Chapter One

Introduction

This book is about the design of numerical algorithms for computational problems posed on smooth search spaces. The work is motivated by matrix optimization problems characterized by symmetry or invariance properties in the cost function or constraints. Such problems abound in algorithmic questions pertaining to linear algebra, signal processing, data mining, and statistical analysis. The approach taken here is to exploit the special structure of these problems to develop efficient numerical procedures.

An illustrative example is the eigenvalue problem. Because of their scale invariance, eigenvectors are not isolated in vector spaces. Instead, each eigendirection defines a linear subspace of eigenvectors. For numerical computation, however, it is desirable that the solution set consist only of isolated points in the search space. An obvious remedy is to impose a norm equality constraint on iterates of the algorithm. The resulting spherical search space is an embedded submanifold of the original vector space. An alternative approach is to “factor” the vector space by the scale-invariant symmetry operation such that any subspace becomes a single point. The resulting search space is a quotient manifold of the original vector space. These two approaches provide prototype structures for the problems considered in this book.

Scale invariance is just one of several symmetry properties regularly encountered in computational problems. In many cases, the underlying symmetry property can be exploited to reformulate the problem as a nondegenerate optimization problem on an embedded or quotient manifold associated with the original matrix representation of the search space. These constraint sets carry the structure of nonlinear matrix manifolds. This book provides the tools to exploit such structure in order to develop efficient matrix algorithms in the underlying total vector space.

Working with a search space that carries the structure of a nonlinear manifold introduces certain challenges in the algorithm implementation. In their classical formulation, iterative optimization algorithms rely heavily on the Euclidean vector space structure of the search space; a new iterate is generated by adding an update increment to the previous iterate in order to reduce the cost function. The update direction and step size are generally computed using a local model of the cost function, typically based on (approximate) first and second derivatives of the cost function, at each step. In order to define algorithms on manifolds, these operations must be translated into the language of differential geometry. This process is a significant research program that builds upon solid mathematical foundations. Advances
in that direction have been dramatic over the last two decades and have led
to a solid conceptual framework. However, generalizing a given optimization
algorithm on an abstract manifold is only the first step towards the objective
of this book. Turning the algorithm into an efficient numerical procedure is a
second step that ultimately justifies or invalidates the first part of the effort.
At the time of publishing this book, the second step is more an art than a
theory.

Good algorithms result from the combination of insight from differential
geometry, optimization, and numerical analysis. A distinctive feature of this
book is that as much attention is paid to the practical implementation of
the algorithm as to its geometric formulation. In particular, the concrete
aspects of algorithm design are formalized with the help of the concepts of
retraction and vector transport, which are relaxations of the classical geometi-
cal concepts of motion along geodesics and parallel transport. The proposed
approach provides a framework to optimize the efficiency of the numerical
algorithms while retaining the convergence properties of their abstract geo-
metric counterparts.

The geometric material in the book is mostly confined to Chapters 3 and 5.
Chapter 3 presents an introduction to Riemannian manifolds and tangent
spaces that provides the necessary tools to tackle simple gradient-descent op-
timization algorithms on matrix manifolds. Chapter 5 covers the advanced
material needed to define higher-order derivatives on manifolds and to build
the analog of first- and second-order local models required in most optimiza-
tion algorithms. The development provided in these chapters ranges from
the foundations of differential geometry to advanced material relevant to
our applications. The selected material focuses on those geometric concepts
that are particular to the development of numerical algorithms on embed-
ded and quotient manifolds. Not all aspects of classical differential geometry
are covered, and some emphasis is placed on material that is nonstandard
or difficult to find in the established literature. A newcomer to the field of
differential geometry may wish to supplement this material with a classical
text. Suggestions for excellent texts are provided in the references.

A fundamental, but deliberate, omission in the book is a treatment of the
group structure of Lie groups and homogeneous spaces. Lie theory is
derived from the concepts of symmetry and seems to be a natural part of
a treatise such as this. However, with the purpose of reaching a community
without an extensive background in geometry, we have omitted this material
in the present book. Occasionally the Lie-theoretic approach provides an
elegant shortcut or interpretation for the problems considered. An effort
is made throughout the book to refer the reader to the relevant literature
whenever appropriate.

The algorithmic material of the book is interlaced with the geometric ma-
terial. Chapter 4 considers gradient-descent line-search algorithms. These
simple optimization algorithms provide an excellent framework within which
to study the important issues associated with the implementation of practi-
cal algorithms. The concept of retraction is introduced in Chapter 4 as a key
INTRODUCTION

step in developing efficient numerical algorithms on matrix manifolds. The
later chapters on algorithms provide the core results of the book: the
development of Newton-based methods in Chapter 6 and of trust-region methods
in Chapter 7, and a survey of other superlinear methods such as conjugate
gradients in Chapter 8. We attempt to provide a generic development of
each of these methods, building upon the material of the geometric chapters.
The methodology is then developed into concrete numerical algorithms
on specific examples. In the analysis of superlinear and second-order meth­ods,
the concept of vector transport (introduced in Chapter 8) is used to
provide an efficient implementation of methods such as conjugate gradient
and other quasi-Newton methods. The algorithms obtained in these sections
of the book are competitive with state-of-the-art numerical linear algebra
algorithms for certain problems.

The running example used throughout the book is the calculation of in­
variant subspaces of a matrix (and the many variants of this problem). This
example is by far, for variants of algorithms developed within the proposed
framework, the problem with the broadest scope of applications and the
highest degree of achievement to date. Numerical algorithms, based on a ge­
ometric formulation, have been developed that compete with the best avail­
able algorithms for certain classes of invariant subspace problems. These
algorithms are explicitly described in the later chapters of the book and,
in part, motivate the whole project. Because of the important role of this
class of problems within the book, the first part of Chapter 2 provides a
detailed description of the invariant subspace problem, explaining why and
how this problem leads naturally to an optimization problem on a matrix
manifold. The second part of Chapter 2 presents other applications that can
be recast as problems of the same nature. These problems are the subject
of ongoing research, and the brief exposition given is primarily an invitation
for interested researchers to join with us in investigating these problems and
expanding the range of applications considered.

The book should primarily be considered a research monograph, as it
reports on recently published results in an active research area that is ex­
pected to develop significantly beyond the material presented here. At the
same time, every possible effort has been made to make the book accessible
to the broadest audience, including applied mathematicians, engineers, and
computer scientists with little or no background in differential geometry. It
could equally well qualify as a graduate textbook for a one-semester course in
advanced optimization. More advanced sections that can be readily skipped
at a first reading are indicated with a star. Moreover, readers are encouraged
to visit the book home page\footnote{1http://press.princeton.edu/titles/8586.html}
where supplementary material is available.

The book is an extension of the first author’s Ph.D. thesis [Abs03], itself a
project that drew heavily on the material of the second author’s Ph.D. the­
sis [Mah94]. It would not have been possible without the many contributions
of a quickly expanding research community that has been working in the area

---

\footnote{1http://press.princeton.edu/titles/8586.html}
over the last decade. The Notes and References section at the end of each chapter is an attempt to give proper credit to the many contributors, even though this task becomes increasingly difficult for recent contributions. The authors apologize for any omission or error in these notes. In addition, we wish to conclude this introductory chapter with special acknowledgments to people without whom this project would have been impossible. The 1994 monograph [HM94] by Uwe Helmke and John Moore is a milestone in the formulation of computational problems as optimization algorithms on manifolds and has had a profound influence on the authors. On the numerical side, the constant encouragement of Paul Van Dooren and Kyle Gallivan has provided tremendous support to our efforts to reconcile the perspectives of differential geometry and numerical linear algebra. We are also grateful to all our colleagues and friends over the last ten years who have crossed paths as coauthors, reviewers, and critics of our work. Special thanks to Ben Andrews, Chris Baker, Alan Edelman, Michiel Hochstenbach, Knut Hüper, Jonathan Manton, Robert Orsi, and Jochen Trumpf. Finally, we acknowledge the useful feedback of many students on preliminary versions of the book, in particular, Mariya Ishteva, Michel Journée, and Alain Sarlette.