

Springer Series in the Data Sciences

Nickolay Trendafilov
Michele Gallo

Multivariate Data Analysis on Matrix Manifolds

(with Manopt)

 Springer

Springer Series in the Data Sciences

Series Editors

David Banks, Duke University, Durham, NC, USA

Jianqing Fan, Department of Financial Engineering, Princeton University,
Princeton, NJ, USA

Michael Jordan, University of California, Berkeley, CA, USA

Ravi Kannan, Microsoft Research Labs, Bangalore, India

Yurii Nesterov, CORE, Universite Catholique de Louvain, Louvain-la-Neuve,
Belgium

Christopher Ré, Department of Computer Science, Stanford University, Stanford,
USA

Ryan J. Tibshirani, Department of Statistics, Carnegie Mellon University,
Pittsburgh, PA, USA

Larry Wasserman, Department of Statistics, Carnegie Mellon University,
Pittsburgh, PA, USA

Springer Series in the Data Sciences focuses primarily on monographs and graduate level textbooks. The target audience includes students and researchers working in and across the fields of mathematics, theoretical computer science, and statistics. Data Analysis and Interpretation is a broad field encompassing some of the fastest-growing subjects in interdisciplinary statistics, mathematics and computer science. It encompasses a process of inspecting, cleaning, transforming, and modeling data with the goal of discovering useful information, suggesting conclusions, and supporting decision making. Data analysis has multiple facets and approaches, including diverse techniques under a variety of names, in different business, science, and social science domains. Springer Series in the Data Sciences addresses the needs of a broad spectrum of scientists and students who are utilizing quantitative methods in their daily research. The series is broad but structured, including topics within all core areas of the data sciences. The breadth of the series reflects the variation of scholarly projects currently underway in the field of machine learning.


More information about this series at <http://www.springer.com/series/13852>


Nickolay Trendafilov · Michele Gallo

Multivariate Data Analysis on Matrix Manifolds

(with Manopt)

 Springer

Nickolay Trendafilov 
School of Mathematics and Statistics
Open University
Milton Keynes, Buckinghamshire, UK

Michele Gallo 
Department of Human and Social Sciences
University of Naples “L’Orientale”
Naples, Italy

ISSN 2365-5674 ISSN 2365-5682 (electronic)
Springer Series in the Data Sciences
ISBN 978-3-030-76973-4 ISBN 978-3-030-76974-1 (eBook)
<https://doi.org/10.1007/978-3-030-76974-1>

Mathematics Subject Classification: 58A05, 58C05, 62–07, 62H12, 62H25, 62H30, 62H99, 65C60, 65Fxx, 65K99, 90C26, 90C51

© Springer Nature Switzerland AG 2021

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, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To the memory of
my mother, Zdravka, and my father, Trendafil
and to my family
wife, Irina, son, Iassen,
and grandchildren, Veronica and Christian

Preface

We want to start with few remarks predating considerably the emerging of the idea for writing this book and our collaboration in general. They are related to the first author's own experience which made him explore matrix manifolds in data analysis problems.

NTT was in an early stage of his research career, after his Ph.D. was completed on a completely different topic. He was assigned to study factor analysis (FA) and do some programming for a particular software product. While working on FA, NTT realized that the most interesting part for him is the FA interpretation and the so-called rotation methods (see Section 4.5). NTT recognized that the main goal of FA is to produce simple for interpretation results, which is achieved by orthogonal rotation of some initial, usually difficult to interpret, FA solution Λ (known as factor loadings matrix). However, how we can define what is "simple for interpretation results"? The problem is really tough, especially if you try to capture its meaning in a single mathematical expression/formula. Because of that, a huge number of different formulas were proposed each of them claiming to approximate in some sense the idea for "simple for interpretation results". In FA, these formulas are called simple structure criteria. They are supposed to measure/quantify the simplicity of a certain FA solution. Let f be such a criterion. Then, if Λ_2 is simpler than Λ_1 , and if Λ_3 is even simpler than Λ_2 , then we should have $f(\Lambda_1) > f(\Lambda_2) > f(\Lambda_3)$, or $f(\Lambda_1) < f(\Lambda_2) < f(\Lambda_3)$ depending on the sign of f . Clearly, one would be interested to find the solution Λ , for which f reaches its smallest/largest value, and the problem smells of optimization.

For reasons to become clear in Section 4.5, the rotation methods transform the initial (difficult to interpret) solution Λ_0 into another, simpler, one $\Lambda_0 Q$, where Q is an orthogonal matrix, which gives the name “rotation methods”. Thus, in formal terms, the rotation methods look for an orthogonal matrix Q , such that $f(\Lambda_0) > f(\Lambda_0 Q)$. The problem stays the same no matter what rotation criterion f is considered. Thus, one needs to solve

$$\min_Q f(\Lambda_0 Q), \tag{1}$$

where Q is orthogonal and f is an arbitrary criterion.

NTT was surprised to see in the psychometric literature, e.g. Harman (1976); Mulaik (1972), that the introduction of any new simplicity criterion, f , was inevitably connected to a new method for solving (1). At that moment his knowledge in optimization was quite limited, but intuitively NTT felt that this cannot be right: a new optimization method for every new function f . It sounded ridiculous!

And then, NTT started to look for a general approach that can work for every f , as far as the variable Q is an orthogonal matrix. In this way, NTT realized that the rotational problem can be naturally defined as optimization on the matrix manifold of all orthogonal matrices. The understanding that, in fact, all MDA problems can be seen defined on specific matrix manifold(s) followed then straight away.

Milton Keynes, UK
Naples, Italy

Nickolay Trendafilov
Michele Gallo

Contents

Preface	vii
1 Introduction	1
2 Matrix analysis and differentiation	9
2.1 Matrix algebra	10
2.2 Vector spaces, bases, and linear maps	18
2.3 Metric, normed and inner product spaces	22
2.4 Euclidean spaces of matrices and their norms	27
2.5 Matrix differentials and gradients	32
2.6 Conclusion	35
2.7 Exercises	35
3 Matrix manifolds in MDA	45
3.1 Several useful matrix sets	46
3.1.1 $\mathcal{GL}(n)$	46

3.1.2	$\mathcal{O}(n)$	46
3.1.3	$\mathcal{O}(n, p)$	48
3.1.4	$\mathcal{O}_0(n, p)$	49
3.1.5	$\mathcal{OB}(n)$	49
3.1.6	$\mathcal{OB}(n, p)$	52
3.1.7	$\mathcal{G}(n, p)$	52
3.2	Differentiable manifolds	53
3.3	Examples of matrix manifolds in MDA	57
3.3.1	$\mathcal{GL}(n)$	57
3.3.2	$\mathcal{O}(n)$	57
3.3.3	$\mathcal{O}(n, p)$	59
3.3.4	$\mathcal{O}_0(n, p)$	59
3.3.5	$\mathcal{OB}(n)$	59
3.3.6	$\mathcal{G}(n, p)$	60
3.4	Tangent spaces	61
3.4.1	$\mathcal{GL}(n)$	63
3.4.2	$\mathcal{O}(n)$	63
3.4.3	$\mathcal{O}(n, p)$	64
3.4.4	$\mathcal{O}_0(n, p)$	66
3.4.5	$\mathcal{OB}(n)$	66
3.4.6	$\mathcal{OB}(n, p)$	67
3.4.7	$\mathcal{G}(n, p)$	67

3.5	Optimization on matrix manifolds	69
3.5.1	Dynamical systems	69
3.5.2	Iterative schemes	72
3.5.3	Do it yourself!	75
3.6	Optimization with Manopt	77
3.6.1	Matrix manifolds in Manopt	78
3.6.2	Solvers	79
3.6.3	Portability	81
3.7	Conclusion	81
3.8	Exercises	82
4	Principal component analysis (PCA)	89
4.1	Introduction	90
4.2	Definition and main properties	91
4.3	Correspondence analysis (CA)	96
4.4	PCA interpretation	98
4.5	Simple structure rotation in PCA (and FA)	100
4.5.1	Simple structure concept	100
4.5.2	Simple structure criteria and rotation methods	102
4.5.3	Case study: PCA interpretation via rotation methods	105
4.5.4	Rotation to independent components	107

4.6	True simple structure: sparse loadings	109
4.6.1	SPARSIMAX: rotation-like sparse loadings	110
4.6.2	Know-how for applying SPARSIMAX	113
4.6.3	Manopt code for SPARSIMAX	115
4.7	Sparse PCA	118
4.7.1	Sparse components: genesis, history and present times	118
4.7.2	Taxonomy of PCA subject to ℓ_1 constraint (LASSO)	120
4.8	Function-constrained sparse components	121
4.8.1	Orthonormal sparse component loadings	121
4.8.2	Uncorrelated sparse components	123
4.8.3	Numerical Example	124
4.8.4	Manopt code for weakly correlated sparse components with nearly orthonormal loadings	124
4.9	New generation dimension reduction	127
4.9.1	Centroid method	127
4.9.2	Randomized SVD	130
4.9.3	CUR approximations	131
4.9.4	The Nystrom Method	132
4.10	Conclusion	133
4.11	Exercises	134

5	Factor analysis (FA)	141
5.1	Introduction	142
5.2	Fundamental equations of EFA	143
5.2.1	Population EFA definition	143
5.2.2	Sample EFA definition	144
5.3	EFA parameters estimation	145
5.3.1	Classical EFA estimation	146
5.3.2	EFA estimation on manifolds.....	148
5.4	ML exploratory factor analysis	151
5.4.1	Gradients.....	151
5.4.2	Optimality conditions	152
5.5	LS and GLS exploratory factor analysis.....	153
5.5.1	Gradients.....	153
5.5.2	Optimality conditions	154
5.6	Manopt codes for classical ML-, LS-, and GLS-EFA	156
5.6.1	Standard case: $\Psi^2 \geq O_p$	156
5.6.2	Avoiding Heywood cases: $\Psi^2 > O_p$	158
5.7	EFA as a low-rank-plus-sparse matrix decomposition	164

5.8	Sparse EFA	171
5.8.1	Introduction	171
5.8.2	Sparse factor loadings with penalized EFA	172
5.8.3	Implementing sparseness	173
5.8.4	Numerical examples	174
5.9	Comparison to other methods	177
5.10	Conclusion	181
5.11	Exercises	182
6	Procrustes analysis (PA)	187
6.1	Introduction	188
6.2	Orthonormal PA	189
6.2.1	Orthogonal Penrose regression (OPR)	189
6.2.2	Projected Hessian	190
6.2.3	Orthonormal Penrose regression (OnPR)	193
6.2.4	Ordinary orthonormal PA	195
6.3	Oblique PA	196
6.3.1	Basic formulations and solutions	196
6.4	Robust PA	198
6.4.1	Some history remarks	198
6.4.2	Robust OnPR	199

6.4.3	Robust oblique PA	200
6.4.4	PA with M-estimator	202
6.5	Multi-mode PA	204
6.5.1	PCA and one-mode PCA	204
6.5.2	Multi-mode PCA and related PA problems	204
6.5.3	Global minima on $\mathcal{O}(p)$	206
6.6	Some other PA problems	208
6.6.1	Average of rotated matrices: generalized PA	208
6.6.2	Mean rotation	209
6.7	PA application to EFA of large data	215
6.7.1	The classical case $n > p$	216
6.7.2	The modern case $p \gg n$	217
6.7.3	EFA and RPCA when $p \gg n$	219
6.7.4	Semi-sparse PCA (well-defined EFA)	220
6.8	Conclusion	225
6.9	Exercises	225
7	Linear discriminant analysis (LDA)	229
7.1	Introduction	230
7.2	LDA of vertical data ($n > p$)	233
7.2.1	Standard canonical variates (CVs)	235
7.2.2	Orthogonal canonical variates (OCVs)	240

7.3	Sparse CVs and sparse OCVs	241
7.4	LDA of horizontal data ($p > n$)	248
7.4.1	LDA through GSVD	249
7.4.2	LDA and pattern recognition	250
7.4.3	Null space LDA (NLDA)	251
7.4.4	LDA with CPC, PLS and MDS	253
7.4.5	Sparse LDA with diagonal W	254
7.4.6	Function-constrained sparse LDA	255
7.4.7	Sparse LDA based on minimization of the classification error	256
7.4.8	Sparse LDA through optimal scoring (SLDA)	258
7.4.9	Multiclass sparse discriminant analysis	259
7.4.10	Sparse LDA through GEVD	260
7.4.11	Sparse LDA without sparse-inducing penalty	262
7.5	Conclusion	263
7.6	Exercises	264
8	Canonical correlation analysis (CCA)	269
8.1	Introduction	270
8.2	Classical CCA Formulation and Solution	270
8.3	Alternative CCA Definitions	272
8.4	Singular Scatter Matrices C_{11} and/or C_{22}	273

8.5	Sparse CCA	273
8.5.1	Sparse CCA Through Sparse GEVD.....	274
8.5.2	LS Approach to Sparse CCA	276
8.6	CCA Relation to LDA and PLS	277
8.6.1	CCA and LDA	277
8.6.2	CCA and PLS.....	277
8.7	More Than Two Groups of Variables.....	279
8.7.1	CCA Generalizations	280
8.7.2	CCA Based on CPC	282
8.8	Conclusion.....	285
8.9	Exercises	286
9	Common principal components (CPC)	289
9.1	Introduction	290
9.2	CPC estimation problems.....	293
9.3	ML- and LS-CPC.....	294
9.3.1	Gradients and optimality conditions.....	294
9.3.2	Example: Fisher's Iris data	296
9.3.3	Appendix: MATLAB code for FG algorithm.....	297
9.4	New procedures for CPC estimation.....	299
9.4.1	Classic numerical solutions of ML- and LS-CPC.....	299

9.4.2	Direct calculation of individual eigenvalues/variances	300
9.4.3	CPC for known individual variances	302
9.5	CPC for dimension reduction	303
9.6	Proportional covariance matrices	309
9.6.1	ML and LS proportional principal components	309
9.6.2	Dimension reduction with PPC	313
9.7	Some relations between CPC and ICA	315
9.7.1	ICA formulations	315
9.7.2	ICA by contrast functions	316
9.7.3	ICA methods based on diagonalization	320
9.8	Conclusion	323
9.9	Exercises	324
10	Metric multidimensional scaling (MDS) and related methods	325
10.1	Introduction	326
10.2	Proximity measures	327
10.3	Metric MDS	328
10.3.1	Basic identities and classic solution	328
10.3.2	MDS that fits distances directly	330
10.3.3	Some related/adjacent MDS problems	332
10.4	INDSCAL – Individual Differences Scaling	336

10.4.1	The classical INDSCAL solution and some problems	337
10.4.2	Orthonormality-constrained INDSCAL	338
10.5	DINDSCAL – Direct INDSCAL	340
10.5.1	DINDSCAL model	341
10.5.2	DINDSCAL solution	342
10.5.3	Manopt code for DINDSCAL	343
10.6	DEDICOM	345
10.6.1	Introduction	345
10.6.2	Alternating DEDICOM	346
10.6.3	Simultaneous DEDICOM	348
10.7	GIPSCAL	348
10.7.1	GIPSCAL model	348
10.7.2	GIPSCAL solution	351
10.7.3	Three-way GIPSCAL	353
10.8	Tensor data analysis	353
10.8.1	Basic notations and definitions	355
10.8.2	CANDECOMP/PARAFAC (CP)	358
10.8.3	Three-mode PCA (TUCKER3)	360
10.8.4	Multi-mode PCA	363
10.8.5	Higher order SVD (HOSVD)	364
10.9	Conclusion	366
10.10	Exercises	368

11	Data analysis on simplexes	373
11.1	Archetypal analysis (AA)	374
11.1.1	Introduction	374
11.1.2	Definition of the AA problem	375
11.1.3	AA solution on multinomial manifold	376
11.1.4	AA solution on oblique manifolds	378
11.1.5	AA as interior point flows	381
11.1.6	Conclusion	384
11.1.7	Exercises	385
11.2	Analysis of compositional data (CoDa)	388
11.2.1	Introduction	388
11.2.2	Definition and main properties	389
11.2.3	Geometric clr structure of the data simplex	391
11.2.4	PCA and sparse PCA for compositions	394
11.2.5	Case study	396
11.2.6	Manopt code for sparse PCA of CoDa	397
11.2.7	Conclusion	401
11.2.8	Exercises	401
	Bibliography	403
	Index	441