

Ulf Grenander and Michael Miller

Pattern Theory

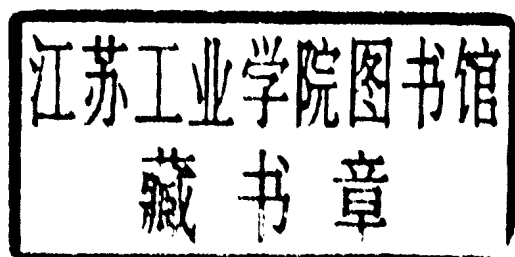
From Representation to Inference



OXFORD

PATTERN THEORY: FROM REPRESENTATION TO INFERENCE

Ulf Grenander and Michael I. Miller



OXFORD
UNIVERSITY PRESS

OXFORD

UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

© Oxford University Press, 2007

The moral rights of the authors have been asserted
Database right Oxford University Press (maker)

First published 2007

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
without the prior permission in writing of Oxford University Press,
or as expressly permitted by law, or under terms agreed with the appropriate
reprographics rights organization. Enquiries concerning reproduction
outside the scope of the above should be sent to the Rights Department,
Oxford University Press, at the address above

You must not circulate this book in any other binding or cover
and you must impose the same condition on any acquirer

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloging in Publication Data
Data available

Typeset by Newgen Imaging Systems (P) Ltd., Chennai, India
Printed in Great Britain
on acid-free paper by
CPI Bath

ISBN 0-19-850570-1 978-0-19-850570-9
ISBN 0-19-929706-1 978-0-19-929706-1

1 3 5 7 9 10 8 6 4 2

PATTERN THEORY

CONTENTS

1	Introduction	1
1.1	Organization	3
2	The Bayes Paradigm, Estimation and Information Measures	5
2.1	Bayes Posterior Distribution	5
2.1.1	Minimum Risk Estimation	6
2.1.2	Information Measures	7
2.2	Mathematical Preliminaries	8
2.2.1	Probability Spaces, Random Variables, Distributions, Densities, and Expectation	8
2.2.2	Transformations of Variables	10
2.2.3	The Multivariate Normal Distribution	10
2.2.4	Characteristic Function	11
2.3	Minimum Risk Hypothesis Testing on Discrete Spaces	12
2.3.1	Minimum Probability of Error via Maximum A Posteriori Hypothesis Testing	13
2.3.2	Neyman–Pearson and the Optimality of the Likelihood Ratio Test	14
2.4	Minimum Mean-Squared Error Risk Estimation in Vector Spaces	16
2.4.1	Normed Linear and Hilbert Spaces	17
2.4.2	Least-Squares Estimation	20
2.4.3	Conditional Mean Estimation and Gaussian Processes	22
2.5	The Fisher Information of Estimators	24
2.6	Maximum-Likelihood and its consistency	26
2.6.1	Consistency via Uniform Convergence of Empirical Log-likelihood	27
2.6.2	Asymptotic Normality and \sqrt{n} Convergence Rate of the MLE	28
2.7	Complete–Incomplete Data Problems and the EM Algorithm	30
2.8	Hypothesis Testing and Model Complexity	38
2.8.1	Model-Order Estimation and the $d/2$ log Sample-Size Complexity	38
2.8.2	The Gaussian Case is Special	41
2.8.3	Model Complexity and the Gaussian Case	42
2.9	Building Probability Models via the Principle of Maximum Entropy	43
2.9.1	Principle of Maximum Entropy	44
2.9.2	Maximum Entropy Models	45
2.9.3	Conditional Distributions are Maximum Entropy	47
3	Probabilistic Directed Acyclic Graphs and Their Entropies	49
3.1	Directed Acyclic Graphs (DAGs)	49
3.2	Probabilities on Directed Acyclic Graphs (PDAGs)	51
3.3	Finite State Markov Chains	54
3.4	Multi-type Branching Processes	56
3.4.1	The Branching Matrix	59
3.4.2	The Moment-Generating Function	60
3.5	Extinction for Finite-State Markov Chains and Branching Processes	62
3.5.1	Extinction in Markov Chains	62
3.5.2	Extinction in Branching Processes	63
3.6	Entropies of Directed Acyclic Graphs	64
3.7	Combinatorics of Independent, Identically Distributed Strings via the Asymptotic Equipartition Theorem	65

3.8	Entropy and Combinatorics of Markov Chains	66
3.9	Entropies of Branching Processes	68
3.9.1	Tree Structure of Multi-Type Branching Processes	69
3.9.2	Entropies of Sub-Critical, Critical, and Super-Critical Processes	70
3.9.3	Typical Trees and the Equipartition Theorem	71
3.10	Formal Languages and Stochastic Grammars	74
3.11	DAGs for Natural Language Modelling	81
3.11.1	Markov Chains and m -Grams	81
3.11.2	Context-Free Models	82
3.11.3	Hierarchical Directed Acyclic Graph Model	84
3.12	EM Algorithms for Parameter Estimation in Hidden Markov Models	87
3.12.1	MAP Decoding of the Hidden State Sequence	88
3.12.2	ML Estimation of HMM parameters via EM Forward/Backward Algorithm	89
3.13	EM Algorithms for Parameter Estimation in Natural Language Models	92
3.13.1	EM Algorithm for Context-Free Chomsky Normal Form	93
3.13.2	General Context-Free Grammars and the Trellis Algorithm of Kupiec	94
4	Markov Random Fields on Undirected Graphs	95
4.1	Undirected Graphs	95
4.2	Markov Random Fields	96
4.3	Gibbs Random Fields	101
4.4	The Splitting Property of Gibbs Distributions	104
4.5	Bayesian Texture Segmentation: The log-Normalizer Problem	110
4.5.1	The Gibbs Partition Function Problem	110
4.6	Maximum-Entropy Texture Representation	112
4.6.1	Empirical Maximum Entropy Texture Coding	113
4.7	Stationary Gibbs Random Fields	116
4.7.1	The Dobrushin/Lanford/Ruelle Definition	116
4.7.2	Gibbs Distributions Exhibit Multiple Laws with the Same Interactions (Phase Transitions): The Ising Model at Low Temperature	117
4.8	1D Random Fields are Markov Chains	119
4.9	Markov Chains Have a Unique Gibbs Distribution	120
4.10	Entropy of Stationary Gibbs Fields	121
5	Gaussian Random Fields on Undirected Graphs	123
5.1	Gaussian Random Fields	123
5.2	Difference Operators and Adjoints	124
5.3	Gaussian Fields Induced via Difference Operators	126
5.4	Stationary Gaussian Processes on \mathbb{Z}^d and their Spectrum	133
5.5	Cyclo-Stationary Gaussian Processes and their Spectrum	134
5.6	The log-Determinant Covariance and the Asymptotic Normalizer	137
5.6.1	Asymptotics of the Gaussian processes and their Covariance	138
5.6.2	The Asymptotic Covariance and log-Normalizer	142
5.7	The Entropy Rates of the Stationary Process	142
5.7.1	Burg's Maximum Entropy Auto-regressive Processes on \mathbb{Z}^d	143
5.8	Generalized Auto-Regressive Image Modelling via Maximum-Likelihood Estimation	144
5.8.1	Anisotropic Textures	147

6	The Canonical Representations of General Pattern Theory	154
6.1	The Generators, Configurations, and Regularity of Patterns	154
6.2	The Generators of Formal Languages and Grammars	158
6.3	Graph Transformations	162
6.4	The Canonical Representation of Patterns: DAGs, MRFs, Gaussian Random Fields	166
6.4.1	Directed Acyclic Graphs	167
6.4.2	Markov Random Fields	169
6.4.3	Gaussian Random Fields: Generators induced via difference operators	170
7	Matrix Group Actions Transforming Patterns	174
7.1	Groups Transforming Configurations	174
7.1.1	Similarity Groups	174
7.1.2	Group Actions Defining Equivalence	175
7.1.3	Groups Actions on Generators and Deformable Templates	177
7.2	The Matrix Groups	177
7.2.1	Linear Matrix and Affine Groups of Transformation	177
7.2.2	Matrix groups acting on \mathbb{R}^d	179
7.3	Transformations Constructed from Products of Groups	181
7.4	Random Regularity on the Similarities	184
7.5	Curves as Submanifolds and the Frenet Frame	190
7.6	2D Surfaces in \mathbb{R}^3 and the Shape Operator	195
7.6.1	The Shape Operator	196
7.7	Fitting Quadratic Charts and Curvatures on Surfaces	198
7.7.1	Gaussian and Mean Curvature	198
7.7.2	Second Order Quadratic Charts	200
7.7.3	Isosurface Algorithm	201
7.8	Ridge Curves and Crest Lines	205
7.8.1	Definition of Sulcus, Gyrus, and Geodesic Curves on Triangulated Graphs	205
7.8.2	Dynamic Programming	207
7.9	Bijections and Smooth Mappings for Coordinatizing Manifolds via Local Coordinates	210
8	Manifolds, Active Models, and Deformable Templates	214
8.1	Manifolds as Generators, Tangent Spaces, and Vector Fields	214
8.1.1	Manifolds	214
8.1.2	Tangent Spaces	215
8.1.3	Vector Fields on M	217
8.1.4	Curves and the Tangent Space	218
8.2	Smooth Mappings, the Jacobian, and Diffeomorphisms	219
8.2.1	Smooth Mappings and the Jacobian	219
8.2.2	The Jacobian and Local Diffeomorphic Properties	221
8.3	Matrix Groups are Diffeomorphisms which are a Smooth Manifold	222
8.3.1	Diffeomorphisms	222
8.3.2	Matrix Group Actions are Diffeomorphisms on the Background Space	223
8.3.3	The Matrix Groups are Smooth Manifolds (Lie Groups)	224
8.4	Active Models and Deformable Templates as Immersions	226
8.4.1	Snakes and Active Contours	226
8.4.2	Deforming Closed Contours in the Plane	226
8.4.3	Normal Deformable Surfaces	227
8.5	Activating Shapes in Deformable Models	229
8.5.1	Likelihood of Shapes Partitioning Image	229
8.5.2	A General Calculus for Shape Activation	229

8.5.3	Active Closed Contours in \mathbb{R}^2	232
8.5.4	Active Unclosed Snakes and Roads	234
8.5.5	Normal Deformation of Circles and Spheres	236
8.5.6	Active Deformable Spheres	236
8.6	Level Set Active Contour Models	237
8.7	Gaussian Random Field Models for Active Shapes	240
9	Second Order and Gaussian Fields	244
9.1	Second Order Processes (SOP) and the Hilbert Space of Random Variables	244
9.1.1	Measurability, Separability, Continuity	244
9.1.2	Hilbert space of random variables	247
9.1.3	Covariance and Second Order Properties	249
9.1.4	Quadratic Mean Continuity and Integration	251
9.2	Orthogonal Process Representations on Bounded Domains	252
9.2.1	Compact Operators and Covariances	253
9.2.2	Orthogonal Representations for Random Processes and Fields	257
9.2.3	Stationary Periodic Processes and Fields on Bounded Domains	258
9.3	Gaussian Fields on the Continuum	262
9.4	Sobolev Spaces, Green's Functions, and Reproducing Kernel Hilbert Spaces	264
9.4.1	Reproducing Kernel Hilbert Spaces	265
9.4.2	Sobolev Normed Spaces	266
9.4.3	Relation to Green's Functions	267
9.4.4	Gradient and Laplacian Induced Green's Kernels	267
9.5	Gaussian Processes Induced via Linear Differential Operators	271
9.6	Gaussian Fields in the Unit Cube	274
9.6.1	Maximum Likelihood Estimation of the Fields: Generalized ARMA Modelling	278
9.6.2	Small Deformation Vector Fields Models in the Plane and Cube	280
9.7	Discrete Lattices and Reachability of Cyclo-Stationary Spectra	283
9.8	Stationary Processes on the Sphere	285
9.8.1	Laplacian Operator Induced Gaussian Fields on the Sphere	289
9.9	Gaussian Random Fields on an Arbitrary Smooth Surface	293
9.9.1	Laplace-Beltrami Operator with Neumann Boundary Conditions	293
9.9.2	Smoothing an Arbitrary Function on Manifolds by Orthonormal Bases of the Laplace-Beltrami Operator	297
9.10	Sample Path Properties and Continuity	299
9.11	Gaussian Random Fields as Prior Distributions in Point Process Image Reconstruction	303
9.11.1	The Need for Regularization in Image Reconstruction	304
9.11.2	Smoothness and Gaussian Priors	304
9.11.3	Good's Roughness as a Gaussian Prior	305
9.11.4	Exponential Spline Smoothing via Good's Roughness	306
9.12	Non-Compact Operators and Orthogonal Representations	309
9.12.1	Cramer Decomposition for Stationary Processes	311
9.12.2	Orthogonal Scale Representation	312
10	Metrics Spaces for the Matrix Groups	316
10.1	Riemannian Manifolds as Metric Spaces	316
10.1.1	Metric Spaces and Smooth Manifolds	316
10.1.2	Riemannian Manifold, Geodesic Metric, and Minimum Energy	317
10.2	Vector Spaces as Metric Spaces	319
10.3	Coordinate Frames on the Matrix Groups and the Exponential Map	320

10.3.1	Left and Right Group Action	320
10.3.2	The Coordinate Frames	321
10.3.3	Local Optimization via Directional Derivatives and the Exponential Map	323
10.4	Metric Space Structure for the Linear Matrix Groups	324
10.4.1	Geodesics in the Matrix Groups	324
10.5	Conservation of Momentum and Geodesic Evolution of the Matrix Groups via the Tangent at the Identity	326
10.6	Metrics in the Matrix Groups	327
10.7	Viewing the Matrix Groups in Extrinsic Euclidean Coordinates	329
10.7.1	The Frobenius Metric	329
10.7.2	Comparing intrinsic and extrinsic metrics in $SO(2,3)$	330
11	Metrics Spaces for the Infinite Dimensional Diffeomorphisms	332
11.1	Lagrangian and Eulerian Generation of Diffeomorphisms	332
11.1.1	On Conditions for Generating Flows of Diffeomorphisms	333
11.1.2	Modeling via Differential Operators and the Reproducing Kernel Hilbert Space	335
11.2	The Metric on the Space of Diffeomorphisms	336
11.3	Momentum Conservation for Geodesics	338
11.4	Conservation of Momentum for Diffeomorphism Splines Specified on Sparse Landmark Points	340
11.4.1	An ODE for Diffeomorphic Landmark Mapping	343
12	Metrics on Photometric and Geometric Deformable Templates	346
12.1	Metrics on Dense Deformable Templates: Geometric Groups Acting on Images	346
12.1.1	Group Actions on the Images	346
12.1.2	Invariant Metric Distances	347
12.2	The Diffeomorphism Metric for the Image Orbit	349
12.3	Normal Momentum Motion for Geodesic Connection Via Inexact Matching	350
12.4	Normal Momentum Motion for Temporal Sequences	354
12.5	Metric Distances Between Orbits Defined Through Invariance of the Metric	356
12.6	Finite Dimensional Landmarked Shape Spaces	357
12.6.1	The Euclidean Metric	357
12.6.2	Kendall's Similitude Invariant Distance	359
12.7	The Diffeomorphism Metric and Diffeomorphism Splines on Landmark Shapes	361
12.7.1	Small Deformation Splines	361
12.8	The Deformable Template: Orbits of Photometric and Geometric Variation	365
12.8.1	Metric Spaces for Photometric Variability	365
12.8.2	The Metrics Induced via Photometric and Geometric Flow	366
12.9	The Euler Equations for Photometric and Geometric Variation	369
12.10	Metrics between Orbits of the Special Euclidean Group	373
12.11	The Matrix Groups (Euclidean and Affine Motions)	374
12.11.1	Computing the Affine Motions	376
13	Estimation Bounds for Automated Object Recognition	378
13.1	The Communications Model for Image Transmission	378
13.1.1	The Source Model: Objects Under Matrix Group Actions	379
13.1.2	The Sensing Models: Projective Transformations in Noise	379
13.1.3	The Likelihood and Posterior	379

13.2	Conditional Mean Minimum Risk Estimation	381
13.2.1	Metrics (Risk) on the Matrix Groups	381
13.2.2	Conditional Mean Minimum Risk Estimators	382
13.2.3	Computation of the HSE for $SE(2,3)$	384
13.2.4	Discrete integration on $SO(3)$	385
13.3	MMSE Estimators for Projective Imagery Models	385
13.3.1	3D to 2D Projections in Gaussian Noise	385
13.3.2	3D to 2D Synthetic Aperture Radar Imaging	389
13.3.3	3D to 2D LADAR Imaging	392
13.3.4	3D to 2D Poisson Projection Model	393
13.3.5	3D to 1D Projections	395
13.3.6	3D(2D) to 3D(2D) Medical Imaging Registration	397
13.4	Parameter Estimation and Fisher Information	398
13.5	Bayesian Fusion of Information	402
13.6	Asymptotic Consistency of Inference and Symmetry Groups	405
13.6.1	Consistency	405
13.6.2	Symmetry Groups and Sensor Symmetry	406
13.7	Hypothesis Testing and Asymptotic Error-Exponents	407
13.7.1	Analytical Representations of the Error Probabilities and the Bayesian Information Criterion	408
13.7.2	m-ary Multiple Hypotheses	412
14	Estimation on Metric Spaces with Photometric Variation	414
14.1	The Deformable Template: Orbits of Signature and Geometric Variation	414
14.1.1	The Robust Deformable Templates	414
14.1.2	The Metric Space of the Robust Deformable Template	415
14.2	Empirical Covariance of Photometric Variability via Principle Components	416
14.2.1	Signatures as a Gaussian Random Field Constructed from Principle Components	417
14.2.2	Algorithm for Empirical Construction of Bases	418
14.3	Estimation of Parameters on the Conditionally Gaussian Random Field Models	422
14.4	Estimation of Pose by Integrating Out EigenSignatures	424
14.4.1	Bayes Integration	427
14.5	Multiple Modality Signature Registration	429
14.6	Models for Clutter: The Transported Generator Model	431
14.6.1	Characteristic Functions and Cumulants	432
14.7	Robust Deformable Templates for Natural Clutter	438
14.7.1	The Euclidean Metric	439
14.7.2	Metric Space Norms for Clutter	439
14.7.3	Computational Scheme	442
14.7.4	Empirical Construction of the Metric from Rendered Images	444
14.8	Target detection/identification in EO imagery	445
15	Information Bounds for Automated Object Recognition	447
15.1	Mutual Information for Sensor Systems	447
15.1.1	Quantifying Multiple-Sensor Information Gain Via Mutual Information	447
15.1.2	Quantifying Information Loss with Model Uncertainty	449
15.1.3	Asymptotic Approximation of Information Measures	452

15.2	Rate-Distortion Theory	456
15.2.1	The Rate-Distortion Problem	456
15.3	The Blahut Algorithm	457
15.4	The Remote Rate Distortion Problem	459
15.4.1	Blahut Algorithm extended	460
15.5	Output Symbol Distribution	465
16	Computational Anatomy: Shape, Growth and Atrophy Comparison via Diffeomorphisms	468
16.1	Computational Anatomy	468
16.1.1	Diffeomorphic Study of Anatomical Submanifolds	469
16.2	The Anatomical Source Model of CA	470
16.2.1	Group Actions for the Anatomical Source Model	472
16.2.2	The Data Channel Model	473
16.3	Normal Momentum Motion for Large Deformation Metric Mapping (LDDMM) for Growth and Atrophy	474
16.4	Christensen Non-Geodesic Mapping Algorithm	478
16.5	Extrinsic Mapping of Surface and Volume Submanifolds	480
16.5.1	Diffeomorphic Mapping of the Face	481
16.5.2	Diffeomorphic Mapping of Brain Submanifolds	481
16.5.3	Extrinsic Mapping of Subvolumes for Automated Segmentation	481
16.5.4	Metric Mapping of Cortical Atlases	483
16.6	Heart Mapping and Diffusion Tensor Magnetic Resonance Imaging	484
16.7	Vector Fields for Growth	488
16.7.1	Growth from Landmarked Shape Spaces	488
17	Computational Anatomy: Hypothesis Testing on Disease	494
17.1	Statistics Analysis for Shape Spaces	494
17.2	Gaussian Random Fields	495
17.2.1	Empirical Estimation of Random Variables	496
17.3	Shape Representation of the Anatomical Orbit Under Large Deformation Diffeomorphisms	496
17.3.1	Principal Component Selection of the Basis from Empirical Observations	497
17.4	The Momentum of Landmarked Shape Spaces	498
17.4.1	Geodesic evolution equations for landmarks	498
17.4.2	Small Deformation PCA Versus Large Deformation PCA	499
17.5	The Small Deformation Setting	502
17.6	Small Deformation Gaussian Fields on Surface Submanifolds	502
17.7	Disease Testing of Automorphic Pathology	503
17.7.1	Hypothesis Testing on Disease in the Small Noise Limit	503
17.7.2	Statistical Testing	505
17.8	Distribution Free Testing	510
17.9	Heteromorphic Tumors	511
18	Markov Processes and Random Sampling	514
18.1	Markov Jump Processes	514
18.1.1	Jump Processes	515
18.2	Random Sampling and Stochastic Inference	516
18.2.1	Stationary or Invariant Measures	517
18.2.2	Generator for Markov Jump Processes	519
18.2.3	Jump Process Simulation	520
18.2.4	Metropolis–Hastings Algorithm	521

18.3	Diffusion Processes for Simulation	523
18.3.1	Generators of 1D Diffusions	525
18.3.2	Diffusions and SDEs for Sampling	527
18.4	Jump-Diffusion Inference on Countable Unions of Spaces	528
18.4.1	The Basic Problem	529
19	Jump Diffusion Inference in Complex Scenes	532
19.1	Recognition of Ground Vehicles	533
19.1.1	CAD Models and the Parameter Space	533
19.1.2	The FLIR Sensor Model	534
19.2	Jump Diffusion for Sampling the Target Recognition Posterior	536
19.2.1	The Posterior distribution	536
19.2.2	The Jump Diffusion Algorithms	536
19.2.3	Jumps via Gibbs' Sampling	539
19.2.4	Jumps via Metropolis–Hastings Acceptance/Rejection	541
19.3	Experimental Results for FLIR and LADAR	543
19.3.1	Detection and Removal of Objects	543
19.3.2	Identification	543
19.3.3	Pose and Identification	544
19.3.4	Identification and recognition via High Resolution Radar (HRR)	546
19.3.5	The Dynamics of Pose Estimation via the Jump–Diffusion Process	546
19.3.6	LADAR Recognition	548
19.4	Powerful Prior Dynamics for Airplane Tracking	549
19.4.1	The Euler-Equations Inducing the Prior on Airplane Dynamics	550
19.4.2	Detection of Airframes	552
19.4.3	Pruning via the Prior distribution	552
19.5	Deformable Organelles: Mitochondria and Membranes	553
19.5.1	The Parameter Space for Contour Models	553
19.5.2	Stationary Gaussian Contour Model	554
19.5.3	The Electron Micrograph Data Model: Conditional Gaussian Random Fields	555
19.6	Jump–Diffusion for Mitochondria	556
19.6.1	The jump parameters	557
19.6.2	Computing gradients for the drifts	557
19.6.3	Jump Diffusion for Mitochondria Detection and Deformation	558
19.6.4	Pseudolikelihood for Deformation	560
	References	563
	Index	581

1 INTRODUCTION

This book is to be an accessible book on patterns, their representation, and inference. There are a small number of ideas and techniques that, when mastered, make the subject more accessible. This book has arisen from ten years of a research program which the authors have embarked upon, building on the more abstract developments of metric pattern theory developed by one of the authors during the 1970s and 1980s. The material has been taught over multiple semesters as part of a second year graduate-level course in pattern theory, essentially an introduction for students interested in the representation of patterns which are observed in the natural world. The course has attracted students studying biomedical engineering, computer science, electrical engineering, and applied mathematics interested in speech recognition and computational linguistics, as well as areas of image analysis, and computer vision.

Now the concept of *patterns* pervades the history of intellectual endeavor; it is one of the eternal followers in human thought. It appears again and again in science, taking on different forms in the various disciplines, and made rigorous through mathematical formalization. But the concept also lives in a less stringent form in the humanities, in novels and plays, even in everyday language. We use it all the time without attributing a formal meaning to it and yet with little risk of misunderstanding. So, what do we really mean by a pattern? Can we define it in strictly logical terms? And if we can, what use can we make of such a definition?

These questions were answered by *General Pattern Theory*, a discipline initiated by Ulf Grenander in the late 1960s [1–5]. It has been an ambitious effort with the only original sketchy program having few if any practical applications, growing in mathematical maturity with a multitude of applications having appeared in biology/medicine and in computer vision, in language theory and object recognition, to mention but a few. Pattern theory attempts to provide an algebraic framework for describing patterns as structures regulated by rules, essentially a finite number of both local and global combinatory operations. Pattern theory takes a *compositional* view of the world, building more and more complex structures starting from simple ones. The basic rules for combining and building complex patterns from simpler ones are encoded via graphs and rules on transformation of these graphs.

In contrast to other dominating scientific themes, in Pattern Theory we start from the belief that *real world patterns are complex*: Galilean simplification that has been so successful in physics and other natural sciences will not suffice when it comes to explaining other regularities, for example in the life sciences. If one accepts this belief it follows that complexity must be allowed in the ensuing representations of knowledge. For this, probabilities naturally enter, superimposed on the graphs so as to express the variability of the real world by describing its fluctuations as randomness. Take as a goal the development of algorithms which assist in the ambitious task of *image understanding* or *recognition*. Imagine an expert studying a natural scene, trying to understand it in terms of the awesome body of knowledge that is informally available to humans about the context of the scene: identify components, relate them to each other, make statements about the fine structure as well as the overall appearance. If it is truly the goal to create algorithmic tools which assist experts in carrying out the time-consuming labor of pattern analysis, while leaving the final decision to their judgment, to arrive at more than ad hoc algorithms the subject matter knowledge must be expressed precisely and as compactly as possible.

This is the central focus of the book: ‘How can such empirical knowledge be represented in mathematical form, including both structure and the all important variability?’ This task of presenting an organized and coherent view of the field of Pattern theory seems bewildering at best. But what are today’s challenges in signal, data and pattern analysis? With the advent of

geometric increases in computational and storage resources, there has been a dramatic increase in the solution of highly complex pattern representation and recognition problems. Historically books on pattern recognition present a diverse set of problems with diverse methods for building recognition algorithms, each approach handcrafted to the particular task. The complexity and diversity of patterns in the world presents one of the most significant challenges to the pedagogical approach to the teaching of Pattern theory. Real world patterns are often the results of evolutionary change, and most times cannot be selected by the practitioner to have particular properties. The representations require models using mathematics which span multiple fields in *algebra, geometry, statistics and statistical communications*.

Contrasting this to the now classical field of statistical communications, it might appear that the task seems orders of magnitude bigger than modelling signal ensembles in the communication environment. Thinking historically of the now classical field of statistical communications, the discipline can be traced back far, to Helmholtz and earlier, but here we are thinking of its history in the twentieth century. For the latter a small number of parameters may be needed, means, covariances, for Gaussian noise, or the spectral density of a signal source, and so on. The development of communication engineering from the 1920s on consisted in part of formalizing the observed, more or less noisy, signals. Statistical signal processing is of course one of the great success stories of statistics/engineering. It is natural to ask why. We believe that it was because the pioneers in the field managed to construct representations of signal ensembles, models that were realistic and at the same time tractable both analytically and computationally (by analog devices at the time). The classical signalling models: choose $s_0(t), s_1(t)$ to be orthogonal elements in L^2 , with the noise model additive stationary noise with covariance representation via a complete orthonormal basis. Such a beautiful story, utilizing ideas from Fourier analysis, stationary stochastic processes, Toeplitz forms, and Bayesian inference! Eventually this resulted in more or less automated procedures for the detection and understanding of noisy signals: matched filters, optimal detectors, and the like. Today these models are familiar, they look simple and natural, but in a historical perspective the phenomena must have appeared highly complex and bewildering.

We believe the same to be true for pattern theory. The central challenge is the formalization of a small set of ideas for constructing the *representations of the patterns themselves* which accommodate variability and structure simultaneously. This is the point of view from which this book is written. Even though the field of pattern theory has grown considerably over the past 30 years, we have striven to emphasize its coherence. There are essentially two overarching principles. The first is the representation of regularity via graphs which essentially encode the rules of combination which allow for the generation of complex structures from simpler ones. The second is the application of transformations to generate from the exemplars entire orbits. To represent typicality probabilistic structures are superimposed on the graphs and the rules of transformation. Naturally then the conditional probabilities encode the regularity of the patterns, and become the central tool for studying pattern formation.

We have been drawn to the field of pattern theory from backgrounds in communication theory, probability theory and statistics. The overall framework fits comfortably within the source-channel view of Shannon. The underlying deep regular structures are descriptions of the source, which are hidden via the sensing channel. We believe that the principle challenge is the representation of the source of patterns, and for that reason the majority of the book is focused precisely on this topic. A multiplicity of channels or sensor models will be used throughout the book, those appropriate for the pattern class being studied. They are however studied superficially, drawn from the engineering literature and taken as given, but certainly studied more deeply elsewhere. The channel sensor models of course shape the overall performance of the inference algorithms; but the major focus of our work is on the development of stochastic models for the structural understanding of the variabilities of the patterns at the source. This also explains the major deviation of this pattern theory from that which has come to be known as pattern recognition. Only in the final chapters will pattern recognition algorithms be studied, attempting to answer the question of how well the algorithm can estimate (recognize) the source when seen through the noisy sensor channel.

1.1 Organization

Throughout this book we use methods from estimation, stochastic processes and information theory. Chapter 2 includes the basic stalwarts of statistics and estimation theory which should be familiar to the reader, including minimum-risk estimation, Fisher Information, hypothesis testing and maximum-likelihood, consistency, model order estimation, and entropies.

Chapters 3–6 bring into central prominence the role of representation of patterns via conditioning structure. Chapter 3 examines discrete patterns represented via probabilistic directed acyclic graphs (DAGs) emphasizing the pivoting properties and conditional factorizations of DAGs which are familiar for Markov chains and random branching processes. This provides ample opportunity to construct and study the syntactic theory of Chomsky Languages via the classical formulation of graphs and grammatical transformation. Chapter 4 relaxes away from the pivoting property of directed graphs to the conditioning structure of neighborhoods in Markov random fields. Chapter 5 brings the added structure of the Gaussian law for representing real-valued patterns via Gaussian fields. In this context entropy and maximum entropy distributions are examined in these three chapters as a means of representing conditioning information for representing patterns of regularity for speech, language, and image analysis. Chapter 6 presents the abstract representation of patterns via generators and probabilistic structures on the generators. The generator representation is explored as it provides a unified way of dealing with DAGs and random fields.

Chapters 7 and 8 begin examining in their own right the second central component of pattern theory, groups of geometric transformation applied to the representation of geometric objects. The patterns and shapes are represented as submanifolds of \mathbb{R}^n , including points, curves, surfaces and subvolumes. They are enriched via the actions of the linear matrix groups, studying the patterns as orbits defined via the group actions. In this context active models and deformable templates are studied. The basic fundamentals of groups and matrix groups are explored assuming that the typical engineering graduate student will not be familiar with their structure.

Chapter 9 makes the first significant foray into probabilistic structures in the continuum, studying random processes and random fields indexed over subsets of \mathbb{R}^n . Classical topics are examined in some detail including second order processes, covariance representation, and Karhunen-Loeve transforms. This is the first chapter where more significant understanding is required for understanding signal and patterns as functions in a Hilbert space.

Chapters 10 and 11 continue the major thrust into transformations and patterns indexed over the continuum. In this context, the finite dimensional matrix groups are studied as diffeomorphic actions on \mathbb{R}^n , as well their infinite dimensional analogues are established. It is in these chapters in which the substantial bridge between Pattern theory, mechanics, and differential geometry is established. The links come through the study of flows of diffeomorphisms. Chapter 10 focuses on the study of the finite dimensional matrix groups, and Chapter 11 on the infinite dimensional diffeomorphisms acting on manifolds of \mathbb{R}^n as a Riemannian metric space. The metric is induced by the geodesic length between elements in the space defined through the Riemannian length of the flow connecting one point to another. Herein the classical equations of motion for the geodesics in the finite dimensional case are expanded to include the Euler formulation of the infinite dimensional case.

Chapter 12 expands this view to examine the orbit of imagery as a deformable template under diffeomorphic action; the orbit is endowed with the metric structure through the length minimizing geodesics connecting images. In this chapter the photometric orbit is studied as well, adding notions from transport to define a metric on the product space of geometric and photometric variation.

Chapters 13–15 extend from the pure representations of shapes to the Bayes estimation of shapes and their parametric representation. Classical minimum-mean-squared error and maximum a-posteriori estimators of shapes are explored in these chapters as viewed through various remote sensing models. Chapter 13 focuses on estimating the pose of rigid objects; chapter 14 focuses on accommodating photometric variability superimposed on the geometric variability of

rigid pose. Chapter 15 focuses on information bounds for quantifying estimation accuracy of the matrix group, comparing mean-squared error bounds with capacity and rate-distortion bounds for codebooks.

Chapters 16 and 17 turn from the estimation of finite dimensional matrix groups to the study of the estimation of infinite dimensional shape in the newly emergent field of Computational Anatomy. Chapter 15 focuses on estimating landmark and image based shape metrics in volumes, with Chapter 16 focusing on submanifolds and on the inference of disease and hypothesis testing in Computational Anatomy.

The last two Chapters 18 and 19 conclude on inference, exploring random sampling approaches for estimation of model order and parametric representing of shapes. Chapter 18 reviews jump and diffusion processes and their use in random sampling of discrete and continuum spaces. Chapter 19 examines a series of problems in object recognition.

We have made an attempt to keep the theory at a consistent level. The mathematical level is a reasonably high one, first-year graduate level, with a background of at least one good semester course in probability and a solid background in mathematics. We have, however, been able to avoid the use of measure theory.

Appendices outlining proofs, theorems and solutions to exercises together with a comprehensive list of figures, tables and plates are freely available on an accompanying website www.oup.com/uk/academic/companion/mathematics/patterntheory

In this book plates 1–16 appear between pages 180–181, plates 17–34 appear between pages 372–373, and plates 35–53 appear between 564–565.

2 THE BAYES PARADIGM, ESTIMATION AND INFORMATION MEASURES

ABSTRACT The basic paradigm is the Bayesian setup, given is the source of parameters $X \in \mathcal{X}$ which are seen through a noisy channel giving observations $Y \in \mathcal{Y}$. The posterior distribution determines the bounds on estimation of X given Y , the risk associated with estimating it, as well as a characterization of the information in the observation in Y about X .

2.1 Bayes Posterior Distribution

The basic set up throughout is we are given a model of the *source* of possible objects $X \in \mathcal{X}$. These are observed through a *noisy channel* giving observations $Y \in \mathcal{Y}$. The source $X \in \mathcal{X}$ is modelled with distribution and density $P_X(dx) = p(x)dx$, $\int_{\mathcal{X}} p(x)dx = 1$. Generally the source can only be observed with loss of information due to observational noise or limited accuracy in the sensor. The mapping from the input source $X \in \mathcal{X}$ to the observed output $Y \in \mathcal{Y}$ expresses the physics of the sensing channels; the data $Y \in \mathcal{Y}$ will in general contain multiple components corresponding to several sensors $Y = (Y_1, Y_2, \dots)$. The observation process is characterized via a statistical transition law, transferring $X \rightarrow Y$ $P_{Y|X}(\cdot|\cdot) : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}^+$, summarizing completely the transition law mapping the input model parameters X to the output Y , the likelihood of Y given X .

This Bayesian paradigm, separating the source from the channel is what has become the modern view of communications developed out of Shannon's theory of communications [6]. Figure 2.1, clearly delineates the separation of the source of information and the channel through which the messages are observed. To infer the transmitted message at the output of the channel, the observation Y must be processed optimally. The inference engine is a decoder, working to recover properties of the original message from the source.

Given such a communications *source/channel* decomposition, we shall be interested in both specifying optimal procedures for inferring properties of pattern generation systems given the observable measurements, and quantifying the information content and information gain of the observation system. Pattern deduction becomes an enterprise consisting of constructing the source and channel models, and essentially has several parts: (i) selection and fitting of parameters parametrizing the models representing the patterns, and (ii) construction of the family of probability models representing the knowledge about the pattern classes. For this we shall examine classical minimum-risk estimators, such as minimum-mean-squared-error (MMSE) estimators, maximum-a-posteriori and likelihood (MAP, MLE). For constructing the models we shall examine various forms and principles of entropy and mutual information.

At the most fundamental level, the posterior distribution represents the information contained in the observables about the underlying imagery. All provably optimal structured methods of inference and information gathering fundamentally involve the posterior density or distribution of the random variables $X \in \mathcal{X}$ given the observed deformed image $Y \in \mathcal{Y}$.

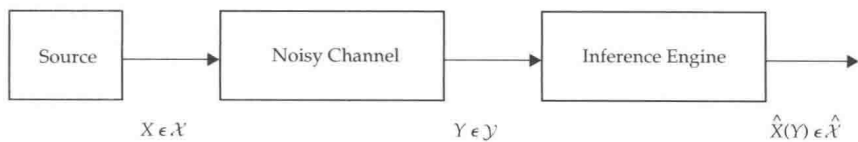


Figure 2.1 Shannon's source channel model for communications systems