Springer Series in Statistics

Series editors

Peter Bickel, CA, USA

Peter Diggle, Lancaster, UK

Stephen E. Fienberg, Pittsburgh, PA, USA

Ursula Gather, Dortmund, Germany

Ingram Olkin, Stanford, CA, USA

Scott Zeger, Baltimore, MD, USA

More information about this series at <http://www.springer.com/series/692>

Anuj Srivastava • Eric P. Klassen

Functional and Shape Data Analysis

Anuj Srivastava
Department of Statistics
Florida State University
Tallahassee, FL, USA

Eric P. Klassen
Department of Mathematics
Florida State University
Tallahassee, FL, USA

ISSN 0172-7397

ISSN 2197-568X (electronic)

Springer Series in Statistics

ISBN 978-1-4939-4018-9

ISBN 978-1-4939-4020-2 (eBook)

DOI 10.1007/978-1-4939-4020-2

Library of Congress Control Number: 2016942552

© Springer-Verlag New York 2016

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature

The registered company is Springer Science+Business Media LLC New York

1923–2016

In memory of Prof. Ulf Grenander, who led the way

Preface

Function and shape data analysis are old topics in statistics, studied off and on over the last several decades. However, the early years of the new millennium saw a renewed focus and energy in these areas. This focus was both exciting, because it sought new directions and resources, and productive, because it was application oriented and data driven. While the new interest was fueled by many factors, the most prominent amongst them was increasing availability of large datasets involving function and curve data, especially in the fields of computer vision and medical imaging. It was also propelled by increases in computational power and storage, a growing interest in Riemannian methods, and a favorable atmosphere for the confluence of ideas from geometry and statistics. Despite a long history of research and methods in function and shape analysis, several researchers took a fresh look at shape analysis during this period. As a result, they came up with novel approaches, based on mathematical tools that were new to this community, and made them practical using elegant computational solutions. This book started as a research monograph, as an exposition of these contemporary ideas in shape analysis of curves. It was heavily motivated by our desire to provide a self-contained treatment of this new generation of methods in shape analysis of curves, with a focus on statistical modeling and inference. However, like many other book projects, this project also grew beyond its original plan. Not only did it become a textbook with appendices, background material, and exercises, but also its scope grew to include a detailed treatment on function data analysis. Not surprisingly, it took much longer than expected to finish this manuscript. This delay allowed us to offer several graduate courses in statistics on the basis of this textbook.

The main topic area of this book is shape analysis of functions and curves—in one, two, and higher dimensions—both closed and open. What differentiates this material from past approaches is that it integrates the *registration problem* into shape analysis. Registration is concerned with matching of points across objects when their shapes are being compared and quantified. The past methods mostly treated registration as a pre-processing step, handled using an arbitrary off-the-shelf technique, followed by an unrelated metric for shape comparison. Instead, this textbook seeks a unified, comprehensive solution. It develops elegant Riemannian frameworks that provide both quantification of shape differences and registration of curves at the same time. Additionally, these methods are used for statistically summarizing given curve data, performing dimension reduction, and

modeling the observed variability. This textbook investigates different mathematical representations and associated (invariant) Riemannian metrics that play a role in facilitating shape analysis. The focus is largely on certain square-root representations that *flatten* shape spaces and allow more traditional vector-space-based statistical analyses to become applicable.

Our main objective in writing this book is to familiarize graduate students with a broad array of tools that are relevant in developing computational solutions for shape and related analyses. These tools, gleaned from geometry, algebra, statistics, and computational science, are traditionally scattered across different courses and departments, which makes it difficult for graduate students to learn them in a reasonable time. While we wish to introduce readers to a diverse array of topics that are fast becoming important in current research, this book is not intended to be an elaborate mathematical treatment of those topics. Naturally, one cannot expect to become an expert in such multidisciplinary areas solely on basis of a single book. Besides, there are already many wonderful books that cover these items individually in great detail. Indeed, in this textbook, we have frequently been lax in presenting in-depth technical details, focusing instead on intuitive explanations and implementable solutions. Our hope is to provide a *working knowledge* of relevant material that is often present across different disciplines and better prepare graduate students for handling future scientific challenges.

This textbook is intended for graduate students who are interested in improving their skills in this broad interdisciplinary problem area. These students can be from statistics, engineering, applied mathematics, neuroscience, biology, bioinformatics, and other related areas. It seems necessary for a reader to have a background in calculus, linear algebra, and numerical analysis and computations. Of course, one can selectively read chapters of his/her interest, and a sequential reading of the entire chapter is not a necessity. This book covers a broad range of ideas, from introductory theory to algorithmic implementations and some statistical case studies. It is self-contained in terms of the background material needed for understanding methods described here. The appendices and background material are highly recommended for students without prior coursework in geometry and algebra.

The textbook starts with a motivation for development of knowledge in this area, by citing a number of current applications that are primarily dependent on analyzing shapes of functions and curves. In particular, it makes a case of using *elastic* functional representations of these objects, rather than sampling them with points in a pre-determined manner. Chapter 2 summarizes some current techniques in shape data analysis. These techniques mostly rely on point-based representations of objects and lack a natural system for registering points across objects. In order to understand the theory presented in this textbook, a reader needs a working knowledge of some relevant parts of algebra and geometry, and that is provided in Chapter 3. Chapter 4 deals with functional data analysis, starting with some standard tools from Hilbert space theory. Then, it raises the problem of pairwise registration of functions, highlights the shortcomings of what may be considered a natural solution (based on the L^2 norm), and proposes a better solution using the SRSF (square-root slope function) representation of functions and extending the Fisher-Rao Riemannian metric to general function spaces. We take this opportunity to formally introduce the concepts of amplitude and relative phase of functional variables. This chapter ends with impositions of the Fisher-Rao metric on some spaces of interest, such as the sets of probability density functions and warping functions.

We present methods for shape analysis of planar curves in Chapter 5, with an emphasis on the elastic Riemannian metric and the square-root velocity function (SRVF) representation. This chapter describes the elastic metric and uses the SRVF mapping to convert the elastic metric into the standard $\mathbb{L}^2$ inner product. This allows for simultaneous registration of points and deformation of curves or *geodesics* using efficient numerical techniques. Chapter 6 focuses on *closed curves* by imposing an additional constraint of closure on the allowed curves. This nonlinear constraint results in representation spaces becoming nonlinear manifolds, and we introduce two general approaches—*shooting method* and *path-straightening method* for computing geodesic paths on such manifolds. In each case, we provide step-by-step algorithms and many examples to illustrate the ideas.

Chapter 7 serves as a general discussion on defining and computing summary statistics on sets that are not vector spaces. Here one uses ideas from differential geometry to specify the notions of sample means and covariances, and some associated parametric probability density functions on these sets. This is followed by tools for modeling functional data in Chapter 8, where an important idea is to separate functions into their amplitude and phase components, followed by modeling these components individually, but not independently. This chapter makes repeated use of functional principal component analysis (FPCA) for reaching finite-dimensional Euclidean representations of complex functional data. These representations are then used in analyzing, testing, classifying, and clustering given functions. Chapter 9 extends these ideas to modeling and analysis of planar curves using the representations developed in Chapters 5 and 6. It illustrates the use of geometries of shape spaces to locally linearize and to perform FPCA to reach tractable representations of planar shapes. These representations are then used in modeling and classifications of shapes using statistical shape analysis. Chapter 10 extends these ideas to curves in arbitrary Euclidean spaces and illustrates some of these ideas using examples from interesting application areas. The textbook ends with a collection of miscellaneous topics relating to shape analysis of curves.

This textbook can also serve as a textbook for an advanced topics course with interdisciplinary flavor. There are several possibilities in terms of teaching a course from this textbook. We list some of these ideas, but of course one can mix and match these ideas as well:

- A one-semester course in function data analysis using Appendices A and B, Chaps. 3, 4, 7, and 8, in conjunction with a more classical text on FDA.
- A one-semester course in shape data analysis using Appendices A and B, Chaps. 2, 3, 5, 6, 7, and 9.
- A two-semester course in function and shape data analysis using Appendices A and B, Chaps. 2–11, possibly in conjunction with a more classical text on FDA.

Tallahassee, FL, USA

Anuj Srivastava
Eric P. Klassen

Acknowledgments

We were able to finish this textbook due to significant contributions from a large number of people. Indeed, this book uses a body of knowledge that was developed in collaboration with a number of wonderful colleagues. Without their contributions, this material would have never matured into this form. Therefore, we are deeply in debt to these people for their contributions and support.

Several of the ideas presented here were developed in collaboration with Prof. Ian H. Jermyn of Durham University, Durham UK. Indeed, he was the first one to introduce us to the literature on square-root representation of probability density functions and its connections with the Fisher-Rao metric. In some sense, this whole effort on elastic shape analysis of curves started for us at that moment in the summer of 2006. We thank him for all his shared wisdom and collaboration. We also acknowledge the efforts of Dr. Shantanu Joshi of UCLA, who was the first of many graduate students who helped transform conceptual ideas into implementations and demonstrated their impact on real data! We are thankful to all of our graduate students—Darshan Bryner, Wade Henning, David Kaziska, Sebastian Kurtek, Jose Laborde, Sayani Lahiri, Wei Liu, Sentibaleng Ncube, Dan Robinson, Michael Rosenthal, Chafik Samir, Jingyong Su, J. Derek Tucker, Linda Crystal White, Qian Xie, and Zhengwu Zhang—all of them chose to work on and develop topics covered in this book for their PhD dissertations. We also received continuous encouragement, comments, and feedbacks from many colleagues during the writing of this book. Of those, we point out Boulbaba Benamor, Rama Chellappa, Mohamed Daoudi, Zhaohua Ding, Hassen Drira, Hamid Laga, Steve Marron, Pavan Turaga, Wei Wu, and Jinfeng Zhang. We are very thankful to them for all their support.

This book exhibits, in its own small way, the beauty of general pattern theory pioneered by Prof. Ulf Grenander of Brown University. The most wonderful aspect of this theory is the rich set of tools it uses: algebra, geometry, statistics, and, of course, computational science. Our wanderings beyond classical multivariate and Euclidean statistics followed paths previously illuminated by Grenander's excursions. “*Discretize as late as possible*”—This was Prof. Grenander's advice to students in a class on pattern theory at Washington University in 1995 and became a source of inspiration for the material presented here. The first author is forever grateful to Prof. Grenander for his teachings, guidance, and mentorship. He also gratefully acknowledges training and guidance received from Prof. Michael Miller of Johns Hopkins University.

We also acknowledge support from the Army Research Office for our early work in shape analysis, especially with a focus on geometry and statistics. This effort was also supported in part by grants from the Office of Naval Research, the National Science Foundation, and the National Institutes of Health. The initialization and culmination of this writing project coincided with sabbatical leaves for AS during 2007–2008 and 2014–2015. He, therefore, acknowledges support from the University of Lille 1, Lille, France, and the Franco-American Fulbright commission in Paris, France, for supporting his long stays in France. AS also acknowledges support from the Statistics Division at the National Institute of Standards and Technology, Gaithersburg, MD, in completing this project.

We are thankful for a highly supportive environment for interdisciplinary research and creativity in our respective departments, Statistics and Mathematics, and the College of Arts and Sciences at FSU. We are thankful to our colleagues in these units for encouraging us to take up these challenges.

Finally, and most importantly, we thank our families for their support and encouragement in accomplishing this project. AS is very thankful of his family—wife Elise, parents Govind and Indraprabha, sons Alex and Neel, and parents-in-law André and Marie-Cécile—for their continuous support and cheer. EK thanks his wife Anna and daughter Rosemary because the happiness they bring to his life makes math so much more fun.

Contents

1	Motivation for Function and Shape Analysis	1
1.1	Motivation	2
1.1.1	Need for Function and Shape Data Analysis Tools	2
1.1.2	Why Continuous Shapes?	3
1.2	Important Application Areas	5
1.3	Specific Technical Goals	11
1.4	Issues and Challenges	17
1.5	Organization of this Textbook	19
2	Previous Techniques in Shape Analysis	21
2.1	Principal Component Analysis	23
2.2	Point-Based Shape-Analysis Methods	24
2.2.1	ICP: Point Cloud Analysis	24
2.2.2	Active Shape Models	29
2.2.3	Kendall's Landmark-Based Shape Analysis	30
2.2.4	Issue of Landmark Selection	32
2.3	Domain-Based Shape Representations	33
2.3.1	Level-Set Methods	33
2.3.2	Deformation-Based Shape Analysis	35
2.4	Exercises	36
2.5	Bibliographic Notes	37
3	Background: Relevant Tools from Geometry	39
3.1	Equivalence Relations	40
3.2	Riemannian Structure and Geodesics	41
3.3	Geodesics in Spaces of Curves on Manifolds	48
3.4	Parallel Transport of Vectors	50
3.5	Lie Group Actions on Manifolds	52
3.5.1	Actions of Single Groups	53
3.5.2	Actions of Product Groups	55
3.6	Quotient Spaces of Riemannian Manifolds	56
3.7	Quotient Spaces as Orthogonal Sections	60
3.8	General Quotient Spaces	65
3.9	Distances in Quotient Spaces: A Summary	67
3.10	Center of an Orbit	67

3.11	Exercises	69
3.11.1	Theoretical Exercises	69
3.11.2	Computational Exercises	72
3.12	Bibliographic Notes	72
4	Functional Data and Elastic Registration	73
4.1	Goals and Challenges	74
4.2	Estimating Function Variables from Discrete Data	75
4.3	Geometries of Some Function Spaces	78
4.3.1	Geometry of Hilbert Spaces	78
4.3.2	Unit Hilbert Sphere	82
4.3.3	Group of Warping Functions	83
4.4	Function Registration Problem	85
4.5	Use of L^2 -Norm and Its Limitations	88
4.6	Square-Root Slope Function Representation	91
4.7	Definition of Phase and Amplitude Components	93
4.7.1	Amplitude of a Function	94
4.7.2	Relative Phase Between Functions	96
4.7.3	A Convenient Approximation	97
4.8	SRSF-Based Registration	98
4.8.1	Registration Problem	98
4.8.2	SRSF Alignment Using Dynamic Programming	99
4.8.3	Examples of Functional Alignments	100
4.9	Connection to the Fisher-Rao Metric	104
4.10	Phase and Amplitude Distances	107
4.10.1	Amplitude Space and a Metric Structure	107
4.10.2	Phase Space and a Metric Structure	109
4.11	Different Warping Actions and PDFs	112
4.11.1	Listing of Different Actions	112
4.11.2	Probability Density Functions	113
4.12	Exercises	118
4.12.1	Theoretical Exercises	118
4.12.2	Computational Exercises	121
4.13	Bibliographic Notes	122
5	Shapes of Planar Curves	125
5.1	Goals and Challenges	125
5.2	Parametric Representations of Curves	126
5.3	General Framework	128
5.3.1	Mathematical Representations of Curves	128
5.3.2	Shape-Preserving Transformations	132
5.4	Pre-shape Spaces	133
5.4.1	Riemannian Structure	135
5.4.2	Geodesics in Pre-shape Spaces	138
5.5	Shape Spaces	141
5.5.1	Removing Parameterization	141
5.6	Motivation for SRVF Representation	144
5.6.1	What Is an Elastic Metric?	144
5.6.2	Significance of the Square-Root Representation	148

5.7	Geodesic Paths in Shape Spaces	149
5.7.1	Optimal Re-Parameterization for Curve Matching	152
5.7.2	Geodesic Illustrations	152
5.8	Gradient-Based Optimization Over Re-Parameterization Group	156
5.9	Summary	160
5.10	Exercises	160
5.10.1	Theoretical Exercises	160
5.10.2	Computational Exercises	163
5.11	Bibliographic Notes	164
6	Shapes of Planar Closed Curves	167
6.1	Goals and Challenges	167
6.2	Representations of Closed Curves	168
6.2.1	Pre-shape Spaces	170
6.2.2	Riemannian Structures	171
6.2.3	Removing Parameterization	174
6.3	Projection on a Manifold	178
6.4	Geodesic Computation	179
6.5	Geodesic Computation: Shooting Method	180
6.5.1	Example 1: Geodesics on $\mathbb{S}^2$	182
6.5.2	Example 2: Geodesics in Non-elastic Pre-shape Space	185
6.6	Geodesic Computation: Path-Straightening Method	189
6.6.1	Theoretical Background	189
6.6.2	Numerical Implementation	194
6.6.3	Example 1: Geodesics on $\mathbb{S}^2$	197
6.6.4	Example 2: Geodesics in Elastic Pre-shape Space	197
6.7	Geodesics in Shape Spaces	204
6.7.1	Geodesics in Non-elastic Shape Space	204
6.7.2	Geodesics in Elastic Shape Space	208
6.8	Examples of Elastic Geodesics	215
6.8.1	Elastic Matching: Gradient Versus Dynamic Programming Algorithm	217
6.8.2	Fast Approximate Elastic Matching of Closed Curves	218
6.9	Elastic Versus Non-elastic Deformations	218
6.10	Parallel Transport of Shape Deformations	219
6.10.1	Prediction of Silhouettes from Novel Views	223
6.10.2	Classification of 3D Objects Using Predicted Silhouettes	224
6.11	Symmetry Analysis of Planar Shapes	225
6.12	Exercises	228
6.12.1	Theoretical Exercises	228
6.12.2	Computational Exercises	229
6.13	Bibliographic Notes	231
7	Statistical Modeling on Nonlinear Manifolds	233
7.1	Goals and Challenges	233
7.2	Basic Setup	234
7.3	Probability Densities on Manifolds	235
7.4	Summary Statistics on Manifolds	236
7.4.1	Intrinsic Statistics	236
7.4.2	Extrinsic Statistics	241

7.5	Examples on Some Useful Manifolds	242
7.5.1	Statistical Analysis on $\mathbb{S}^1$	242
7.5.2	Statistical Analysis on $\mathbb{S}^2$	248
7.5.3	Space of Probability Density Functions	255
7.5.4	Space of Warping Functions	259
7.6	Statistical Analysis on a Quotient Space M/G	261
7.6.1	Quotient Space as Orthogonal Section	262
7.6.2	General Case: Without Using Sections	263
7.7	Exercises	265
7.7.1	Theoretical Exercises	265
7.7.2	Computational Exercises	266
7.8	Bibliographic Notes	267
8	Statistical Modeling of Functional Data	269
8.1	Goals and Challenges	270
8.2	Template-Based Alignment and $\mathbb{L}^2$ Metric	271
8.3	Elastic Phase-Amplitude Separation	273
8.3.1	Karcher Mean of Amplitudes	273
8.3.2	Template: Center of the Mean Orbit	274
8.3.3	Phase-Amplitude Separation Algorithm	276
8.4	Alternate Interpretation as Estimation of Model Parameters	281
8.5	Phase-Amplitude Separation After Transformation	282
8.6	Penalized Function Alignment	284
8.7	Function Components, Alignment, and Modeling	286
8.8	Sequential Approach	288
8.8.1	FPCA of Amplitude Functions: A-FPCA	289
8.8.2	FPCA of Phase Functions: P-FPCA	290
8.8.3	Joint Modeling of Principle Coefficients	292
8.9	Joint Approach: Elastic FPCA	294
8.9.1	Model-Based Elastic FPCA in Function Space $\mathcal{F}$	294
8.9.2	Elastic FPCA Using SRSF Representation	297
8.10	Exercises	301
8.10.1	Theoretical Exercises	301
8.10.2	Computational Exercises	302
8.11	Bibliographic Notes	303
9	Statistical Modeling of Planar Shapes	305
9.1	Goals and Challenges	306
9.2	Clustering in Shape Spaces	307
9.2.1	Hierarchical Clustering	307
9.2.2	A Minimum-Dispersion Clustering	309
9.3	A Finite Representation of Planar Shapes	311
9.3.1	Shape Representation: A Brief Review	311
9.3.2	Finite Shape Representation: Planar Curves	313
9.3.3	Finite Representation: Planar Closed Curves	316
9.4	Models for Planar Curves as Elements of $\mathcal{S}_2$	317
9.4.1	Truncated Wrapped-Normal (TWN) Model	317
9.4.2	Learning TWN Model from Training Shapes in $\mathcal{S}_2$	318
9.5	Models for Planar Closed Curves	322
9.6	Beyond TWN Shape Models	327

9.7	Modeling Nuisance Variables	329
9.7.1	Modeling Re-Parameterization Function	330
9.7.2	Modeling Shape Orientations	333
9.8	Classification of Shapes With Contour Data	334
9.8.1	Nearest-Neighbor Classification	335
9.8.2	Probabilistic Classification	335
9.9	Detection/Classification of Shapes in Cluttered Point Clouds	339
9.9.1	Point Process Models for Cluttered Data	341
9.9.2	Maximum Likelihood Estimation of Model Parameters	343
9.10	Problems	345
9.10.1	Theoretical Problems	345
9.10.2	Computational Problems	346
9.11	Bibliographic Notes	347
10	Shapes of Curves in Higher Dimensions	349
10.1	Goals and Challenges	349
10.2	Mathematical Representations of Curves	350
10.3	Elastic and Non-elastic Metrics	351
10.4	Shape Spaces of Curves in $\mathbb{R}^n$	353
10.4.1	Direction Function Representation	353
10.4.2	Under SRVF Representation	356
10.4.3	Hierarchical Clustering of Elastic Curves	361
10.4.4	Sample Statistics and Modeling of Elastic Curves in $\mathbb{R}^n$	362
10.5	Registration of Curves	365
10.5.1	Pairwise Registration of Curves in $\mathbb{R}^n$	366
10.5.2	Registration of Multiple Curves	368
10.6	Shapes of Closed Curves in $\mathbb{R}^n$	369
10.6.1	Non-elastic Closed Curves	369
10.6.2	Elastic Closed Curves	372
10.7	Shape Analysis of Augmented Curves	376
10.7.1	Joint Representation of Augmented Curves	378
10.7.2	Invariances and Equivalence Classes	379
10.8	Problems	382
10.8.1	Theoretical Problems	382
10.8.2	Computational Problems	384
10.9	Bibliographic Notes	384
11	Related Topics in Shape Analysis of Curves	385
11.1	Goals and Challenges	385
11.2	Joint Analysis of Shape and Other Features	386
11.2.1	Geodesics and Geodesic Distances on Feature Spaces	387
11.2.2	Feature-Based Clustering	389
11.3	Affine-Invariant Shape Analysis of Planar Curves	391
11.3.1	Global Section Under the Affine Action	393
11.3.2	Geodesics Using Path-Straightening Algorithm	396
11.4	Registration of Trajectories on Nonlinear Manifolds	401
11.4.1	Transported SRVF for Trajectories	405
11.4.2	Analysis of Trajectories on $\mathbb{S}^2$	411

11.5 Problems	413
11.5.1 Theoretical Problems	413
11.5.2 Computational Problems	415
11.6 Bibliographic Notes	415
Background Material	417
A.1 Basic Differential Geometry	417
A.1.1 Tangent Spaces on a Manifold	421
A.1.2 Submanifolds	425
A.2 Basic Algebra	427
A.3 Basic Geometry of Function Spaces	431
A.3.1 Hilbert Manifolds and Submanifolds	432
The Dynamic Programming Algorithm	435
B.1 Theoretical Setup	435
B.2 Computer Implementation	436
References	439
Index	445