)}, x_{t-1}^{(j)} | y_{1:t-1}) w_t^{(i)} w_{t-1}^{(j)} \tag{5.3.12}$$

$$\hat{C}_t = \sum_{i=1}^N \sum_{j=1}^N \hat{q}(x_t^{(i)}, x_{t-1}^{(j)} | y_{1:t}) \tag{5.3.13}$$

$$\hat{p}(x_t^{(i)}, x_{t-1}^{(j)} | y_{1:t}) = \hat{C}_t^{-1} \hat{q}(x_t^{(i)}, x_{t-1}^{(j)} | y_{1:t}), \tag{5.3.14}$$

for $i, j = 1, \dots, N$.

The form of the density functions $p(y_t|x_t, x_{t-1})$ (equation (5.3.12)) will vary for the different measurements models. Extension to higher-order processes is straightforward. A modification is needed for the TAR model to cope with the discontinuity at θ . The interval (A_t, B_t) must now be split into two pieces, i.e. (A_t, θ) and (θ, B_t) , and the Gauss-Legendre abscissas applied to each part (if $\theta \notin (A_t, B_t)$ only one of the pieces is needed). The algorithm is in all other respects as above.

Several strategies can be used for selecting A_t and B_t . The simplest is to keep them unchanged, i.e. $A_t = A_0$ and $B_t = B_0$. The grids are then exactly the same everywhere, and the first two lines of the algorithm disappear. One advantage is that the transition probabilities $p(x_t^{(i)}|x_{t-1}^{(j)}, x_{t-2}^{(k)})$ do not change with t and can be pre-stored. This is our preferred method when dealing with the very noisy observations (5.3.4) and (5.3.5), but it is a possibility only because $\{x_t\}$ is stationary. In reality, many of the particles $(x_t^{(i)}, x_{t-1}^{(j)})$ have very low probabilities a-priori and can be removed with no noticeable effect. When $q = 3$, this trick can lead to huge reductions, up to several hundred times. When the data are more informative, as under the Poisson model (5.3.3), this strategy of fixed particles is not effective, and it is better to adapt (A_t, B_t) for each t . There are many ways to accomplish this. A rough way is first to compute, from the observed y_t , an interval where x_t is highly likely to be located, then a similar interval from the prediction density $p(x_t|y_{1:t-1})$ and use their intersection as (A_t, B_t) . This worked well in all our examples. Notice that this must be done before to finding updated means and variances of x_t , and it is inadvisable to allow the computer to spend too much time on this detail.

5.3.3 Numerical experiments

Different methods will be used in this section to evaluate likelihoods according to the models in Section 5.3.1. Results produced by quadrature filters and Monte Carlo filters will be compared, and we shall for linear autoregression (5.3.1) also investigate a simple approximate technique proposed in (Schnatter 1992), which is a special case of the so-called Gaussian sum filter in control engineering (Sorenson and Alspach 1971). Instead of working with particle approximations in the scheme described in Section 5.2.2, we use Gaussian ones instead. See (Schnatter 1992) for details.

In Figure 5.3 we have computed likelihood functions from simulated data based on the *linear* model (5.3.1) under the three scenarios shown in Table 5.1, all of them taken from estimates obtained from real data. All data are annual. Scenarios I and II represent lemmings or other small rodents. The period of the population models (corresponding to the maximum of the spectral density function) is a little over three years and the length of the series (50 years) representative for real data. The third scenario is constructed from real observations for the Canadian snow shoe hare;

Scenario	Population			Observations				
	a_1	a_2	σ_ε	Type	α	β	T	S
I	-0.5	-0.5	1.0	Poisson (5.3.3)	1.0	1.0	50	1
II	-0.5	-0.5	1.0	Logistic (5.3.4)	0.0	2.0	50	1
III	1.5	-0.9	1.0	Categorical (5.3.5)	-0.8	-0.8	17	100

Table 5.1. Scenarios for the linear auto-regression used in Figure 5.3. S is the number of time series and T is the length of each time series.

see (Bølviken, Glöckner and Stenseth 1999). The period is now much longer (about 9.6 years), and the research question is whether it is possible to recover the dynamics of the animal fluctuations from a large number of short series of 17 years.

It is fairly clear that the categorical data must contain information about the frequencies of oscillations of the underlying populations, but perhaps not so obvious that these data can reflect other information about the dynamics. We have therefore in Figure 5.3 plotted likelihood profiles along curves of constant periods in the (a_1, a_2) -space. The values of the nuisance parameters in the measurement models (5.3.3-5.3.5) are the correct ones, except on the lower right, where the horizontal axis represents variations of β . The results are encouraging. In scenario III (lower half of Figure 5.3), where much data support the computations, it is shown that the parameters can be consistently estimated, in accordance with the theoretical results in (Bickel, Ritov and Ryden 1998). The Monte Carlo effects there are much dampened as averages of 100 independent likelihoods, one for each 17-year series. The similarity between methods is great in all cases considered, but the Monte Carlo versions (based on 1000 replications) for scenario I and II are too erratic to optimise by numerical software. By contrast, the likelihood surfaces for the two other methods are smooth. The closeness in shape between the results from the Schnatter approximation and the more accurate quadrature filter is noteworthy, and most impressive because Schnatter's method is extremely fast. However, there are numerical errors in the *level* of the likelihood curves produced by her method. If model fit is to be judged through AIC or some other likelihood-based criterion, her approximation is not good enough. The quadrature filter was run with $m = 20$ abscissas in each direction in scenarios I and II, but in scenario III, where the parameters are much closer to the boundary of the stationary region, it was necessary to raise the number to $m = 50$. Computing time for the Poisson case was about 14 seconds for each likelihood evaluation on an Intel Pentium 200MHz MMX running on Linux 2.0.

A second round of experiments is presented for the TAR model (5.3.2) in Figure 5.4 under the scenarios shown in Table 5.2. Notice that the two regimes are defined at lag $d = 2$ in accordance with empirical studies (Stenseth, Bjørnstad and Falck 1996). For the snow shoe hare (Scenario III'), the model is the superposition of two stationary regimes, correspond-

Scenario	Population				
	a_1	a_2	b_1	b_2	σ_ε
I'	-0.5	0.0	-0.5	-0.7	1.0
III'	1.0	-0.4	1.0	-0.9	1.0

Table 5.2. Scenarios for the TAR auto-regression used in Figure 5.4. S is the number of time series and T is the length of each time series. The observation models are equal to scenarios I and III in Table 5.1.

ing to periods of about 12 (a -regime) and 6 (b -regime) years. This reflects prevailing ecological theory, deduced from studies of the lynx, of a slow growth phase and a faster decline, but no empirical studies of this nature have as yet been undertaken for the hare itself. The results displayed in Figure 5.4 are, however, encouraging and indicate that statistical studies may be able to estimate underlying nonlinear effects from categorical data.

It should be noted that for likelihood calculations, Monte Carlo estimates obtained by running the filter independently for each parameter set is probably not recommended. Some reduction in variability is obtained by using common random numbers, but in our situation this did not suffice to give smooth enough likelihood surfaces. A more promising alternative is the simulated likelihood ratio method (Billio, Monfort and Robert 1998), in which the filter (a smoothed version) is run only for one parameter set (the null set) and the likelihood ratio can be estimated at all other parameter sets by ratios of full likelihoods (including the underlying state process). This approach is also not really satisfactory because, although a smooth surface was obtained, the bias was too high for parameter sets insufficiently close to the null set.

5.4 Concluding remarks

In this paper we have been concerned with filtering for problems with low-dimensional state vectors. In such situations, *deterministic* filters may be preferable. Some results on the errors involved are given. It is shown that the errors in likelihood calculations propagate linearly in time. However, since the errors at each step can be made very small with a moderate number of particles, the linear increase is in many situations not a problem.

Constructing quadrature filters is a matter of selecting particles and their associated weights. In contrast to Monte Carlo filters, whereby the particles are automatically laid out in the high-probable regions, particles now have to be user-specified. Although many sophisticated designs are possible, we prefer to select the particles marginally on each variable. This results in many particles with weights close to zero, but this is outweighed by a much smaller computational burden and easier implementation. Further, for situations in which the information in the observations is sparse, parti-

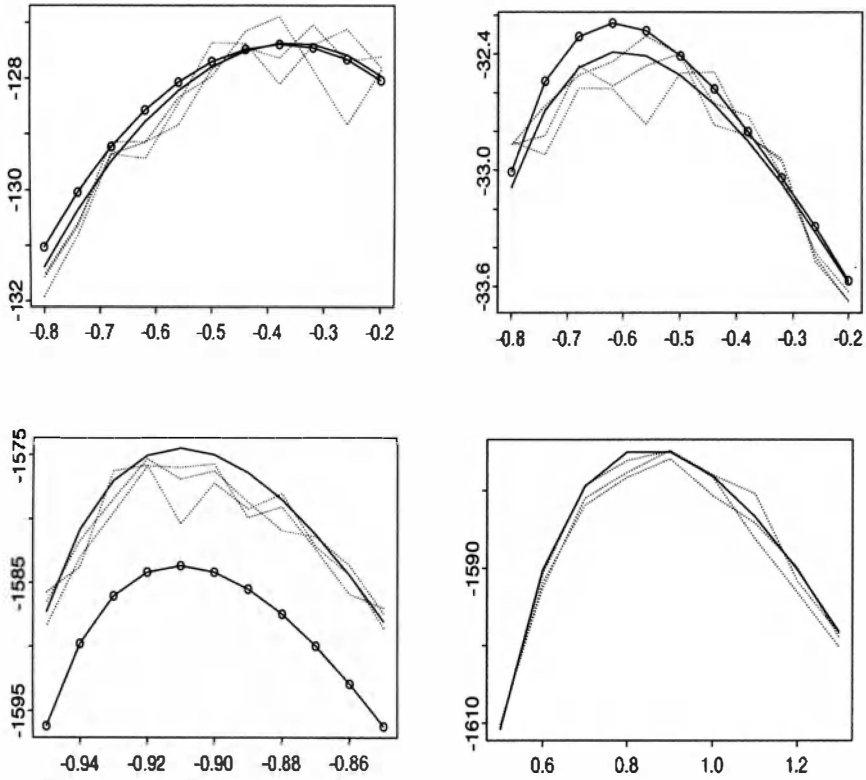


Figure 5.3. Estimated profile (log-)likelihood curves for the AR(2) process with Poisson (upper left), binary (upper right) and categorical (lower left and right) observations. Scenarios are given in Table 5.1. The x -axis displays the a_2 values used, except in the last plot, where the x -axis represents β -values. The quadrature filter estimate is given as a solid curve, and the estimate based on the Schnatter filter is given as a solid curve with dots (not given in the last plot). The three dashed dotted curves are different Monte Carlo estimates.

cles can successfully be specified directly from properties of the stationary distribution.

Deterministic filters are compared with Monte Carlo filters on some simulated data based on models representing animal populations. For these problems, likelihood-based inference is the primary concern, and large variabilities in these functions are problematic if the likelihood is to be maximised. In such cases, deterministic filters are superior because of their high accuracy.

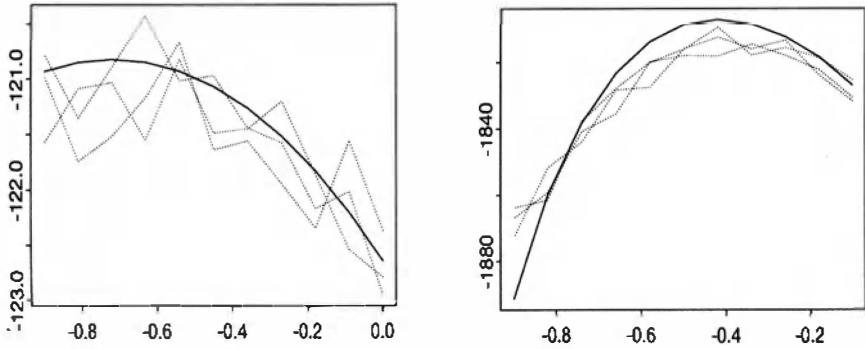


Figure 5.4. Estimated profile (log-)likelihood curves for the TAR(2) process with Poisson (left) and categorical (right) observations. Scenarios are given in Table 5.2. The x -axis displays the a_2 (b_2 in right plot) values used. The quadrature filter estimate is given as a solid curve. The three dashed dotted curves are different Monte Carlo estimates.

5.5 Appendix: Derivation of numerical errors

Equation (5.2.18) is verified by first noting that by (5.2.5) and (5.2.19)

$$\exp(\delta_t) \hat{p}(\mathbf{x}_{t+1} | \mathbf{y}_{1:t}) = \exp(\delta_t) \int p(\mathbf{x}_{t+1} | \mathbf{x}_t) \tilde{p}(\mathbf{x}_t | \mathbf{y}_{1:t}) d\mathbf{x}_t + \eta(\mathbf{x}_{t+1}).$$

Hence,

$$\varepsilon(\mathbf{x}_{t+1} | \mathbf{y}_{1:t}) = \int p(\mathbf{x}_{t+1} | \mathbf{x}_t) [\exp(\delta_t) \tilde{p}(\mathbf{x}_t | \mathbf{y}_{1:t}) - p(\mathbf{x}_t | \mathbf{y}_{1:t})] d\mathbf{x}_t + \eta(\mathbf{x}_{t+1}),$$

or, by using (5.2.3) and (5.2.6),

$$\begin{aligned} \varepsilon(\mathbf{x}_{t+1} | \mathbf{y}_{1:t}) &= \int p(\mathbf{x}_{t+1} | \mathbf{x}_t) p(\mathbf{y}_t | \mathbf{x}_t) [\exp(\delta_t) \tilde{C}_t^{-1} \hat{p}(\mathbf{x}_t | \mathbf{y}_{1:t-1}) \\ &\quad - C_t^{-1} p(\mathbf{x}_t | \mathbf{y}_{1:t-1})] d\mathbf{x}_t + \eta(\mathbf{x}_{t+1}), \end{aligned}$$

which reduces to (5.2.18).

The proof of (5.2.25) starts by noting that, using (5.2.14), we may write

$$\begin{aligned} \varepsilon(\mathbf{x}_{t+1} | \mathbf{y}_{1:t}) &= \exp(\delta_t) \sum_{i=1}^{N_t} p(\mathbf{x}_{t+1} | \mathbf{x}_t^{(i)}) \hat{p}(\mathbf{x}_t^{(i)} | \mathbf{y}_{1:t}) - p(\mathbf{x}_{t+1} | \mathbf{y}_{1:t}) \\ &= \sum_{i=1}^{N_t} p(\mathbf{x}_{t+1} | \mathbf{x}_t^{(i)}) [\exp(\delta_t) \hat{p}(\mathbf{x}_t^{(i)} | \mathbf{y}_{1:t}) - w_t^{(i)} p(\mathbf{x}_t^{(i)} | \mathbf{y}_{1:t})] + \\ &\quad \sum_{i=1}^{N_t} p(\mathbf{x}_{t+1} | \mathbf{x}_t^{(i)}) p(\mathbf{x}_t^{(i)} | \mathbf{y}_{1:t}) - p(\mathbf{x}_{t+1} | \mathbf{y}_{1:t}). \end{aligned}$$

Now, using (5.2.2) and (5.2.14), the first term is equal to

$$\begin{aligned} & \sum_{i=1}^{N_t} p(\mathbf{x}_{t+1}|\mathbf{x}_t^{(i)})p(\mathbf{y}_t|\mathbf{x}_t^{(i)})w_t^{(i)}[\exp(\delta_t)\widehat{C}_t^{-1}\hat{p}(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t-1})-C_t^{-1}p(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t-1})] \\ &= C_t^{-1}\sum_{i=1}^{N_t} p(\mathbf{x}_{t+1}|\mathbf{x}_t^{(i)})p(\mathbf{y}_t|\mathbf{x}_t^{(i)})w_t^{(i)}[\exp(\delta_{t-1})\hat{p}(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t-1})-p(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t-1})] \end{aligned}$$

by noticing that $\exp(\delta_t) = \exp(\delta_{t-1})\widehat{C}_t C_t^{-1}$. The expression in the bracket is now equal to $\varepsilon(\mathbf{x}_t|\mathbf{y}_{1:t-1})$, showing that

$$\begin{aligned} \varepsilon(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}) &= C_t^{-1}\sum_{i=1}^{N_t} p(\mathbf{x}_{t+1}|\mathbf{x}_t^{(i)})p(\mathbf{y}_t|\mathbf{x}_t^{(i)})w_t^{(i)}\varepsilon(\mathbf{x}_t|\mathbf{y}_{1:t-1}) + \\ & \quad \sum_{i=1}^{N_t} p(\mathbf{x}_{t+1}|\mathbf{x}_t^{(i)})p(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t}) - p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}). \end{aligned} \quad (5.5.15)$$

By (5.2.20) an upper limit for the first term on the right side of (5.5.15) is

$$r_t C_t^{-1} \sum_{i=1}^{N_t} w_t^{(i)} p(\mathbf{x}_{t+1}|\mathbf{x}_t^{(i)})p(\mathbf{y}_t|\mathbf{x}_t^{(i)})p(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t-1}),$$

which can be simplified to

$$r_t \sum_{i=1}^{N_t} w_t^{(i)} p(\mathbf{x}_{t+1}|\mathbf{x}_t^{(i)})p(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t}),$$

using (5.2.2) again. By (5.2.23), this is less than or equal to $r_t(1 + \varepsilon)p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t})$. By the same condition, the absolute value of the last two terms on the right hand side of (5.5.15) is less or equal to $\varepsilon p(\mathbf{x}_{t+1}^{(i)}|\mathbf{y}_t)$. Combining these upper limits gives

$$\varepsilon(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}) \leq [r_t(1 + \varepsilon) + \varepsilon]p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}),$$

showing that

$$r_{t+1} \leq r_t(1 + \varepsilon) + \varepsilon,$$

from which (5.2.25) can be proven easily by recursion. Now,

$$\begin{aligned}
|\varepsilon(\mathbf{x}_{t+1}|\mathbf{y}_{1:t})| &= |\exp(\delta_t)\hat{p}(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}) - p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t})| \\
&= |\exp(\delta_t) \sum_{i=1}^{N_t} p(\mathbf{x}_{t+1}|\mathbf{x}_t^{(i)})\hat{p}(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t}) - p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t})| \\
&\leq \left| \sum_{i=1}^{N_t} p(\mathbf{x}_{t+1}|\mathbf{x}_t^{(i)})[\exp(\delta_t)\hat{p}(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t}) - w_t^{(i)}p(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t})] \right| \\
&\quad + \left| \sum_{i=1}^{N_t} p(\mathbf{x}_{t+1}|\mathbf{x}_t^{(i)})p(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t}) - p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}) \right| \\
&\leq \left| \sum_{i=1}^{N_t} p(\mathbf{x}_{t+1}|\mathbf{x}_t^{(i)})p(\mathbf{y}_t|\mathbf{x}_t^{(i)})w_t^{(i)}[\exp(\delta_t)\hat{C}_t^{-1}\hat{p}(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t-1}) \right. \\
&\quad \left. - C_t^{-1}p(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t-1})] \right| + \varepsilon p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}) \\
&= C_t^{-1} \left| \sum_{i=1}^{N_t} p(\mathbf{x}_{t+1}|\mathbf{x}_t^{(i)})p(\mathbf{y}_t|\mathbf{x}_t^{(i)})w_t^{(i)}[\exp(\delta_{t-1})\hat{p}(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t-1}) \right. \\
&\quad \left. - p(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t-1})] \right| + \varepsilon p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}).
\end{aligned}$$

Hence,

$$\begin{aligned}
|\varepsilon(\mathbf{x}_{t+1}|\mathbf{y}_{1:t})| &\leq C_t^{-1} \left| \sum_{i=1}^{N_t} p(\mathbf{x}_{t+1}|\mathbf{x}_t^{(i)})p(\mathbf{y}_t|\mathbf{x}_t^{(i)})w_t^{(i)}\varepsilon(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t-1}) \right| \\
&\quad + \varepsilon p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}) \\
&\leq r_t C_t^{-1} \sum_{i=1}^{N_t} p(\mathbf{x}_{t+1}|\mathbf{x}_t^{(i)})p(\mathbf{y}_t|\mathbf{x}_t^{(i)})p(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t-1})w_t^{(i)} + \varepsilon p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}) \\
&= r_t \left| \sum_{i=1}^{N_t} w_t^{(i)} p(\mathbf{x}_{t+1}|\mathbf{x}_t^{(i)})p(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t}) \right| + \varepsilon p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}) \\
&\leq r_t(1 + \varepsilon)p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}) + \varepsilon p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}).
\end{aligned}$$

Part IV

Applications

Statistical Models of Visual Shape and Motion

Andrew Blake
Michael Isard
John MacCormick

16.1 Introduction

This paper¹ addresses some problems in the interpretation of visually observed shapes, both planar and three-dimensional, in motion. Mumford (1996), interpreting the Pattern Theory developed over a number of years by Grenander (1976), views images as pure patterns that have been distorted by a combination of four kinds of degradations. This view applies naturally to the analysis of static, two-dimensional images. The four degradations are given here, together with comments on how they need to be extended to take account of three-dimensional objects in motion.

- (i) *Domain warping*, in which the domain of an image I is transformed by a mapping g :

$$I(\mathbf{r}) \rightarrow I(g(\mathbf{r})).$$

The three-dimensional nature of the world means that the warp g may be composed largely of projective or affine transformations. The dynamic nature of the problems addressed here will require time-varying warps $g(\mathbf{r}, t)$.

- (ii) *Superposition*: objects may overlap, and in certain forms of imaging this may produce linear combinations. This is fortunate because such combinations can be analysed by linear spectral decomposition. In images of opaque, three dimensional objects, however, far surfaces are obscured by near ones.

¹This is a modified version of (Blake, Bascle, Isard and MacCormick 1998), and we are grateful to the Royal Society for permission to reprint this material.

- (iii) *Distortion and noise*: image measurements are corrupted by noise and blur:

$$I(\mathbf{r}) \rightarrow f(I(\mathbf{r}), \mathbf{n}).$$

Image degradations may be most effectively modelled as applying to certain image features obtained by suitable pre-processing of an image, rather than directly to an image itself.

- (iv) *Observation failure*: disturbance of the observation process; often caused, in the work described here, by distracting background clutter.

A key idea in pattern theory is recognition by synthesis, in which predictions following from particular hypotheses play an important role. The predictions are generated and tested against the products of analysis of an image. Bayesian frameworks, which have gained significant influence in modelling perception (Knill, Kersten and Yuille 1996), seem to be a natural vehicle for this combination of analysis and synthesis. In the context of machine perception of shapes, we can state the problem as one of interpreting a *posterior* density function $p(\mathbf{X}|\mathbf{Z})$ for a shape $\mathbf{X}$ in some appropriate *shape-space* $\mathcal{S}$, given data $\mathbf{Z}$ from an image (or data $(\mathbf{Z}_1, \mathbf{Z}_2, \dots)$ from a sequence of images). The posterior density must be computed in terms of *prior* knowledge about $\mathbf{X}$ and inference about $\mathbf{X}$ based on the *observations* $\mathbf{Z}$. Bayes' formula expresses this as

$$p(\mathbf{X}|\mathbf{Z}) \propto p(\mathbf{Z}|\mathbf{X})p_0(\mathbf{X}), \quad (16.1.1)$$

in which $p_0(\mathbf{X})$ is the prior density for $\mathbf{X}$ and the conditional density $p(\mathbf{Z}|\mathbf{X})$ conveys the range of likely observations to arise from a given shape $\mathbf{X}$. All this connects directly to the four degradations above. In particular, type 1 (warping) is represented in the prior p_0 . Types 3 and 4 (noise and observation failure) are incorporated into the observation density $p(\mathbf{Z}|\mathbf{X})$.

The framework for Bayesian inference of visual shape and motion that forms the basis of this paper is set out in detail in (Blake and Isard 1998). Here, we aim to summarise that framework and introduce several new ideas. The organisation of this chapter is summarised by section, as follows.

- 16.2. *Statistical modelling of shape* — how to choose a suitable shape-space $\mathcal{S}$ and a prior p_0 , or to learn them from a set of examples.
- 16.3. *Statistical modelling of image observations* — how to construct an effective observation density $p(\mathbf{Z}|\mathbf{X})$ that takes into account image intensities, both within the shape of interest and in the background.
- 16.4. *Sampling methods* — using random sample generation to construct an approximate representation of the posterior for $\mathbf{X}$, given that the complexity of $p(\mathbf{Z}|\mathbf{X})$ can render exact representation of the posterior infeasible.

- 16.5. *Modelling dynamics* — extending the Bayesian framework to deal with sequences of images demands priors for temporal sequences $\mathbf{X}_1, \mathbf{X}_2, \dots$. These can either be constructed by hand or learned from examples.
- 16.6. *Learning dynamics* — the most effective way to set up dynamic models is to learn them from training sets.
- 16.7. *Particle filtering* — applied to the interpretation of shapes in motion.
- 16.8. *Dynamics with discrete states* — extending the dynamical repertoire to modelling of motion with several modes, for example walk–trot–canter–gallop.

16.2 Statistical modelling of shape

This section addresses the construction of a prior model $p_0(\mathbf{X})$ for a shape. This can be done in a somewhat general way if the dimensionality of the shape-space $\mathcal{S}$ is fixed in advance to be small, for example just translations in the plane. Then extended observation of the positions of moving objects in some area can be summarised as a histogram which serves as an approximate representation of the prior p_0 (Fernyhough, Cohn and Hogg 1996). In higher dimensional shape-spaces, involving three-dimensional rigid motion and deformation of shape, histograms are less practical. Here we focus on Gaussian distributions.

A Gaussian distribution is specified by its mean and covariance, and these can be estimated from a training sequence $\mathbf{X}_1, \mathbf{X}_2, \dots$ of shapes by taking the sample mean $\bar{\mathbf{X}}$ and the sample covariance

$$\Sigma = \frac{1}{M} \sum_{k=1}^M (\mathbf{X}_k - \bar{\mathbf{X}})(\mathbf{X}_k - \bar{\mathbf{X}})^T.$$

Moreover, Principal Components Analysis (PCA) (Rao 1973) can be used to restrict the shape-space $\mathcal{S}$ to explain most of the variance in the training set while keeping the dimension of $\mathcal{S}$ small, in the interests of computational efficiency (Cootes, Taylor, Lanitis, Cooper and Graham 1993, Baumberg and Hogg 1994, Lanitis, Taylor and Cootes 1995, Beymer and Poggio 1995, Baumberg and Hogg 1995a, Vetter and Poggio 1996, Bascle and Blake 1998). An example is given in figure 16.1.

However, the resulting shape-space, though economical, is not especially easy to interpret because Principal Components need not be meaningful. More meaningful constructive shape-spaces can be generated by acknowledging three-dimensional projective effects and constructing affine spaces, for instance, whose components are directly related to rigid body transformations (Ullman and Basri 1991, Koenderink and van Doorn 1991). In

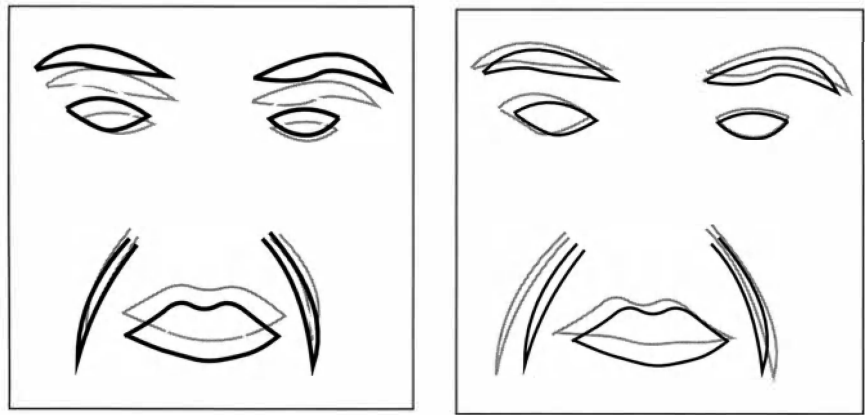


Figure 16.1. **PCA for faces.** A shape-space of facial expressions is reduced here by PCA to the two-dimensional space that best covers the expressions in a certain training sequence.

addition, named deformations can be included in a basis for $\mathcal{S}$ as key-frames (Blake and Isard 1994), as in figure 16.2.

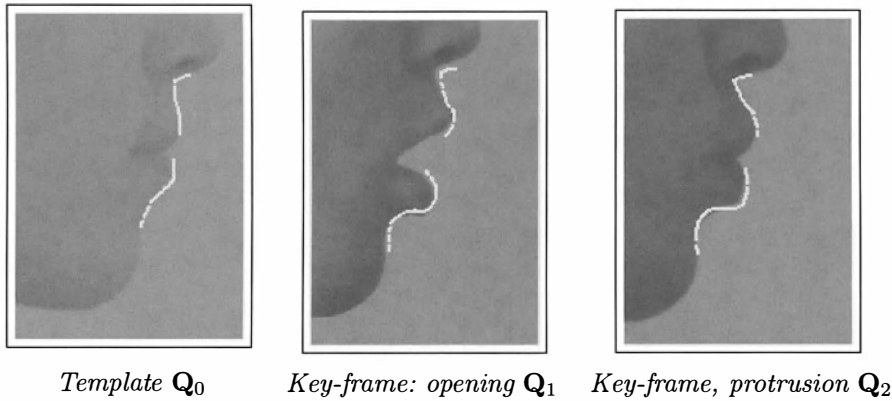


Figure 16.2. **Key-frames.** Lips template followed by two key-frames, representing interactively tracked lips in characteristic positions. The key-frames are combined linearly with appropriate rigid degrees of freedom, to give a shape-space suitable for use in a tracker for non-rigid motion.

A constructive shape-space $\mathcal{S}^c$ can be combined with PCA to give the best of both worlds. ‘Residual PCA operates on a constructive shape-space that does not totally cover a certain dataset, supplying missing components by PCA. Then the constructive subspace retains its interpretation and only the residual components, covered by PCA, cannot be directly interpreted. This is done by constructing a projection operator E^c that maps $\mathcal{S}$ to $\mathcal{S}^c$

and applying PCA to the residual training-set vectors $\mathbf{X}_1^r, \mathbf{X}_2^r, \dots$ where

$$\mathbf{X}^r = \mathbf{X} - E^c \mathbf{X}.$$

Full details of the algorithm are given in (Blake and Isard 1998) and an example of its application is shown in figure 16.3.

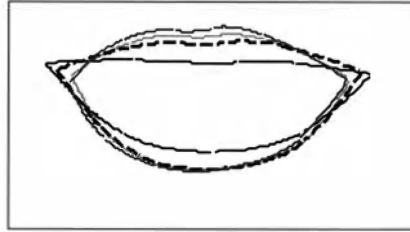


Figure 16.3. Sampling from a prior for lip-shape, excluding translation
Random sampling illustrates how a learned prior represents plausible lip configurations. Any rigid translations in the training set, due to head-motion, are separated out as a constructive shape-space in residual PCA.

16.3 Statistical modelling of image observations

Gaussian distributions may often be acceptable as models of prior shape, but they are adequate as observation distributions only in the clutter-free case. Typically, in our framework, observations are made along a series of splines, normal to the hypothesised shape, as in figure 16.4. Consider the one-dimensional problem of observing, along a single spline, the set of feature positions $\{\mathbf{z} = (z_1, z_2, \dots, z_M)\}$. Assuming a uniform distribution of background clutter, and a Gaussian model for error in measurement of the position of the true object edge, leads (Isard and Blake 1996) to the multi-modal observation density $p(\mathbf{z}|\mathbf{X})$ depicted in figure 16.5. The multiple peaks in the density are generated by clutter and cannot possibly be modelled by a single Gaussian. A mixture of Gaussians might be feasible, but a very efficient alternative is to use random sampling (next section).

The observation model above was based on the assumption that the observable contour is a “bent wire” resting on a cluttered background. This is not very realistic. It is highly desirable in practice to allow for object opacity and to distinguish between the textured interior of an object and its cluttered exterior. A probabilistic model that reflects this is based on the following assumptions.

Feature localisation error. It is assumed that the feature detector reports object outline position with an error whose density is $\mathcal{E}(\cdot)$, taken usually to be Gaussian.

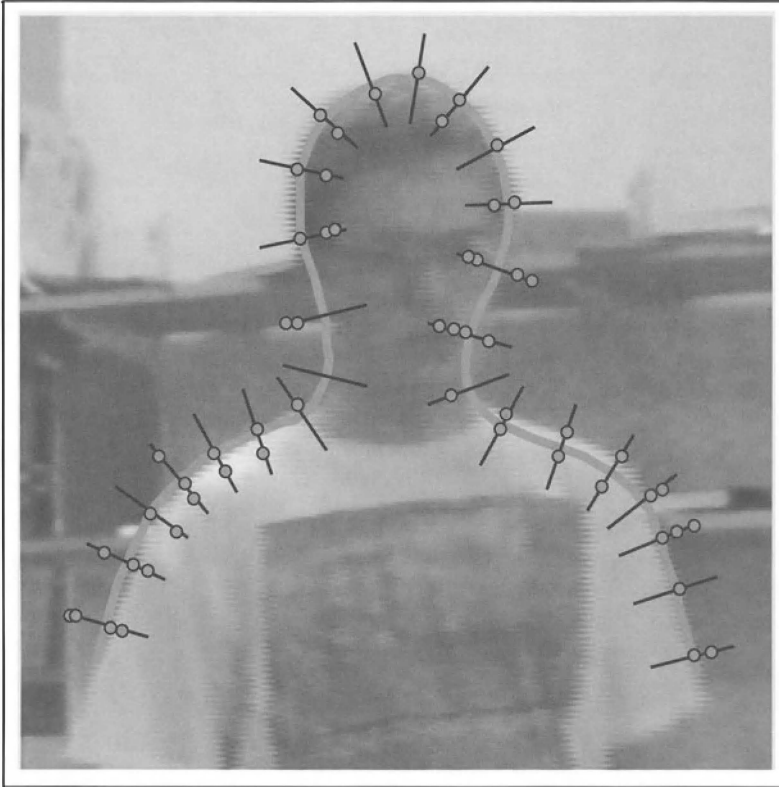


Figure 16.4. **Observation process.** The thick line is a hypothesised shape, represented as a parametric spline curve. The spines are curve normals along which high-contrast features (white circles) are sought.

Occlusion probability. The possibility that the outline is missed by the feature detector is allowed, with probability q .

Clutter model. Detection of clutter features is regarded as an i.i.d. random process on the portion of each measurement line that lies outside the object. The probability $\pi(n)$ that n clutter features are detected on a normal is generally taken to be uniform.

Interior model. Interior features on a measurement line are modelled as uniformly distributed along the interior portion of the normal. The distribution $\rho(m)$ for the number m of interior features observed is taken to be Poisson with a known density parameter, which is actually learned by observing instances of the object.

A density $p(\mathbf{z}|\mathbf{X})$ based on these assumptions can be constructed and expressed as $p = \lambda D$ where λ is a constant and

$$D(\mathbf{X}) = p(\mathbf{z}|\mathbf{X})/p(\mathbf{z}|\text{no object present})$$

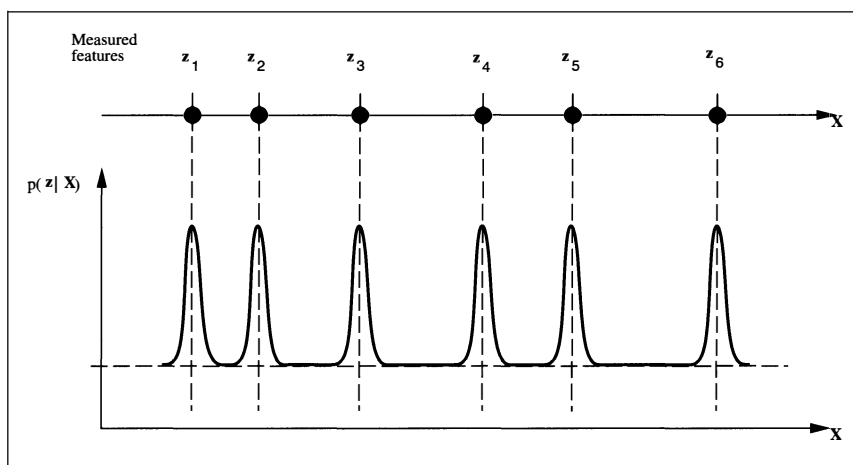


Figure 16.5. **Multi-modal observation density** (*one-dimensional illustration*). A probabilistic observation model allowing for clutter and the possibility of missing the target altogether is specified here as a conditional density $p(\mathbf{z}|\mathbf{X})$. It has a peak corresponding to each observed feature.

— the *contour discriminant*. This is a discriminant function (Duda and Hart 1973) in the form of a *likelihood ratio*. It has the attraction that, in addition to conveying the relative values of $p(\mathbf{z}|\mathbf{X})$, its absolute value is meaningful: $D(\mathbf{X}) > 1$ implies that the observed features $\mathbf{z}$ are more likely to have arisen from the object in location $\mathbf{X}$ than from clutter.

Lastly, densities $p(\mathbf{z}|\mathbf{X})$ for each normal need to be combined into a grand observation density $p(\mathbf{Z}|\mathbf{X})$, and this raises some issues about independence of measurements along an object contour. Details of the form and computation of the full observation density are given in (MacCormick and Blake 1998). Results of the evaluation of the contour discriminant on a real image are shown in figure 16.6. Analysis of the same image using the simpler bent wire observation model degrades the results, failing altogether to locate the leftmost of the three people. The explicit modelling of object opacity has clearly brought significant benefits.

16.4 Sampling methods

The next stage of the pattern recognition problem is to construct the posterior density $p(\mathbf{X}|\mathbf{Z})$ by applying Bayes' rule (16.1.1). In the previous section, it became plain that the observation density has a complex form in clutter. This means that direct evaluation of $p(\mathbf{X}|\mathbf{Z})$ is infeasible. However, iterative sampling techniques can be used (Geman and Geman 1984, Ripley and Sutherland 1990, Grenander, Chow and Keenan 1991, Storvik 1994). The *factored sampling* algorithm (Grenander et al. 1991) — also known

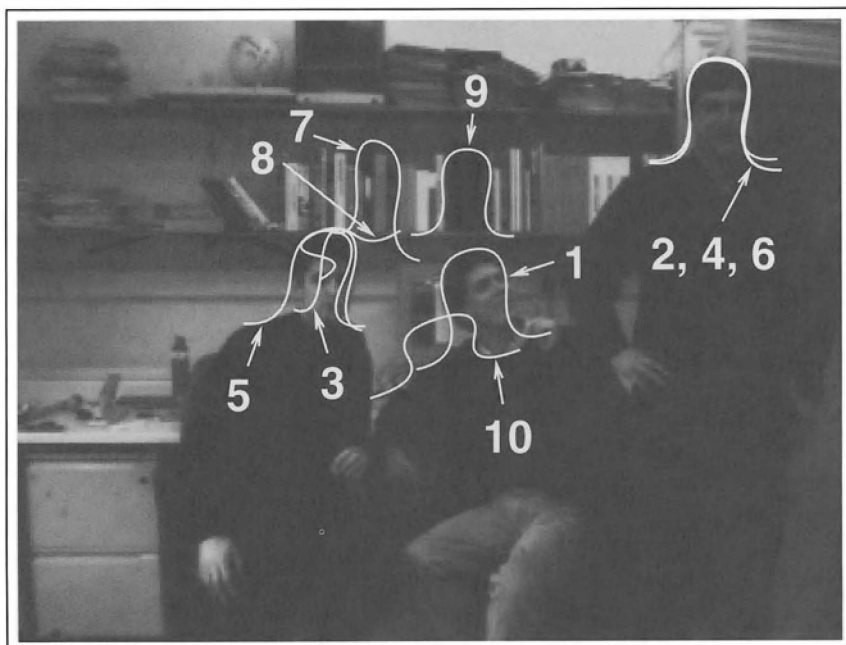


Figure 16.6. **Finding head-and-shoulders outlines in an office scene.** *The results of a sample of 1,000 configurations are shown ranked by value of their contour discriminant. The figure displays the cases in which $D > 1$, indicating a configuration that is more target-like than clutter-like.*

as sampling importance resampling, or SIR (Rubin 1988) — generates a random variate $\mathbf{X}$ from a distribution $\tilde{p}(\mathbf{X})$ that approximates the posterior $p(\mathbf{X}|\mathbf{Z})$. First a sample-set $\{\mathbf{s}^{(1)}, \dots, \mathbf{s}^{(N)}\}$ is generated from the prior density $p(\mathbf{x})$ and then a sample $\mathbf{X} = \mathbf{X}_i$, $i \in \{1, \dots, N\}$ is chosen with probability

$$\pi_i = \frac{p(\mathbf{Z}|\mathbf{X} = \mathbf{s}^{(i)})}{\sum_{j=1}^N p(\mathbf{Z}|\mathbf{X} = \mathbf{s}^{(j)})}.$$

Sampling methods have proved remarkably effective for recovering static objects from cluttered images. For such problems, $\mathbf{X}$ is a multi-dimensional set of parameters for curve position and shape. In that case the sample-set $\{\mathbf{s}^{(1)}, \dots, \mathbf{s}^{(N)}\}$ is drawn from the posterior distribution of $\mathbf{X}$ -values, as illustrated in figure 16.7.

16.5 Modelling dynamics

In order to be able to interpret moving shapes in sequences of images, it is necessary to supply a prior distribution, not just for shape, but also for

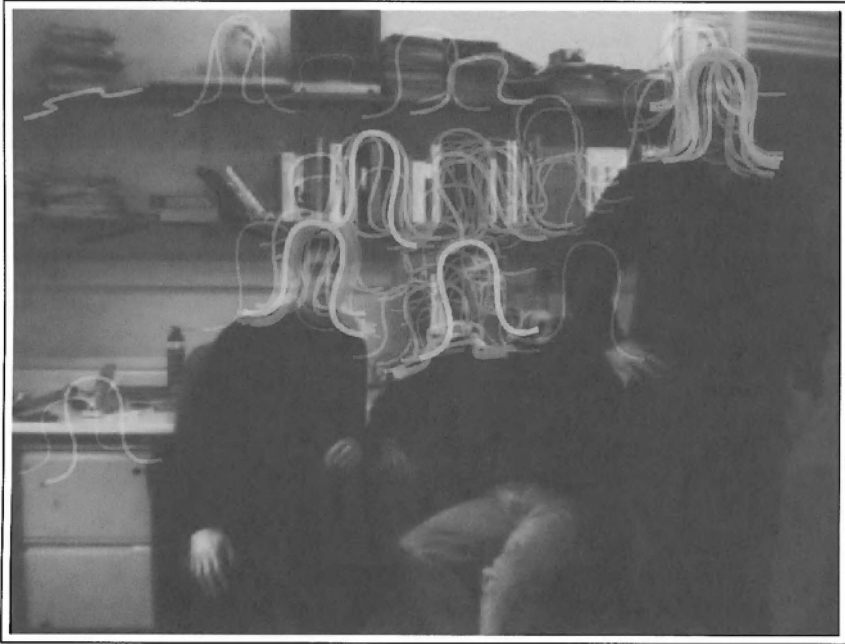


Figure 16.7. **Sample-set representation of posterior shape distribution** for a curve with parameters $\mathbf{X}$, modelling a head outline. Each sample $\mathbf{s}^{(n)}$ is shown as a curve (of varying position and shape) with a mass proportional to the weight $\pi^{(n)}$. The prior is uniform over translation, with some constrained Gaussian variability in the remainder of its affine shape-space.

the motion of that shape. Consider the problem of building an appropriate prior model for the position of a hand-mouse engaged in an interactive graphics task. A typical trace in the x - y plane of a finger drawing letters is given in figure 16.8. If the entire trajectory were treated as a training set, the methods discussed earlier could be applied to learn a Gaussian prior distribution for finger position. The learned prior is broad, spanning a sizeable portion of the image area, and places little constraint on the measured position at any given instant. Nonetheless, it is quite clear from the figure that successive positions are tightly constrained. Although the prior covariance ellipse spans about 300×50 pixels, successive sampled positions are seldom more than 5 pixels apart.

For sequences of images, then, a global prior $p_0(\mathbf{X})$ is not enough. What is needed is a conditional distribution $p(\mathbf{X}_k | \mathbf{X}_{k-1})$ giving the distributions of possibilities for the shape $\mathbf{X}_k$ at time $t = k\tau$ given the shape $\mathbf{X}_{k-1}$ at time $t = (k-1)\tau$ (where τ is the time-interval between successive images). This amounts to a first order Markov chain model in shape space in which, although in principle $\mathbf{X}_k$ may be correlated with all of $\mathbf{X}_1 \dots \mathbf{X}_{k-1}$, only correlation with the immediate predecessor is explicitly acknowledged.

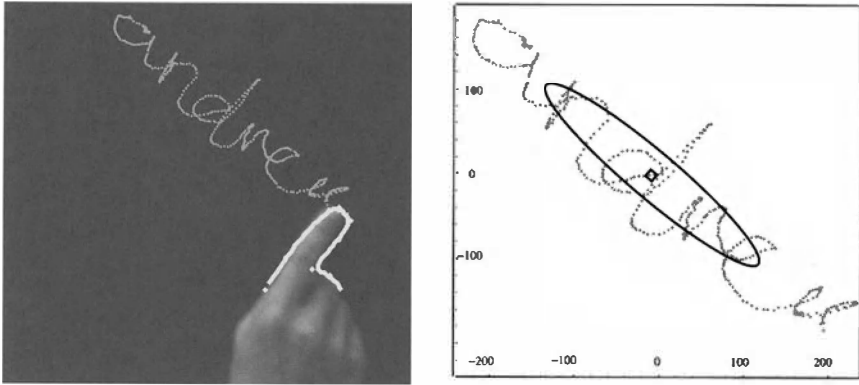


Figure 16.8. **The moving finger writes.** The finger trajectory (left), which has a duration of about 10 seconds, executes a broad sweep over the plane. When the trajectory is treated as a training set, the learned Gaussian prior is broad, as the covariance ellipse (right) shows. Clearly, though, successive positions (individual dots represent samples captured every 20ms) are much more tightly constrained.

For the sake of tractability, it is reasonable to restrict Markov modelling to linear processes. In principle and in practice it turns out that a first order Markov chain is not quite enough, generally, but second order suffices. The detailed arguments for this, addressing such issues as capacity to represent oscillatory signals and trajectories of inertial bodies, can be found in (Blake and Isard 1998, Chapter 9). Figure 16.9 illustrates the point for a practical example. A second order, Auto-Regressive Process (ARP) is most concisely expressed by defining a state vector

$$\mathcal{X}_k = \begin{pmatrix} \mathbf{X}_{k-1} \\ \mathbf{X}_k \end{pmatrix}, \quad (16.5.2)$$

and then specifying the conditional probability density $p(\mathcal{X}_k | \mathcal{X}_{k-1})$. In the case of a linear model, this can be done constructively as follows:

$$\mathcal{X}_k - \overline{\mathcal{X}} = A(\mathcal{X}_{k-1} - \overline{\mathcal{X}}) + B\mathbf{w}_k, \quad (16.5.3)$$

where

$$A = \begin{pmatrix} 0 & I \\ A_2 & A_1 \end{pmatrix}, \quad \overline{\mathcal{X}} = \begin{pmatrix} \overline{\mathbf{X}} \\ \overline{\mathbf{X}} \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 \\ B_0 \end{pmatrix}. \quad (16.5.4)$$

Each $\mathbf{w}_k$ is a vector of N_X independent random $\mathcal{N}(0,1)$ variables and $\mathbf{w}_k, \mathbf{w}_{k'}$ are independent for $k \neq k'$. This specifies the probable temporal evolution of the shape $\mathbf{X}$ in terms of parameters A, B , and covers multiple oscillatory modes and/or constant velocity motion. The constructive form is attractive because it is amenable to direct simulation, simply by supplying a realisation of the succession of random variates $\mathbf{w}_k$.

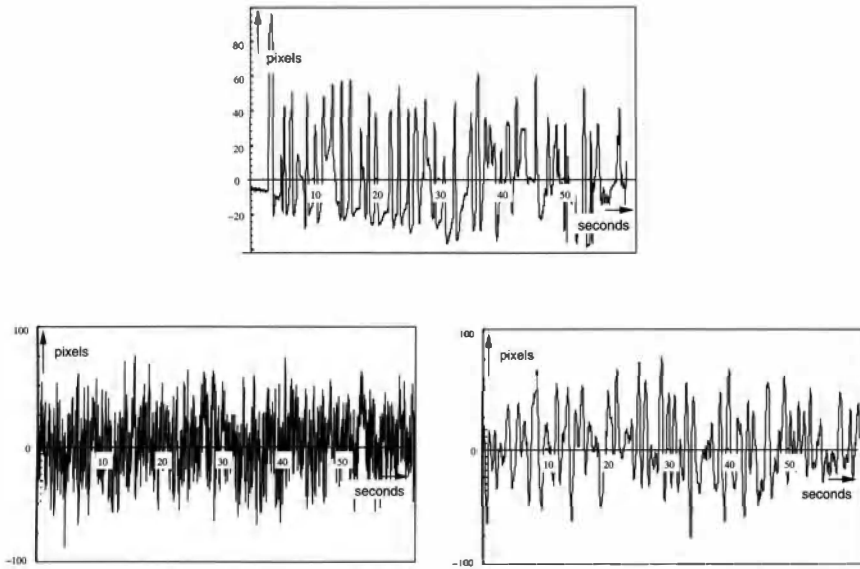


Figure 16.9. **Motion data from talking lips.** *Training sequence of 60 seconds duration (top). Random simulations of learned models — 1st order (left) and 2nd order (right). Only the 2nd order model captures the natural periodicity (around 1 Hz) of the training set, and spectrogram analysis confirms this.*

16.6 Learning dynamics

Motion parameters (A, B in this paper) can be set by hand to obtain desired effects, and a logical approach to this has been developed (Blake and Isard 1998, chapter 9). Experimentation allows these parameters to be refined by hand for improved tracking, but this is a difficult and unsystematic business. It is far more attractive to learn dynamic models on the basis of training sets. A number of alternative approaches have been proposed for learning dynamics, with a view to gesture-recognition — see for instance (Mardia, Ghali, Howes, Hainsworth and Sheehy 1993, Campbell and Bobick 1995, Bobick and Wilson 1995). The requirement there is to learn models that are sufficiently finely tuned to discriminate among similar motions. In our context of the problem of motion tracking, rather different methods are required so as to learn models that are sufficiently coarse to encompass all likely motions.

Initially, a hand-built model is used in a tracker to follow a training sequence that must be not be too hard to track. This can be achieved by allowing only motions that are not too fast and limiting background clutter, or eliminating it using background subtraction (Baumberg and Hogg 1994, Murray and Basu 1994, Koller, Weber and Malik 1994, Rowe and Blake 1996). Once a new dynamic model has been learned, it can be used to build a more competent tracker, one that is specifically tuned to

the sort of motions it is expected to encounter. This can be used either to track the original training sequence more accurately, or to track a new and more demanding training sequence involving greater agility of motion. The cycle of learning and tracking is illustrated in figure 16.10. Typically two

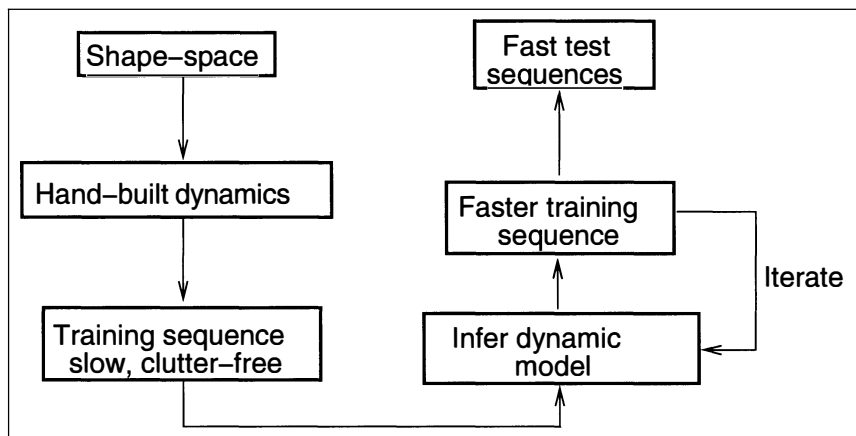


Figure 16.10. **Iterative learning of dynamics.** *The model acquired in one cycle of learning is installed in a tracker to interpret the training sequence for the next cycle. The process is initialised with a tracker whose prior is based on a hand-built dynamic model.*

or three cycles suffice to learn an effective dynamic model.

In mathematical terms, the general problem is to estimate the coefficients A_1 , A_2 , $\bar{\mathbf{X}}$ and B from a training sequence of shapes $\mathbf{X}_1, \dots, \mathbf{X}_M$, gathered at the image sampling frequency. Known algorithms to do this are based on the Maximum Likelihood principle (Rao 1973, Kendall and Stuart 1979) and use variants of Yule-Walker equations for estimation of the parameters of auto-regressive models (Gelb 1974, Goodwin and Sin 1984, Ljung 1987). Suitable adaptations for multidimensional shape-spaces are given by (Blake and Isard 1994, Baumberg and Hogg 1995b, Blake, Isard and Reynard 1995), with a number of useful extensions in (Reynard, Wildenberg, Blake and Marchant 1996). One example is the scribble in figure 16.11, learned from the training-sequence in figure 16.8.

A more complex example consists of learning the motions of an actor's face, using the shape-space described earlier that covers both rigid and non-rigid motion. Figure 16.12 illustrates how much more accurately realistic facial motion can be represented by a dynamic model actually learned from examples.

The learning algorithms referred to above treat the training set as exact, whereas in fact it is inferred from noisy observations. Dynamics can be learned directly from the observations using Expectation-Maximisation (EM) (Dempster, Laird and Rubin 1977). Learning dynamics by EM is

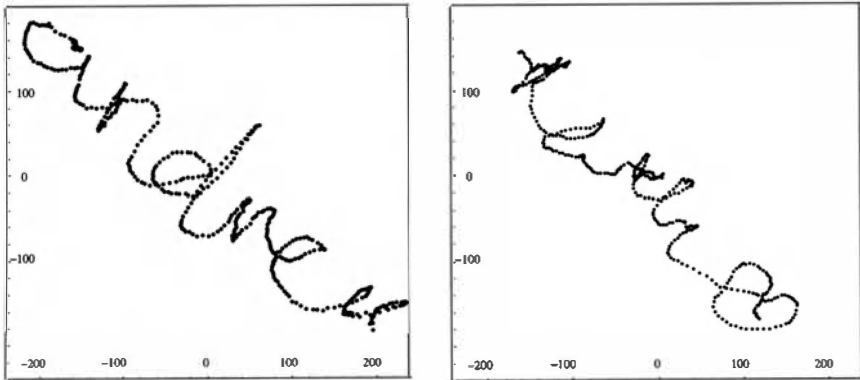


Figure 16.11. **Scribbling: simulating a learned model for finger-writing.** A training set (left) consisting of six handwritten letters is used to learn a dynamical model for finger motion. A random simulation from the model (right) exhibits reasonable gross characteristics.

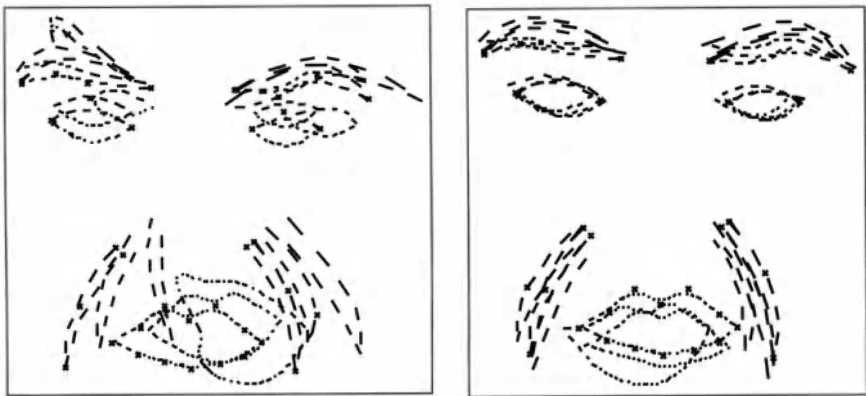


Figure 16.12. **Trained dynamics for facial motion.** Hand-built dynamics, exhibited here by random simulation (left) are just good enough, when used in tracking, to gather a training sequence. Trained dynamics (right) however, capture more precisely the constraints of realistic facial motion.

suggested by Ljung (1987) and the detailed algorithm is given in (North and Blake 1997). It is related to the Baum-Welch algorithm used to learn speech models (Huang, Arika and Jack 1990, Rabiner and Bing-Hwang 1993) but with additional complexity because the state-space is continuous rather than discrete. In practice, accuracy of the learned dynamics is significantly improved when EM is used, especially in the case of more coherent oscillations.

An extension of the basic algorithm for *classes* of objects, dealing independently with motion and with variability of mean shape/position over

the class, is described in (Reynard et al. 1996). The same algorithm is also used for modular learning — the aggregation of training sets for which a joint dynamic model is to be constructed.

16.7 Particle filtering

The principles of particle filtering have been developed earlier in the book. Particle filtering has been introduced into the practice of Computer Vision (Isard and Blake 1996, Isard and Blake 1998a) over the past few years, where it is known as the CONDENSATION algorithm, and has been applied to great effect for tracking moving objects in image sequences.

Each time-step of the particle filter generates a weighted, time-stamped sample-set, denoted $\mathbf{s}_k^{(n)}$, $n = 1, \dots, N$ with weights $\pi_k^{(n)}$, representing approximately the conditional state-density $p(\mathcal{X}_k | \underline{\mathbf{Z}}_k)$ at time $t = k\tau$, where $\underline{\mathbf{Z}}_k = (\mathbf{Z}_1, \dots, \mathbf{Z}_k)$, the history of observations. How is this sample-set obtained? Clearly, the process must begin with a prior density, and the effective prior for time-step k should be $p(\mathcal{X}_k | \underline{\mathbf{Z}}_{k-1})$. This prior is of course multi-modal in general and no functional representation of it is available. It is derived from the representation as a sample set $\{(\mathbf{s}_{k-1}^{(n)}, \pi_{k-1}^{(n)})\}$, $n = 1, \dots, N$ of $p(\mathcal{X}_{k-1} | \underline{\mathbf{Z}}_{k-1})$, the output from the previous time-step, to which prediction must then be applied.

The iterative process applied to the sample-sets is depicted in figure 16.13. At the top of the diagram, the output from time-step $k-1$ is the weighted sample-set $\{(\mathbf{s}_{k-1}^{(n)}, \pi_{k-1}^{(n)})\}$, $n = 1, \dots, N$. The aim is to maintain, at successive time-steps, sample sets of fixed size N , so that the algorithm can be guaranteed to run within a given computational resource. The first operation therefore is to sample (with replacement) N times from the set $\{\mathbf{s}_{k-1}^{(n)}\}$, choosing a given element with probability $\pi_{k-1}^{(n)}$. Some elements, especially those with high weights, may be chosen several times, leading to identical copies of elements in the new set. Others with relatively low weights may not be chosen at all.

Each element chosen from the set is now subjected to a predictive step, using an ARP dynamic model, as in equation (16.5.3). This involves sampling a value of $\mathcal{X}_k$ randomly from the conditional density $p(\mathcal{X}_k | \mathcal{X}_{k-1})$ to form a new set member $\mathbf{s}_k^{(n)}$. Since the predictive step includes a random component, identical elements may now split as each undergoes its own independent random motion step. At this stage, the sample set $\{\mathbf{s}_k^{(n)}\}$ for the new time-step has been generated but, as yet, without its weights; it is approximately a fair random sample from the effective prior density $p(\mathcal{X}_k | \underline{\mathbf{Z}}_{k-1})$ for time-step $t = k\tau$. Finally, the observation step from factored sampling is applied, generating weights from the observation density $p(\mathbf{Z}_k | \mathcal{X}_k)$ to obtain the sample-set representation $\{(\mathbf{s}_k^{(n)}, \pi_k^{(n)})\}$ of state-density for time t .

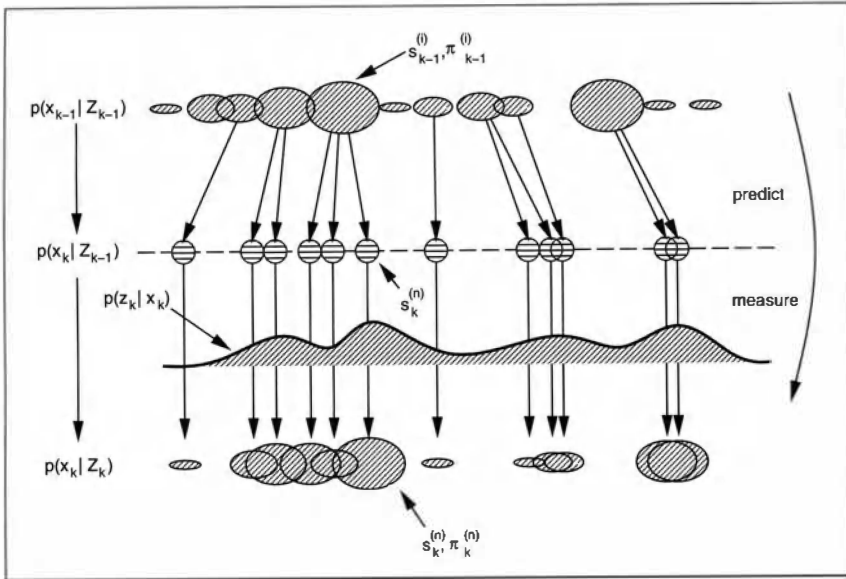


Figure 16.13. **One time-step in the CONDENSATION algorithm.** *Blob centres represent sample values and sizes depict sample weights.*

In general sequential importance sampling, almost any “importance distribution” can be used to propagate particles from one time-step to the next, provided the weights are calculated so as to counteract the effects of the importance distribution. The dynamical model $p(\mathcal{X}_k|\mathcal{X}_{k-1})$ used to propagate members of the sample set here is a particular choice of importance distribution that is effective for the visual tracking problems to be addressed. Of course this distribution cannot guarantee that the sample weights have low variance — indeed, the optimal distribution for this purpose is $p(\mathcal{X}_k|\mathbf{Z}_k, \mathcal{X}_{k-1})$ (Doucet, Godsill and Andrieu 2000), but this distribution is unknown and intractable in the problems considered here. Using $p(\mathcal{X}_k|\mathcal{X}_{k-1})$ as the importance distribution has two advantages: it can be learnt by the methods described in the last section, and samples can be drawn from it easily.

A good deal of experimentation has been conducted in applying the CONDENSATION algorithm to the tracking of visual motion, including moving hands and dancing figures. Perhaps one of the most stringent tests involved the tracking of a leaf on a bush, in which the foreground leaf is effectively camouflaged against the background. Results are shown in figure 16.14 and experimental details can be found in (Isard and Blake 1996).

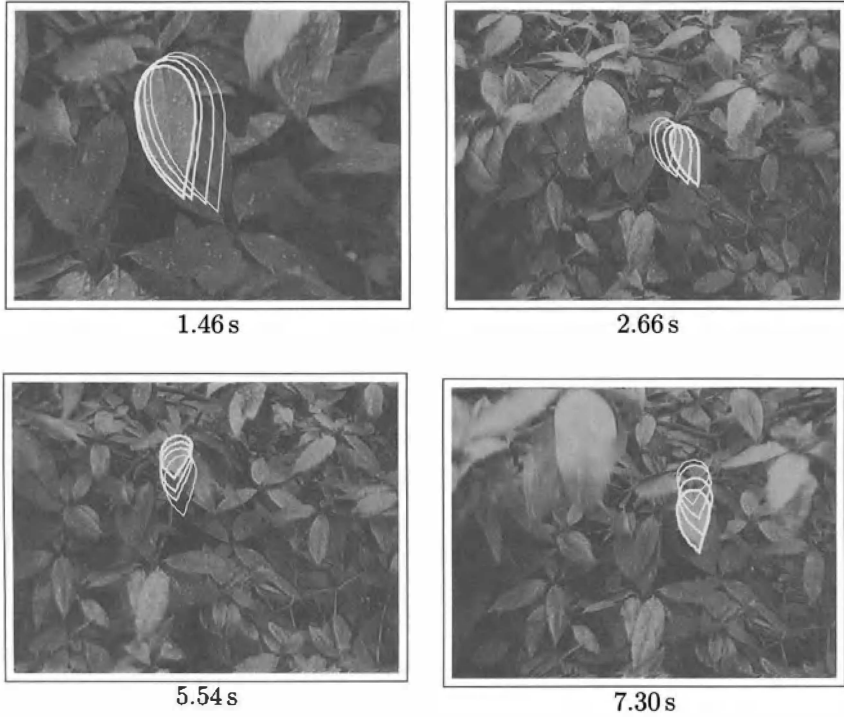


Figure 16.14. **Tracking with camouflage.** Stills depict mean contour configurations, with preceding tracked leaf positions plotted at 40 ms intervals to indicate motion.

16.8 Dynamics with discrete states

A recent development of the dynamic models already described is to append to the state variable $\mathcal{X}$ a discrete state y_k to make a mixed state

$$\mathcal{X}_k^+ = \begin{pmatrix} \mathcal{X}_k \\ y_k \end{pmatrix}, \quad (16.8.5)$$

where $y_k \in \{1, \dots, N_S\}$ is drawn from a finite set of discrete states with integer labels. Each discrete state represents a mode of motion such as stroke, rest and shade for a hand engaged in drawing. Corresponding to each state $y_{k-1} = i$ there is a dynamical model $p_i(\mathcal{X}_k | \mathcal{X}_{k-1})$, which, in the case of the drawing hand, is likely to be an ARP as in (16.5.3). The stroke model, for instance, might represent constant velocity motion, whereas shading would be oscillatory. In addition, and independently, state transitions are governed by

$$P(y_k = j | y_{k-1} = i) = T_{i,j},$$

a transition matrix following usual practice for Markov chains. More generally, transition probabilities could be made sensitive to the context $\mathcal{X}_{k-1}$ in state-space, so that

$$P(y_k = j | y_{k-1} = i, \mathcal{X}_{k-1}) = T_{i,j}(\mathcal{X}_{k-1}).$$

For example this could be used to express an enhanced probability of transition into the resting state when the hand is moving slowly.

The incorporation of mixed states into the CONDENSATION algorithm is straightforward. It involves using the extended state $\mathcal{X}_k^+$ in place of the original $\mathcal{X}_k$, so that a sample $\mathbf{s}_k^{(n)}$ is now a value of the extended state. The prediction step, which generates a new sample $\mathbf{s}_k^{(n)}$ from an old one $\mathbf{s}_{k-1}^{(n)}$, requires a discrete step and a continuous one. First, the discrete state y_k for the new sample is $y_k = j$, chosen randomly, with probability $T_{i,j}$, where i is the discrete state of the old sample. Then the continuous state is chosen by sampling randomly from a continuous density, as in the original algorithm, but now one of several possible densities $p_i(\mathcal{X}_k | \mathcal{X}_{k-1})$, where again i is the discrete state of the old sample.

Experiments with a three-state model for drawing have been described in detail elsewhere (Isard and Blake 1998b). In addition to enhancing tracking performance, there is the bonus that the current discrete state y_k can be estimated at each time $t = k\tau$, effectively carrying out gesture recognition as a side-effect. One interesting variation on the mixed-state theme uses continuous conditional densities $p_i(\mathcal{X}_k | \mathcal{X}_{k-1})$ which are not ARP models. Consider the example of a moving ball, which may occasionally bounce. This could be represented using two states $\{1, 2\}$ in which $i = 1$ stands for the free ballistic motion of the ball, and $i = 2$ is the bounce event. A suitable transition matrix would be

$$T = \begin{pmatrix} 1 - \epsilon & \epsilon \\ 1 & 0 \end{pmatrix},$$

in which $0 < \epsilon \ll 1$ so that ballistic motion has a mean duration τ/ϵ between bounces. The fact that $T_{2,2} = 0$ ensures that the model always returns to ballistic motion after a bounce — bouncing at each of two consecutive time-steps is disallowed. Now $p_1(\dots | \dots)$ is an ARP for ballistic motion but $p_2(\dots | \dots)$ models the instantaneous reversal of velocity normal to the reflecting surface. Details of experiments with such a model are in (Isard and Blake 1998b), but the results are illustrated in figure 16.15. The use of mixed discrete/continuous states in a particle filter has been proposed independently by (Semerdjiev, Jilkov and Angelove 1998).

16.9 Conclusions

A high-speed tour has been given of a framework for probabilistic modelling of shapes in motion, and of their visual observation. The key points are that

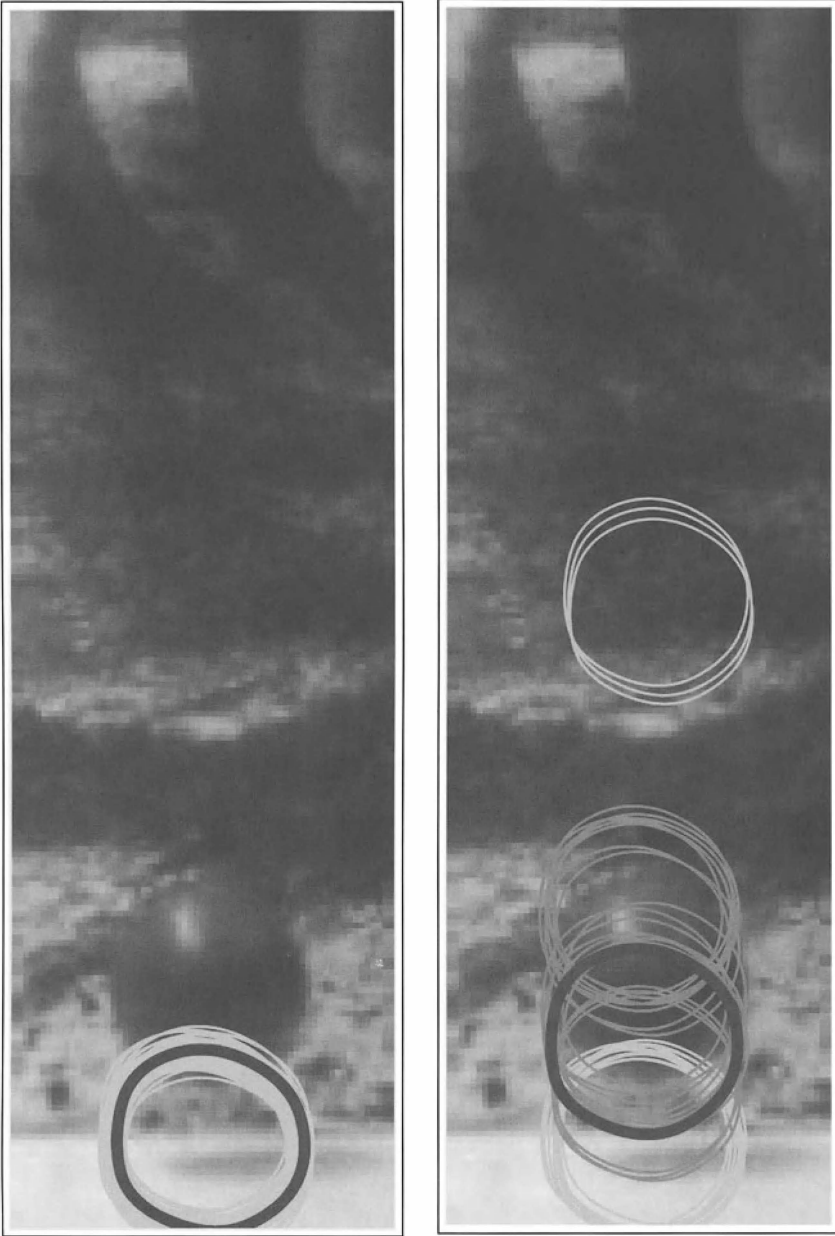


Figure 16.15. **Mixed states tighten constraints in dynamic models.** *A conventional, continuous-state ARP model (left) used to track ballistic motion fails unrecoverably as the ball bounces. Introducing an explicit discrete state for the bounce allows the sample set to split, so that a significant proportion are able to track the bounce.*

visual clutter makes motion analysis difficult, and demands sophisticated probabilistic mechanisms to handle the resulting uncertainty. Further, prior models of motion and observation provide powerful constraints, especially so when the models are learned. A more detailed treatment is given in (Blake and Isard 1998). Since that account, several new modelling tools have been developed. First, the contour discriminant is a new observational model expressed as a likelihood ratio taking opacity of objects into account. Second, complex models for combined rigid and nonrigid motion have been constructed, with a new algorithm for decomposing the two components. Third, extending dynamic states to include discrete labels can significantly enhance their power to constrain perceptual interpretation of shape.

Many interesting questions remain to be addressed. One of these is whether sampling methods for object localisation can be fused elegantly with the CONDENSATION algorithm, to allow robust handling of birth and death (Grenander and Miller 1994) processes in which objects enter and leave the scene. A second is to extend mixed-state models to give reliable gesture recognition on the fly, in a manner that is integrated with the tracking process. A third is to develop algorithms, based on EM, to learn dynamical models from sequences tracked by CONDENSATION, using the full richness of its probabilistic representation, both for continuous and mixed state systems. Finally, another important goal is the extension of particle filtering to allow its use in many dimensions and complex configuration spaces. Such techniques are emerging all the time, and recent contributions include the partitioned sampling of (MacCormick 1999) and the layered sampling of (Sullivan, Blake, Isard and MacCormick 1999).

Acknowledgements

The authors would like to acknowledge the support of the EPSRC.

Bibliography

- Ackerson, G. A. and Fu, K. S. (1970). On state estimation in switching environments, *IEEE Transactions on Automatic Control* **15**(1): 10–17.
- Aguilar, O. and West, M. (2000). Bayesian dynamic factor models and portfolio allocation, *Journal of Business and Economic Statistics*.
- Alspach, D. L. and Sorenson, H. W. (1972). Nonlinear Bayesian estimation using Gaussian sum approximations, *IEEE Transactions on Automatic Control* **AC-17**(4): 439–448.
- Anderson, B. D. and Moore, J. B. (1979). *Optimal Filtering*, Prentice-Hall, New Jersey.
- Andersson, P. (1985). Adaptive forgetting in recursive identification through multiple models, *International Journal of Control* **42**(5): 1175–1193.
- Andrade Netto, M. L., Gimeno, L. and Mendes, M. J. (1978). On the optimal and suboptimal nonlinear filtering problem for discrete-time systems, *IEEE Transactions on Automatic Control* **23**: 1062–1067.
- Andrieu, C., de Freitas, J. F. G. and Doucet, A. (1999a). Sequential Bayesian estimation and model selection applied to neural networks, *Technical Report CUED/F-INFENG/TR 341*, Cambridge University Engineering Department.
- Andrieu, C., de Freitas, J. F. G. and Doucet, A. (1999b). Sequential MCMC for Bayesian model selection, *IEEE Higher Order Statistics Workshop*, Caesarea, Israel, pp. 130–134.
- Aoki, M. (1967). *Optimisation of stochastic systems*, Academic Press.
- Athreya, K. B. and Ney, P. E. (1972). *Branching Processes*, Springer-Verlag.
- Averbuch, A., Itzikowitz, S. and Kapon, T. (1991). Radar target tracking - Viterbi versus IMMA, *IEEE Transactions on Aerospace and Electronic Systems* **27**(3): 550–563.
- Avitzour, D. (1995). Stochastic simulation Bayesian approach to multitarget tracking, *IEE Proceedings-Radar Sonar and Navigation* **142**(2): 41–44.
- Azzam, R. M. A. and Bashara, N. M. (1987). *Ellipsometry and Polarized Light*, North-Holland Physics Publishing.
- Baker, J. (1987). Reducing bias and inefficiency in the selection algorithm, *International Conference on Genetic Algorithms and Their Applications*.
- Bally, V. and Talay, D. (1996). The law of the Euler scheme for stochastic differential equations: 2. Approximation of the density, Monte Carlo methods and applications, *Probab. Th. Related Fields* **2**: 93–128.

- Bar-Shalom, Y. (1981). Stochastic dynamic programming: caution and probing, *IEEE Transactions on Automatic Control* **AC-26**(5): 1184–1195.
- Bar-Shalom, Y. and Fortman, T. E. (1987). *Tracking and Data Association*, Academic Press, New York.
- Bar-Shalom, Y. and Tse, E. (1974). Dual effect, certainty equivalence and separation in stochastic control, *IEEE Transactions on Automatic Control* **AC-19**(5): 494–500.
- Bar-Shalom, Y., Chang, K. C. and Blom, H. A. P. (1985). Tracking a maneuvering target using input estimation versus the interacting multiple model algorithm, *IEEE Transactions on Aerospace and Electronic Systems* **25**(2): 296–300.
- Basile, B. and Blake, A. (1998). Separability of pose and expression in facial tracking and animation, *Proceedings of the 6th International Conference on Computer Vision*, pp. 323–328.
- Baumberg, A. and Hogg, D. (1994). Learning flexible models from image sequences, in J.-O. Eklundh (ed.), *Proceedings of the 3rd European Conference on Computer Vision*, Springer-Verlag, pp. 299–308.
- Baumberg, A. and Hogg, D. (1995a). An adaptive eigenshape model, *Proceedings of the British Machine Vision Conference* pp. 87–96.
- Baumberg, A. and Hogg, D. (1995b). Generating spatiotemporal models from examples, *Proceedings of the British Machine Vision Conference*, Vol. 2, pp. 413–422.
- Bayard, D. S. (1991). A forward method for optimal stochastic nonlinear and adaptive control, *IEEE Transactions on Automatic Control* **36**(9): 1046–1055.
- Bellman, R. (1961). *Adaptive control processes: a guided tour*, Princeton University Press.
- Bentley, J. L. (1980). Multidimensional divide and conquer, *Communications of the ACM* **23**(4): 214–229.
- Bergman, N. (1999). *Recursive Bayesian Estimation: Navigation and Tracking Applications*, PhD thesis, Department of Electrical Engineering. Linköping Studies in Science and Technology. Dissertations No. 579.
- Bergman, N., Ljung, L. and Gustafsson, F. (1999). Terrain navigation using Bayesian statistics, *IEEE Control Systems Magazine* **19**(3): 33–40.
- Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*, Wiley Series in Applied Probability and Statistics.
- Bertsekas, D. P. (1987). *Dynamic programming: deterministic and stochastic models*, Prentice-Hall.
- Bertsekas, D. P. (1995). *Dynamic programming and optimal control (Volumes 1 and 2)*, Athena Scientific.
- Berzuini, C., Best, N. G., Gilks, W. R. and Larizza, C. (1997). Dynamic conditional independence models and Markov Chain Monte Carlo methods, *Journal of the American Statistical Association* **92**(440): 1403–1412.
- Besag, J. and Green, P. J. (1993). Spatial statistics and bayesian computation, *Journal of the Royal Statistical Society B* **55**: 25–38.

- Beymer, D. and Poggio, T. (1995). Face recognition from one example view, *Proceedings of the 5th International Conference on Computer Vision*, pp. 500–507.
- Bickel, P. J., Ritov, Y. and Ryden, T. (1998). Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models, *Annals of Statistics* **26**(4): 1614–1635.
- Billio, M., Monfort, A. and Robert, C. P. (1998). The simulated likelihood ratio (SLR) method, *Technical report*, University Cà Foscari of Venice.
- Birmiwal, K. and Bar-Shalom, Y. (1984). Dual control guidance for simultaneous identification and interception, *Automatica* **20**(6): 737–749.
- Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*, Oxford University Press.
- Blair, W. D. and Watson, G. A. (1994). IMM algorithm for solution to benchmark problem for tracking maneuvering targets, *Proceedings of the SPIE* **2221**: 476–488.
- Blake, A. and Isard, M. (1998). *Active contours*, Springer-Verlag.
- Blake, A. and Isard, M. A. (1994). 3D position, attitude and shape input using video tracking of hands and lips, *Proceedings of Siggraph*, ACM, pp. 185–192.
- Blake, A., Bascle, B., Isard, M. and MacCormick, J. (1998). Statistical models of visual shape and motion, *Philosophical Transactions of the Royal Society A* **356**: 1283–1302.
- Blake, A., Isard, M. A. and Reynard, D. (1995). Learning to track the visual motion of contours, *Journal of Artificial Intelligence* **78**: 101–134.
- Blom, H. A. P. and Bar-Shalom, Y. (1988). The interacting multiple model algorithm for systems with Markovian switching coefficients, *IEEE Transactions on Automatic Control* **AC-33**(8): 780–783.
- Bobick, A. F. and Wilson, A. D. (1995). A state-based technique for the summarisation and recognition of gesture, *Proceedings of the 5th International Conference on Computer Vision*, pp. 382–388.
- Bollerslev, T., Engle, R. F. and Nelson, D. B. (1994). ARCH models, in R. F. Engle and D. McFadden (eds), *The Handbook of Econometrics, Volume 4*, North-Holland, Amsterdam, pp. 2959–3038.
- Bølviken, E., Glöckner, F. and Stenseth, N. C. (1999). Estimating parameters in autoregressive processes of low order from noisy, binary data, manuscript under preparation.
- Boothby, W. M. (1986). *An Introduction to Differential Manifolds and Riemannian Geometry*, Academic Press, Inc.
- Borenstein, J., Everett, B. and Feng, L. (1996). *Navigating Mobile Robots: Systems and Techniques*, A K Peters, Ltd., Wellesley, MA.
- Borowsky, B. Z. and Zakai, M. (1975). A lower bound on the estimation error for Markov processes, *IEEE Transactions on Automatic Control* **20**(6): 785–788.
- Boutilier, C., Friedman, N., Goldszmidt, M. and Koller, D. (1996). Context-specific independence in Bayesian networks, *Proceedings of the 12th Conference on Uncertainty in AI*, Morgan Kauffmann, pp. 115–123.

- Boyen, X. and Koller, D. (1998). Tractable inference for complex stochastic processes, *Proceedings of the Fourteenth Annual Conference on Uncertainty in AI (UAI)*, pp. 33–42.
- Boyen, X. and Koller, D. (1999). Exploiting the architecture of dynamic systems, *Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI)*, pp. 313–320.
- Brindle, A. (1981). *Genetic algorithm for function optimization*, PhD thesis, University of Alberta.
- Bruckner, J. M., Scott, H. R. W. and Rea, G. R. (1973). Analysis of multimodal systems, *IEEE Transactions on Aerospace and Electronic Systems* **9**(6): 883–888.
- Bucy, R. S. and Senne, K. D. (1971). Digital synthesis of Nonlinear Filters, *Automatica* **7**: 287–298.
- Burgard, W., Fox, D., Hennig, D. and Schmidt, T. (1996). Estimating the absolute position of a mobile robot using position probability grids, *Proceedings of the Thirteenth National Conference on Artificial Intelligence, AAAI, AAAI Press/MIT Press, Menlo Park*.
- Campbell, L. W. and Bobick, A. F. (1995). Recognition of human body motion using phase space constraints, *Proceedings of the 5th International Conference on Computer Vision*, pp. 624–630.
- Carlin, B. P., Polson, N. G. and Stoffer, D. S. (1992). A Monte Carlo approach to nonnormal and nonlinear state-space modeling, *Journal of the American Statistical Association* **87**(418): 493–500.
- Carpenter, J., Clifford, P. and Fearnhead, P. (1999a). Building robust simulation-based filters for evolving data sets, unpublished, Department of Statistics, Oxford University.
- Carpenter, J., Clifford, P. and Fearnhead, P. (1999b). An improved particle filter for nonlinear problems, *IEE Proceedings-Radar Sonar and Navigation* **146**(1): 2–7.
- Carter, C. K. and Kohn, R. (1994). On Gibbs sampling for state space models, *Biometrika* **81**(3): 541–553.
- Carter, C. K. and Kohn, R. (1996). Markov-Chain Monte-Carlo in conditionally Gaussian state-space models, *Biometrika* **83**(3): 589–601.
- Casella, G. and Robert, C. P. (1996). Rao-Blackwellisation of sampling schemes, *Biometrika* **83**(1): 81–94.
- Cavers, J. K. (1991). An analysis of pilot symbol assisted modulation for rayleigh fading channels, *IEEE Transactions on Vehicle Technology* **40**: 686–693.
- Chan, K. S. and Ledolter, J. (1995). Monte Carlo EM estimation for time series models involving counts, *Journal of the American Statistical Association* **90**(429): 242–252.
- Chen, R. and Liu, J. S. (2000). The mixture Kalman filter, *Journal of the Royal Statistical Society, Ser. B*.
- Clapp, T. C. (2000). *Statistical Methods in the Processing of Communications Data*, PhD thesis, Cambridge University Engineering Department.

- Clapp, T. C. and Godsill, S. J. (1999). Fixed-lag smoothing using sequential importance sampling, in J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (eds), *Bayesian Statistics 6*, Oxford University Press, Oxford, pp. 743–52.
- Cleveland, W. S. and Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting, *Journal of the American Statistical Association* **83**: 596–610.
- Collings, I. B. and Moore, J. B. (1994). An HMM approach to adaptive demodulation of QAM signals in fading channels, *International Journal of Adaptive Control and Signal Processing* **8**(5): 457–474.
- Collings, I. B. and Moore, J. B. (1995). An adaptive hidden Markov model approach to FM and M-Ary DPSK demodulation in noisy fading channels, *Signal Processing* **47**(1): 71–84.
- Cooper, M. L., Grenander, U., Miller, M. I. and Srivastava, A. (1997). Accommodating geometric and thermodynamic variability for forward-looking infrared sensors, *Proceedings of the SPIE, Algorithms for Synthetic Aperture Radar IV*, Vol. 3070.
- Cooper, M. L., Lanterman, A. D., Joshi, S. C. and Miller, M. I. (1996). Representing the variation of thermodynamic state via principal components analysis, *Proceedings of the Third Workshop On Conventional Weapon ATR*.
- Cootes, T. F., Taylor, C. J., Lanitis, A., Cooper, D. H. and Graham, J. (1993). Building and using flexible models incorporating grey-level information, *Proceedings of the 4th International Conference on Computer Vision*, pp. 242–246.
- Cover, T. and Thomas, J. (1991). *Elements of Information Theory*, Wiley.
- Cowell, R. G., Dawid, A. P., Lauritzen, S. L. and Spiegelhalter, D. J. (1999). *Probabilistic Networks and Expert Systems*, Springer-Verlag.
- Cox, I. J. (1991). Blanche—an experiment in guidance and navigation of an autonomous robot vehicle, *IEEE Transactions on Robotics and Automation* **7**(2): 193–204.
- Cox, I. J. and Wilfong, G. T. (eds) (1990). *Autonomous Robot Vehicles*, Springer-Verlag.
- Crisan, D. and Doucet, A. (2000). Convergence of sequential Monte Carlo methods, *Technical Report 381*, CUED-F-INFENG.
- Crisan, D. and Grunwald, M. (1999). Large deviation comparison of branching algorithms versus resampling algorithms: Application to discrete time stochastic filtering, *Technical Report 1999-9*, Cambridge University Statistical Laboratory.
- Crisan, D. and Lyons, T. (1999). Minimal entropy approximations and optimal algorithms for the filtering problem, Preprint.
- Crisan, D. and Lyons, T. J. (1997). Nonlinear filtering and measure-valued processes, *Probability Theory and Related Fields* **109**(2): 217–244.
- Crisan, D., Del Moral, P. and Lyons, T. (1999). Discrete filtering using branching and interacting particle systems, *Markov Processes and Related Fields* **5**(3): 293–318.

- Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function, *Mathematics of Control, Signals, and Systems* **2**(4): 303–314.
- Dagum, P. and Galper, A. (1995). Forecasting sleep apnea with dynamic network models, *Proc. Ninth Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 64–71.
- Davis, L. D. (1991). *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, New York.
- de Freitas, J. F. G. (1999). *Bayesian Methods for Neural Networks*, PhD thesis, University of Cambridge, Cambridge, UK.
- de Freitas, J. F. G., Macleod, I. M. and Maltz, J. S. (1999). Neural networks for pneumatic actuator fault detection, *Transactions of the South African IEE* **90**: 28–34.
- de Freitas, J. F. G., Niranjani, M., Gee, A. H. and Doucet, A. (2000). Sequential Monte Carlo methods to train neural network models, *Neural Computation* **12**(4): 955–993.
- Dean, T. and Kanazawa, K. (1989). A model for reasoning about persistence and causation, *Computational Intelligence* **5**(3): 142–150.
- Dean, T. L. and Boddy, M. (1988). An analysis of time-dependent planning, *Proceedings of the Seventh National Conference on Artificial Intelligence AAAI-92*, AAAI, AAAI Press/The MIT Press, Menlo Park, CA, pp. 49–54.
- DeGroot, M. (1970). *Optimal Statistical Decisions*, McGraw-Hill, New York, NY.
- Del Moral, P. (1996). Nonlinear filtering: interacting particle solution, *Markov Processes and Related Fields* **2**(4): 555–580.
- Del Moral, P. (1998). Measure valued processes and interacting particle systems, application to non linear filtering problems, *Annals of Applied Probability* **8**: 438–495.
- Del Moral, P. and Guionnet, A. (1998a). Large deviation for interacting particle systems. Application to nonlinear filtering problems, *Stochastic Processes and their Applications* **78**(1): 69–95.
- Del Moral, P. and Guionnet, A. (1998b). On the stability of measure valued processes. Applications to nonlinear filtering and interacting particle systems, *Publication du Laboratoire de Statistique et Probabilités 3-98*, Université Paul Sabatier, Toulouse.
- Del Moral, P. and Guionnet, A. (1999). A central limit theorem for non linear filtering using interacting particle systems, *Annals of Applied Probability* **9**: 275–297.
- Del Moral, P. and Jacod, J. (1998). The Monte-Carlo method for filtering with discrete time observations, *Publications 453 du Laboratoire de Probabilités Paris 6*.
- Del Moral, P. and Jacod, J. (1999). The Monte-Carlo method for filtering with discrete time observations: Central limit theorems, *Publications 515 du Laboratoire de Probabilités Paris 6*.
- Del Moral, P. and Ledoux, M. (2000). Convergence of empirical processes for interacting particle systems with applications to non-linear filtering, *Journal of Theoretical Probability*.

- Del Moral, P. and Miclo, L. (2000). Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non linear filtering, in M. Emery, M. Ledoux and M. Yor (eds), *Séminaire de Probabilités XXXIV*, Lecture Notes in Mathematics, Springer-Verlag, pp. 1–145.
- Dellaert, F., Burgard, W., Fox, D. and Thrun, S. (1999). Using the condensation algorithm for robust, vision-based mobile robot localization, *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, IEEE, Fort Collins, CO.
- Dellaert, F., Fox, D., Burgard, W. and Thrun, S. (1999). Monte Carlo localization for mobile robots, *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*.
- Dellaert, F., Thorpe, C. and Thrun, S. (1999). Mosaicing a large number of widely dispersed, noisy, and distorted images: A Bayesian approach, *Technical Report CMU-RI-TR-99-34*, Carnegie Mellon University, Pittsburgh, PA.
- Delyon, B., Lavielle, M. and Moulines, E. (1999). Convergence of a stochastic approximation version of the EM algorithm, *Annals of Statistics* pp. 94–128.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via EM algorithm, *Journal of the Royal Statistical Society Series B-Methodological* **39**(1): 1–38.
- Devroye, L. (1986). *Non-Uniform Random Variate Generation*, Springer-Verlag.
- Devroye, L. (1987). *A Course on Density Estimation*, Vol. 14 of *Progress in Probability and Statistics*, Birkhäuser, Boston.
- Devroye, L. and Györfi, L. (1985). *Nonparametric Density Estimation : The L_1 View*, Wiley series in Probability and Mathematical Statistics, John Wiley & Sons, New York.
- Diaconis, P. and Shahshahani, M. (1987). The subgroup algorithm for generating uniform random variables, *Probability in the Engineering and Informational Sciences* **1**: 15–32.
- Diebold, F. X. and Nerlove, M. (1989). The dynamics of exchange rate volatility: a multivariate latent factor ARCH model, *Journal of Applied Econometrics* **4**: 1–21.
- Djurić, P. M. (1999). Monitoring and selection of dynamic models by Monte Carlo sampling, *IEEE Higher Order Statistics Workshop*, Caesarea, Israel, pp. 191–194.
- Doerschuk, P. C. (1995). Cramér-Rao bounds for discrete-time nonlinear filtering problems, *IEEE Transactions on Automatic Control* **40**(8): 1465–1469.
- Doss, H., Sethuraman, J. and Athreya, K. B. (1996). On the convergence of the Markov chain simulation, *Annals of Statistics* **24**: 69–100.
- Doucet, A. (1998). On sequential simulation-based methods for Bayesian filtering, *Technical Report CUED/F-INFENG/TR 310*, Department of Engineering, Cambridge University.
- Doucet, A., Andrieu, C. and Fitzgerald, W. J. (1998). Bayesian estimation of HMM via Monte Carlo methods, *IEEE Workshop on Neural Networks and Signal Processing*, Cambridge.

- Doucet, A., Barat, E. and Duvaut, P. (1995). Monte Carlo approach to recursive Bayesian state estimation, *Proceedings of the IEEE Signal Processing/Athos Workshop on Higher Order Statistics*, Girona, Spain, pp. 12–14.
- Doucet, A., de Freitas, J. F. G., Murphy, K. and Russell, S. (2000). Rao Blackwellised particle filtering for dynamic Bayesian networks, *Uncertainty in Artificial Intelligence*.
- Doucet, A., Godsill, S. and Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering, *Statistics and Computing* **10**(3): 197–208.
- Doucet, A., Gordon, N. J. and Krishnamurthy, V. (1999). Particle filters for state estimation of jump Markov linear systems, *Technical Report CUED/F-INFENG/TR 359*, Cambridge University Engineering Department.
- Doyle, R. S. and Harris, C. J. (1996). Multi-sensor data fusion for helicopter guidance using neuro-fuzzy estimation algorithms, *The Aeronautical Journal* **100**(996): 241–251.
- Du, J. and Vucetic, B. (1991). New 16QAM trellis codes for fading channels, *Electronics Letters* **27**(7): 1009–1010.
- Duda, R. O. and Hart, P. E. (1973). *Pattern Classification and Scene Analysis*, John Wiley & Sons.
- Dudgen, D. E. and Lacoss, R. T. (1993). An overview of automatic target recognition, *MIT Lincoln Laboratory Journal* **6**(1): 3–10.
- Durbin, J. and Koopman, S. J. (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models, *Biometrika* **84**: 1403–1412.
- Engelson, S. and McDermott, D. (1992). Error correction in mobile robot map learning, *Proceedings of the 1992 IEEE International Conference on Robotics and Automation*, Nice, France, pp. 2555–2560.
- Ethier, S. N. and Kurtz, T. G. (1986). *Markov Processes*, John Wiley & Sons.
- Fahrmeir, L. (1992). Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalized linear model, *Journal of the American Statistical Association* **87**(418): 501–509.
- Fahrmeir, L. and Kaufmann, H. (1991). On Kalman filtering, posterior mode estimation and Fisher-scoring in dynamic exponential family regression, *Metrika* **38**: 37–60.
- Feldbaum, A. A. (1965). *Optimal control systems*, Academic Press.
- Fernyhough, J. H., Cohn, A. G. and Hogg, D. C. (1996). Generation of semantic regions from image sequences, *Proceedings of the 4th European Conference on Computer Vision*, pp. 475–484.
- Fisher, N. I. (1993). *Statistical Analysis of Circular Data*, Cambridge University Press, Cambridge.
- Fox, D., Burgard, W. and Thrun, S. (1998). Active Markov localization for mobile robots, *Robotics and Autonomous Systems* **25**: 195–207.
- Fox, D., Burgard, W. and Thrun, S. (1999). Markov localization for mobile robots in dynamic environments, *Journal of Artificial Intelligence Research (JAIR)* **11**: 391–427.

- Fox, D., Burgard, W., Dellaert, F. and Thrun, S. (1999). Monte Carlo localization: Efficient position estimation for mobile robots, *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, AAAI, Orlando, FL.
- Fox, D., Burgard, W., Kruppa, H. and Thrun, S. (2000). A probabilistic approach to collaborative multi-robot localization, *Autonomous Robots*.
- Frühwirth-Schnatter, S. (1994). Applied state space modelling of non-Gaussian time series using integration-based Kalman filtering, *Statistics and Computing* **4**: 259–269.
- Fukuda, T., Ito, S., Oota, N., Arai, F., Abe, Y., Tanake, K. and Tanaka, Y. (1993). Navigation system based on ceiling landmark recognition for autonomous mobile robot, *Proceedings of the International Conference on Industrial Electronics Control and Instrumentation (IECON'93)*, Vol. 1, pp. 1466 – 1471.
- Fung, R. and Del Favero, B. (1994). Backward simulation in Bayesian networks, *Proceedings of the Conference on Uncertainty in AI (UAI)*.
- Galdos, J. I. (1980). A Cramér-Rao bound for discrete-time nonlinear filtering problems, *IEEE Transactions on Automatic Control* **25**(1): 117–119.
- Gelb, A. (ed.) (1974). *Applied Optimal Estimation*, MIT Press, Cambridge, MA.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images., *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**(6): 721–741.
- Geman, S. and Hwang, C.-R. (1987). Diffusions for global optimization, *SIAM Journal of Control and Optimization* **24**(24): 1031–1043.
- Georghiades, C. N. and Han, J. C. (1997). Sequence estimation in the presence of random parameters via the EM algorithm, *IEEE Transactions on Communications* **45**(3): 300–308.
- Gerlach, R., Carter, C. and Kohn, R. (1999). Diagnostics for time series analysis, *Journal Time Series Analysis* **21**: 309–330.
- Gersch, W. and Kitagawa, G. (1988). Smoothness priors in time series, *Bayesian Analysis of Time Series and Dynamic Models*, New York, pp. 431–476.
- Gertsman, M. J. and Lodge, J. H. (1997). Symbol-by-symbol MAP demodulation of CPM and PSK signals on rayleigh flat-fading channels, *IEEE Transactions on Communications* **45**(7): 788–799.
- Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration, *Econometrica* **24**: 1317–1399.
- Geweke, J. F. and Zhou, G. (1996). Measuring the pricing error of the arbitrage pricing theory, *The Review of Financial Studies* **9**: 557–587.
- Geyer, C. J. (1993). Estimating normalizing constants and reweighting mixtures in Markov chain Monte Carlo, *Technical report*, University of Minnesota.
- Ghahramani, Z. and Jordan, M. I. (1997). Factorial hidden Markov models, *Machine Learning* **29**(2–3): 245–273.
- Gikhman, I. I. and Skorokhod, A. V. (1969). *Introduction to the Theory of Random Processes*, Saunders Company.

- Gilks, W. R. and Berzuini, C. (1999). Following a moving target – Monte Carlo inference for dynamic Bayesian models, *submitted to Journal of the Royal Statistical Society, Series B*.
- Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (eds) (1996). *Markov Chain Monte Carlo in Practice*, Chapman and Hall, Suffolk.
- Gill, R. D. and Levit, B. Y. (1995). Application of the van Trees inequality: a Bayesian Cramér-Rao bound, *Bernoulli* **1**(1/2): 59–79.
- Givens, G. H. and Raftery, A. E. (1996). Local adaptive importance sampling for multivariate densities with strong nonlinear relationships, *Journal of the American Statistical Association* **91**: 132–141.
- Godsill, S. J. (1997). Bayesian enhancement of speech and audio signals which can be modelled as ARMA processes, *International Statistical Review* **65**(1): 1–21.
- Godsill, S. J. and Rayner, P. J. W. (1998a). *Digital Audio Restoration: A Statistical Model-Based Approach*, Springer-Verlag.
- Godsill, S. J. and Rayner, P. J. W. (1998b). Robust reconstruction and analysis of autoregressive signals in impulsive noise using the Gibbs sampler, *IEEE Transactions on Speech and Audio Processing* **6**(4): 352–372.
- Godsill, S. J., Doucet, A. and West, M. (2000). Methodology for Monte Carlo smoothing with application to time-varying autoregressions, *Proceedings of the International Symposium on Frontiers of Time Series Modelling*. Institute of Statistical Mathematics, Tokyo.
- Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley.
- Goodwin, C. G. and Sin, K. S. (1984). *Adaptive filtering prediction and control*, Prentice-Hall.
- Gordon, N. (1993). *Bayesian methods for tracking*, PhD thesis, University of London.
- Gordon, N. (1997). A hybrid bootstrap filter for target tracking in clutter, *IEEE Transactions on Aerospace and Electronic Systems* **33**: 353–358.
- Gordon, N. J., Marrs, A. D. and Salmond, D. J. (2001). Sequential analysis of nonlinear dynamic systems using particles and mixtures, in W. J. Fitzgerald, A. Walden, P. C. Young and R. L. Smith (eds), *Nonlinear and Nonstationary Signal Processing*, Cambridge University Press.
- Gordon, N. J., Salmond, D. J. and Ewing, C. (1995). Bayesian state estimation for tracking and guidance using the bootstrap filter, *Journal of Guidance, Control and Dynamics* **18**(6): 1434–1443.
- Gordon, N. J., Salmond, D. J. and Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *IEE Proceedings-F* **140**(2): 107–113.
- Grassberger, P. (1997). Pruned-enriched Rosenbluth method: simulations of q polymers of chain length up to 1000000., *Physical Review E* **56**: 3682–93.
- Green, P. J. (1994). Markov chain monte carlo in image analysis, *Markov chain Monte Carlo in Practice*.

- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika* **82**(4): 711–32.
- Grenander, U. (1976–1981). *Lectures in Pattern Theory I, II and III*, Springer-Verlag.
- Grenander, U. (1993). *General Pattern Theory*, Oxford University Press.
- Grenander, U. and Miller, M. I. (1994). Representations of knowledge in complex systems, *Journal of the Royal Statistical Society* **56**(3): 549–603.
- Grenander, U. and Srivastava, A. (2000). Probability models for background clutter in natural images, *Monograph of Department of Statistics, Florida State University*.
- Grenander, U., Chow, Y. and Keenan, D. M. (1991). *HANDS. A Pattern Theoretical Study of Biological Shapes*, Springer-Verlag.
- Grenander, U., Miller, M. I. and Srivastava, A. (1998). Hilbert-Schmidt lower bounds for estimators on matrix Lie groups for ATR, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**(8): 790–802.
- Gunther, J., Beard, R., Wilson, J., Oliphant, T. and Stirling, W. (1997). Fast nonlinear filtering via Galerkins method, *American Control Conference*, Albuquerque, USA, pp. 2815–2819.
- Haeb, R. and Meyr, H. (1989). A systematic approach to carrier recovery and detection of digitally phase modulated signals on fading channels, *IEEE Transactions on Communications* **COM-37**: 748–759.
- Hammersley, J. M. and Morton, K. W. (1954). Poor man's Monte Carlo, *Journal of the Royal Statistical Society B* **16**: 23–38.
- Handschin, J. E. and Mayne, D. Q. (1969). Monte Carlo techniques to estimate the conditional expectation in multi-stage non-linear filtering, *International Journal of Control* **9**(5): 547–559.
- Harrison, P. J. and Stevens, C. F. (1976). Bayesian forecasting (with discussion), *Journal of the Royal Statistical Society, Series B* **38**: 205–247.
- Harvey, A. C., Ruiz, E. and Sentana, E. (1992). Unobserved component time series models with ARCH disturbances, *Journal of Econometrics* **52**: 129–158.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* **57**: 97–109.
- Hesterberg, T. (1995). Weighted average importance sampling and defensive mixture distributions, *Technometrics* **37**(2): 185–194.
- Hewish, M. and Gourley, S. R. (1999). Ultra-wideband technology opens up new horizons, *Jane's International Defence Review* p. 20.
- Higuchi, T. (1997). Monte Carlo filter using the genetic algorithm operators, *Journal of Statistical Computation and Simulation* **59**(1): 1–23.
- Higuchi, T. (1999). Applications of quasi-periodic oscillation models to seasonal small count time series, *Computational Statistics & Data Analysis* **30**: 281–301.
- Hinkel, R. and Knieriemen, T. (1988). Environment perception with a laser radar in a fast moving robot, *Proceedings of Symposium on Robot Control*, Karlsruhe, Germany, pp. 68.1–68.7.

- Holland, J. H. (1975). *Adaption in natural and artificial systems*, University of Michigan Press, Ann Arbor.
- Holmes, C. C. and Mallick, B. K. (1998). Bayesian radial basis functions of variable dimension, *Neural Computation* **10**(5): 1217–1233.
- Holst, U. and Lindgren, G. (1991). Recursive estimation in mixture models with Markov regime, *IEEE Transactions on Information Theory* **37**: 1683–1690.
- Horn, L. J., Showalter, M. R. and Russell, C. T. (1996). Detection and behavior of pan wakes in Saturn's ring, *Icarus* **124**: 643–676.
- Huang, T., Koller, D., Malik, J., Ogasawara, G., Rao, B., Russell, S. J. and Weber, J. (1994). Automatic symbolic traffic scene analysis using belief networks, *Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI)*, pp. 966–972.
- Huang, X. D., Arika, Y. and Jack, M. A. (1990). *Hidden Markov Models for Speech Recognition*, Edinburgh University Press.
- Hull, D. G., Speyer, J. L. and Burris, D. B. (1990). Linear-quadratic guidance law for dual-control of homing missiles, *Journal of Guidance, Control and Dynamics* **13**(1): 137–144.
- Hull, J. and White, A. (1987). The pricing of options on assets with stochastic volatilities, *Journal of Finance* **42**: 281–300.
- Hürzeler, M. (1998). *Statistical Methods for General State-Space Models*, Ph.D. Thesis, Department of Mathematics, ETH Zürich, Zürich.
- Hürzeler, M. and Künsch, H. R. (1998). Monte Carlo approximations for general state space models, *Journal of Computational and Graphical Statistics* **7**(2): 175–193.
- Ikoma, N. (1996). Estimation of time varying peak of power spectrum based on non-Gaussian nonlinear state space modeling, *Journal of Signal Processing* **49**: 85–95.
- Irwing, M., Cox, N. and Kong, A. (1994). Sequential imputation for multilocus linkage analysis, *Proceedings of the National Academy of Science, USA* **91**: 11684–11688.
- Isard, M. and Blake, A. (1996). Contour tracking by stochastic propagation of conditional density, *European Conference on Computer Vision*, Cambridge, UK, pp. 343–356.
- Isard, M. and Blake, A. (1998a). Condensation – conditional density propagation for visual tracking, *International Journal of Computer Vision* **28**(1): 5–28.
- Isard, M. and Blake, A. (1998b). A mixed-state condensation tracker with automatic model switching, *Proceedings of the 6th International Conference on Computer Vision*, pp. 107–112.
- Jacquier, E., Polson, N. G. and Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models (with discussion), *Journal of Business and Economic Statistics* **12**: 371–417.
- Jazwinski, A. H. (1970). *Stochastic Processes and Filtering Theory*, Academic Press.

- Jensen, F. V., Kjærulff, U., Olesen, K. G. and Pedersen, J. (1989). An expert system for control of waste water treatment – a pilot project, *Technical report*, Judex Datasystemer A/S, Aalborg, Denmark.
- Julier, S. J. and Uhlmann, J. K. (1997). A new extension of the Kalman filter to nonlinear systems, *Proc. of AeroSense: The 11th International Symposium on Aerospace/Defence Sensing, Simulation and Controls, Orlando, Florida*, Vol. Multi Sensor Fusion, Tracking and Resource Management II.
- Kaelbling, L. P., Cassandra, A. R. and Kurien, J. A. (1996). Acting under uncertainty: Discrete Bayesian models for mobile-robot navigation, *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems, *Journal of Basic Engineering* **82**: 35–45.
- Kam, P. Y. and Ching, H. M. (1992). Sequence estimation over the slow nonselective rayleigh fading channel with diversity reception and its application to viterbi decoding, *IEEE J. Select. Areas Commun.* **COM-10**(3): 562–570.
- Kanazawa, K., Koller, D. and Russell, S. J. (1995). Stochastic simulation algorithms for dynamic probabilistic networks, *Proceedings of the Eleventh Annual Conference on Uncertainty in AI (UAI '95)*, pp. 346–351.
- Kashiwagi, N. and Yanagimoto, T. (1992). Smoothing serial count data through a state-space model, *Biometrics* **48**: 1187–1194.
- Kendall, M. G. and Stuart, A. (1979). *The advanced theory of statistics, vol 2, inference and relationship*, Charles Griffing and Co Ltd, London.
- Kerr, T. (1989). Status of CR-like lower bounds for nonlinear filtering, *IEEE Transactions on Aerospace and Electronic Systems* **25**: 590–601.
- Kim, S., Shephard, N. and Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models, *Review of Economic Studies* **65**: 361–393.
- King, M., Sentana, E. and Wadhwani, S. (1994). Volatility and links between national stock markets, *Econometrica* **62**: 901–933.
- Kitagawa and Sato (2000). Nonlinear state space model approach to financial time series with time-varying variance, *Proceedings of the Hong Kong International Workshop on Statistics in Finance*, Imperial College Press, London.
- Kitagawa, G. (1987). Non-Gaussian state-space modeling of nonstationary time series, *Journal of the American Statistical Association* **82**(400): 1032–1063.
- Kitagawa, G. (1991). A nonlinear smoothing method for time series analysis, *Statistica Sinica* **1**(2): 371–388.
- Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, *Journal of Computational and Graphical Statistics* **5**(1): 1–25.
- Kitagawa, G. (1998). Self-organising state space model, *Journal of the American Statistical Association* **93**(443): 1203–1215.
- Kitagawa, G. and Gersch, W. (1985). A smoothness priors time varying AR coefficient modeling of nonstationary time series, *IEEE Transactions on Automatic Control* **AC-30**: 48–56.

- Kitagawa, G. and Gersch, W. (1996). *Smoothness Priors Analysis of Time Series*, Vol. 116 of *Lecture Notes In Statistics*, Springer-Verlag.
- Knill, D. C., Kersten, D. and Yuille, A. L. (1996). A Bayesian formulation of visual perception, in D. C. Knill and W. Richard (eds), *Perception as Bayesian inference*, Cambridge University Press, pp. 1–22.
- Koenderink, J. J. and van Doorn, A. J. (1991). Affine structure from motion, *Journal of the Optical Society of America A*. **8**(2): 337–385.
- Koenig, S. and Simmons, R. (1996). Passive distance learning for robot navigation, in L. Saitta (ed.), *Proceedings of the Thirteenth International Conference on Machine Learning*.
- Koller, D. and Fratkina, R. (1998). Using learning for approximation in stochastic processes, *Proceedings of the International Conference on Machine Learning (ICML)*.
- Koller, D., Lerner, U. and Angelov, D. (1999). A general algorithm for approximate inference and its application to hybrid Bayes nets, *Proceedings of the Fifteenth Annual Conference on Uncertainty in AI (UAI)*, pp. 324–333.
- Koller, D., Weber, J. and Malik, J. (1994). Robust multiple car tracking with occlusion reasoning, *Proceedings of the 3rd European Conference on Computer Vision*, Springer-Verlag, pp. 189–196.
- Kong, A., Liu, J. S. and Wong, W. H. (1994). Sequential imputations and Bayesian missing data problems, *Journal of the American Statistical Association* **89**(425): 278–288.
- Kortenkamp, D., Bonasso, R. P. and Murphy, R. (eds) (1998). *AI-based Mobile Robots: Case studies of successful robot systems*, MIT Press.
- Kotecha, J. H. and Djurić, P. M. (2000). Sequential Monte Carlo sampling detector for Rayleigh fast-fading channels, *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Istanbul, Turkey.
- Kramer, S. C. and Sorensen, H. W. (1988). Recursive Bayesian estimation using piece-wise constant approximations, *Automatica* **24**(6): 789–801.
- Kremer, K. and Binder, K. (1988). Monte Carlo simulation of lattice models for macromolecules, *Computer Physics Reports*.
- Künsch, H. R. (2000). State space and hidden Markov models, in C. Klüppelberg (ed.), *Complex Stochastic Systems*, Chapman and Hall.
- Lanitis, A., Taylor, C. J. and Cootes, T. F. (1995). A unified approach to coding and interpreting face images, *Proceedings of the 5th International Conference on Computer Vision*, pp. 368–373.
- Lanternman, A. (1998). *Modeling Clutter and Target Signatures for Pattern-Theoretic Understanding of Infrared Scenes*, D. Sc. Thesis, Washington University, St. Louis, Missouri.
- Lanternman, A., Miller, M. and Snyder, D. (1997). General Metropolis-Hastings jump diffusions for automatic target recognition in infrared scenes, *Optical Engineering* **36**(4): 1123–1137.
- Lawrence, R. V. (1998). Interceptor line of sight rate steering: necessary conditions for a direct hit, *Journal of Guidance, Control and Dynamics* **21**(3): 471–476.

- Le Gland, F., Musso, C. and Oudjane, N. (1998). An analysis of regularized interacting particle methods for nonlinear filtering, *Proceedings of the IEEE European Workshop on Computer-Intensive Methods in Control and Data Processing, Prague*, pp. 167–174.
- Lenser, S. and Veloso, M. (2000). Sensor resetting localization for poorly modelled mobile robots, *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, San Francisco, CA.
- Leonard, J., Durrant-Whyte, H. F. and Cox, I. J. (1992). Dynamic map building for an autonomous mobile robot, *International Journal of Robotics Research* **11**(4): 89–96.
- Liu, J. S. (1996a). Metropolized independent sampling with comparisons to rejection sampling and importance sampling, *Statistics and Computing* **6**(2): 113–119.
- Liu, J. S. (1996b). Nonparametric hierarchical Bayes via sequential imputations, *Annals of Statistics* **24**(3): 911–930.
- Liu, J. S. and Chen, R. (1995). Blind deconvolution via sequential imputations, *Journal of the American Statistical Association* **90**(430): 567–576.
- Liu, J. S. and Chen, R. (1998). Sequential Monte Carlo methods for dynamic systems, *Journal of the American Statistical Association* **93**(443): 1032–1044.
- Liu, J. S., Chen, R. and Wong, W. H. (1998). Rejection control and sequential importance sampling, *Journal of the American Statistical Association* **93**(443): 1022–1031.
- Liu, J. S., Wong, W. H. and Kong, A. (1994). Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes, *Biometrika* **81**(1): 27–40.
- Liu, Y. and Blosstein, S. D. (1995). Identification of frequency non-selective fading channels using decision feedback and adaptive linear prediction, *IEEE Transactions on Communications* **43**: 1484–1492.
- Ljung, L. (1979). Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems, *IEEE Transactions on Automatic Control* **AC-24**(1): 36–50.
- Ljung, L. (1987). *System identification: theory for the user*, Prentice-Hall.
- Lodge, J. H. and Moher, M. L. (1990). Maximum likelihood sequence estimation of CPM signals transmitted over rayleigh flat-fading channels, *IEEE Transactions on Communications* **38**: 787–794.
- Loizeaux, M., Srivastava, A. and Miller, M. I. (1999). Pose/location estimation of ground targets, *Proceedings of SPIE Conference on Signal processing, sensor fusion, and target recognition*, Orlando, FL, pp. 140–151.
- MacCormick, J. (1999). Partitioned sampling, *Accepted to ICCV'99*.
- MacCormick, J. and Blake, A. (1998). A probabilistic contour discriminant for object localisation, *Proceedings of the 6th International Conference on Computer Vision*, pp. 390–395.
- MacEachern, S. N., Clyde, M. and Liu, J. S. (1999). Sequential importance sampling for nonparametric Bayes models: the next generation, *Canadian Journal of Statistics* **27**(2): 251–267.

- Mardia, K. V., Ghali, N. M., Howes, M., Hainsworth, T. J. and Sheehy, N. (1993). Techniques for online gesture recognition, *Journal of Image and Vision Computing* **11**(5): 283–294.
- Marshall, A. W. (1956). The use of multi-stage sampling schemes in Monte Carlo computations, in M. Meyer (ed.), *Symposium on Monte Carlo Methods*, Wiley, New York, pp. 123–140.
- Mazor, E., Averbuch, A., Bar-Shalom, Y. and Dayan, J. (1998). Interacting multiple model methods in target tracking: a survey, *IEEE Transactions on Aerospace and Electronic Systems* **34**(1): 103–123.
- McGinnity, S. (1998). *Nonlinear estimation techniques for target tracking*, PhD thesis, The Queen's University of Belfast, Department of Electrical and Electronic Engineering.
- McGinnity, S. and Irwin, G. W. (1998). A multiple model bootstrap filter for manoeuvring target tracking, *Submitted to IEEE Transactions on Aerospace and Electronic Systems*.
- Meirovitch, H. (1985). The scanning method with a mean-field parameter: Computer-simulation study of critical exponents of self-avoiding walks on a square lattice, *Macromolecules* **18**(3): 563–569.
- Melvin, D. G. (1996). A comparison of statistical and connectionist techniques for liver transplant monitoring, *Technical Report CUED/F-INFENG/TR 282*, Cambridge University Engineering Department.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equations of state calculations by fast computing machines, *Journal of Chemical Physics* **21**: 1087–1091.
- Meyn, S. P. and Tweedie, R. L. (1993). *Markov chains and stochastic stability*, Springer-Verlag.
- Michigan Technological University (1987). *Prism 3.1 User's Manual*, Keweenaw Research Center.
- Miller, M. I., Srivastava, A. and Grenander, U. (1995). Conditional-expectation estimation via jump-diffusion processes in multiple target tracking/recognition, *IEEE Transactions on Signal Processing* **43**(11): 2678–2690.
- Monahan, J. and Genz, A. (1997). Spherical-radial integration rules for Bayesian computation, *Journal of the American Statistical Association* **92**(438): 664–674.
- Moore, A. W. (1990). *Efficient Memory-based Learning for Robot Control*, PhD thesis, Trinity Hall College, Cambridge University, England.
- Moore, A. W., Schneider, J. and Deng, K. (1997). Efficient locally weighted polynomial regression predictions, *Proceedings of the International Conference on Machine Learning (ICML)*.
- Moravec, H. and Elfes, A. (1985). High resolution maps from wide angle sonar, *Proceedings of the IEEE International Conference on Robotics & Automation (ICRA)*, pp. 116–121.
- Mosterman, P. J. and Biswas, G. (1999). Diagnosis of continuous valued systems in transient operating regions, *IEEE Transactions on Systems, Man, and Cybernetics, Part A* **29**(6): 554–565.

- Mumford, D. (1996). Pattern theory: a unifying perspective, in D. C. Knill and W. Richard (eds), *Perception as Bayesian inference*, Cambridge University Press, pp. 25–62.
- Murphy, K. P. (1999). A variational approximation for Bayesian networks with discrete and continuous latent variables, *Proceedings of the 14th Conference on Uncertainty in AI*, Morgan Kauffmann, pp. 457–466.
- Murphy, K. P. (2000). Bayesian map learning in dynamic environments, *Advances in Neural Information Processing Systems 12*, MIT Press, pp. 1015–1021.
- Murray, D. W. and Basu, A. (1994). Motion tracking with an active camera, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **16**(5): 449–459.
- Musso, C. and Oudjane, N. (1998). Regularisation schemes for branching particle systems as a numerical solving method of the nonlinear filtering problem, *Proceedings of the Irish Signals Systems Conference, Dublin*.
- Nardone, S. C. and Aidala, V. J. (1981). Observability criteria for bearings-only target motion analysis, *IEEE Transactions on Aerospace and Electronic Systems* **AES-17**(2): 162–166.
- Narendra, K. S. and Parthasarathy, K. (1990). Identification and control of dynamical systems using neural networks, *IEEE Transactions on Neural Networks* **1**: 4–27.
- Neal, R. M. (1998). Annealed importance sampling, *Technical Report 9805*, Department of Statistics, University of Toronto.
- Nicholson, A. E. and Brady, J. M. (1994). Dynamic belief networks for discrete monitoring, *IEEE Transactions on Systems, Man and Cybernetics* **24**(11): 1593–1610.
- North, B. and Blake, A. (1997). Using expectation-maximisation to learn dynamical models from visual data, *Proceedings of the British Machine Vision Conference*, pp. 669–678.
- Nourbakhsh, I., Powers, R. and Birchfield, S. (1995). DERVISH, an office-navigating robot, *AI Magazine* **16**(2): 53–60.
- Oehlert, G. W. (1998). Faster adaptive importance sampling in low dimensions, *Journal of Computational and Graphical Statistics* **7**: 158–174.
- O'Hagan, A. (1994). *Kendall's Advanced Theory of Statistics: Bayesian Inference, Volume 2B*, Halsted Press, New York.
- Omohundro, S. M. (1991). Bumptrees for efficient function, constraint, and classification learning, in R. P. Lippmann, J. E. Moody and D. S. Touretzky (eds), *Advances in Neural Information Processing Systems 3*, Morgan Kaufmann.
- Oudjane, N. and Musso, C. (1999). Multiple model particle filter, *Actes du 17ème Colloque GRETSI, Vannes*, pp. 681–684.
- Parthasarathy, P. (1967). *Probability Measures on Metric Spaces*, Academic Press, New York.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann.
- Penny, W. D., Roberts, S. J., Curran, E. and Stokes, M. (1999). EEG-based communication: a pattern recognition approach, To appear in *IEEE Transactions on Rehabilitation Engineering*.

- Pickering, C. (1998). Complementary in-situ and post-deposition diagnostics of thin film semiconductor structures, *Thin Solid Films* pp. 313–314.
- Pitt, M. and Shephard, N. (1999a). Analysis of time varying covariances: A factor stochastic volatility approach, in J. O. Berger, J. M. Bernardo, A. P. Dawid and A. F. M. Smith (eds), *Bayesian Statistics 5*, Oxford University Press, pp. 547–570.
- Pitt, M. K. and Shephard, N. (1999b). Filtering via simulation: Auxiliary particle filters, *Journal of the American Statistical Association* **94**(446): 590–599.
- Pitt, M. K. and Shephard, N. (1999c). Time varying covariances: a factor stochastic volatility approach (with discussion), in J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (eds), *Bayesian Statistics 6*, Oxford University Press, pp. 547–570.
- Poggio, T. and Girosi, F. (1990). Networks for approximation and learning, *Proceedings of the IEEE* **78**(9): 1481–1497.
- Pole, A. (1988). Transfer response models: A numerical approach, in J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith (eds), *Bayesian Statistics 3*, Oxford University Press, pp. 733–746.
- Pole, A. and West, M. (1988). Efficient numerical integration in dynamic models, *Technical Report Warwick Research Report 136*, Department of Statistics, University of Warwick, Coventry.
- Pole, A. and West, M. (1990). Efficient Bayesian learning in non-linear dynamic models, *Journal of Forecasting* **9**(2): 119–136.
- Pole, A., West, M. and Harrison, P. J. (1988). Non-normal and non-linear dynamic Bayesian modelling, in J. C. Spall (ed.), *Bayesian Analysis of Time Series and Dynamic Models*, Marcel Dekker, New York, pp. 167–198.
- Pontryagin, L. S. (1957). Some mathematical properties arising in connection with the theory of optimal automatic control systems, *Proceedings of the Conference on Basic Problems in Automatic Control and Regulation*, Academy of Sciences, Moscow.
- Prado, R., West, M. and Krystal, A. D. (1999). Multi-channel EEG analyses via dynamic regression models with time-varying lag/lead structure, *ISDS Discussion paper 99-05*, Duke University.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992). *Numerical Recipes in C - The Arts of Scientific Computing*, 2nd edn, Cambridge University Press.
- Priestley, M. B. (1981). *Spectral Analysis and Time Series*, Probability and Mathematical Statistics, Academic Press, Inc.
- Proakis, J. G. (1995). *Digital Communications*, 3rd edn, McGraw-Hill.
- Punskaya, E., Andrieu, C., Doucet, A. and Fitzgerald, W. J. (2000). Particle filtering for demodulation in fading channels, *Technical Report 384*, CUED-F-INFENG.
- Puskorius, G. V. and Feldkamp, L. A. (1994). Neurocontrol of nonlinear dynamical systems with Kalman filter trained recurrent networks, *IEEE Transactions on Neural Networks* **5**(2): 279–297.

- Rabiner, L. and Bing-Hwang, J. (1993). *Fundamentals of speech recognition*, Prentice-Hall.
- Rabiner, L. and Juang, B. (1986). An introduction to hidden Markov models, *IEEE Acoustics, Speech & Signal Processing Magazine* **3**(1): 4–16.
- Rajan, J. J., Rayner, P. J. W. and Godsill, S. J. (1997). A Bayesian approach to parameter estimation and interpolation of time-varying autoregressive processes using the Gibbs sampler, *IEE Proceedings on Vision, Image and Signal Processing*.
- Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, John Wiley & Sons, New York.
- Rencken, W. D. (1993). Concurrent localisation and map building for mobile robots using ultrasonic sensors, *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Yokohama, Japan, pp. 2129–2197.
- Reno, A. and Booth, D. (1998). Deformable models for object recognition in aerial image, *F. Sadjadi (ed.), Automatic Target Recognition VIII SPIE Proc.* **3371**: 322–333.
- Reynard, D., Wildenberg, A. P., Blake, A. and Marchant, J. (1996). Learning dynamics of complex motions from image sequences, *Proceedings of the 4th European Conference on Computer Vision*, Cambridge, England, pp. 357–368.
- Ripley, B. D. (1996). *Pattern Recognition and Neural Networks*, Cambridge University Press.
- Ripley, B. D. and Sutherland, A. L. (1990). Finding spiral structures in images of galaxies, *Philosophical Transactions of the Royal Society of London A* **332**(1627): 477–485.
- Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*, Springer-Verlag.
- Roberts, G. O. and Tweedie, R. L. (September, 1995). Exponential convergence of Langevin's diffusions and their discrete approximations, *MCMC website*.
- Rosenbluth, M. N. and Rosenbluth, A. W. (1955). Monte Carlo calculation of the average extension of molecular chains, *Journal of Chemical Physics* **23**: 356–359.
- Rowe, S. M. and Blake, A. (1996). Statistical mosaics for tracking, *Journal of Image and Vision Computing* **14**: 549–564.
- Royama, T. (1992). *Analytical Population Dynamics*, Vol. 10 of *Population and community biology series*, Chapman & Hall, London.
- Rubin, D. B. (1987a). *Multiple imputation for nonresponse in surveys*, John Wiley & Sons, New York.
- Rubin, D. B. (1987b). A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: the SIR algorithm, *Journal of the American Statistical Association*.

- Rubin, D. B. (1988). Using the SIR algorithm to simulate posterior distributions, in J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith (eds), *Bayesian Statistics 3*, Oxford University Press, pp. 395–402.
- Rydén, T. (1995). On recursive estimation for hidden Markov models, *Stochast. Proc. Applic.* **66**: 79–96.
- Sage, A. P. and Melsa, J. L. (1971). *Estimation Theory with Applications to Communication and Control*, McGraw-Hill.
- Salmond, D. J. (1990). Mixture reduction for target tracking in clutter, *SPIE Vol. 1305, Signal and Data Processing of Small Targets*, pp. 434–445.
- Salmond, D. J., Fisher, D. and Gordon, N. J. (1998). Tracking in the presence of intermittent spurious objects and clutter, in O. E. Drummond (ed.), *Signal and Data Processing of Small Targets 1998*, Vol. 3373, SPIE, pp. 460–474.
- Sampei, S. and Sunaga, T. (1993). Rayleigh fading compensation for QAM in land mobile communications, *IEEE Transactions on Vehicle Technology* **42**: 137–147.
- Schnatter, S. (1992). Integration-based Kalman-filtering for a dynamic generalized linear trend model, *Computational Statistics & Data Analysis* **13**(13): 447–459.
- Semerdjiev, T., Jilkov, V. and Angelove, D. (1998). Target tracking using Monte Carlo simulation, *Mathematics and Computers in Simulation* **47**: 441–447.
- Seymour, J. P. and Fitz, M. P. (1995). Near-optimal symbol-by-symbol detection schemes for flat rayleigh fading, *IEEE Transactions on Communications* **43**(2/3/4): 1525–1533.
- Shachter, R. and Kenley, C. (1989). Gaussian influence diagrams, *Management Science* **35**: 527–550.
- Shan, L. and Doerschuk, P. C. (1997). Performance bounds for nonlinear filters, *IEEE Transactions on Aerospace and Electronic Systems* **33**(1): 316–318.
- Shephard, N. (1994). Partial non-Gaussian state space, *Biometrika* **81**: 115–131.
- Shephard, N. and Pitt, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series, *Biometrika* **84**: 653–67.
- Shiryayev, A. N. (1996). *Probability*, Graduate Texts in Mathematics, 2nd edn, Springer-Verlag.
- Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Vol. 26 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, London.
- Simmons, R. and Koenig, S. (1995). Probabilistic robot navigation in partially observable environments, *Proceedings of IJCAI-95*, IJCAI, Inc., Montreal, Canada, pp. 1080–1087.
- Simmons, R., Goodwin, R., Haigh, K., Koenig, S. and O’Sullivan, J. (1997). A layered architecture for office delivery robots, *Proceedings of the First International Conference on Autonomous Agents*, Marina del Rey, CA.
- Smith, A. and Roberts, G. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods, *J. Royal Statistical Society B* **55**(1): 3–37.

- Smith, A. F. M. and Gelfand, A. E. (1992). Bayesian statistics without tears: a sampling-resampling perspective, *American Statistician* **46**(4): 84–88.
- Smith, A. F. M. and West, M. (1983). Monitoring renal transplants: An application of the multi-process Kalman filter, *Biometrics* **39**: 867–878.
- Smith, R., Self, M. and Cheeseman, P. (1988). Estimating uncertain spatial relationships in robotics, in Lemmer and Kanal (eds), *Uncertainty in Artificial Intelligence*, Vol. 2, Elsevier, pp. 435–461.
- Smyth, P., Heckerman, D. and Jordan, M. I. (1997). Probabilistic independence networks for hidden Markov probability models, *Neural Computation* **9**(2): 227–269.
- Snyder, D. L., Hammoud, A. M. and White, R. L. (1993). Image recovery from data acquired with a charge-coupled-device camera, *Journal of the Optical Society of America A* **10**(5): 1014–1023.
- Snyder, P. G., Woollam, J. A., Alterovitz, S. A. and Johs, B. (1990). Modelling $\text{Al}_x\text{Ga}_{1-x}\text{As}$ optical constants as functions of composition, *Journal of Applied Physics*. **68**(11): 5925–5926.
- Solo, V. (1980). Some aspect of recursive parameter estimation, *International Journal of Control* **32**: 395–410.
- Sorensen, H. W. (1966). Kalman filtering techniques, *Advanced Control Systems* **3**: 219–292.
- Sorensen, H. W. (1974). On the development of practical nonlinear filters, *Information Sciences* **7**: 253–270.
- Sorenson, H. W. and Alspach, D. L. (1971). Recursive Bayesian estimation using Gaussian sums, *Automatica* **7**: 465–479.
- Srivastava, A. (1996). *Inferences on Transformation Groups Generating Patterns on Rigid Motions*, D. Sc. Thesis, Washington University, St. Louis, Missouri.
- Srivastava, A. (2000). Bayesian filtering for tracking pose and location of rigid targets, *Proceedings of SPIE*.
- Srivastava, A., Grenander, U., Jensen, G. R. and Miller, M. I. (December, 1999). Jump-diffusion Markov processes on orthogonal groups for object recognition, *accepted for publication by Journal of Statistical Planning and Inference*.
- Srivastava, A., Miller, M. I. and Grenander, U. (1999). Bayesian automated target recognition, *Handbook of Image and Video Processing*.
- Stavropoulos, P. and Titterton, D. M. (1998). Computational Bayesian filtering, *Technical report 98-13, Department of Statistics, University of Glasgow*.
- Stenseth, N. C., Bjørnstad, O. N. and Falck, W. (1996). Is spacing behaviour coupled with predation causing the microtine density cycle? A synthesis of current process-oriented and pattern-oriented studies, *Proceedings of the Royal Society of London, B* **263**: 1423–1435.
- Stenseth, N. C., Bjørnstad, O. N. and Saitoh, T. (1996). A gradient from stable to cyclic populations of *Clethrionomys rufocanus* in Hokkaido, *Proceedings of the Royal Society of London, B* **263**: 1117–1126.

- Stenseth, N. C., Falck, W., Bjørnstad, O. N. and Krebs, C. J. (1997). Population regulation in snowshoe hare and Canadian lynx: Asymmetric food web configurations between hare and lynx, *Proceedings of National Academy of Science, Washington* **94**: 5147–5152.
- Stenseth, N. C., Falck, W., Chain, K. S., Bjørnstad, O. N., Donoghue, M. O. and Tong, H. (1998). From ecological patterns to ecological processes: Phase- and density-dependencies in the Canadian lynx cycle, *Proceedings of National Academy of Science, Washington* **95**: 15430–15435.
- Storvik, G. (1994). A Bayesian approach to dynamic contours through stochastic sampling and simulated annealing, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **16**(10): 976–986.
- Stuber, G. L. (1996). *Principles of Mobile Communication*, Kluwer.
- Sullivan, J., Blake, A., Isard, M. and McCormick, J. (1999). Bayesian correlation, *Accepted to ICCV'99*.
- Sutherland, A. I. and Titterton, D. M. (1994). Bayesian analysis of image sequences, *Technical report 94-3, Department of Statistics, University of Glasgow*.
- Tanizaki, H. (1993). *Nonlinear Filters*, Springer-Verlag, New York.
- Tanner, G. L. (1995). Missile control against multiple targets using non-quadratic cost functions, *American Control Conference*, Vol. 1, pp. 515–519.
- Tanner, M. A. (1993). *Tools for Statistical Inference*, 2nd edn, Springer Verlag.
- Taylor, J. H. (1979). Cramer-Rao estimation error lower bound analysis for nonlinear systems with unknown deterministic variables, *IEEE Transactions on Automatic Control* **24**(2): 343–344.
- Thrun, . (1998). Bayesian landmark learning for mobile robot localization, *Machine Learning*.
- Thrun, S., Bennewitz, M., Burgard, W., Cremers, A. B., Dellaert, F., Fox, D., Hähnel, D., Rosenberg, C., Roy, N., Schulte, J. and Schulz, D. (1999). MIN-ERVA: A second generation mobile tour-guide robot, *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*.
- Thrun, S., Burgard, W. and Fox, D. (1998). A probabilistic approach to concurrent mapping and localization for mobile robots, *Machine Learning* **31**: 29–53.
- Thrun, S., Fox, D. and Burgard, W. (2000). Monte Carlo localization with mixture proposal distribution, *Proceedings of the AAAI National Conference on Artificial Intelligence*, AAAI, Austin, TX.
- Tichavský, P., Muravchik, C. and Nehorai, A. (1998). Posterior Cramér-Rao bounds for discrete-time nonlinear filtering, *IEEE Transactions on Signal Processing* **46**(5): 1386–1396.
- Tierney, L. (1994). Markov chains for exploring posterior distributions, *Annals of Statistics* **22**: 1701–1728.
- Titterton, D. M. (1973). A method of extremum adaptation, *Journal of the Institute of Mathematics and its Applications* **11**: 297–315.
- Titterton, D. M., Smith, A. F. M. and Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*, John Wiley & Sons, New York.

- Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*, Oxford statistical science series, Oxford University Press.
- Torrance, J. M. and Hanzo, L. (1995). Comparative study of pilot symbol assisted modem schemes, *Proceedings of Radio Receivers and Associated Systems*, Bath, England, pp. 36–41.
- Ullman, S. and Basri, R. (1991). Recognition by linear combinations of models, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13**(10): 992–1006.
- van der Merwe, R., Doucet, A., de Freitas, J. F. G. and Wan, E. (2000). The unscented particle filter, *Technical Report CUED/F-INFENG/TR 380*, Cambridge University Engineering Department.
- Van Trees, H. L. (1968). *Detection, Estimation and Modulation Theory*, John Wiley & Sons, New York.
- Vermaak, J., Andrieu, C., Doucet, A. and Godsill, S. J. (1999). On-line Bayesian modelling and enhancement of speech signals, *Technical Report CUED/F-INFENG/TR 361*, Cambridge University Engineering Department.
- Vetter, T. and Poggio, T. (1996). Image synthesis from a single example image, *Proceedings of the 4th European Conference on Computer Vision*, Cambridge, England, pp. 652–659.
- Vidoni, P. (1999). Exponential family state space models based on a conjugate latent process, *Journal of the Royal Statistical Society B* **61**: 213–221.
- Vitetta, G. M. and Taylor, D. P. (1995). Maximum likelihood decoding of uncoded and coded PSK signal sequences transmitted over rayleigh flat-fading channels, *IEEE Transactions on Communications* **43**(11): 2750–2758.
- von Neumann, J. (1951). Various techniques used in connection with random digits, *National Bureau of Standards Applied Mathematics Series* **12**: 36–38.
- Wall, F. T. and Erpenbeck, J. (1959). New method for the statistical computation of polymer dimensions, *Journal of Chemical Physics* **30**(3): 634–37.
- Wang, A. H. and Klein, R. L. (1978). Optimal quadrature formula nonlinear estimators, *Information Sciences* **16**: 169–184.
- Wei, G. C. G. and Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms, *Journal of the American Statistical Association* **85**: 699–704.
- Weiß, G., Wetzler, C. and von Puttkamer, E. (1994). Keeping track of position and orientation of moving indoor systems by correlation of range-finder scans, *Proceedings of the International Conference on Intelligent Robots and Systems*, pp. 595–601.
- West, M. (1986). Bayesian model monitoring, *Journal of the Royal Statistical Society (Ser. B)* **48**: 70–78.
- West, M. (1993a). Approximating posterior distributions by mixtures, *Journal of Royal Statistical Society* **55**: 409–422.
- West, M. (1993b). Mixture models, Monte Carlo, Bayesian updating and dynamic models, in J. H. Newton (ed.), *Computing Science and Statistics: Proceedings of the 24th Symposium on the Interface*, Interface Foundation of North America, Fairfax Station, Virginia, pp. 325–333.

- West, M. and Harrison, J. (1997). *Bayesian forecasting and dynamic models*, Springer Series in Statistics, 2nd edn, Springer-Verlag, New York.
- West, M. and Harrison, P. J. (1986). Monitoring and adaptation in Bayesian forecasting models, *Journal of the American Statistical Association* **81**: 741–750.
- West, M. and Harrison, P. J. (1989). Subjective intervention in formal models, *Journal of Forecasting* **8**: 33–53.
- West, M., Harrison, P. J. and Migon, H. S. (1985). Dynamic generalized linear models and Bayesian forecasting (with discussion), *Journal of the American Statistical Association* **80**: 73–97.
- West, M., Prado, R. and Krystal, A. (1999). Evaluation and comparison of EEG traces: Latent structure in non-stationary time series, *Journal of the Acoustical Society America* **94**: 1083–1095.
- Whitley, D. (1994). A genetic algorithm tutorial, *Statistics and Computing* **4**: 63–85.
- Williams, D. (1991). *Probability with Martingales*, Cambridge University Press.
- Zeger, S. L. (1988). A regression model for time series of counts, *Biometrika* **75**: 621–629.
- Zilberstein, S. and Russell, S. (1995). Approximate reasoning using anytime algorithms, in S. Natarajan (ed.), *Imprecise and Approximate Computation*, Kluwer.
- Zweig, G. and Russell, S. J. (1998). Speech recognition with dynamic Bayesian networks, *Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI)*, pp. 173–180.

Index

A

Acceptance sampling 274
Adaption 275
Adaptive importance sampling 300
Aircraft navigation 329
Alternative dynamic models 80
ARCH models 283
ARMA models 89
Astrophysics 442
ATR 533
Audio processing 139
Automated target recognition (ATR)
 533
Autoregressive models 104, 145
Autoregressive processes 348, 438
Autoregressive moving average
 (ARMA) models 89
Auxiliary particle filter 84, 139, 149,
 197, 273

B

BAT network 457
Bayes' theorem 6
Bayesian filtering 4, 79, 97
Bayesian forecasting 197
Bayesian networks 445
Beam pointing control 488
Bearings-only tracking 129, 264, 279
Bellman equation 520
Belief 402
Belief state 447
Biology 104
BK algorithm 464
Black-Scholes 285
Boltzmann distribution 225

Bootstrap filter 10, 139, 148
Bounds 322
Branching mechanisms 28
Bridging densities 144
Butterworth filter 89

C

CAD models 533
Central limit theorem 8, 125
Certainty equivalence 521
Classification 359
Closed-loop control 523
Clutter 344, 525, 533
Coding 89
Computer vision 339, 533
Condensation 4, 352
Condition monitoring 373, 458
Conditional dynamic linear systems
 (CDLMs) 236
Conditional expectations 35
Conditional independence 445
Conditional linear Gaussian 449
Conditional probability distribution
 (CPD) 449
Conditional probabilities 35
Control 308, 517
Continuous-time control 309
Convergence 17, 43, 86
Convergence of particle filters 17
Convolutional coding 89
Counting models 106
CPD 449
Cramér-Rao bound 322

D

Data association 525
DBNs 445, 499
Decoding 95
Demodulation 88
Density estimation 253, 296
Density trees 417
Deterministic filtering 97
Deterministic particle filters 97
Deterministic probability measure 19
Deterministic selection 84
Diagnosis 458
Digital communications 79
Discrete time filtering 43
Dual effect 520
Dynamic Bayesian networks (DBNs) 445, 499
Dynamic generalised linear models 430
Dynamic programming 520

E

Ecology 104
Econometrics 186, 208, 284
Effective sample size 333, 473
Ellipsometry 469
EM algorithm 167, 350
Epanechnikov kernel 253, 303
Exponential forgetting 87
Extended Kalman filter (EKF) 4

F

Factor stochastic volatility modelling 208
Factored dynamic systems 445
Factored sampling 345
Fading channels 88
Fault detection 373
Fault diagnosis 458
Feedback control 522
Feller transition kernel 19
Feynman-Kac formulae 46
Filtering distribution 5
Finance 186, 197, 283, 375

Financial options 375
Fixed lag filtering 290
Fixed lag smoothing 95
FLIR models 534
Forecasting 197, 375
Forward looking infrared (FLIR) 534
Fresnel reflection 469
Function approximation 359

G

GAs 432
Gaussian kernel 253
Gaussian mixtures 479
Gaussian quadrature 100
Generalised dynamic linear models 177
Generalised pseudo Bayes (GBP) algorithm 479
Generalised state-space model (GSSM) 429
Genetic algorithm filter 432
Genetic algorithms 432
Gibbs sampler 92, 124, 143, 369
Grid-based filters 4
Grid-based localization 410
Guidance 517

H

Hidden Markov models (HMMs) 3
Hilbert-Schmidt bounds 538
HMM filter 3, 366
Hybrid networks 449

I

Image analysis 339
Image processing 533
Image tracking 339
IMM 269, 479, 483
Importance sampling 8, 82, 144
Importance sampling distribution 8
Importance Sampling/Resampling 117
Importance weights 8, 83

Imputation 228
 Industrial process monitoring 465
 Information state 519
 In-situ ellipsometry 465
 Interacting multiple model (IMM)
 algorithm 269, 479
 Interacting particle approximations 4
 Interacting particle filter 250
 Interacting particle system 58
 Ion mass spectroscopy 465

J

Jump-diffusion processes 537
 Jump-diffusion sampling 539
 Junction tree algorithm 366, 510

K

Kalman filter 3, 236, 479
 Kernel density estimation 298, 487
 Kernel filter 248
 Kernel methods 162, 247
 Kernel sampler 302
 Kernel smoothing 203, 253
 Kidnapped robot problem 401

L

Lagrange 263
 Langevin's stochastic differential
 equation 537
 Layered sampling 355
 Learning dynamics 349
 Lebesgue measure 44
 Lie groups 535
 Likelihood 98, 159, 189, 431
 Linear-Quadratic-Gaussian (LQG)
 517
 Log-concave densities 284
 LQG 517

M

M-ary modulation 88

Malmquist theorem 251
 Manoeuvring targets 479
 Map learning 499
 Maps 329
 Marine diesel engines 373
 Markov chain Monte Carlo (MCMC)
 8, 85, 139, 208, 274, 369, 538
 Markov chains 8, 18, 43, 138, 248
 Markov process 5, 19
 Markovian switching 486
 Maximum likelihood 159, 350
 MCMC 8, 85, 139, 208, 274, 369, 538
 Metropolis-Hastings 124, 143, 369
 Minimum variance sampling 84
 Missile systems 517
 Mixed discrete-continuous
 state-spaces 348
 Mixture densities 383
 Mixture Kalman filter (MKF) 236
 Mixture of Gaussians 94
 Mixture proposal distributions 414
 Mixtures 296, 369, 383, 517
 Mixtures of normals 287
 MKF 236
 Mobile robot localization 401
 Mobile robots 508
 Model averaging 117, 362, 381, 539
 Model selection 117, 362, 381, 471,
 479, 539
 Model uncertainty 117, 126, 381
 Molecular simulation 225
 Monitoring freeway traffic 457
 Monte Carlo EM (MCEM) 167
 Monte Carlo filtering 177, 430
 Monte Carlo Localization 401
 Monte Carlo particle filters 97
 Monte Carlo Root Mean Square error
 327
 Monte Carlo smoothing 177
 Motion 339
 Multi-layer perceptrons 359
 Multi-robot localization 401
 Multinomial branching 28
 Multinomial sampling 84
 Multiple imputation 228
 Multiple-model bootstrap filter 479
 Multiple model particle filter
 (MMPF) 269
 Multiple models 471, 479

Mutation 49

N

Navigation 329, 401, 499

Neural networks 359

Noisy fading channels 88

Nonlinear filtering problem 43

Numerical integration 336

O

Object detection 533

Object localisation 355

Object recognition 533

Occlusion 344

On-line classification 371

One step ahead prediction 182

Open-loop feedback control 524

Optics 469

Optimal control 517

Optimal estimation 322

Optimal filtering 4, 79, 97

Optimal importance sampling 335

Optimisation 432

Options' pricing 285

Outliers 290, 362

P

Parallel processing 432

Parameter estimation 81, 159, 189,
197, 362, 381

Partially smooth bootstrap 303

Particle filtering 4, 79, 99, 119, 139,
249, 274, 306

Particle smoothing 306

Partitioned sampling 355

Pattern theory 339, 533

PCA 341

Phase shift keyed (PSK) signals 88

Phased array radar 488

Photopolarimeters 442

Perfect Monte Carlo sampling 7

Polymers 225

Point-mass filter 336

Population-based optimisation 432

Population biology 98

Position tracking 401

Posterior Cramér-Rao bound 322

Posterior distribution 5

Potential energy 225

Power spectra 438

Prediction 182

Prediction density 274

Principal Components Analysis
(PCA) 341

Probabilistic dynamic system 229

Proposal distribution 8, 124

Q

Quadrature filtering 101, 197

R

Radar 488, 525

Radial bases 359

Radon-Nikodym derivative 83

Random probability measure 19

Rao-Blackwellisation 91, 117, 146,
237, 366, 394, 499

Rayleigh fading channels 88

Real-time processing 465

Real-time spectroscopic ellipsometry
465

Recursive Bayesian inference 295

Reduced random sampling 289

Regression 359

Regularization 247

Regularised particle filters 251

Reinforcement learning 518

Rejection control 231

Rejection sampling 231, 256, 276

Resample-move algorithm 117

Resampling 10, 84, 230, 312, 368, 435

Residual PCA 341

Residual resampling 84

Reversible jump MCMC 128, 369

Robot localization 401, 508

Robot motion and perception 404

Robotics 401, 499

Roughening 305

S

Sample augmentation 305
 Sampling/importance resampling
 (SIR) 148, 274
 Scene analysis 533
 Selection 10, 49, 84, 368
 Self-avoid (random) walk (SAW) 225
 Self-organizing state-space models
 177
 Self-organizing time series models 429
 Semiconductor design 465
 Semiconductors 465
 Sequential estimation 79
 Sequential importance sampling (SIS)
 9, 225
 Sequential Monte Carlo (SMC) 4
 Sequential Monte Carlo methods 82
 Shape and motion 339
 Signal detection 95, 381
 SIR 148, 275, 390
 Small count data 438
 Smooth bootstrap 296
 Smoothing 139, 161, 177, 296, 430
 Softmax CPD 449
 Speech processing 139
 Splines 344
 Spurious objects 525
 Statistical modelling of shape 340
 Stochastic control 518
 Stochastic differential equations 71
 Stochastic EM 168
 Stochastic volatility models 186, 208,
 284
 Strong law of large numbers 8, 297
 Structured probabilistic models 448
 Survival of the fittest algorithm 4
 Switching matrix 486
 Switching models 486
 Switching state-space models 471

T

Target recognition 533
 Target tracking 239, 264, 467, 479,
 517
 Templates 533
 Terrain navigation 329

Time series analysis 177, 194, 375,
 438
 Time-varying coefficient AR model
 438
 Time-varying frequencies 439
 Time-varying autoregressive (TVAR)
 models 145
 Tracking 129, 264, 339, 479, 517, 533
 Tracking range and bearing 264
 Traffic monitoring 447
 Transition kernel 18, 85, 121,
 Tree based branching 28
 Trend analysis 194
 TVAR models 145
 Two tank network 458, 503

U

Unknown variances 311

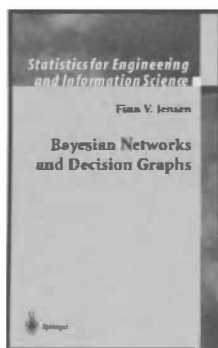
V

Volatility 186, 197, 208, 284
 Voyager 442
 Voyager PPS data 442

Y

Yule Walker equations 350

ALSO AVAILABLE FROM SPRINGER!



FINN V. JENSEN

BAYESIAN NETWORKS AND DECISION GRAPHS

Bayesian networks and decision graphs are formal graphical languages for representation and communication of decision scenarios requiring reasoning under uncertainty. The book emphasizes both the human and the computer side. Part I gives a thorough introduction to Bayesian networks as well as decision trees and influence diagrams, and through examples and exercises, the reader is instructed in building graphical models from domain knowledge. This part is self-contained. Part II is devoted to the presentation of algorithms and complexity issues.

2001/280 PAGES/HARDCOVER

ISBN 0-387-95259-4

STATISTICS FOR ENGINEERING AND INFORMATION SCIENCE

CHRISTIAN P. ROBERT

THE BAYESIAN CHOICE

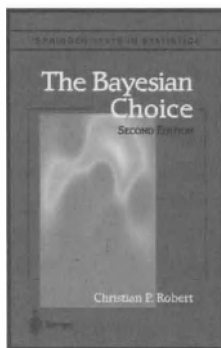
*From Decision-Theoretic Foundations
to Computational Implementation*
Second Edition

This book introduces Bayesian statistics and decision theory. Its scope covers both the basic ideas of statistical theory, and also some of the more modern and advanced topics of Bayesian statistics such as complete class theorems, the Stein effect, Bayesian model choice, hierarchical and empirical Bayes modeling, Monte Carlo integration including Gibbs sampling, and other MCMC techniques.

2001/624 PAGES/HARDCOVER

ISBN 0-387-95231-4

SPRINGER TEXTS IN STATISTICS



JUN S. LIU

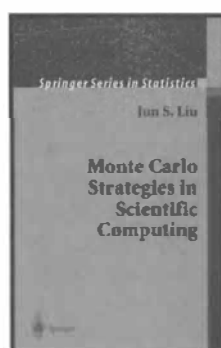
MONTE CARLO STRATEGIES IN SCIENTIFIC COMPUTING

This book provides a self-contained and up-to-date treatment of the Monte Carlo method and develops a common framework under which various Monte Carlo techniques can be "standardized" and compared. The Monte Carlo method is a computer-based statistical sampling approach for solving numerical problems concerned with a complex system. Given the interdisciplinary nature of the topics and a moderate prerequisite for the reader, this book should be of interest to a broad audience of quantitative researchers such as computational biologists, computer scientists, econometricians, engineers, probabilists, and statisticians.

2001/352 PAGES/HARDCOVER

ISBN 0-387-95230-6

SPRINGER SERIES IN STATISTICS



To Order or for Information:

In the Americas: CALL: 1-800-SPRINGER or FAX: (201) 348-4505 • WRITE: Springer-Verlag New York, Inc., Dept. S2569, P.O. Box 2485, Secaucus, NJ 07096-2485 • VISIT: Your local technical bookstore • E-MAIL: orders@springer-ny.com

Outside the Americas: CALL: +49 (0) 6221 345-217/8 • FAX: +49 (0) 6221 345-229 • WRITE: Springer Customer Service, Haberstrasse 7, 69126, Heidelberg, Germany • E-MAIL: orders@springer.de

PROMOTION: S2569



Springer

www.springer-ny.com