Optimization On Manifolds: Methods And Applications

P.-A. Absil[1], R. Mahony[2], and R. Sepulchre[3]

[1] Department of Mathematical Engineering, Université catholique de Louvain,
    B-1348 Louvain-la-Neuve, Belgium http://www.inma.ucl.ac.be/~absil/
[2] Australian National University, ACT, 0200, Australia
[3] Department of Electrical Engineering and Computer Science, Université de
    Liège, Belgium

Summary. This paper provides an introduction to the topic of optimization on manifolds. The approach taken uses the language of differential geometry, however, we choose to emphasise the intuition of the concepts and the structures that are important in generating practical numerical algorithms rather than the technical details of the formulation. There are a number of algorithms that can be applied to solve such problems and we discuss the steepest descent and Newton's method in some detail as well as referencing the more important of the other approaches. There are a wide range of potential applications that we are aware of, and we briefly discuss these applications, as well as explaining one or two in more detail.

1 Introduction

This paper is written as an invitation for the reader to the area of optimization on manifolds. It follows quite closely the structure of the plenary talk given by the first author at the 14th Belgian-French-German Conference on Optimization, Leuven, 14–18 September 2009. The style is rather on the informal side, and there is a definite bias towards the exposition given in the monograph [5], to which we refer for more details and for a larger bibliography. When we cite [5], we do not imply that it is the original reference for the topic in question and we refer the reader to the "Notes and References" sections of [5] for details of the history.

   The general problem of optimization on manifolds is introduced in Section 2. A motivation for considering the problem in its most abstract form is given in Section 3. Manifolds are defined in more technical terms in Section 4. Several specific manifolds are presented in Section 5, along with pointers to applications where they are involved. Section 6 describes a steepest-descent optimization scheme on Riemannian manifolds. Its application to a simple problem is worked out in Section 7. Section 8 is dedicated to Newton's
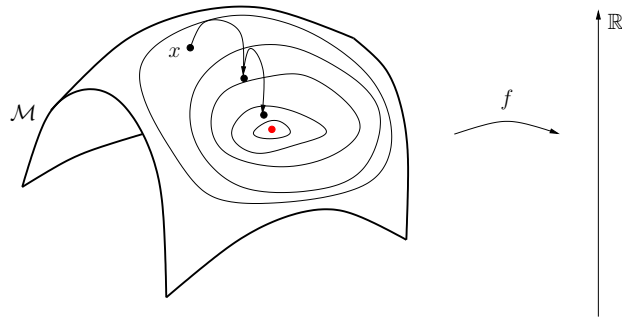
Fig. 1. Optimization on manifolds in one picture.

method on Riemannian manifolds. Other optimization methods on manifolds are briefly discussed in Section 9. Section 10 provides some conclusions.

2 Optimization on manifolds in one picture

The archetypal problem in optimization on manifolds is pictured in Figure 1. The set of feasible points is a manifold $\mathcal{M}$ that, for the sake of developing intuition, can be viewed as a smooth surface. We will argue in Section 3, however, that it is beneficial to depart from this restrictive representation. Anticipating Section 4, one can think of $\mathcal{M}$ as a collection of points, endowed with a yet-to-be-defined *manifold structure* that turns $\mathcal{M}$ into a topological set—so we can talk about neighborhoods—and that makes it possible to declare whether a real-valued function on $\mathcal{M}$ is smooth or not. The reader who cannot wait to get a more precise definition of the concept of a manifold is invited to take a peek at Section 4. It may also be reassuring to have a look at the list of specific manifolds in Section 5.

The smooth real-valued function f on the set $\mathcal{M}$ that defines the goal of the optimization problem is termed the *objective function*. A few of its level curves $f^{-1}(c)$, $c \in \mathbb{R}$, are represented in Figure 1. The dot inside the level curves is an optimal point of f, say, a minimizer of f.

Computing minimizers of f is our goal. More precisely, the problem is as follows:

Problem 1 (optimization on manifolds).
Given: a manifold $\mathcal{M}$ and a smooth function $f : \mathcal{M} \to \mathbb{R}$.
Sought: an element x_* of $\mathcal{M}$ such that there is a neighborhood V of x_* in $\mathcal{M}$ with $f(x_*) \leq f(x)$ for all $x \in V$.

Such an x_* is termed a *local minimizer* of f.

The methods we are interested in for solving Problem 1 are iterative algorithms on the manifold $\mathcal{M}$. Given a starting point $x_0 \in \mathcal{M}$, such an algorithm

produces a sequence $(x_k)_{k \geq 0}$ in $\mathcal{M}$ that converges to x_* whenever x_0 is in a certain neighborhood, or basin of attraction, of x_*. As in classical optimization algorithms, the following properties are desirable: (i) the set of points x_0 for which convergence to x_* occurs should be large; (ii) convergence to x_* should be fast; (iii) the numerical effort required to compute each new iterate should be reasonable.

3 Why consider general manifolds?

A motivation for considering general manifolds—and not only manifolds that come to us as subsets of Euclidean spaces—is that they offer an adequate common framework for dealing with the following two problems.

Problem 2. Given a matrix $A = A^T \in \mathbb{R}^{n \times n}$ and a diagonal $n \times n$ matrix $N = \mathrm{diag}(1, \ldots, p)$ with $p \leq n$, solve

$$\min f(X) = \mathrm{trace}(X^T A X N)$$
$$\text{subj. to } X \in \mathbb{R}^{n \times p}, X^T X = I.$$

Solving this problem yields the p "leftmost" eigenvectors of A, i.e., those associated with the p algebraically smallest eigenvalues of A; see Section 7 or [5, §4.8] for details.

The optimization domain in Problem 2 is the set

$$\mathrm{St}(p, n) = \{X \in \mathbb{R}^{n \times p} : X^T X = I\}, \tag{1}$$

which is a subset of the Euclidean space $\mathbb{R}^{n \times p}$. If a subset of a Euclidean space can be locally smoothly straightened—the *submanifold property*—, then it admits one and only one "natural" manifold structure [5, Prop. 3.3.2]; see Figure 2 for an illustration. The set $\mathrm{St}(p, n)$ happens to be such a subset [5, §3.3.2]. Endowed with its natural manifold structure, $\mathrm{St}(p, n)$ is termed the *Stiefel manifold* of orthonormal p-frames in $\mathbb{R}^n$.[4]

Problem 3. Given a matrix $A = A^T \in \mathbb{R}^{n \times n}$, solve

$$\min f(Y) = \mathrm{trace}\left((Y^T Y)^{-1} Y^T A Y\right)$$
$$\text{subj. to } Y \in \mathbb{R}_*^{n \times p},$$

where $\mathbb{R}_*^{n \times p}$ denotes the set of all full-rank $n \times p$ matrices $(p < n)$.

---

[4] The Stiefel manifold is named in honor of Eduard L. Stiefel who studied its topology in [55]. Stiefel is perhaps better known for proposing with M. R. Hestenes the conjugate gradient method [25]. Incidentally, he was born 100 years ago.
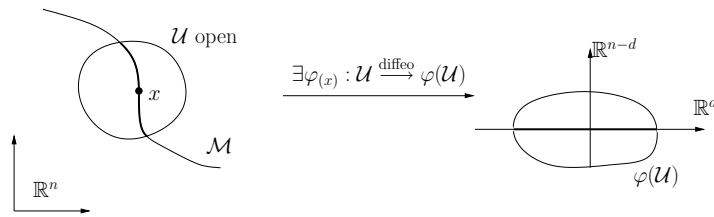
Fig. 2. The *set* $\mathcal{M} \subset \mathbb{R}^n$ is termed a *submanifold* of $\mathbb{R}^n$ if the situation described above holds for all $x \in \mathcal{M}$. Charts for $\mathcal{M}$ are obtained by extracting the d first coordinates.

The function f in Problem 3 has the following invariance property:

$$f(YM) = f(Y), \quad \text{for all } Y \in \mathbb{R}_*^{n \times p} \text{ and all } M \in \mathbb{R}_*^{p \times p}. \tag{2}$$

In other words, f is constant on each equivalence class

$$[Y] = \{YM : M \in \mathbb{R}_*^{p \times p}\}, \tag{3}$$

$Y \in \mathbb{R}_*^{n \times p}$. The equivalence class $[Y]$ is precisely the set of all $n \times p$ matrices that have the same column space as Y, and Y is a minimizer of f if and only if the column space of Y is a p-dimensional minor eigenspace of A (i.e., associated with the smallest eigenvalues); see [5, §2.1.1]. It is thus tempting to reconsider Problem 3 on a search space whose elements are the equivalence classes $[Y]$, $Y \in \mathbb{R}_*^{n \times p}$, and optimize the function

$$\check{f} : \{[Y] : Y \in \mathbb{R}_*^{n \times p}\} \to \mathbb{R} : [Y] \mapsto f(Y), \tag{4}$$

which is well defined in view of (2) and (3).

A major advantage of this reformulation of Problem 3 is that, for generic A, the minimizers of $\check{f}$ are isolated, while they are never isolated in the original formulation in view of the invariance property (2). The apparent downside is that the new search space, the quotient space

$$\mathrm{Gr}(p, n) = \{[Y] : Y \in \mathbb{R}_*^{n \times p}\}, \tag{5}$$

is no longer a Euclidean space. However, it turns out (see [5, §3.4.4]) that (5) admits one and only one "natural" manifold structure, which is inherited from the fact that, around every element of $\mathbb{R}_*^{n \times p}$, the bundle of equivalence classes can be smoothly straightened; see Figure 3 for an illustration. Endowed with this natural manifold structure, the set $\mathrm{Gr}(p, n)$ is termed the *Grassmann manifold*[5] of p-planes in $\mathbb{R}^{n \times p}$. (The set (5) is identified with the set of all p-dimensional subspaces of $\mathbb{R}^n$ because $[Y]$ is the set of all $n \times p$ matrices whose

---

[5] The Grassmann manifold is named in honor of Hermann Günther Graßmann who proposed a representation of the manifold known as *Plücker coordinates*. Graßmann is perhaps better known for his Sanskrit dictionary and his translation of the Rgveda [21]. Incidentally, he was born 200 years ago.
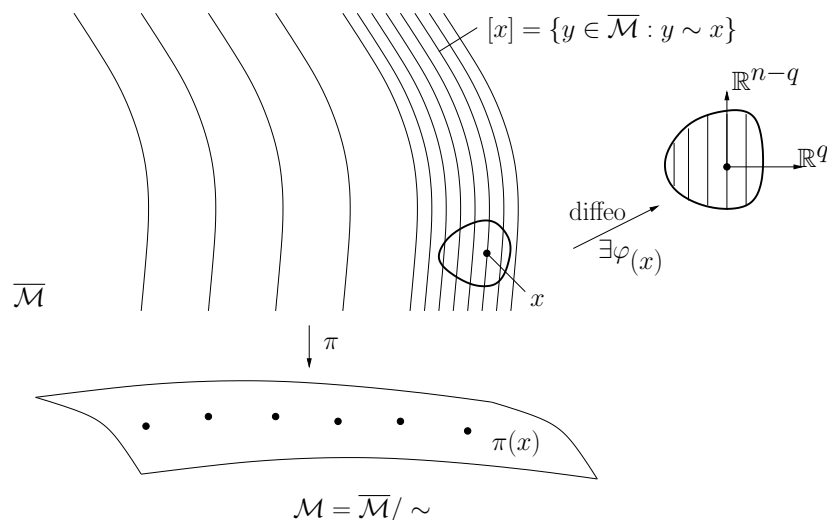
Fig. 3. The set $\overline{\mathcal{M}}/\sim \; := \{[x] : x \in \overline{\mathcal{M}}\}$ is termed a *quotient manifold* if the situation described above holds for all $x \in \overline{\mathcal{M}}$. Charts for $\overline{\mathcal{M}}/\sim$ are obtained by extracting the q first coordinates.

columns form a basis of the same p-dimensional subspace of $\mathbb{R}^n$.) Dealing with optimization problems such as the minimization of (4) is precisely what optimization on (quotient) manifolds is all about.

In summary, Problem 2 and the reformulated Problem 3 have the following properties in common: (i) their search space admits a natural manifold structure; (ii) in the sense of the manifold structure, the objective function is smooth, as a consequence of [33, Prop. 8.22] and [33, Prop. 7.17]. In the next section, we explain more technically what a manifold structure is, and what it means for a objective function on a manifold to be smooth.

4 Manifolds and smooth objective functions

The time has come to give an informal, application-driven, definition of a manifold structure. Details can be found in [5, §3.1.1] or in any textbook on differential geometry.

The intuition can be obtained from Figure 4. We are given a set $\mathcal{M}$, which initially is just a collection of points without any particular structure, and we are given a real-valued function f on the set $\mathcal{M}$. Since $\mathcal{M}$ does not have a vector space structure, the classical definition of differentiability of a function $f : \mathcal{M} \to \mathbb{R}$ at a point $x \in \mathcal{M}$ does not apply. The remedy is to consider a one-to-one correspondence φ between a subset $\mathcal{U}$ of $\mathcal{M}$ containing x and an open subset $\varphi(\mathcal{U})$ of some $\mathbb{R}^d$. Then f is declared to be differentiable at $x \in \mathcal{M}$ when
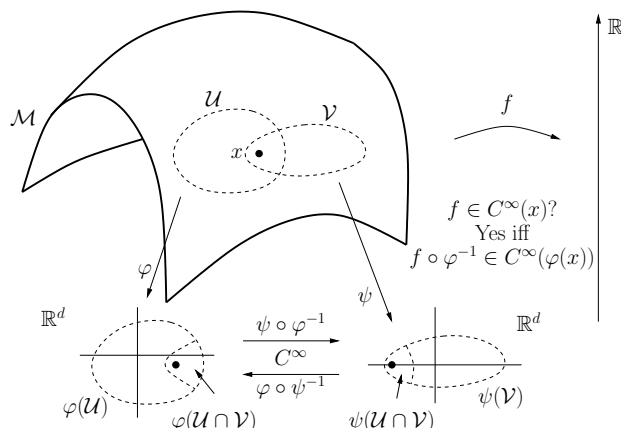
Fig. 4. Manifold structures and smoothness of objective functions.

the function $f \circ \varphi^{-1} : \varphi(\mathcal{U}) \to \mathbb{R} : y \mapsto f(\varphi^{-1}(y))$ is differentiable at $\varphi(x)$. Since $\varphi(\mathcal{U})$ is an open subset of $\mathbb{R}^d$, the usual definition of differentiability applies.

For this procedure to be applicable to every point of the set $\mathcal{M}$, we need to provide a collection of φ's such that the union of their domains is the whole set $\mathcal{M}$. Moreover, whenever the domains $\mathcal{U}$ and $\mathcal{V}$ of two correspondences φ and ψ overlap on a point $x \in \mathcal{M}$, we must require that, for all $f : \mathcal{M} \to \mathbb{R}$, $f \circ \varphi^{-1}$ is differentiable at $\varphi(x)$ if and only if $f \circ \psi^{-1}$ is differentiable at $\psi(x)$; otherwise differentiability of f at x is not well defined. This goal is achieved by imposing that the charts *overlap smoothly*, i.e., $\psi \circ \varphi^{-1}$ is a diffeomorphism—a smooth bijection with smooth inverse—between $\varphi(\mathcal{U} \cap \mathcal{V})$ and $\psi(\mathcal{U} \cap \mathcal{V})$. The collection of correspondences is then called an *atlas*, and the correspondences are called *charts*. The *maximal atlas* generated by an atlas is the collection of all charts that overlap smoothly with those of the given atlas. Finally, a *manifold* is a pair $(\mathcal{M}, \mathcal{A}^+)$, where $\mathcal{M}$ is a set and $\mathcal{A}^+$ is a maximal atlas on the set $\mathcal{M}$. In other words, a maximal atlas uniquely specifies a manifold structure on $\mathcal{M}$. For brevity, it is common to say "the manifold $\mathcal{M}$" when the maximal atlas is clear from the context or irrelevant.

Let us work out an example. When $p = 1$ and $n = 2$, the Stiefel manifold (1) reduces to the unit circle in $\mathbb{R}^2$. Let $\mathcal{U} = \mathrm{St}(1,2) \setminus \{(0,1),(0,-1)\}$, $\varphi : \mathcal{U} \to \mathbb{R} : x \mapsto x_2/x_1$, $\mathcal{V} = \mathrm{St}(1,2) \setminus \{(1,0),(-1,0)\}$, $\psi : \mathcal{V} \to \mathbb{R} : x \mapsto x_1/x_2$. Then $\{\varphi, \psi\}$ is an atlas of the set $\mathrm{St}(1,2)$. Moreover, it can be shown that this atlas induces the natural manifold structure mentioned in the previous section.

Let us show that the objective function f defined in Problem 2 is smooth. To this end, pick $x \in \mathrm{St}(1,2)$, and assume that $x \in \mathcal{U} \cap \mathcal{V}$. Observe that $\varphi^{-1}(y) = \frac{1}{\sqrt{1+y^2}} \begin{bmatrix} 1 & y \end{bmatrix}^T$ for all $y \in \mathbb{R}$. Hence $f \circ \varphi^{-1}(y) = \frac{1}{\sqrt{1+y^2}} \begin{bmatrix} 1 & y \end{bmatrix} A \begin{bmatrix} 1 \\ y \end{bmatrix}$, and we see that f is a smooth function on $\mathcal{U}$. A similar reasoning shows that f

is a smooth function on $\mathcal{V}$. Hence f is smooth on the whole manifold $\mathcal{M}$. (An alternate way of obtaining this result is by invoking the fact [33, Prop. 8.22] that the restriction of a smooth function to a submanifold is smooth.)

Looking back at the original Problem 1, we see that all the concepts involved therein are now well defined, except for "neighborhood". The notion of neighborhood in $\mathcal{M}$ is directly inherited from its manifold structure: a *neighborhood* of a point x in a manifold $\mathcal{M}$ is a subset of $\mathcal{M}$ that contains a set of the form $\varphi^{-1}(\Omega)$, where φ is a chart of the manifold $\mathcal{M}$ whose domain contains x and Ω is an open subset that contains $\varphi(x)$.

If all the charts of the maximal atlas are into the same $\mathbb{R}^d$, then d is called the *dimension* of the manifold. In particular, when the manifold is connected, its dimension is well defined.

Finally, we point out that the notion of smoothness extends to functions between two manifolds: the definition relies on expressing the function in charts and checking whether this expression is smooth. Note also that the Cartesian product of two manifolds admits a manifold structure in a natural way.

5 Specific manifolds, and where they appear

In this section, we present a few specific manifolds, and we discuss their use in science and engineering applications.

5.1 Stiefel manifold

The (compact) Stiefel manifold $\mathrm{St}(p,n)$ is the set of all p-tuples $(x_1,\ldots,x_p)$ of orthonormal vectors in $\mathbb{R}^n$. The notation $V_{n,p}$ or $V_p(\mathbb{R}^n)$ is also frequently encountered in the literature.

If we view $\mathbb{R}^n$ as the space of length-n column vectors and turn the p-tuples into $n \times p$ matrices,

$$(x_1,\ldots,x_p) \mapsto \begin{bmatrix} x_1 \cdots x_p \end{bmatrix},$$

we obtain the definition (1), i.e.,

$$\mathrm{St}(p,n) = \{X \in \mathbb{R}^{n \times p} : X^T X = I\}.$$

To relate this definition with the illustration in Figure 1, imagine that each point of $\mathcal{M}$ stands for an orthonormal p-frame $(x_1,\ldots,x_p)$, and that the objective function f assigns a real value to each orthonormal p-frame. We have already encountered such an f in Problem 2.

Here are a few domains of application for optimization methods on the Stiefel manifold, along with related references, which are by no means exhaustive: principal component analysis and the singular value decomposition [24, 18]; independent component analysis and the related problem of joint diagonalization of matrices [8, 42, 26, 56]; more generally, several applications

related to machine learning [46, 57, 15]; Procrustes problems [20, 38]; computer vision [34, 59]; Lyapunov exponent computation for dynamical systems [14].

5.2 Sphere

When $p = 1$, the Stiefel manifold $\mathrm{St}(p, n)$ reduces to the unit sphere S^{n-1}, a particularly simple nonlinear manifold.

5.3 Orthogonal group

When $p = n$, the Stiefel manifold $\mathrm{St}(p, n)$ admits a group structure, where the group operation is the matrix product. This group is termed the *orthogonal group*, often denoted by O_n or $\mathrm{O}(n)$. Moreover, the group operation and its inverse are smooth in the sense of the manifold structure of $\mathrm{St}(p, n)$. This makes O_n a *Lie group*. For more information on Lie groups at an introductory level, see, e.g., [62].

The orthogonal group O_n has two connected components. The component that contains the identity matrix is called the *special orthogonal group* $\mathrm{SO}(n)$. The set $\mathrm{SO}(3)$ corresponds to the set of rotations.

5.4 Grassmann manifold

The Grassmann manifold $\mathrm{Gr}(p, n)$ is the set of all p-dimensional subspaces of $\mathbb{R}^n$. Most applications bear some relation with dimensionality reduction: [24, 18, 40, 53, 4, 39, 54, 12, 23, 52, 59, 15, 27].

5.5 Set of fixed-rank positive-semidefinite matrices

The differential geometry of the set

$$\mathrm{S}_+(p, n) = \{X \in \mathbb{R}^{n \times n} : X \succeq 0, \mathrm{rk}(X) = p\}$$

is a topic of interest, in view of its application in rank reduction of positive-definite matrices [13, 30, 60, 61].

5.6 Shape manifold

A quotient geometry arises because the notion of shape is invariant by rotation and by reparameterization; see, e.g., [31, 32, 29].

5.7 Oblique manifold and products of spheres

The oblique manifold $\{Y \in \mathbb{R}_*^{n \times p} : \mathrm{diag}(YY^T) = I_p\}$—where $\mathrm{diag}(YY^T) = I_p$ means that the rows of Y belong to the unit sphere—and Cartesian products of spheres appear, e.g., in the oblique Procrustes problem [58], in nonorthogonal joint diagonalization [3], and in time-varying system identification [49].

5.8 Flag manifold

Given $0 < p_1 < \ldots < p_k$, the flag manifold of type $(p_1, \ldots, p_k)$ is the collection of all k-tuples of linear subspaces of $\mathbb{R}^{p_k}$ $(V_1, \ldots, V_k)$ with $\dim(V_i) = p_i$ and V_i subspace of V_{i+1}. Flag manifolds are useful in the analysis of eigenvalue methods [9, 28] and in independent subspace analysis [47].

5.9 Essential manifold

An essential matrix is the product $E = \Omega R$ of a skew-symmetric matrix Ω and a rotation matrix R. The essential manifold appears in stereo vision processing [41, 22].

5.10 Other products of manifolds

Various Cartesian products of manifolds appear in applications. For example, the Euclidean group SE(3), an important manifold in computer vision and robotics, can be identified with $\mathrm{SO}(3) \times \mathbb{R}^3$. A product of 16 copies of SO(3) was used in [7] to specify the position of a human spine.

The next step is to consider products of infinitely many copies of a manifold, which brings us to curve fitting on manifolds; see [50] and references therein. See also [51] where the problem consists in finding a curve in the Euclidean group SE(3).

5.11 Other quotient manifolds

Quotient manifolds appear in several applications where the objective function has an invariance property that induces a regular equivalence relation; a characterization of regular equivalence relations can be found in [5, Prop. 3.4.2]. In fact, most of the manifolds above admit well-known quotient representations. For example, $\mathrm{St}(p,n)$ can be identified with $\mathrm{O}(n)/\mathrm{O}(n-p)$; see [18] for details.

6 Steepest descent: from $\mathbb{R}^n$ to manifolds

We now turn to optimization algorithms on manifolds. Amongst optimization methods on manifolds that exploit the smoothness of the cost function, the steepest-descent scheme is arguably the most basic.

The next table, where $\boldsymbol{\nabla} f(x) = \begin{bmatrix} \partial_1 f(x) & \cdots & \partial_n f(x) \end{bmatrix}^T$ denotes the classical Euclidean gradient, sketches a comparison between steepest-descent in $\mathbb{R}^n$ and its generalization to manifolds. An illustration is given in Figure 5.
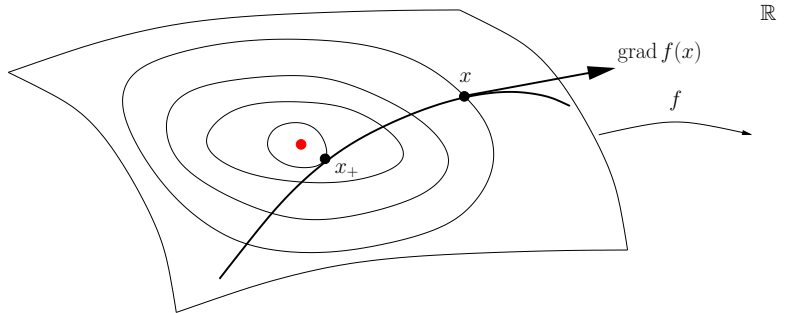
Fig. 5. Steepest descent on Riemannian manifolds.

	$\mathbb{R}^n$	Manifold
Search direction	Vector at x	Tangent vector at x
Steepest-desc. dir.	$-\boldsymbol{\nabla} f(x)$	$-\mathrm{grad}\, f(x)$
Search curve	$\gamma : t \mapsto x - t\,\boldsymbol{\nabla} f(x)$	γ s.t. $\gamma(0) = x$ and $\dot{\gamma}(0) = -\mathrm{grad}\, f(x)$

Figure 5 corresponds to a submanifold of a Euclidean space. However, we are interested in a theory that subsumes both submanifolds and quotient manifolds, for which we will need definitions of tangent vectors and gradients that are rather abstract. Nevertheless, the reader is invited to keep Figure 5 in mind, because it helps in developing the intuition.

The particularization of the abstract steepest-descent scheme to submanifolds of Euclidean spaces is rather simple and will be covered in this paper. For quotient manifolds, the situation is a bit more complicated, and we refer to [5] for details.

6.1 Tangent vectors and tangent spaces

The notion of a tangent vector at a point $x \in \mathcal{M}$ is intuitively clear when $\mathcal{M}$ is a submanifold of a Euclidean space $\mathcal{E}$. To obtain a tangent vector at x, take a smooth curve $\gamma : \mathbb{R} \to \mathcal{M}$ with $\gamma(0) = x$; then $\dot{\gamma}(0)$—the derivative of γ at $t = 0$—is a tangent vector to $\mathcal{M}$ at x. Here the derivative is the usual derivative: since $\mathcal{M}$ is a subset of the Euclidean space $\mathcal{E}$, γ can be viewed as a curve in $\mathcal{E}$, and the derivative of γ is understood in this sense. The set of all tangent vectors at x is termed the *tangent space* to $\mathcal{M}$ at x and denoted by $T_x\mathcal{M}$. Given $\xi_x \in T_x\mathcal{M}$, we say that a curve γ on $\mathcal{M}$ *realizes* ξ_x if $\gamma(0) = x$ and $\dot{\gamma}(0) = \xi_x$.

A tangent vector ξ_x can be paired with any smooth real-valued function f on $\mathcal{M}$ to yield the real number

$$\mathrm{D}f(x)[\xi_x] = \left.\frac{\mathrm{d}}{\mathrm{d}t} f(\gamma(t))\right|_{t=0}, \qquad (6)$$

where γ is any curve that realizes ξ_x. This property is the key to generalizing tangent vectors to abstract manifolds. A mapping $\xi_x : f \mapsto \xi_x(f)$ is a *tangent vector* to $\mathcal{M}$ at x is there exists a curve γ on $\mathcal{M}$ such that $\gamma(0) = x$ and $\xi_x(f) = \frac{\mathrm{d}}{\mathrm{d}t} f(\gamma(t))\big|_{t=0}$ for all smooth real-valued functions f on $\mathcal{M}$. Again, the curve γ is said to *realize* ξ_x. An alternate notation for $\xi_x(f)$ is $\mathrm{D}f(x)[\xi_x]$, but one should bear in mind that it is only for $\mathcal{M}$ submanifold of a Euclidean space that $\mathrm{D}f(x)[\xi_x]$ is equal to $\lim_{t \to 0} \frac{\bar{f}(x+t\xi_x)-\bar{f}(x)}{t}$ for any smooth extension $\bar{f}$ of f.

The above is a curve-based definition of tangent vectors; several equivalent definitions can be found in the literature. We also point out that the disjoint union of the tangent spaces admits a natural manifold structure. This manifold is called the *tangent bundle* and denoted by $T\mathcal{M}$. This concept will reappear below when we introduce the notion of retraction.

6.2 Descent directions

With the notion of a tangent vector at hand, we can define a *descent direction* for an objective function f on a manifold $\mathcal{M}$ at a point x to be a tangent vector ξ_x at x such that $\mathrm{D}f(x)[\xi_x] < 0$. In this case, for any curve γ that realizes ξ_x, we have $\frac{\mathrm{d}}{\mathrm{d}t} f(\gamma(t))\big|_{t=0} < 0$. Hence, for all t positive and sufficiently small, $f(\gamma(t)) < f(x)$.

6.3 Steepest-descent direction and the gradient

By definition, the steepest ascent direction is along

$$\arg\max_{\substack{\xi_x \in T_x\mathcal{M} \\ \|\xi_x\|=1}} \mathrm{D}f(x)[\xi_x].$$

For this expression to be well-defined, we need a norm on $T_x\mathcal{M}$. The most convenient way of introducing such a norm is via an inner product. For all $x \in \mathcal{M}$, let g_x be an inner product in $T_x\mathcal{M}$, and define

$$\|\xi_x\| := \sqrt{g_x(\xi_x, \xi_x)}.$$

When g_x smoothly depends on x, $(\mathcal{M}, g)$ is termed a *Riemannian manifold*. As was the case with the maximal atlas, the notation $(\mathcal{M}, g)$ is often replaced by $\mathcal{M}$ when no confusion arises.

There is a unique element of $T_x\mathcal{M}$, called the *gradient* of f at x and denoted by $\operatorname{grad} f(x)$, such that

$$\begin{cases} \operatorname{grad} f(x) \in T_x\mathcal{M} \\ g_x(\operatorname{grad} f(x), \xi_x) = \mathrm{D}f(x)[\xi_x], \quad \forall \xi_x \in T_x\mathcal{M}. \end{cases}$$

The gradient of f at x, whose definition depends on the Riemannian metric, is along the steepest-ascent direction of f at x, whose definition also depends on the Riemannian metric:

$$\frac{\operatorname{grad} f(x)}{\|\operatorname{grad} f(x)\|} = \underset{\substack{\xi_x \in T_x \mathcal{M} \\ \|\xi_x\| = 1}}{\arg \max} \, \mathrm{D} f(x)[\xi_x].$$

Hence, the steepest-descent direction is along $-\operatorname{grad} f(x)$.

Moreover, the norm of the gradient of f at x is equal to the slope at $t = 0$ of $t \mapsto f(\gamma(t))$, where γ is any curve that realizes $\frac{\operatorname{grad} f(x)}{\|\operatorname{grad} f(x)\|}$:

$$\|\operatorname{grad} f(x)\| = \mathrm{D} f(x) \left[\frac{\operatorname{grad} f(x)}{\|\operatorname{grad} f(x)\|} \right].$$

6.4 Gradient on submanifolds

Let $(\overline{\mathcal{M}}, \overline{g})$ be a Riemannian manifold and $\mathcal{M}$ be a submanifold of $\overline{\mathcal{M}}$. Then

$$g_x(\xi_x, \zeta_x) := \overline{g}_x(\xi_x, \eta_x), \ \forall \xi_x, \zeta_x \in T_x \mathcal{M}$$

defines a Riemannian metric g on $\mathcal{M}$. With this Riemannian metric, $\mathcal{M}$ is a *Riemannian submanifold* of $\overline{\mathcal{M}}$. Let $T_x^\perp \mathcal{M}$ denote the orthogonal complement of $T_x \mathcal{M}$ in $T_x \overline{\mathcal{M}}$ in the sense of $\overline{g}$. Every $z \in T_x \overline{\mathcal{M}}$ admits a decomposition $z = \mathrm{P}_x z + \mathrm{P}_x^\perp z$, where $\mathrm{P}_x z$ belongs to $T_x \mathcal{M}$ and $\mathrm{P}_x^\perp z$ to $T_x^\perp \mathcal{M}$. If $\bar{f} : \overline{\mathcal{M}} \to \mathbb{R}$ and $f = \bar{f}|_{\mathcal{M}}$, then

$$\operatorname{grad} f(x) = \mathrm{P}_x \operatorname{grad} \bar{f}(x). \tag{7}$$

6.5 Gradient on quotient manifolds

For the case of quotient manifolds, see [5, §3.6.2].

6.6 Choice of the search curve

The next task is to choose a curve γ through x at $t = 0$ such that

$$\dot{\gamma}(0) = -\operatorname{grad} f(x).$$

The curve selection process can be specified by a retraction. A *retraction* on $\mathcal{M}$ is a smooth mapping $R : T\mathcal{M} \to \mathcal{M}$ such that, for all $x \in \mathcal{M}$ and all $\xi_x \in T_x \mathcal{M}$,

1. $R(0_x) = x$, where 0_x denotes the origin of $T_x \mathcal{M}$;
2. $\frac{\mathrm{d}}{\mathrm{d}t} R(t\xi_x)\big|_{t=0} = \xi_x$.

Given a retraction R on $\mathcal{M}$, the curve $\gamma : t \mapsto R(-t\operatorname{grad} f(x))$ is a descent curve at $t = 0$ provided that $\operatorname{grad} f(x) \neq 0$.

Note that, in topology, a continuous map from a topological space X to a subspace A is termed a retraction if the restriction of the map to domain A is the identity map on A. In view of the property $R(0_x) = x$ and the natural inclusion of $\mathcal{M}$ in $T\mathcal{M}$, the differential-geometric retractions are topological retractions.

6.7 Line-search procedure

It remains to find t_* such that $f(\gamma(t_*))$ is sufficiently smaller than $f(\gamma(0))$. Since $t \mapsto f(\gamma(t))$ is simply a function from $\mathbb{R}$ to $\mathbb{R}$, we can use the step selection techniques that are available for classical line-search methods, e.g., exact minimization or Armijo backtracking.

The next iterate of the steepest-descent method is defined to be $x_+ = \gamma(t_*)$. Observe that the method can be tuned by modifying the Riemannian metric and the retraction.

7 A steepest-descent method for Problem 2

As an illustration, we apply the steepest-descent method of the previous section to Problem 2.

Let $A = A^T \in \mathbb{R}^{n \times n}$ with (unknown) eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$. The goal is to compute the p dominant eigenvectors of A, i.e., those associated to $\lambda_1, \ldots, \lambda_p$, which are uniquely defined (up to sign reversal, assuming a unit-norm constraint) if $\lambda_1 > \cdots > \lambda_p$. To this end, we define $N = \mathrm{diag}(p, p - 1, \cdots, 1)$ and solve

$$\max_{X^T X = I_p} \mathrm{trace}(X^T A X N). \tag{8}$$

The columns of the solution X (unique up to sign reversal) are the p dominant eigenvectors or A; see [24] or [5, §4.8].

Let us sketch the derivation of a steepest-ascent method on $\mathrm{St}(p, n) = \{X \in \mathbb{R}^{n \times p} : X^T X = I\}$ for solving (8). Details can be found in [5, §4.8]. Define $\bar{f} : \mathbb{R}^{n \times p} \to \mathbb{R} : X \mapsto \mathrm{trace}(X^T A X N)$ and $f = \bar{f}|_{\mathrm{St}(p,n)}$. We have $\frac{1}{2}\mathrm{grad}\,\bar{f}(X) = AXN$. Thus, in view of (7), $\frac{1}{2}\mathrm{grad}\,f(X) = \mathrm{P}_{T_X \mathrm{St}(p,n)}(AXN) = AXN - X\mathrm{sym}(X^T AXN)$, where $\mathrm{sym}(Z) := (Z + Z^T)/2$. This is the gradient in the sense of the Riemannian metric inherited from the embedding of $\mathrm{St}(p, n)$ in $\mathbb{R}^{n \times p}$. Possible choices for the retraction are given in [5, Ex. 4.1.3]. For example, the mapping given by $R(\xi_X) = \mathrm{qf}(X + \xi_X)$ is a retraction, where qf returns the Q factor of the QR decomposition of A.

This basic steepest-descent algorithm is given as an illustration; it is not meant to be competitive with state-of-the-art algorithms for eigenvalue computation. Competitive algorithms that stem from a Riemannian optimization approach can be found in [11, 10].

8 Newton's method on manifolds

We first present Newton's method on general manifolds. Then we particularize the algorithm to obtain an algorithm for Problem 2 with $p = 1$.

8.1 Newton on abstract manifolds

The central equation for Newton's method in $\mathbb{R}^n$ is

$$\mathrm{D}(\operatorname{grad} f)(x)[\eta_x] = -\operatorname{grad} f(x),$$

a linear equation in the update vector η_x. On a Riemannian manifold, it is clear that η_x becomes a tangent vector at x, and that $\operatorname{grad} f$ becomes the gradient vector field defined in Section 6.3. It remains to define the directional derivative of a vector field such as $\operatorname{grad} f$. A thoughtless extension of (6) would yield the formula $\lim_{t \to 0} \frac{\operatorname{grad} f(\gamma(t)) - \operatorname{grad} f(x)}{t}$, which is inapplicable to abstract manifolds since $\operatorname{grad} f(\gamma(t))$ and $\operatorname{grad} f(x)$ belong to $T_{\gamma(t)}\mathcal{M}$ and $T_x\mathcal{M}$, which are two different vector spaces. The remedy is given by endowing $\mathcal{M}$ with an object called an affine connection and denoted by ∇, that takes as argument a vector field and a tangent vector and returns the (covariant) derivative of the vector field along the tangent vector.

The Riemannian Newton method given below is formulated as in [7] (or see [5, §6.2]).

Required: Riemannian manifold $\mathcal{M}$; retraction R on $\mathcal{M}$; affine connection ∇ on $\mathcal{M}$; real-valued function f on $\mathcal{M}$.

Iteration $x_k \in \mathcal{M} \mapsto x_{k+1} \in \mathcal{M}$ defined by

1. Solve the Newton equation

$$\operatorname{Hess} f(x_k)\eta_k = -\operatorname{grad} f(x_k)$$

   for the unknown $\eta_k \in T_{x_k}\mathcal{M}$, where

$$\operatorname{Hess} f(x_k)\eta_k := \nabla_{\eta_k}\operatorname{grad} f.$$

2. Set

$$x_{k+1} := R_{x_k}(\eta_k).$$

The algorithm has convergence properties akin to those of Newton's algorithm in $\mathbb{R}^n$ [5, §6.3].

8.2 Newton on submanifolds of $\mathbb{R}^n$

If $\mathcal{M}$ is a submanifold of $\mathbb{R}^n$, it naturally inherits a Riemannian metric by the restriction of the standard inner product of $\mathbb{R}^n$. If moreover the so-called Levi-Civita connection is chosen for ∇, the algorithm below is obtained.

Required: Riemannian submanifold $\mathcal{M}$ of $\mathbb{R}^n$; retraction R on $\mathcal{M}$; real-valued function f on $\mathcal{M}$.

Iteration $x_k \in \mathcal{M} \mapsto x_{k+1} \in \mathcal{M}$ defined by

1. Solve the Newton equation

$$\operatorname{Hess} f(x_k)\eta_k = -\operatorname{grad} f(x_k)$$

for the unknown $\eta_k \in T_{x_k}\mathcal{M}$, where

$$\text{Hess } f(x_k)\eta_k := \text{P}_{T_{x_k}\mathcal{M}}\text{Dgrad } f(x_k)[\eta_k].$$

2. Set

$$x_{k+1} := R_{x_k}(\eta_k).$$

8.3 Newton on the unit sphere S^{n-1}

Let us now particularize the algorithm to the case where $\mathcal{M}$ is the unit sphere S^{n-1}, viewed as a Riemannian submanifold of $\mathbb{R}^n$, with a particular choice for the retraction. We obtain a numerical algorithm that can be formulated without any reference to differential-geometric objects, and that inherits the desirable convergence properties of the abstract Riemannian Newton method.

Required: real-valued function f on S^{n-1}.

Iteration $x_k \in S^{n-1} \mapsto x_{k+1} \in S^{n-1}$ defined by

1. Solve the Newton equation

$$\begin{cases} \text{P}_{x_k}\text{D}(\text{grad } f)(x_k)[\eta_k] = -\text{grad } f(x_k) \\ x^T\eta_k = 0, \end{cases}$$

for the unknown $\eta_k \in \mathbb{R}^n$, where $\text{P}_{x_k} = (I - x_k x_k^T)$.

2. Set

$$x_{k+1} := \frac{x_k + \eta_k}{\|x_k + \eta_k\|}.$$

In the algorithm above, $\text{grad } f(x) = (I - xx^T)\text{grad } \bar{f}(x)$, where $\bar{f}(x)$ is any smooth extension of f.

8.4 Newton for Rayleigh quotient optimization on unit sphere

Finally, if we apply the above algorithm to a specific objective function, such as the one given in Problem 2 with $p = 1$, we obtain a concrete numerical algorithm.

Iteration $x_k \in S^{n-1} \mapsto x_{k+1} \in S^{n-1}$ defined by

1. Solve the Newton equation

$$\begin{cases} \text{P}_{x_k}A\text{P}_{x_k}\eta_k - \eta_k x_k^T A x_k = -\text{P}_{x_k}A x_k, \\ x_k^T\eta_k = 0, \end{cases}$$

for the unknown $\eta_k \in \mathbb{R}^n$, where $\text{P}_{x_k} = (I - x_k x_k^T)$.

2. Set

$$x_{k+1} := \frac{x_k + \eta_k}{\|x_k + \eta_k\|}.$$

Not surprisingly for such a fundamental problem, we fall back on a known eigenvalue algorithm, the Rayleigh quotient iteration. On several other problems, the Riemannian Newton method has led to novel numerical algorithms; see, e.g., [41, 7, 40, 58, 22, 19].

9 Other optimization methods on manifolds

Besides steepest descent and Newton, several other classical methods for unconstrained optimization admit a generalization to manifolds. Chapter 8 in [5] briefly mentions approximate Newton methods and conjugate gradient schemes. A Riemannian trust-region method was proposed in [2] (or see [5, Ch. 7]), which led to competitive algorithms for symmetric eigenvalue problems [11, 10]. For a Riemannian BFGS method, see [48] and references therein.

The relation between optimization methods on manifolds and feasible methods for equality-constrained optimization is investigated in [6]. This concerns in particular the theory of $\mathcal{U}$-Lagrangians, and the related $\mathcal{VU}$-decompositions and fast tracks [35, 43], as well as the theory of partly smooth functions [36], both of which coincide in the convex case [44, Th. 2.9]. The concepts of $\mathcal{U}$-Lagrangian and partly smooth functions led to several Newton-like algorithms whose iterates are constrained to a submanifold $\mathcal{M}$ such that the restriction $f_{|\mathcal{M}}$ is smooth. These algorithms are unified in [16] under a common two-step, predictor-corrector form, and connections with SQP and Riemannian Newton are studied in [44].

We also mention the literature on proximal point algorithms on Hadamard manifolds; see [37] and references therein.

10 Conclusion

We have proposed an introduction to the area of optimization on manifolds, written as a digest of [5] enhanced with references to the most recent literature. In summary, optimization on manifolds is about exploiting tools of differential geometry to build optimization schemes on abstract manifolds, then turning these abstract geometric algorithms into practical numerical methods for specific manifolds, with applications to problems that can be rephrased as optimizing a differentiable function over a manifold. This research program has shed new light on existing algorithms and produced novel numerical methods backed by a strong convergence analysis.

We close by pointing out that optimization of real-valued functions on manifolds, as formulated in Problem 1, is not the only place where numerical optimization and differential geometry meet. Noteworthy are the Riemannian geometry of the central path in linear programming [17, 45], and an intriguing continuous-time system on the Grassmann manifold associated with linear programs [63, 1].

Acknowledgements

References

1. P.-A. Absil. Numerical representations of a universal subspace flow for linear programs. *Communications in Information and Systems*, 8(2):71–84, 2009.
2. P.-A. Absil, C. G. Baker, and K. A. Gallivan. Trust-region methods on Riemannian manifolds. *Found. Comput. Math.*, 7(3):303–330, July 2007.
3. P.-A. Absil and K. A. Gallivan. Joint diagonalization on the oblique manifold for independent component analysis. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 5, pages V–945–V–948, 2006.
4. P.-A. Absil, R. Mahony, and R. Sepulchre. Riemannian geometry of Grassmann manifolds with a view on algorithmic computation. *Acta Appl. Math.*, 80(2):199–220, January 2004.
5. P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, NJ, 2008.
6. P.-A. Absil, Jochen Trumpf, Robert Mahony, and Ben Andrews. All roads lead to Newton: Feasible second-order methods for equality-constrained optimization. Technical Report UCL-INMA-2009.024, UCLouvain, 2009.
7. Roy L. Adler, Jean-Pierre Dedieu, Joseph Y. Margulies, Marco Martens, and Mike Shub. Newton's method on Riemannian manifolds and a geometric model for the human spine. *IMA J. Numer. Anal.*, 22(3):359–390, July 2002.
8. Bijan Afsari and P. S. Krishnaprasad. Some gradient based joint diagonalization methods for ICA. In Springer LCNS Series, editor, *Proceedings of the 5th International Conference on Independent Component Analysis and Blind Source Separation*, 2004.
9. Gregory Ammar and Clyde Martin. The geometry of matrix eigenvalue methods. *Acta Appl. Math.*, 5(3):239–278, 1986.
10. C. G. Baker, P.-A. Absil, and K. A. Gallivan. An implicit trust-region method on Riemannian manifolds. *IMA J. Numer. Anal.*, 28(4):665–689, 2008.
11. Christopher G. Baker. *Riemannian manifold trust-region methods with applications to eigenproblems*. PhD thesis, School of Computational Science, Florida State University, Summer Semester 2008.
12. M. Baumann and U. Helmke. Riemannian subspace tracking algorithms on Grassmann manifolds. In *Proceedings of the 46th IEEE Conference on Decision and Control*, 2007.

13. Silvère Bonnabel and Rodolphe Sepulchre. Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank. *SIAM J. Matrix Anal. Appl.*, 31(3):1055–1070, 2009.
14. Thomas J. Bridges and Sebastian Reich. Computing Lyapunov exponents on a Stiefel manifold. *Phys. D*, 156(3-4):219–238, 2001.
15. Hasan Ertan Çetingül and René Vidal. Intrinsic mean shift for clustering on Stiefel and Grassmann manifolds. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'09)*, 2009.
16. Aris Daniilidis, Warren Hare, and Jérôme Malick. Geometrical interpretation of the predictor-corrector type algorithms in structured optimization problems. *Optimization*, 55(5-6):481–503, 2006.
17. Jean-Pierre Dedieu, Gregorio Malajovich, and Mike Shub. On the curvature of the central path of linear programming theory. *Found. Comput. Math.*, 5(2):145–171, 2005.
18. Alan Edelman, Tomás A. Arias, and Steven T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.*, 20(2):303–353, 1998.
19. L. Eldén and B. Savas. A Newton–Grassmann method for computing the best multi-linear rank-(r_1, r_2, r_3) approximation of a tensor. *SIAM J. Matrix Anal. Appl.*, 31:248–271, 2009.
20. Lars Eldén and Haesun Park. A Procrustes problem on the Stiefel manifold. *Numer. Math.*, 82(4):599–619, 1999.
21. Hermann Grassmann. Wörterbuch zum Rig-Veda, 1873. Leipzig.
22. Uwe Helmke, Knut Hüper, Pei Yean Lee, and John B. Moore. Essential matrix estimation using Gauss-Newton iterations on a manifold. *Int. J. Computer Vision*, 74(2):117–136, 2007.
23. Uwe Helmke, Knut Hüper, and Jochen Trumpf. Newton's method on Grassmann manifolds, September 2007. arXiv:0709.2205v2.
24. Uwe Helmke and John B. Moore. *Optimization and Dynamical Systems.* Communications and Control Engineering Series. Springer-Verlag London Ltd., London, 1994. With a foreword by R. Brockett.
25. Magnus R. Hestenes and Eduard Stiefel. Methods of conjugate gradients for solving linear systems. *J. Research Nat. Bur. Standards*, 49:409–436 (1953), 1952.
26. Knut Hüper, Hao Shen, and Abd-Krim Seghouane. Local convergence properties of FastICA and some generalisations. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 5, pages V–1009–V–1012, 2006.
27. M. Ishteva, L. De Lathauwer, P.-A. Absil, and S. Van Huffel. Best low multilinear rank approximation of higher-order tensors, based on the Riemannian trust-region scheme. Technical Report 09-142, ESAT-SISTA, K.U.Leuven, Belgium, 2009.
28. Jens Jordan and Uwe Helmke. Controllability of the QR-algorithm on Hessenberg flags. In David S. Gilliam and Joachim Rosenthal, editors, *Proceeding of the Fifteenth International Symposium on Mathematical Theory of Network and Systems (MTNS 2002)*, 2002.
29. Shantanu H. Joshi, Eric Klassen, Anuj Srivastava, and Ian Jermyn. A novel representation for Riemannian analysis of elastic curves in R^n. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.

30. M. Journée, F. Bach, P.-A. Absil, and R. Sepulchre. Low-rank optimization for semidefinite convex problems, 2008. arXiv:0807.4423.

31. E. Klassen, A. Srivastava, M. Mio, and S.H. Joshi. Analysis of planar shapes using geodesic paths on shape spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(3):372–383, 2004.

32. Eric Klassen and Anuj Srivastava. Geodesics between 3D closed curves using path-straightening. In A. Leonardis, H. Bischof, and A. Pinz, editors, *ECCV 2006, Part I,*, volume 3951 of *LNCS*, pages 95–106. Springer-Verlag, Berlin Heidelberg, 2006.

33. John M. Lee. *Introduction to smooth manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2003.

34. Pei Yean Lee and John B. Moore. Pose estimation via a Gauss-Newton-on-manifold approach. In *Proceedings of the 16th International Symposium on Mathematical Theory of Network and System (MTNS), Leuven*, 2004.

35. Claude Lemaréchal, François Oustry, and Claudia Sagastizábal. The $\mathcal{U}$-Lagrangian of a convex function. *Trans. Amer. Math. Soc.*, 352(2):711–729, 2000.

36. A. S. Lewis. Active sets, nonsmoothness, and sensitivity. *SIAM J. Optim.*, 13(3):702–725 (electronic) (2003), 2002.

37. Chong Li, Genaro López, and Victoria Martín-Márquez. Monotone vector fields and the proximal point algorithm on Hadamard manifolds. *J. London Math. Soc.*, 79(3):663–683, 2009.

38. R. Lippert and A. Edelman. Nonlinear eigenvalue problems with orthogonality constraints (Section 9.4). In Zhaojun Bai, James Demmel, Jack Dongarra, Axel Ruhe, and Henk van der Vorst, editors, *Templates for the Solution of Algebraic Eigenvalue Problems*, pages 290–314. SIAM, Philadelphia, 2000.

39. Xiuwen Liu, Anuj Srivastava, and Kyle Gallivan. Optimal linear representations of images for object recognition. *IEEE Pattern Anal. and Mach. Intell.*, 26(5):662–666, May 2004.

40. Eva Lundström and Lars Eldén. Adaptive eigenvalue computations using Newton's method on the Grassmann manifold. *SIAM J. Matrix Anal. Appl.*, 23(3):819–839, 2001/02.

41. Yi Ma, Jana Kosecka, and Shankar S. Sastry. Optimization criteria and geometric algorithms for motion and structure estimation. *Int. J. Computer Vision*, 44(3):219–249, 2001.

42. Jonathan H. Manton. A centroid (Karcher mean) approach to the joint approximate diagonalization problem: The real symmetric case. *Digital Signal Processing*, 16(5):468–478, 2005.

43. Robert Mifflin and Claudia Sagastizábal. On $\mathcal{VU}$-theory for functions with primal-dual gradient structure. *SIAM J. Optim.*, 11(2):547–571 (electronic), 2000.

44. Scott A. Miller and Jérôme Malick. Newton methods for nonsmooth convex minimization: connections among $\mathcal{U}$-Lagrangian, Riemannian Newton and SQP methods. *Math. Program.*, 104(2-3, Ser. B):609–633, 2005.

45. Y. Nesterov and A. Nemirovski. Primal central paths and Riemannian distances for convex sets. *Found. Comput. Math.*, 8(5):533–560, 2008.

46. Yasunori Nishimori and Shotaro Akaho. Learning algorithms utilizing quasi-geodesic flows on the Stiefel manifold. *Neurocomputing*, 67:106–135, 2005.

47. Yasunori Nishimori, Shotaro Akaho, and Mark D. Plumbley. Natural conjugate gradient on complex flag manifolds for complex independent subspace analysis. In Vera Kurkova-Pohlova, Roman Neruda, and Jan Koutnik, editors, *Artificial Neural Networks - ICANN 2008*, volume 5163 of *LNCS*, pages 165–174. Springer, Berlin Heidelberg, 2008.

48. Chunhong Qi, Kyle A. Gallivan, and P.-A. Absil. Riemannian BFGS algorithm with applications. In *Recent Advances in Optimization and its Applications in Engineering*. Springer, 2010. To appear.

49. Quentin Rentmeesters, P.-A. Absil, and Paul Van Dooren. Identification method for time-varying ARX models. Submitted, 2009.

50. Chafik Samir, P.-A. Absil, Anuj Srivastava, and Eric Klassen. A gradient-descent method for curve fitting on Riemannian manifolds. Technical Report UCL-INMA-2009.023, UCLouvain, 2009.

51. Oliver Sander. Geodesic finite elements for Cosserat rods. submitted, 2009.

52. Berkant Savas and Lek-Heng Lim. Best multilinear rank approximation of tensors with quasi-Newton methods on Grassmannians. Technical Report LITH-MAT-R-2008-01-SE, Department of Mathematics, Linköpings Universitet, 2008.

53. Anuj Srivastava and Eric Klassen. Bayesian and geometric subspace tracking. *Adv. in Appl. Probab.*, 36(1):43–56, 2004.

54. Anuj Srivastava and Xiuwen Liu. Tools for application-driven linear dimension reduction. *Neurocomputing*, 67:136–160, 2005.

55. E. Stiefel. Richtungsfelder und Fernparallelismus in n-dimensionalen Mannigfaltigkeiten. *Comment. Math. Helv.*, 8(1):305–353, 1935.

56. Fabian J. Theis, Thomas P. Cason, and P.-A. Absil. Soft dimension reduction for ICA by joint diagonalization on the Stiefel manifold. In *Proc. ICA 2009*, volume 5441 of *LNCS*, pages 354–361, Paraty, Brazil, 2009. Springer.

57. Frank Tompkins and Patrick J. Wolfe. Bayesian filtering on the Stiefel manifold. In *2nd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMPSAP 2007)*, 2007.

58. Nickolay T. Trendafilov and Ross A. Lippert. The multimode Procrustes problem. *Linear Algebra Appl.*, 349:245–264, 2002.

59. Pavan Turaga, Ashok Veeraraghavan, and Rama Chellappa. Statistical analysis on Stiefel and Grassmann manifolds with applications in computer vision. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

60. Bart Vandereycken, P.-A. Absil, and Stefan Vandewalle. Embedded geometry of the set of symmetric positive semidefinite matrices of fixed rank. In *Proceedings of the IEEE 15th Workshop on Statistical Signal Processing*, pages 389–392, 2009.

61. Bart Vandereycken and Stefan Vandewalle. A Riemannian optimization approach for computing low-rank solutions of Lyapunov equations. Submitted, 2009.

62. Frank W. Warner. *Foundations of differentiable manifolds and Lie groups*, volume 94 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1983. Corrected reprint of the 1971 edition.

63. Gongyun Zhao. Representing the space of linear programs as the Grassmann manifold. *Math. Program.*, 121(2, Ser. A):353–386, 2010.