

An Introduction to Optimization on Smooth Manifolds

NICOLAS BOUMAL

École Polytechnique Fédérale de Lausanne



CAMBRIDGE
UNIVERSITY PRESS

Contents

	<i>Preface</i>	<i>page xi</i>
	<i>Notation</i>	<i>xvi</i>
1	Introduction	1
2	Simple examples	4
	2.1 Sensor network localization from directions: an affine subspace	4
	2.2 Single extreme eigenvalue or singular value: spheres	5
	2.3 Dictionary learning: products of spheres	6
	2.4 Principal component analysis: Stiefel and Grassmann	8
	2.5 Synchronization of rotations: special orthogonal group	11
	2.6 Low-rank matrix completion: fixed-rank manifold	12
	2.7 Gaussian mixture models: positive definite matrices	13
	2.8 Smooth semidefinite programs	14
3	Embedded geometry: first order	16
	3.1 Reminders of Euclidean space	19
	3.2 Embedded submanifolds of a linear space	23
	3.3 Smooth maps on embedded submanifolds	32
	3.4 The differential of a smooth map	33
	3.5 Vector fields and the tangent bundle	36
	3.6 Moving on a manifold: retractions	38
	3.7 Riemannian manifolds and submanifolds	39
	3.8 Riemannian gradients	41
	3.9 Local frames*	45
	3.10 Notes and references	48
4	First-order optimization algorithms	51
	4.1 A first-order Taylor expansion on curves	52
	4.2 First-order optimality conditions	53
	4.3 Riemannian gradient descent	54
	4.4 Regularity conditions and iteration complexity	57
	4.5 Backtracking line-search	59

4.6	Local convergence*	62
4.7	Computing gradients*	68
4.8	Numerically checking a gradient*	75
4.9	Notes and references	77
5	Embedded geometry: second order	79
5.1	The case for another derivative of vector fields	81
5.2	Another look at differentials of vector fields in linear spaces	81
5.3	Differentiating vector fields on manifolds: connections	82
5.4	Riemannian connections	84
5.5	Riemannian Hessians	90
5.6	Connections as pointwise derivatives*	93
5.7	Differentiating vector fields on curves	96
5.8	Acceleration and geodesics	101
5.9	A second-order Taylor expansion on curves	102
5.10	Second-order retractions	104
5.11	Special case: Riemannian submanifolds*	106
5.12	Special case: metric projection retractions*	110
5.13	Notes and references	112
6	Second-order optimization algorithms	115
6.1	Second-order optimality conditions	115
6.2	Riemannian Newton's method	117
6.3	Computing Newton steps: conjugate gradients	120
6.4	Riemannian trust regions	126
6.5	The trust-region subproblem: truncated CG	140
6.6	Local convergence of RTR with tCG*	142
6.7	Simplified assumptions for RTR with tCG*	143
6.8	Numerically checking a Hessian*	145
6.9	Notes and references	146
7	Embedded submanifolds: examples	149
7.1	Euclidean spaces as manifolds	149
7.2	The unit sphere in a Euclidean space	152
7.3	The Stiefel manifold: orthonormal matrices	154
7.4	The orthogonal group and rotation matrices	158
7.5	Fixed-rank matrices	160
7.6	The hyperboloid model	168
7.7	Manifolds defined by $h(x) = 0$	171
7.8	Notes and references	174
8	General manifolds	176
8.1	A permissive definition	176
8.2	The atlas topology, and a final definition	182

8.3	Embedded submanifolds are manifolds	185
8.4	Tangent vectors and tangent spaces	187
8.5	Differentials of smooth maps	189
8.6	Tangent bundles and vector fields	191
8.7	Retractions and velocity of a curve	192
8.8	Coordinate vector fields as local frames	193
8.9	Riemannian metrics and gradients	194
8.10	Lie brackets as vector fields	195
8.11	Riemannian connections and Hessians	197
8.12	Covariant derivatives and geodesics	198
8.13	Taylor expansions and second-order retractions	199
8.14	Submanifolds embedded in manifolds	200
8.15	Notes and references	203
9	Quotient manifolds	205
9.1	A definition and a few facts	209
9.2	Quotient manifolds through group actions	212
9.3	Smooth maps to and from quotient manifolds	215
9.4	Tangent, vertical and horizontal spaces	217
9.5	Vector fields	219
9.6	Retractions	224
9.7	Riemannian quotient manifolds	224
9.8	Gradients	227
9.9	A word about Riemannian gradient descent	228
9.10	Connections	230
9.11	Hessians	232
9.12	A word about Riemannian Newton's method	233
9.13	Total space embedded in a linear space	235
9.14	Horizontal curves and covariant derivatives	238
9.15	Acceleration, geodesics and second-order retractions	240
9.16	Grassmann manifold: summary*	243
9.17	Notes and references	246
10	Additional tools	252
10.1	Distance, geodesics and completeness	252
10.2	Exponential and logarithmic maps	256
10.3	Parallel transport	262
10.4	Lipschitz conditions and Taylor expansions	265
10.5	Transporters	276
10.6	Finite difference approximation of the Hessian	283
10.7	Tensor fields and their covariant differentiation	286
10.8	Notes and references	293

11	Geodesic convexity	298
	11.1 Convex sets and functions in linear spaces	298
	11.2 Geodesically convex sets and functions	301
	11.3 Alternative definitions of geodesically convex sets*	305
	11.4 Differentiable geodesically convex functions	307
	11.5 Geodesic strong convexity and Lipschitz continuous gradients	310
	11.6 Example: positive reals and geometric programming	314
	11.7 Example: positive definite matrices	317
	11.8 Notes and references	319
	<i>References</i>	321
	<i>Index</i>	333

Preface

Optimization problems on smooth manifolds arise in science and engineering as a result of natural geometry (e.g., the set of orientations of physical objects in space is a manifold), latent data simplicity (e.g., high-dimensional data points lie close to a low-dimensional linear subspace, leading to low-rank data matrices), symmetry (e.g., observations are invariant under rotation, translation or other group actions, leading to quotients) and positivity (e.g., covariance matrices and diffusion tensors are positive definite). This has led to successful applications notably in machine learning, computer vision, robotics, scientific computing, dynamical systems and signal processing.

Accordingly, optimization on manifolds has garnered increasing interest from researchers and engineers alike. Building on 50 years of research efforts that have recently intensified, it is now recognized as a wide, beautiful and effective generalization of unconstrained optimization on linear spaces.

Yet, engineering programs seldom include training in differential geometry, that is, the field of mathematics concerned with smooth manifolds. Moreover, existing textbooks on this topic usually align with the interests of mathematicians more than with the needs of engineers and applied mathematicians. This creates a significant but avoidable barrier to entry for optimizers.

One of my goals in writing this book is to offer a different, if at times unorthodox, introduction to differential geometry. Definitions and tools are introduced in a need-based order for optimization. We start with a restricted setting—that of embedded submanifolds of linear spaces—which allows us to define all necessary concepts in direct reference to their usual counterparts from linear spaces. This covers a wealth of applications.

In what is perhaps the clearest departure from standard exposition, charts and atlases are not introduced until quite late. The reason for doing so is twofold: pedagogically, charts and atlases are more abstract than what is needed to work on embedded submanifolds; and pragmatically, charts are seldom if ever useful in practice. It would be unfortunate to give them center stage.

Of course, charts and atlases are the right tool to provide a unified treatment of all smooth manifolds in an intrinsic way. They are introduced eventually, at which point it becomes possible to discuss quotient manifolds: a powerful language to understand symmetry in optimization. Perhaps this abstraction is necessary to fully appreciate the depth of optimization on manifolds as more than just a fancy tool for