MATHEMATICS OF COMPUTATION
Volume 78, Number 266, April 2009, Pages 1233–1236
S 0025-5718(08)02236-9
Article electronically published on December 30, 2008

2[90-02, 58-02, 90C30, 58Cxx]—*Optimization algorithms on matrix manifolds*, by P.-A. Absil, R. Mahony and R. Sepulchre, Princeton University Press, Princeton, New Jersey, 2008, xvi + 224 pp., hardcover, ISBN 978-0-691-13298-3, US$49.50

Continuous optimization has been a very vibrant area of research and applications since the 1950s. The area has grown tremendously during the last 60 or so years and has made many strong ties to other disciplines in mathematics and computer science as well as to system and control theory, engineering and the management sciences.

The term *continuous optimization* is generally used to encompass a very wide array of problems. We may define a *continuous optimization problem* as the problem of optimizing (minimizing or maximizing) a **function** (called the *objective function*, without loss of generality, to be minimized) over a **set** (called *the feasible solution set*). The restrictions/conditions defining the feasible solution set are called the *constraints*. Sometimes, one distinguishes between *constrained* (those with nontrivial constraints) and *unconstrained* (those with no constraints) optimization problems. However, in general, from a theoretical viewpoint, unless we further restrict the choice of the function to be optimized, the latter is not any easier than the former, in the worst case.

The above definition of the continuous optimization problem is so general that almost any mathematical problem we encounter can be cast as a continuous optimization problem. The domain of the function is usually restricted to a finite dimensional Hilbert space or a Euclidean space. However, even the exclusion of infinite dimensional problems does not completely tame what we might face as a continuous optimization problem. For example, consider the problem of minimizing a polynomial function of d variables where the polynomial has degree at most four. This amounts to restricting the nonlinearity of the function, specializing to a very well-behaved class of functions and trivializing the constraints to nothing. However, even this problem is known to be $\mathcal{NP}$-hard. Moreover, almost any optimization problem on finite dimensional Euclidean spaces can be formulated (at least in principle) as a continuous optimization problem with a linear objective function and quadratic inequalities (as the constraints). One might be tempted to think that if the number of variables is very small, perhaps then all such continuous optimization problems would be easy. However, this is also false. We can formulate a very simple looking continuous optimization problem with only four variables such that its optimal objective value is zero and with the key property: the optimal objective value of zero is attained if and only if Fermat's Last Theorem is false. Let

$$S := \{x \in \mathbb{R}^4 : x_1 \geq 1, x_2 \geq 1, x_3 \geq 1, x_4 \geq 3\}.$$

Also let

$$f(x) := (x_1^{x_4} + x_2^{x_4} - x_3^{x_4})^2 + (\sin \pi x_1)^2 + (\sin \pi x_2)^2 + (\sin \pi x_3)^2 + (\sin \pi x_4)^2.$$

It is easy to prove (and well known) that the infimum of f over S is zero and (more importantly) that the infimum is attained if and only if Fermat's Last Theorem is false.

So, providing a proof of Fermat's Last Theorem corresponds to proving that the optimal objective value of this continuous optimization problem with four variables is unattained.

We see that some continuous optimization problems can be very hard even if the number of variables is very small or even if the nonlinearity is bounded. One well-behaved class is the class of convex optimization problems (where the objective function to be minimized is a convex function and the feasible solution set is a convex set). Even for this class of optimization problems we must be careful. Consider for instance as the feasible solution set the convex hull of incidence vectors of Hamiltonian paths in a given graph. If the input is just the graph, then the problem of minimizing (even) a linear function over the feasible solution set is $\mathcal{NP}$-hard.

Given all these scary discussions, how is it that many researchers report successfully solving very many nontrivial continuous optimization problems with at least tens of thousands of variables and thousands of constraints? The key is that *they study the structure of the problem and exploit it.* This is perhaps one of the most important lessons of continuous optimization. To have some hope for being successful at solving nontrivial problems, we need to study the problem at hand, extract the special structure (if any) and exploit it.

A classical approach to continuous optimization formulates the problem as

$$\begin{array}{llll} \inf & f(x) & & \\ \text{subject to:} & g(x) & \in & \mathbb{R}^m_+, \\ & h(x) & = & 0, \end{array}$$

where $f : \mathbb{E} \to \mathbb{R}$, $g : \mathbb{E} \to \mathbb{R}^m$, $h : \mathbb{E} \to \mathbb{R}^p$ ($\mathbb{E}$ denotes a Euclidean space, $\mathbb{R}^m_+$ denotes the nonnegative orthant in $\mathbb{R}^m$). As we hinted above, the formulation phase is extremely important. So, when we separate the constraints as above, we usually have (good reasons to hope for the existence of) some

$$\bar{x} \in \operatorname{int}\left\{x \in \mathbb{E} : g(x) \in \mathbb{R}^m_+\right\} \text{ such that } h(\bar{x}) = 0.$$

So, from a local perspective, it is relatively "easy" or "standard" to deal with the constraint $g(x) \in \mathbb{R}^m_+$ in an iterative scheme. However, the equality constraints $h(x) = 0$ are the most problematic in theory as well as in numerical computations (unless $h : \mathbb{E} \to \mathbb{R}^p$ is affine). The book under consideration deals with this issue head on, and it applies the main principle of continuous optimization (or any other general approach to solving a wide class of problems), namely *study the problem, extract useful structure and exploit it.*

The authors consider the situation when the set

$$\{x \in \mathbb{E} : h(x) = 0\}$$

admits an atlas and hence can be considered a manifold. Then, they exploit this manifold structure when they design the algorithms. Their approach to the subject can be called a *differential geometric optimization approach.* Also the authors place a lot of emphasis on preserving the matrix structure (when it exists) in intermediate steps of the development.

The book is made up from eight chapters and an appendix. We briefly discuss each of these nine pieces below.

Chapter 1 is the Introduction, but it is really an extended version of a preface, without getting into the mathematics of the subject. The book starts in Chapter 2 in earnest; it is entitled Motivation and Applications. To motivate the equality constraint structure, the authors begin with the calculation of the Rayleigh Quotient and a generalization of it. Hence the norm constraint and in general a normalizing constraint for functions that are homogeneous of degree zero are introduced. Then, singular value decomposition, the closest matrix problem via singular value decomposition, and rank-constrained matrix approximation problems are discussed. These applications lead to the oblique manifold and the Stiefel manifold. The chapter concludes with (strictly speaking, every chapter except Chapter 1 ends with Notes and References) independent component analysis, pose estimation and motion recovery applications. These lead to the usage of the function $-\ln(\det(\cdot))$ over the cone of Hermitian positive definite matrices as well as to the oblique manifold and to the normalized essential manifold.

Chapter 3 develops the first order geometry necessary for the optimization algorithms on matrix manifolds. After some basic information on manifolds, notions of differentiability and diffeomorphism are defined. Then immersions, submersions and submanifolds are introduced. Quotient manifolds are developed in a self-contained manner. The discussion then moves on to the tangent vectors. These are illustrated on the hypersphere, orthogonal Stiefel manifold, the orthogonal group, real projective space and the Grassmann manifolds. At this stage the reader is armed with a notion of the *search direction*; however, to talk about *steepest-descent* or *step size*, one needs to establish a notion of length (or distance). So, Riemannian metrics are discussed next. Riemannian submanifolds and Riemannian quotient manifolds are defined and illustrated on the above-mentioned, commonly occurring examples.

Now that the reader is provided with the essential notions of *search direction* and *step size*, Chapter 4 presents some line search algorithms on manifolds. After retractions are covered, the first algorithm of the book is presented: accelerated line search. Global convergence results for this algorithm are derived next. Stability of the fixed points of mappings and convergence rates are discussed. The rest of the presentation is centered around the Rayleigh quotient example. The Armijo line search algorithms for the Rayleigh quotient on the hypersphere and the Rayleigh quotient on Grassmann manifolds anchor the rest of the chapter.

Chapter 5 is the next step toward developing the second order methods. It contains the second order geometry on matrix manifolds. After motivating the discussion with Newton's method (so that the reader knows what the next goal is), the authors cover affine connections and the Riemannian connection. Geodesics and the exponential mapping are introduced next and these are illustrated on the hypersphere, orthogonal Stiefel manifold and the Grassmann manifolds. Two notions of a "Hessian," namely the Riemannian Hessian operator and the so-called *second covariant derivative*, nicely round out the chapter.

Analogous to the beginning of Chapter 4, now the readers are ready for the discussion of the second order methods. Chapter 6 is appropriately named: Newton's Method. The algorithms, their specializations to the Rayleigh quotient type problems and local convergence proofs are presented.

A very important approach to solving continuous optimization problems is the trust-region approach. Chapter 7 is Trust-Region Methods. As in the previous chapters, the authors derive the main ingredient of the algorithms first (such as the construction of a local quadratic model for the objective function and the trust-region step), and then design and present the algorithms. A global convergence analysis as well as a local convergence analysis are presented. Matrix/eigenvalue problem applications are highlighted and even the results of some numerical experiments are presented.

The final chapter is called "A Constellation of Superlinear Algorithms." It discusses the *hybrid* algorithms. These are the algorithms that try to keep good global convergence properties while trying to get close to Newton's method (to attain local quadratic convergence locally, under some favorable conditions—hopefully not very strict). After establishing the concepts of vector transport and vector transport by differentiated retraction, the authors present the inexact Newton method, secant methods (or quasi-Newton methods), conjugate gradient methods, least-square, Gauss-Newton and Levenberg-Marquardt methods.

The appendix is brief and to the point. It has five parts covering those very elementary aspects of linear algebra, topology, Taylor's theorem, functions and their differentiation, and the notation for the asymptotic behavior of sequences that are needed in the rest of the text.

The references and the index are adequate. The references include many good pointers to the literature that can help a reader who might be unfamiliar with some of the areas on which the book is built.

I enjoyed reading this book and recommend it. Also, in my view, the book is very suitable to be used as a textbook for a one-semester graduate course.

Levent Tunçel
Department of Combinatorics and Optimization
Faculty of Mathematics
University of Waterloo
Waterloo, Ontario N2L 3G1