An Introduction to Optimization on Smooth Manifolds

Optimization on Riemannian manifolds—the result of smooth geometry and optimization merging into one elegant modern framework—spans many areas of science and engineering, including machine learning, computer vision, signal processing, dynamical systems and scientific computing.

This text introduces the differential geometry and Riemannian geometry concepts that will help students and researchers in applied mathematics, computer science and engineering gain a firm mathematical grounding to use these tools confidently in their research. Its charts-last approach will prove more intuitive from an optimizer's viewpoint, and all definitions and theorems are motivated to build time-tested optimization algorithms. Starting from first principles, the text goes on to cover current research on topics including worst-case complexity and geodesic convexity. Readers will appreciate the tricks of the trade sprinkled throughout the book for conducting research in this area and for writing effective numerical implementations.

Nicolas Boumal is Assistant Professor of Mathematics at the École Polytechnique Fédérale de Lausanne (EPFL) in Switzerland and Associate Editor of the journal *Mathematical Programming*. His current research focuses on optimization, statistical estimation and numerical analysis. Over the course of his career, Boumal has contributed to several modern theoretical advances in Riemannian optimization. He is a lead developer of the award-winning toolbox Manopt, which facilitates experimentation with optimization on manifolds.

An Introduction to Optimization on Smooth Manifolds

NICOLAS BOUMAL
École Polytechnique Fédérale de Lausanne

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

Shaftesbury Road, Cambridge CB2 8EA, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,
New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment,
a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of
education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781009166171

DOI: 10.1017/9781009166164

First published 2023

A catalogue record for this publication is available from the British Library.

A Cataloging-in-Publication data record for this book is available from the Library of Congress.

ISBN 978-1-009-16617-1 Hardback
ISBN 978-1-009-16615-7 Paperback

To my family and mentors

Contents

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 constrained optimization in linear spaces, and truly a mathematically natural setting for *unconstrained* optimization in a wider sense.

Time-tested optimization algorithms are introduced immediately after the early chapters about embedded geometry. Crucially, the design and analysis of these methods remain unchanged whether we are optimizing on a manifold which is embedded in a linear space or not. This makes it possible to get to algorithms early on, without sacrificing generality. It also underlines the conceptual point that the algorithms truly operate on the manifolds intrinsically.

The last two chapters visit more advanced topics that are not typically necessary for simple applications. The first one delves deeper into geometric tools. The second one introduces the basics of *geodesic* convexity: a broad generalization of convexity, which is one of the most fruitful structures in classical optimization.

Intended audience

This book is intended for students and researchers alike. The material has proved popular with applied mathematicians and mathematically inclined engineering and computer science students at the graduate and advanced undergraduate levels.

Readers are assumed to be comfortable with linear algebra and multivariable calculus. Central to the raison d'être of this book, there are no prerequisites in differential geometry or optimization. For computational aspects, it is helpful to have notions of numerical linear algebra, for which I recommend the approachable textbook by Trefethen and Bau [TB97].

Building on these expectations, the aim is to give full proofs and intuition for all concepts that are introduced, at least for submanifolds of linear spaces. The hope is to equip readers to pursue research projects in (or using) optimization on manifolds, involving both mathematical analysis and efficient implementation.

How to use this book

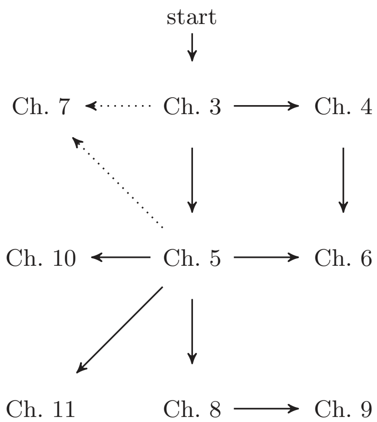
The book is self-contained and should suit both self-learners and instructors.

Chapters 3 and 5 can serve as a standalone introduction to differential and Riemannian geometry. They focus on embedded submanifolds of linear spaces, with proofs. Chapter 7 details examples of manifolds: it is meant for on-and-off reading in parallel with Chapters 3 and 5. These chapters do not involve charts, and they aim to convey the fact that geometric tools are computational tools.

From there, the expected next step is to work through Chapters 4 and 6 about optimization algorithms. Readers may also choose to embark on Chapter 8 to see how embedded manifolds fit into the general theory of smooth manifolds. That is a useful (though not fully necessary) stepping stone toward Chapter 9 about quotient manifolds. Alternatively, they may decide to learn about further geometric tools in Chapter 10 or about a Riemannian notion of convexity in Chapter 11.

Cambridge University Press & Assessment
978-1-009-16617-1 — An Introduction to Optimization on Smooth Manifolds
Nicolas Boumal
Frontmatter
More Information

These chapter dependencies are summarized in the diagram below, where an arrow from A to B means it is preferable to read A before B.



1. Introduction
2. Applications
3. First-order geometry
4. First-order optimization
5. Second-order geometry
6. Second-order optimization
7. Examples of manifolds
8. General manifolds
9. Quotient manifolds
10. Additional geometric tools
11. Geodesic convexity

In a graduate course at Princeton University in 2019 and 2020 (24 lectures of 80 minutes each), I covered much of Chapters 1–6 and select parts of Chapter 7 before the midterm break, then much of Chapters 8–9 and select parts of Chapters 10–11 after the break. At EPFL in 2021, I discussed mostly Chapters 1–8 in 13 lectures of 90 minutes each supplemented with exercise sessions.

The numerous exercises in the book have wide-ranging difficulty levels. Some are included in part as a way to convey information while skipping technicalities.

Starred sections can be skipped safely for a first encounter with the material. Chapters end with references and notes that many readers may find relevant but which would otherwise break the flow. Did the mark in the margin catch your attention? That ⋆
is its purpose. You may see a couple of those in the book.

What is new, or different, or hard to find elsewhere

The de facto reference for optimization on manifolds is the landmark 2008 book *Optimization Algorithms on Matrix Manifolds* by Pierre-Antoine Absil, Robert Mahony and Rodolphe Sepulchre [AMS08]. It is an important source for the present book as well, with significant overlap of topics. In the years since, the field has evolved, and with it the need for an entry point catering to a broader audience. In an effort to address these needs, I aim to:

1. **Provide a different, self-contained introduction to the core concepts.** This includes a "charts last" take on differential geometry with proofs adapted accordingly; a somewhat unusual yet equivalent definition of connections that (I believe) is more intuitive from an optimizer's point of view; and an account of optimization on quotient manifolds which benefits from years of hindsight. This introduction is informed by the pains I had entering the field.

2. **Discuss new topics that have grown in importance since 2008.**
   This includes a replacement of asymptotic convergence results in favor of worst-case, non-asymptotic iteration complexity results; a related take on Lipschitz continuity for Riemannian gradients and Hessians paired with their effect on Taylor expansions on manifolds; an explicit construction of geometric tools necessary for optimization over matrices of fixed rank; a simple study of metric projection retractions; an extrinsic view of the Riemannian Hessian for submanifolds through the Weingarten map and second fundamental form; a discussion of the smooth invertibility of retractions and of the domain of the inverse of the exponential map; transporters as a natural alternative to vector and parallel transports; finite differences of gradients to approximate Hessians; and an introduction to geodesic convexity (not restricted to Hadamard manifolds) with a gradient algorithm for the strongly convex case. Many of these build on research papers referenced in text.
3. **Share tricks of the trade that are seldom, if ever, spelled out.**
   This includes several examples of manifolds worked out in full detail; pragmatic instructions for how to derive expressions for gradients and Hessians of matrix functions, and how to check them numerically; explicit formulas for geometric tools on product manifolds (mostly given as exercises); and a number of comments informed by 10 years of software development in the field.

The main differential geometry references I used are the fantastic books by Lee [Lee12, Lee18], O'Neill [O'N83], and Brickell and Clark [BC70]. Definitions of geometric concepts in this book, though at times stated differently, are fully compatible with Absil et al.'s book. This is also compatible with Lee's textbooks with one exception: Riemannian *sub*manifolds to us are understood to be embedded submanifolds, whereas Lee also allows them to be merely immersed submanifolds. Moreover, we use the word "manifold" to mean "smooth manifold," that is, C^∞. Most results extend to manifolds and functions of class C^k.

There is much to say about the impact of curvature on optimization. This is an active research topic that has not stabilized yet. Therefore, I chose to omit curvature entirely from this book, save for a few brief comments in the last two chapters. Likewise, optimization on manifolds is proving to be a particularly fertile ground for benign non-convexity and related phenomena. There are only a few hints to that effect throughout the book: the research continues.

Software and online resources

Little to no space is devoted to existing software packages for optimization on manifolds, or to numerical experiments. Yet, such packages significantly speed up research and development in the field. The reader may want to experiment with Manopt (Matlab), PyManopt (Python) or Manopt.jl (Julia), all available from manopt.org.

In particular, the Matlab implementations of most manifolds discussed in this book are listed in Table 7.1 on p. 150. Gradient descent (Algorithm 4.1) with backtracking line-search (Algorithm 4.2) is available as `steepestdescent`. The trust-region method (Algorithm 6.3) with the truncated conjugate gradient subproblem solver (Algorithm 6.4) is available as `trustregions`. These implementations include a wealth of tweaks and tricks that are important in practice: many are explained here, some are only documented in the code. The Python and Julia versions offer similar features.

The following webpage collects further resources related to this book:

nicolasboumal.net/book

In particular, teaching and learning material will be listed there, as well as errata.

Thanks

A number of people offered decisive comments. I thank Pierre-Antoine Absil (who taught me much of this) and Rodolphe Sepulchre for their input at the early stages of planning for this book, as well as (in no particular order) Eitan Levin, Chris Criscitiello, Quentin Rebjock, Razvan-Octavian Radu, Joe Kileel, Coralia Cartis, Bart Vandereycken, Bamdev Mishra, Suvrit Sra, Stephen McKeown, John M. Lee and Sándor Z. Németh for numerous conversations that led to direct improvements. Likewise, reviewers offered welcome advice and suggestions for additions (that I was only able to implement partially). Another big thank you to the people involved with the Manopt toolboxes: these efforts are led with Bamdev for the Matlab version; by Jamie Townsend, Niklas Koep and Sebastian Weichwald for the Python version; and by Ronny Bergmann for the Julia version.

I am also indebted to the mathematics departments at Princeton University and EPFL for supporting me while I was writing. Finally, I thank Katie Leach at Cambridge University Press for her enthusiasm and candid advice that helped shape this project into its final form.

Notation

The following lists typical uses of symbols. Local exceptions are documented in place. For example, c typically denotes a curve, but sometimes denotes a real constant. Symbols defined and used locally only are omitted.

$\mathbb{R}, \mathbb{C}$	Real and complex numbers
$\mathbb{R}_+$	Positive reals ($x > 0$)
$\mathbb{R}^{m\times n}$	Real matrices of size $m \times n$
$\mathbb{R}_r^{m\times n}$	Real matrices of size $m \times n$ and rank r
$\mathrm{Sym}(n), \mathrm{Skew}(n)$	Symmetric and skew-symmetric real matrices of size n
$\mathrm{sym}(M), \mathrm{skew}(M)$	Symmetric and skew-symmetric parts of a matrix M
$\mathrm{Sym}(n)^+$	Symmetric positive definite real matrices of size n
$\mathrm{Tr}(M), \det(M)$	Trace, determinant of a square matrix M
$\mathrm{diag}(M)$	Vector of diagonal entries of a matrix M
$\mathrm{diag}(u_1, \ldots, u_n)$	Diagonal matrix of size n with given diagonal entries
$M^\dagger$	Moore–Penrose pseudoinverse of matrix M
I_d	Identity matrix of size d
I	Subset of $\mathbb{R}$ (often open with $0 \in I$) or identity matrix
Id	Identity operator
$\lvert a\rvert$	Modulus of $a \in \mathbb{C}$ (absolute value if $a \in \mathbb{R}$)
$\lvert A\rvert$	Cardinality of a set A
$\mathcal{E}, \mathcal{E}', \mathcal{F}$	Linear spaces, often with a Euclidean structure
$\mathcal{M}, \mathcal{M}', \overline{\mathcal{M}}, \mathcal{N}$	Smooth manifolds, often with a Riemannian structure
S^{d-1}	Unit sphere, in a Euclidean space of dimension d
$\mathrm{OB}(d, n)$	Oblique manifold (product of S^{d-1} copied n times)
$\mathrm{O}(d), \mathrm{SO}(d)$	Orthogonal and special orthogonal groups in $\mathbb{R}^{d\times d}$
$\mathrm{St}(n, p)$	Stiefel manifold embedded in $\mathbb{R}^{n\times p}$
$\mathrm{Gr}(n, p)$	Grassmann manifold as the quotient $\mathrm{St}(n, p)/\mathrm{O}(p)$
$\mathrm{GL}(n)$	General linear group (invertible matrices in $\mathbb{R}^{n\times n}$)
H^n	Hyperbolic space as hyperboloid embedded in $\mathbb{R}^{n+1}$
$\dim \mathcal{M}$	Dimension of $\mathcal{M}$
x, y, z	Points on a manifold
u, v, w, s, ξ, ζ	Tangent vectors
p, q	Integers or polynomials or points on a manifold
$\mathrm{T}\mathcal{M}$	Tangent bundle of $\mathcal{M}$

$\mathrm{T}_x\mathcal{M}$	Tangent space to $\mathcal{M}$ at $x \in \mathcal{M}$
$\mathrm{N}_x\mathcal{M}$	Normal space (orthogonal complement of $\mathrm{T}_x\mathcal{M}$)
$\mathrm{Proj}_x, \mathrm{Proj}_x^\perp$	Orthogonal projector to $\mathrm{T}_x\mathcal{M}$, $\mathrm{N}_x\mathcal{M}$
$\mathrm{H}_x, \mathrm{V}_x$	Horizontal and vertical space at x for a quotient manifold
$\mathrm{Proj}_x^{\mathrm{H}}, \mathrm{Proj}_x^{\mathrm{V}}$	Orthogonal projectors to $\mathrm{H}_x, \mathrm{V}_x$
lift_x	Horizontal lift operator for quotient manifolds
$\mathrm{R}_x(v)$	Retraction R evaluated at $(x, v) \in \mathrm{T}\mathcal{M}$
$\mathrm{Exp}_x(v)$	Exponential map Exp evaluated at $(x, v) \in \mathrm{T}\mathcal{M}$
$\mathrm{Log}_x(y)$	Vector v such that $\mathrm{Exp}_x(v) = y$ (see Definition 10.20)
$\exp, \log$	Scalar or matrix exponential and logarithm
$\mathcal{O}, \mathcal{O}_x$	Domain of Exp (subset of $\mathrm{T}\mathcal{M}$), Exp_x (subset of $\mathrm{T}_x\mathcal{M}$); Can also denote these domains for a non-global retraction
$\mathrm{inj}(\mathcal{M}), \mathrm{inj}(x)$	Injectivity radius of a manifold, at a point
$\langle \cdot, \cdot \rangle_x$	Riemannian inner product on $\mathrm{T}_x\mathcal{M}$
$\langle \cdot, \cdot \rangle$	Euclidean inner product; Sometimes denotes $\langle \cdot, \cdot \rangle_x$ with subscript omitted
$\|\cdot\|, \|\cdot\|_x$	Norms associated to $\langle \cdot, \cdot \rangle$ and $\langle \cdot, \cdot \rangle_x$
$\|\cdot\|$	Also denotes operator norm for linear maps
$[\cdot, \cdot]$	Lie bracket
∇	Affine connection (often Riemannian) on a manifold
$\frac{\mathrm{d}}{\mathrm{d}t}$	Classical derivative with respect to t
$\frac{\mathrm{D}}{\mathrm{d}t}$	Covariant derivative induced by a connection ∇
$\frac{\partial}{\partial x_i}$	Partial derivative with respect to real variable x_i
x_i	Often the ith coordinate of a vector $x \in \mathbb{R}^n$
x_k	Often the kth element of a sequence $x_0, x_1, x_2, \ldots \in \mathcal{M}$
f, g	Real-valued functions
f_{low}	A real number such that $f(x) \geq f_{\mathrm{low}}$ for all x
h	Often a local defining function with values in $\mathbb{R}^k$
$\mathrm{grad} f, \mathrm{Hess} f$	Riemannian gradient and Hessian of f; Euclidean gradient, Hessian if domain of f is Euclidean
$\nabla f, \nabla^2 f$	First and second covariant derivatives of f as tensor fields
c, γ	Curves
$c', \gamma', \dot{c}, \dot{\gamma}$	Velocity vector fields of curves c, γ
c'', γ''	Intrinsic acceleration vector fields of c, γ
$\ddot{c}, \ddot{\gamma}$	Extrinsic acceleration vector fields of c, γ
L, L_g, L_H	Lipschitz constants (nonnegative reals)
$L(c)$	Length of a curve c
$\mathrm{dist}(x, y)$	Distance (often Riemannian) between two points x, y
$B(x, r)$	Open ball $\{v \in \mathrm{T}_x\mathcal{M} : \|v\|_x < r\}$ or $\{y \in \mathcal{E} : \|y - x\| < r\}$
$\bar{B}(x, r)$	Closed ball as above
$\mathcal{A}, \mathcal{L}$	Linear maps
$\mathcal{A}, \mathcal{A}^+$	Atlas, maximal atlas
$\mathrm{im}\, \mathcal{L}, \ker \mathcal{L}$	Range space (image) and null space (kernel) of $\mathcal{L}$
$\mathrm{rank}(M), \mathrm{rank}(\mathcal{L})$	Rank of a matrix or linear map

$M^\top, M^*$	Transpose or Hermitian conjugate-transpose of matrix M
$\mathcal{L}^*$	Adjoint of a linear map $\mathcal{L}$ between Euclidean spaces
$\mathcal{A} \succeq 0, \mathcal{A} \succ 0$	States $\mathcal{A} = \mathcal{A}^*$ is positive semidefinite or positive definite
$\mathrm{span}(u_1, \ldots, u_m)$	Linear subspace spanned by vectors $u_1, \ldots, u_m$
F, G, H	Maps, usually to and from linear spaces or manifolds
$F\colon A \to B$	A map defined on the whole domain A
$F\|_U$	Restriction of the map F to the domain U
$F(\cdot, y)$	For a map $(x, y) \mapsto F(x, y)$, this is the map $x \mapsto F(x, y)$
$F \circ G$	Composition of maps: $(F \circ G)(x) = F(G(x))$
$\mathrm{D}F(x)[v]$	Differential of F at x along v
U, V, W, X, Y, Z	Vector fields on a manifold, or
	Matrices which could be tangent vectors or points on $\mathcal{M}$;
Y, Z	Can also be vector fields along a curve
U, V, W, O	Can also be open sets, usually in a linear space
$\bar{f}, \bar{F}, \bar{V}, \ldots$	Smooth extensions or lifts of $f, F, V, \ldots$
$\bar{u}, \bar{U}$	Can also denote complex conjugation of u, U
T	Tensor field
T_s	Differential of retraction $\mathrm{DR}_x(s)$
$\mathcal{U}, \mathcal{V}$	Open sets in a manifold
$\mathcal{W}$	Weingarten map
II	Second fundamental form
fV	Vector field $x \mapsto f(x)V(x)$ (with f real valued)
Vf	Real function $x \mapsto \mathrm{D}f(x)[V(x)]$
$\mathfrak{F}(\mathcal{M}), \mathfrak{F}(\mathcal{E})$	Set of smooth real-valued functions on $\mathcal{M}$, $\mathcal{E}$
$\mathfrak{X}(\mathcal{M}), \mathfrak{X}(\mathcal{E})$	Set of smooth vector fields on $\mathcal{M}$, $\mathcal{E}$
$\mathfrak{X}(c)$	Set of smooth vector fields along a curve c
$\mathrm{T}_{y \leftarrow x}$	Vector transport from $\mathrm{T}_x\mathcal{M}$ to $\mathrm{T}_y\mathcal{M}$
$\mathrm{PT}^c_{t_1 \leftarrow t_0}$	Parallel transport along curve c from $c(t_0)$ to $c(t_1)$
P_s	Parallel transport along $\gamma(t) = \mathrm{Exp}_x(ts)$ from 0 to 1
$\sim$	Equivalence relation
$A/\sim$	Quotient set of A by the relation $\sim$
$[x]$	Equivalence class of x for some equivalence relation
π	Canonical projection $\pi\colon \mathrm{T}\mathcal{M} \to \mathcal{M}$ or $\pi\colon \overline{\mathcal{M}} \to \overline{\mathcal{M}}/\sim$;
	Occasionally denotes the mathematical constant
$A \subset B$	A is a proper subset of B (the sets are not equal)
$A \subseteq B$	A is a subset of B (the sets may be equal)
$A \cap B$	Intersection of sets A, B
$A \cup B$	Union of sets A, B
$\emptyset$	Empty set