Chapter Four

Line-Search Algorithms on Manifolds

Line-search methods in $\mathbb{R}^n$ are based on the update formula

$$x_{k+1} = x_k + t_k \eta_k, \tag{4.1}$$

where $\eta_k \in \mathbb{R}^n$ is the *search direction* and $t_k \in \mathbb{R}$ is the *step size*. The goal of this chapter is to develop an analogous theory for optimization problems posed on nonlinear manifolds.

The proposed generalization of (4.1) to a manifold $\mathcal{M}$ consists of selecting η_k as a tangent vector to $\mathcal{M}$ at x_k and performing a search along a curve in $\mathcal{M}$ whose tangent vector at $t = 0$ is η_k. The selection of the curve relies on the concept of retraction, introduced in Section 4.1. The choice of a computationally efficient retraction is an important decision in the design of high-performance numerical algorithms on nonlinear manifolds. Several practical examples are given for the matrix manifolds associated with the main examples of interest considered in this book.

This chapter also provides the convergence theory of line-search algorithms defined on Riemannian manifolds. Several example applications related to the eigenvalue problem are presented.

4.1 RETRACTIONS

Conceptually, the simplest approach to optimizing a differentiable function is to continuously translate a test point $x(t)$ in the direction of steepest descent, $-\operatorname{grad} f(x)$, on the constraint set until one reaches a point where the gradient vanishes. Points x where $\operatorname{grad} f(x) = 0$ are called *stationary points* or *critical points* of f. A numerical implementation of the continuous gradient descent approach requires the construction of a curve γ such that $\dot\gamma(t) = -\operatorname{grad} f(\gamma(t))$ for all t. Except in very special circumstances, the construction of such a curve using numerical methods is impractical. The closest numerical analogy is the class of optimization methods that use *line-search* procedures, namely, iterative algorithms that, given a point x, compute a descent direction $\eta := -\operatorname{grad} f(x)$ (or some approximation of the gradient) and move in the direction of η until a “reasonable” decrease in f is found. In $\mathbb{R}^n$, the concept of moving in the direction of a vector is straightforward. On a manifold, the notion of moving in the direction of a tangent vector, while staying on the manifold, is generalized by the notion of a retraction mapping.

Conceptually, a retraction R at x, denoted by R_x, is a mapping from $T_x\mathcal{M}$ to $\mathcal{M}$ with a local rigidity condition that preserves gradients at x; see Figure 4.1.

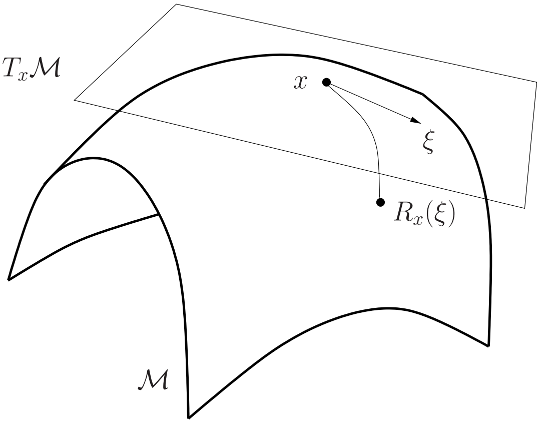

Figure 4.1 Retraction.

Definition 4.1.1 (retraction) *A* retraction *on a manifold $\mathcal{M}$ is a smooth mapping R from the tangent bundle $T\mathcal{M}$ onto $\mathcal{M}$ with the following properties. Let R_x denote the restriction of R to $T_x\mathcal{M}$.*

(i) $R_x(0_x) = x$, where 0_x denotes the zero element of $T_x\mathcal{M}$.

(ii) With the canonical identification $T_{0_x}T_x\mathcal{M} \simeq T_x\mathcal{M}$, R_x satisfies

$$\mathrm{D}R_x(0_x) = \mathrm{id}_{T_x\mathcal{M}}, \tag{4.2}$$

where $\mathrm{id}_{T_x\mathcal{M}}$ denotes the identity mapping on $T_x\mathcal{M}$.

We generally assume that the domain of R is the whole tangent bundle $T\mathcal{M}$. This property holds for all practical retractions considered in this book.

Concerning condition (4.2), notice that, since R_x is a mapping from $T_x\mathcal{M}$ to $\mathcal{M}$ sending 0_x to x, it follows that $\mathrm{D}R_x(0_x)$ is a mapping from $T_{0_x}(T_x\mathcal{M})$ to $T_x\mathcal{M}$ (see Section 3.5.6). Since $T_x\mathcal{M}$ is a vector space, there is a natural identification $T_{0_x}(T_x\mathcal{M}) \simeq T_x\mathcal{M}$ (see Section 3.5.2). We refer to the condition $\mathrm{D}R_x(0_x) = \mathrm{id}_{T_x\mathcal{M}}$ as the *local rigidity* condition. Equivalently, for every tangent vector ξ in $T_x\mathcal{M}$, the curve $\gamma_\xi : t \mapsto R_x(t\xi)$ satisfies $\dot{\gamma}_\xi(0) = \xi$. Moving along this curve γ_ξ is thought of as moving in the direction ξ while constrained to the manifold $\mathcal{M}$.

Besides turning elements of $T_x\mathcal{M}$ into points of $\mathcal{M}$, a second important purpose of a retraction R_x is to transform cost functions defined in a neighborhood of $x \in \mathcal{M}$ into cost functions defined on the vector space $T_x\mathcal{M}$. Specifically, given a real-valued function f on a manifold $\mathcal{M}$ equipped with a retraction R, we let $\widehat{f} = f \circ R$ denote the *pullback* of f through R. For $x \in \mathcal{M}$, we let

$$\widehat{f}_x = f \circ R_x \tag{4.3}$$

denote the restriction of $\widehat{f}$ to $T_x\mathcal{M}$. Note that $\widehat{f}_x$ is a real-valued function on a vector space. Observe that because of the local rigidity condition (4.2), we have (with the canonical identification (3.11)) $\mathrm{D}\widehat{f}_x(0_x) = \mathrm{D}f(x)$. If $\mathcal{M}$ is endowed with a Riemannian metric (and thus $T_x\mathcal{M}$ with an inner product), we have

$$\operatorname{grad} \widehat{f}_x(0_x) = \operatorname{grad} f(x). \tag{4.4}$$

All the main examples that are considered in this book (and most matrix manifolds of interest) admit a Riemannian metric. Every manifold that admits a Riemannian metric also admits a retraction defined by the *Riemannian exponential mapping* (see Section 5.4 for details). The domain of the exponential mapping is not necessarily the whole $T\mathcal{M}$. When it is, the Riemannian manifold is called *complete*. The Stiefel and Grassmann manifolds, endowed with the Riemannian metrics defined in Section 3.6, are complete.

The Riemannian exponential mapping is, in the geometric sense, the most natural retraction to use on a Riemannian manifold and featured heavily in the early literature on the development of numerical algorithms on Riemannian manifolds. Unfortunately, the Riemannian exponential mapping is itself defined as the solution of a nonlinear ordinary differential equation that, in general, poses significant numerical challenges to compute cheaply. In most cases of interest in this book, the solution of the Riemannian exponential can be expressed in terms of classical analytic functions with matrix arguments. However, the evaluation of matrix analytic functions is also a challenging problem and usually computationally intensive to solve. Indeed, computing the exponential may turn out to be more difficult than the original Riemannian optimization problem under consideration (see Section 7.5.2 for an example). These drawbacks are an invitation to consider alternatives in the form of approximations to the exponential that are computationally cheap without jeopardizing the convergence properties of the optimization schemes. Retractions provide a framework for analyzing such alternatives. All the algorithms in this book make use of retractions in one form or another, and the convergence analysis is carried out for general retractions.

In the remainder of this Section 4.1, we show how several structures (embedded submanifold, quotient manifold) and mathematical objects (local coordinates, projections, factorizations) can be exploited to define retractions.

4.1.1 Retractions on embedded submanifolds

Let $\mathcal{M}$ be an embedded submanifold of a vector space $\mathcal{E}$. Recall that $T_x\mathcal{M}$ can be viewed as a linear subspace of $T_x\mathcal{E}$ (Section 3.5.7) which itself can be identified with $\mathcal{E}$ (Section 3.5.2). This allows us, slightly abusing notation, to consider the sum $x+\xi$ of a point x of $\mathcal{M}$, viewed as an element of $\mathcal{E}$, and a tangent vector $\xi \in T_x\mathcal{M}$, viewed as an element of $T_x\mathcal{E} \simeq \mathcal{E}$. In this setting, it is tempting to define a retraction along the following lines. Given x in $\mathcal{M}$ and $\xi \in T_x\mathcal{M}$, compute $R_x(\xi)$ by

1. moving along ξ to get the point $x+\xi$ in the linear embedding space;
2. "projecting" the point $x+\xi$ back to the manifold $\mathcal{M}$.

The issue is to define a projection that (i) turns the procedure into a well-defined retraction and (ii) is computationally efficient. In the embedded submanifolds of interest in this book, as well as in several other cases, the second step can be based on matrix decompositions. Examples of such decompositions include QR factorization and polar decomposition. The purpose of the present section is to develop a general theory of decomposition-based retractions. With this theory at hand, it will be straightforward to show that several mappings constructed along the above lines are well-defined retractions.

Let $\mathcal{M}$ be an embedded manifold of a vector space $\mathcal{E}$ and let $\mathcal{N}$ be an abstract manifold such that $\dim(\mathcal{M}) + \dim(\mathcal{N}) = \dim(\mathcal{E})$. Assume that there is a diffeomorphism

$$\phi : \mathcal{M} \times \mathcal{N} \to \mathcal{E}_* : (F, G) \mapsto \phi(F, G),$$

where $\mathcal{E}_*$ is an open subset of $\mathcal{E}$ (thus $\mathcal{E}_*$ is an open submanifold of $\mathcal{E}$), with a neutral element $I \in \mathcal{N}$ satisfying

$$\phi(F, I) = F, \quad \forall F \in \mathcal{M}.$$

(The letter I is chosen in anticipation that the neutral element will be the identity matrix of a matrix manifold $\mathcal{N}$ in cases of interest.)

Proposition 4.1.2 *Under the above assumptions on ϕ, the mapping*

$$R_X(\xi) := \pi_1(\phi^{-1}(X+\xi)),$$

where $\pi_1 : \mathcal{M} \times \mathcal{N} \to \mathcal{M} : (F, G) \mapsto F$ is the projection onto the first component, defines a retraction on $\mathcal{M}$.

Proof. Since $\mathcal{E}_*$ is open, it follows that $X + \xi$ belongs to $\mathcal{E}_*$ for all ξ in some neighborhood of 0_X. Since ϕ^{-1} is defined on the whole $\mathcal{E}_*$, it follows that $R_X(\xi)$ is defined for all ξ in a neighborhood of the origin of $T_X\mathcal{M}$. Smoothness of R and the property $R_X(0_X) = X$ are direct. For the local rigidity property, first note that for all $\xi \in T_X\mathcal{M}$, we have

$$\mathrm{D}_1\phi(X, I)[\xi] = \mathrm{D}\phi(X, I)[(\xi, 0)] = \xi.$$

Since $\pi_1 \circ \phi^{-1}(\phi(F, I)) = F$, it follows that, for all $\xi \in T_X\mathcal{M}$,

$$\xi = \mathrm{D}(\pi_1 \circ \phi^{-1})(\phi(X, I))\,[\mathrm{D}_1\phi(X, I)[\xi]] = \mathrm{D}(\pi_1 \circ \phi^{-1})(X)[\xi] = \mathrm{D}R_X(0_X)[\xi],$$

which proves the claim that R_X is a retraction. □

Example 4.1.1 ***Retraction on the sphere S^{n-1}***

Let $\mathcal{M} = S^{n-1}$, let $\mathcal{N} = \{x \in \mathbb{R} : x > 0\}$, and consider the mapping

$$\phi : \mathcal{M} \times \mathcal{N} \to \mathbb{R}^n_* : (x, r) \mapsto xr.$$

It is straightforward to verify that ϕ is a diffeomorphism. Proposition 4.1.2 yields the retraction

$$R_x(\xi) = \frac{x+\xi}{\|x+\xi\|},$$

defined for all $\xi \in T_x S^{n-1}$. Note that $R_x(\xi)$ is the point of S^{n-1} that minimizes the distance to $x + \xi$.

Example 4.1.2 ***Retraction on the orthogonal group***

Let $\mathcal{M} = O_n$ be the orthogonal group. The QR decomposition *of a matrix $A \in \mathbb{R}_*^{n\times n}$ is the decomposition of A as $A = QR$, where Q belongs to O_n and R belongs to $\mathcal{S}_{\mathrm{upp+}}(n)$, the set of all upper triangular matrices with strictly positive diagonal elements. The inverse of QR decomposition is the mapping*

$$\phi : O_n \times \mathcal{S}_{\mathrm{upp+}}(n) \to \mathbb{R}_*^{n\times n} : (Q, R) \mapsto QR. \tag{4.5}$$

We let $\mathrm{qf} := \pi_1 \circ \phi^{-1}$ denote the mapping that sends a matrix to the Q factor of its QR decomposition. The mapping qf *can be computed using the* Gram-Schmidt orthonormalization.

We have to check that ϕ satisfies all the assumptions of Proposition 4.1.2. The identity matrix is the neutral element: $\phi(Q, I) = Q$ for all $Q \in O_n$. It follows from the existence and uniqueness properties of the QR decomposition that ϕ is bijective. The mapping ϕ is smooth since it is the restriction of a smooth map (matrix product) to a submanifold. Concerning ϕ^{-1}, notice that its first matrix component Q is obtained by a Gram-Schmidt process, which is C^∞ on the set of full-rank matrices. Since the second component R is obtained as $Q^{-1}M$, it follows that ϕ^{-1} is C^∞. In conclusion, the assumptions of Proposition 4.1.2 hold for (4.5), *and consequently,*

$$R_X(X\Omega) := \mathrm{qf}(X + X\Omega) = \mathrm{qf}(X(I + \Omega)) = X\mathrm{qf}(I + \Omega)$$

is a retraction on the orthogonal group O_n.

A second possibility is to consider the polar decomposition *of a matrix $A = QP$, where $Q \in O_n$ and $P \in \mathcal{S}_{\mathrm{sym+}}(n)$, the set of all symmetric positive-definite matrices of size n. The inverse of polar decomposition is a mapping*

$$\phi : O_n \times \mathcal{S}_{\mathrm{sym+}}(n) \to \mathbb{R}_*^{n\times n} : (Q, P) \mapsto QP.$$

We have $\phi^{-1}(A) = (A(A^TA)^{-1/2}, (A^TA)^{1/2})$. This shows that ϕ is a diffeomorphism, and thus

$$\begin{aligned} R_X(X\Omega) &:= X(I + \Omega)((X(I + \Omega))^T X(I + \Omega))^{-1/2} \\ &= X(I + \Omega)(I - \Omega^2)^{-1/2} \end{aligned} \tag{4.6}$$

is a retraction on O_n. Computing this retraction requires an eigenvalue decomposition of the $n \times n$ symmetric matrix $(I - \Omega^2)$. Note that it does not make sense to use this retraction in the context of an eigenvalue algorithm on O_n since the computational cost of computing a single retraction is comparable to that for solving the original optimization problem.

A third possibility is to use Givens rotations. *For an $n\times n$ skew-symmetric matrix Ω, let $\mathrm{Giv}(\Omega) = \prod_{1\le i<j\le n} G(i, j, \Omega_{ij})$, where the order of multiplication is any fixed order and where $G(i, j, \theta)$ is the Givens rotation of angle θ in the (i, j) plane, namely, $G(i, j, \theta)$ is the identity matrix with the substitutions $e_i^T G(i, j, \theta)e_i = e_j^T G(i, j, \theta)e_j = \cos(\theta)$ and $e_i^T G(i, j, \theta)e_j = -e_j^T G(i, j, \theta)e_i = \sin(\theta)$. Then the mapping $R : TO_n \to O_n$ defined by*

$$R_X(X\Omega) = X\,\mathrm{Giv}(\Omega)$$

is a retraction on O_n*.*

Another retraction on O_n*, based on the* Cayley transform*, is given by*

$$R_X(X\Omega) = X(I - \tfrac{1}{2}\Omega)^{-1}(I + \tfrac{1}{2}\Omega).$$

Anticipating the material in Chapter 5, we point out that the Riemannian exponential mapping on O_n *(viewed as a Riemannian submanifold of* $\mathbb{R}^{n\times n}$*) is given by*

$$\mathrm{Exp}_X(X\Omega) = X \exp(\Omega),$$

where exp *denotes the matrix exponential defined by* $\exp(\Omega) := \sum_{i=0}^{\infty} \frac{1}{i!}\Omega^i$*. Note that Riemannian exponential mappings are always retractions (Proposition 5.4.1). Algorithms for accurately evaluating the exponential have a numerical cost at best similar to those for evaluating* (4.6)*. However, there are several computationally efficient Lie group-based algorithms for approximating the exponential that fit the definition of a retraction (see pointers in Notes and References).*

Example 4.1.3 ***Retraction on the Stiefel manifold***

Consider the Stiefel manifold $\mathrm{St}(p,n) = \{X \in \mathbb{R}^{n\times p} : X^TX = I_p\}$*. The retraction based on the polar decomposition is*

$$R_X(\xi) = (X + \xi)(I_p + \xi^T\xi)^{-1/2}, \tag{4.7}$$

where we have used the fact that ξ*, as an element of* $T_X\,\mathrm{St}(p,n)$*, satisfies* $X^T\xi + \xi^TX = 0$*. When* p *is small, the numerical cost of evaluating* (4.7) *is reasonable since it involves the eigenvalue decomposition of the small* $p \times p$ *matrix* $(I_p + \xi^T\xi)^{-1/2}$ *along with matrix linear algebra operations that require only* $O(np^2)$ *additions and multiplications.*

Much as in the case of the orthogonal group, an alternative to choice (4.7) *is*

$$R_X(\xi) := \mathrm{qf}(X + \xi), \tag{4.8}$$

where $\mathrm{qf}(A)$ *denotes the* Q *factor of the decomposition of* $A \in \mathbb{R}_*^{n\times p}$ *as* $A = QR$*, where* Q *belongs to* $\mathrm{St}(p,n)$ *and* R *is an upper triangular* $n \times p$ *matrix with strictly positive diagonal elements. Computing* $R_X(\xi)$ *can be done in a finite number of basic* arithmetic operations *(addition, subtraction, multiplication, division) and square roots using, e.g., the modified Gram-Schmidt algorithm.*

4.1.2 Retractions on quotient manifolds

We now consider the case of a quotient manifold $\mathcal{M} = \overline{\mathcal{M}}/\sim$. Recall the notation π for the canonical projection and $\overline{\xi}_{\overline{x}}$ for the horizontal lift at $\overline{x}$ of a tangent vector $\xi \in T_{\pi(\overline{x})}\mathcal{M}$.

Proposition 4.1.3 *Let* $\mathcal{M} = \overline{\mathcal{M}}/\sim$ *be a quotient manifold with a prescribed horizontal distribution. Let* $\overline{R}$ *be a retraction on* $\overline{\mathcal{M}}$ *such that for all* $x \in \mathcal{M}$ *and* $\xi \in T_x\mathcal{M}$*,*

$$\pi(\overline{R}_{\overline{x}_a}(\overline{\xi}_{\overline{x}_a})) = \pi(\overline{R}_{\overline{x}_b}(\overline{\xi}_{\overline{x}_b})) \tag{4.9}$$

for all $\overline{x}_a, \overline{x}_b \in \pi^{-1}(x)$. *Then*

$$R_x(\xi) := \pi(\overline{R}_{\overline{x}}(\overline{\xi}_{\overline{x}})) \tag{4.10}$$

defines a retraction on $\mathcal{M}$.

Proof. Equation (4.9) guarantees that R is well defined as a mapping from $T\mathcal{M}$ to $\mathcal{M}$. Since $\overline{R}$ is a retraction, it also follows that the property $R_x(0_x) = x$ is satisfied. Finally, the local rigidity condition holds since, given $\overline{x} \in \pi^{-1}(x)$,

$$\mathrm{D}R_x\left(0_x\right)[\eta] = \mathrm{D}\pi(\overline{x}) \circ \mathrm{D}\overline{R}_{\overline{x}}\left(0_{\overline{x}}\right)[\overline{\eta}_{\overline{x}}] = \mathrm{D}\pi\left(\overline{x}\right)[\overline{\eta}_{\overline{x}}] = \eta$$

for all $\eta \in T_x\mathcal{M}$, by definition of the horizontal lift. □

From now on we consider the case where the structure space $\overline{\mathcal{M}}$ is an open, dense (not necessarily proper) subset of a vector space $\overline{\mathcal{E}}$. Assume that a horizontal distribution $\mathcal{H}$ has been selected that endows every tangent vector to $\mathcal{M}$ with a horizontal lift. The natural choice for $\overline{R}$ is then

$$\overline{R}_y(\zeta) = y + \overline{\zeta}_y.$$

However, this choice does not necessarily satisfy (4.9). In other words, if x and y satisfy $\pi(x) = \pi(y)$, the property $\pi(x + \overline{\xi}_x) = \pi(y + \overline{\xi}_y)$ may fail to hold.

As an example, take the quotient of $\mathbb{R}^2$ for which the graphs of the curves $x_1 = a + a^3 x_2^2$ are equivalence classes, where $a \in \mathbb{R}$ parameterizes the set of all equivalence classes. Define the horizontal distribution as the constant subspace $e_1\mathbb{R}$. Given a tangent vector ξ to the quotient at the equivalence class $e_2\mathbb{R}$ (corresponding to $a = 0$), we obtain that the horizontal lift $\overline{\xi}_{(0,x_2)}$ is a constant $(C, 0)$ independent of x_2. It is clear that the equivalence class of $(0, x_2) + \overline{\xi}_{(0,x_2)} = (C, x_2)$ depends on x_2.

If we further require the equivalence classes to be the orbits of a Lie group acting linearly on $\overline{\mathcal{M}}$, with a horizontal distribution that is invariant by the Lie group action, then condition (4.9) holds. In particular, this is the case for the main examples considered in this book.

Example 4.1.4 ***Retraction on the projective space***

Consider the real projective space $\mathbb{RP}^{n-1} = \mathbb{R}^n_*/\mathbb{R}_*$ *with the horizontal distribution defined in* (3.27). *A retraction can be defined as*

$$R_{\pi(y)}\xi = \pi(y + \overline{\xi}_y),$$

where $\overline{\xi}_y \in \mathbb{R}^n$ *is the horizontal lift of* $\xi \in T_{\pi(y)}\mathbb{RP}^{n-1}$ *at* y.

Example 4.1.5 ***Retraction on the Grassmann manifold***

Consider the Grassmann manifold $\mathrm{Grass}(p, n) = \mathbb{R}^{n\times p}_*/\mathrm{GL}_p$ *with the horizontal distribution defined in* (3.40). *It can be checked using the homogeneity property of horizontal lifts (Proposition 3.6.1) that*

$$R_{\mathrm{span}(Y)}(\xi) = \mathrm{span}(Y + \overline{\xi}_Y) \tag{4.11}$$

is well-defined. Hence (4.11) *defines a retraction on* $\mathrm{Grass}(p, n)$.

Note that the matrix $Y + \overline{\xi}_Y$ is in general not orthonormal. In particular, if Y is orthonormal, then $Y + \overline{\xi}_Y$ is not orthonormal unless $\xi = 0$. In the scope of a numerical algorithm, in order to avoid ill-conditioning, it may be advisable to use $\mathrm{qf}\left(Y + \overline{\xi}_Y\right)$ *instead of $Y + \overline{\xi}_Y$ as a basis for the subspace $R_{\mathrm{span}(Y)}(\xi)$.*

4.1.3 Retractions and local coordinates*

In this section it is shown that every smooth manifold can be equipped with "local" retractions derived from its coordinate charts and that every retraction generates an atlas of the manifold. These operations, however, may pose computational challenges.

For every point x of a smooth manifold $\mathcal{M}$, there exists a smooth map $\mu_x : \mathbb{R}^d \mapsto \mathcal{M}$, $\mu_x(0) = x$, that is a local diffeomorphism around $0 \in \mathbb{R}^d$; the map μ_x is called a *local parameterization around* x and can be thought of as the inverse of a coordinate chart around $x \in \mathcal{M}$. If $\mathcal{U}$ is a neighborhood of a point x_* of $\mathcal{M}$, and $\mu : \mathcal{U} \times \mathbb{R}^d \to \mathcal{M}$ is a smooth map such that $\mu(x, z) = \mu_x(z)$ for all $x \in \mathcal{U}$ and $z \in \mathbb{R}^d$, then $\{\mu_x\}_{x \in \mathcal{M}}$ is called a *locally smooth family of parameterizations around* x_*. Note that a locally smooth parameterization μ around x_* can be constructed from a single chart around x_* by defining $\mu_x(z) = \varphi^{-1}(z + \varphi(x))$.

If $\{\mu_x\}_{x \in \mathcal{M}}$ is a locally smooth family of parameterizations around x_*, then the mappings

$$R_x : T_x\mathcal{M} \to \mathcal{M} : \xi \mapsto \mu_x(\mathrm{D}\mu_x^{-1}\,(x)\,[\xi])$$

define a retraction R whose domain is in general not the whole $T\mathcal{M}$. (It is readily checked that R_x satisfies the requirements in Definition 4.1.1.) Conversely, to define a smooth family of parameterizations around x_* from a retraction R, we can select smooth vector fields $\xi_1, \ldots, \xi_d$ on $\mathcal{M}$ such that, for all x in a neighborhood of x_*, $(\xi_1(x), \ldots, \xi_d(x))$ forms a basis of $T_x\mathcal{M}$, and then define

$$\mu_x(u_1, \ldots, u_d) = R_x(u_1\xi_1(x) + \cdots + u_d\xi_d(x)).$$

Note, however, that such a basis $\xi_1, \ldots, \xi_d$ of vector fields can in general be defined only locally. Moreover, producing the ξ's in practical cases may be tedious. For example, on the unit sphere S^{n-1}, the set T_xS^{n-1} is a vector space of dimension $(n-1)$ identified with $x_\perp := \{y \in \mathbb{R}^n : x^Ty = 0\}$; however, when n is large, generating and storing a basis of $x_\perp$ is impractical, as this requires $(n-1)$ vectors of n components. In other words, even though the $(n-1)$-dimensional vector space T_xS^{n-1} is known to be isomorphic to $\mathbb{R}^{n-1}$, creating an explicit isomorphism is computationally difficult. In comparison, it is computationally inexpensive to generate an element of $x_\perp$ (using the projection onto $x_\perp$) and to perform in $x_\perp$ the usual operations of addition and multiplication by a scalar.

In view of the discussion above, one could anticipate difficulty in dealing with pullback cost functions $\widehat{f}_x := f \circ R_x$ because they are defined on vector spaces $T_x\mathcal{M}$ that we may not want to explicitly represent as $\mathbb{R}^d$. Fortunately, many classical optimization techniques can be defined on abstract vector spaces, especially when the vector space has a structure of Euclidean space, which is the case for $T_x\mathcal{M}$ when $\mathcal{M}$ is Riemannian. We refer the reader to Appendix A for elements of calculus on abstract Euclidean spaces.

4.2 LINE-SEARCH METHODS

Line-search methods on manifolds are based on the update formula

$$x_{k+1} = R_{x_k}(t_k\eta_k),$$

where η_k is in $T_{x_k}\mathcal{M}$ and t_k is a scalar. Once the retraction R is chosen, the two remaining issues are to select the search direction η_k and then the step length t_k. To obtain global convergence results, some restrictions must be imposed on η_k and t_k.

Definition 4.2.1 (gradient-related sequence) *Given a cost function f on a Riemannian manifold $\mathcal{M}$, a sequence $\{\eta_k\}$, $\eta_k \in T_{x_k}\mathcal{M}$, is* gradient-related *if, for any subsequence $\{x_k\}_{k\in\mathcal{K}}$ of $\{x_k\}$ that converges to a non-critical point of f, the corresponding subsequence $\{\eta_k\}_{k\in\mathcal{K}}$ is bounded and satisfies*

$$\limsup_{k\to\infty,\ k\in\mathcal{K}} \langle \operatorname{grad} f(x_k), \eta_k\rangle < 0.$$

The next definition, related to the choice of t_k, relies on Armijo's backtracking procedure.

Definition 4.2.2 (Armijo point) *Given a cost function f on a Riemannian manifold $\mathcal{M}$ with retraction R, a point $x \in \mathcal{M}$, a tangent vector $\eta \in T_x\mathcal{M}$, and scalars $\overline{\alpha} > 0$, $\beta, \sigma \in (0,1)$, the* Armijo point *is $\eta^A = t^A\eta = \beta^m\overline{\alpha}\eta$, where m is the smallest nonnegative integer such that*

$$f(x) - f(R_x(\beta^m\overline{\alpha}\eta)) \geq -\sigma \langle \operatorname{grad} f(x), \beta^m\overline{\alpha}\eta\rangle_x.$$

The real t^A is the Armijo step size.

We propose the accelerated Riemannian line-search framework described in Algorithm 1.

The motivation behind Algorithm 1 is to set a framework that is sufficiently general to encompass many methods of interest while being sufficiently restrictive to satisfy certain fundamental convergence properties (proven in the next sections). In particular, it is clear that the choice $x_{k+1} = R_{x_k}(t_k^A\eta_k)$ in Step 3 of Algorithm 1 satisfies (4.12), but this choice is not mandatory. The loose condition (4.12) leaves a lot of leeway for exploiting problem-related information that may lead to a more efficient algorithm. In particular, the choice $x_{k+1} = R_{x_k}(t_k^*\eta_k)$, where $t_k^* = \arg\min_t f(R_{x_k}(t\eta_k))$, satisfies (4.12) and is a reasonable choice if this exact line search can be carried out efficiently.

Algorithm 1 Accelerated Line Search (ALS)

```
Require: Riemannian manifold $\mathcal{M}$; continuously differentiable scalar field
    $f$ on $\mathcal{M}$; retraction $R$ from $T\mathcal{M}$ to $\mathcal{M}$; scalars $\overline{\alpha} > 0$, $c, \beta, \sigma \in (0,1)$.
Input: Initial iterate $x_0 \in \mathcal{M}$.
Output: Sequence of iterates $\{x_k\}$.
for $k = 0, 1, 2, \ldots$ do
    Pick $\eta_k$ in $T_{x_k}\mathcal{M}$ such that the sequence $\{\eta_i\}_{i=0,1,\ldots}$ is gradient-related
    (Definition 4.2.1).
    Select $x_{k+1}$ such that
        $f(x_k) - f(x_{k+1}) \geq c\,(f(x_k) - f(R_{x_k}(t_k^A \eta_k))),$    (4.12)
    where $t_k^A$ is the Armijo step size (Definition 4.2.2) for the given
    $\overline{\alpha}, \beta, \sigma, \eta_k$.
end for
```

If there exists a computationally efficient procedure to minimize $f \circ R_{x_k}$ on a two-dimensional subspace of $T_{x_k}\mathcal{M}$, then a possible choice for x_{k+1} in Step 3 is $R_{x_k}(\xi_k)$, with ξ_k defined by

$$\xi_k := \arg\min_{\xi \in \mathcal{S}_k} f(R_{x_k}(\xi)), \quad \mathcal{S}_k := \operatorname{span}\left\{\eta_k, R_{x_k}^{-1}(x_{k-1})\right\}, \tag{4.13}$$

where $\operatorname{span}\{u, v\} = \{au + bv : a, b \in \mathbb{R}\}$. This is a minimization over a two-dimensional subspace $\mathcal{S}_k$ of $T_{x_k}\mathcal{M}$. It is clear that $\mathcal{S}_k$ contains the Armijo point associated with η_k, since η_k is in $\mathcal{S}_k$. It follows that the bound (4.12) on x_{k+1} holds with $c = 1$. This "two-dimensional subspace acceleration" is well defined on a Riemannian manifold as long as x_k is sufficiently close to x_{k-1} that $R_{x_k}^{-1}(x_{k-1})$ is well defined. The approach is very efficient in the context of eigensolvers (see Section 4.6).

4.3 CONVERGENCE ANALYSIS

In this section, we define and discuss the notions of convergence and limit points on manifolds, then we give a global convergence result for Algorithm 1.

4.3.1 Convergence on manifolds

The notion of convergence on manifolds is a straightforward generalization of the $\mathbb{R}^n$ case. An infinite sequence $\{x_k\}_{k=0,1,\ldots}$ of points of a manifold $\mathcal{M}$ is said to be *convergent* if there exists a chart $(\mathcal{U}, \psi)$ of $\mathcal{M}$, a point $x_* \in \mathcal{U}$, and a $K > 0$ such that x_k is in $\mathcal{U}$ for all $k \geq K$ and such that the sequence $\{\psi(x_k)\}_{k=K,K+1,\ldots}$ converges to $\psi(x_*)$. The point $\psi^{-1}(\lim_{k\to\infty} \psi(x_k))$ is called the *limit* of the convergent sequence $\{x_k\}_{k=0,1,\ldots}$. Every convergent sequence of a (Hausdorff) manifold has one and only one limit point. (The Hausdorff assumption is crucial here. Multiple distinct limit points are possible for non-Hausdorff topologies; see Section 4.3.2.)

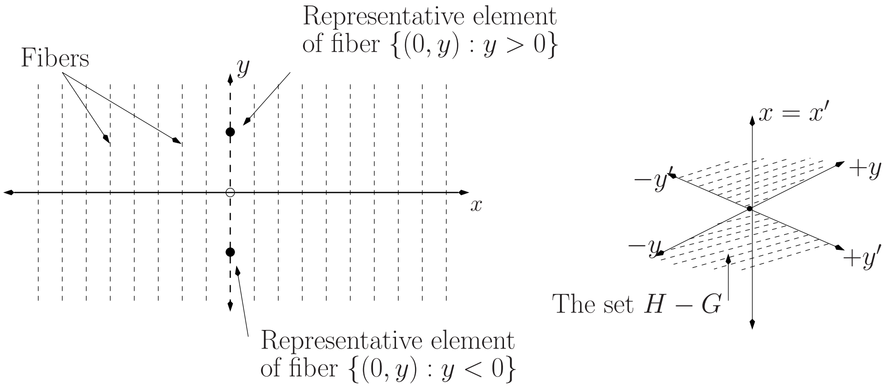
Fibers
y
Representative element of fiber $\{(0,y) : y > 0\}$
x
Representative element of fiber $\{(0,y) : y < 0\}$
$x = x'$
$-y$
$+y$
$-y$
$+y'$
The set $H - G$

Figure 4.2 Left: A few equivalence classes of the quotient defined in Section 4.3.2. Right: The graph $\mathcal{G}$ consists of all the points in $\mathcal{H} \equiv \mathbb{R}^3$ that do not lie on the dashed planes indicated.

An equivalent and more concise definition is that a sequence on a manifold is convergent if it is convergent in the manifold topology, i.e., there is a point x_* such that every neighborhood of x_* contains all but finitely many points of the sequence.

Given a sequence $\{x_k\}_{k=0,1,\ldots}$, we say that x is an *accumulation point* or a *limit point* of the sequence if there exists a subsequence $\{x_{j_k}\}_{k=0,1,\ldots}$ that converges to x. The set of accumulation points of a sequence is called the *limit set* of the sequence.

4.3.2 A topological curiosity*

We present a non-Hausdorff quotient and a convergent sequence with two limit points.

Consider the set $\overline{\mathcal{M}} = \mathbb{R}^2_*$, i.e., the real plane with the origin excerpted. Consider the equivalence relation $\sim$ on $\overline{\mathcal{M}}$, where $(x, y) \sim (x', y')$ if and only if $x = x'$ and the straight line between (x, y) and (x', y') lies wholly in $\mathbb{R}^2_*$. In other words, the equivalence classes of $\sim$ are the two vertical half-lines $\{(0, y) : y > 0\}$ and $\{(0, y) : y < 0\}$ and all the vertical lines $\{(x, y) : y \in \mathbb{R}\}$, $x \neq 0$; see Figure 4.2.

Using Proposition 3.4.3, we show that $\overline{\mathcal{M}}/\sim$ admits a (unique) differentiable structure that makes the natural projection π a submersion, and we show that the topology induced by this differentiable structure is not Hausdorff. Consider the graph $\mathcal{G} = \{((x, y), (x', y')) : (x, y) \sim (x', y')\} \subset \overline{\mathcal{M}} \times \overline{\mathcal{M}}$. Set $\mathcal{H} = \{((x, y), (x', y')) : x = x'\}$ and observe that $\mathcal{G} \subseteq \mathcal{H}$ and $\mathcal{H}$ is an embedded submanifold of $\overline{\mathcal{M}} \times \overline{\mathcal{M}}$. The set $\mathcal{H} - \mathcal{G} = \{((x, y), (x', y')) : x = x' = 0, \operatorname{sign}(y) \neq \operatorname{sign}(y')\}$ is a closed subset of $\mathcal{H}$. It follows that $\mathcal{G}$ is an open submanifold of $\mathcal{H}$ and consequently an embedded submanifold of $\overline{\mathcal{M}} \times \overline{\mathcal{M}}$. It is straightforward to verify that $\pi_1 : \mathcal{G} \to \overline{\mathcal{M}}$ is a submersion. However, $\mathcal{G}$ is open in $\mathcal{H}$, hence $\mathcal{G}$ is not closed in $\overline{\mathcal{M}} \times \overline{\mathcal{M}}$. The conclusion follows from

Proposition 3.4.3.

To help the intuition, we produce a diffeomorphism between $\overline{\mathcal{M}}/\sim$ and a subset of $\overline{\mathcal{M}}$. Let $\mathcal{X}_0 = \{(x, 0) : x \neq 0\}$ denote the horizontal axis of the real plane with the origin excluded. The quotient set $\overline{\mathcal{M}}/\sim$ is in one-to-one correspondence with $\mathcal{N} := \mathcal{X}_0 \cup \{(0,1),(0,-1)\}$ through the mapping Φ that sends each equivalence class to its element contained in $\mathcal{N}$. Let $\mathcal{U}_+ := \mathcal{X}_0 \cup \{(0,1)\}$ and $\mathcal{U}_- := \mathcal{X}_0 \cup \{(0,-1)\}$. Define charts ψ_+ and ψ_- of the set $\mathcal{N}$ into $\mathbb{R}$ with domains $\mathcal{U}_+$ and $\mathcal{U}_-$ by $\psi_\pm((x,0)) = x$ for all $x \neq 0$ and $\psi_+((0,1)) = 0$, $\psi_-((0,-1)) = 0$. These charts form an atlas of the set $\mathcal{N}$ and thus define a differentiable structure on $\mathcal{N}$. It is easy to check that the mapping $\Phi \circ \pi : \overline{\mathcal{M}} \to \mathcal{N}$, where $\pi : \overline{\mathcal{M}} \to \overline{\mathcal{M}}/\sim$ is the natural projection, is a submersion. In view of Proposition 3.4.3, this implies that the sets $\overline{\mathcal{M}}/\sim$ and $\mathcal{N}$, endowed with their differentiable structures, are diffeomorphic.

It is easy to produce a convergent sequence on $\mathcal{N}$ with two limit points. The sequence $\{(1/k, 0)\}_{k=1,2,\ldots}$ converges to $(0,1)$ since $\{\psi_+(1/k,0)\}$ converges to $\psi_+(0,1)$. It also converges to $(0,-1)$ since $\{\psi_-(1/k,0)\}$ converges to $\psi_-(0,-1)$.

4.3.3 Convergence of line-search methods

We give a convergence result for the line-search method defined in Algorithm 1. The statement and the proof are inspired by the classical theory in $\mathbb{R}^n$. However, even when applied to $\mathbb{R}^n$, our statement is more general than the standard results. First, the line search is not necessarily done along a straight line. Second, points other than the Armijo point can be selected; for example, using a minimization over a subspace containing the Armijo point.

Theorem 4.3.1 *Let $\{x_k\}$ be an infinite sequence of iterates generated by Algorithm 1. Then every accumulation point of $\{x_k\}$ is a critical point of the cost function f.*

Proof. By contradiction. Suppose that there is a subsequence $\{x_k\}_{k \in \mathcal{K}}$ converging to some x_* with $\operatorname{grad} f(x_*) \neq 0$. Since $\{f(x_k)\}$ is nonincreasing, it follows that the whole sequence $\{f(x_k)\}$ converges to $f(x_*)$. Hence $f(x_k) - f(x_{k+1})$ goes to zero. By construction of the algorithm,

$$f(x_k) - f(x_{k+1}) \geq -c\sigma\alpha_k \langle \operatorname{grad} f(x_k), \eta_k \rangle_{x_k}.$$

Since $\{\eta_k\}$ is gradient-related, we must have $\{\alpha_k\}_{k \in \mathcal{K}} \to 0$. The α_k's are determined from the Armijo rule, and it follows that for all k greater than some $\overline{k}$, $\alpha_k = \beta^{m_k}\overline{\alpha}$, where m_k is an integer greater than zero. This means that the update $\frac{\alpha_k}{\beta}\eta_k$ did not satisfy the Armijo condition. Hence

$$f(x_k) - f\left(R_{x_k}\left(\frac{\alpha_k}{\beta}\eta_k\right)\right) < -\sigma\frac{\alpha_k}{\beta}\langle \operatorname{grad} f(x_k), \eta_k \rangle_{x_k}, \quad \forall k \in \mathcal{K},\ k \geq \overline{k}.$$

Denoting

$$\tilde{\eta}_k = \frac{\eta_k}{\|\eta_k\|} \quad \text{and} \quad \tilde{\alpha}_k = \frac{\alpha_k \|\eta_k\|}{\beta}, \tag{4.14}$$

the inequality above reads

$$\frac{\widehat{f}_{x_k}(0) - \widehat{f}_{x_k}(\tilde{\alpha}_k\tilde{\eta}_k)}{\tilde{\alpha}_k} < -\sigma\langle \operatorname{grad} f(x_k), \tilde{\eta}_k\rangle_{x_k}, \quad \forall k \in \mathcal{K},\ k \geq \overline{k},$$

where $\widehat{f}$ is defined as in (4.3). The mean value theorem ensures that there exists $t \in [0, \tilde{\alpha}_k]$ such that

$$-\operatorname{D}\widehat{f}_{x_k}(t\tilde{\eta}_k)[\tilde{\eta}_k] < -\sigma\langle \operatorname{grad} f(x_k), \tilde{\eta}_k\rangle_{x_k}, \quad \forall k \in \mathcal{K},\ k \geq \overline{k}, \tag{4.15}$$

where the differential is taken with respect to the Euclidean structure on $T_{x_k}\mathcal{M}$. Since $\{\alpha_k\}_{k\in\mathcal{K}} \to 0$ and since η_k is gradient-related, hence bounded, it follows that $\{\tilde{\alpha}_k\}_{k\in\mathcal{K}} \to 0$. Moreover, since $\tilde{\eta}_k$ has unit norm, it thus belongs to a compact set, and therefore there exists an index set $\tilde{\mathcal{K}} \subseteq \mathcal{K}$ such that $\{\tilde{\eta}_k\}_{k\in\tilde{\mathcal{K}}} \to \tilde{\eta}_*$ for some $\tilde{\eta}_*$ with $\|\tilde{\eta}_*\| = 1$. We now take the limit in (4.15) over $\tilde{\mathcal{K}}$. Since the Riemannian metric is continuous (by definition), and $f \in C^1$, and $D\hat{f}_{x_k}(0)[\tilde{\eta}_k] = \langle \operatorname{grad} f(x_k), \tilde{\eta}_k\rangle_{x_k}$—see (3.31) and (4.4)—we obtain

$$-\langle \operatorname{grad} f(x_*), \tilde{\eta}_*\rangle_{x_*} \leq -\sigma\langle \operatorname{grad} f(x_*), \tilde{\eta}_*\rangle_{x_*}.$$

Since $\sigma < 1$, it follows that $\langle \operatorname{grad} f(x_*), \tilde{\eta}_*\rangle_{x_*} \geq 0$. On the other hand, from the fact that $\{\eta_k\}$ is gradient-related, one obtains that $\langle \operatorname{grad} f(x_*), \tilde{\eta}_*\rangle_{x_*} < 0$, a contradiction. □

More can be said under compactness assumptions using a standard argument.

Corollary 4.3.2 *Let $\{x_k\}$ be an infinite sequence of iterates generated by Algorithm 1. Assume that the level set $\mathcal{L} = \{x \in \mathcal{M} : f(x) \leq f(x_0)\}$ is compact (which holds in particular when $\mathcal{M}$ itself is compact). Then* $\lim_{k\to\infty} \|\operatorname{grad} f(x_k)\| = 0$.

Proof. By contradiction, assume the contrary. Then there is a subsequence $\{x_k\}_{k\in\mathcal{K}}$ and $\epsilon > 0$ such that $\|\operatorname{grad} f(x_k)\| > \epsilon$ for all $k \in \mathcal{K}$. Because f is nonincreasing on $\{x_k\}$, it follows that $x_k \in \mathcal{L}$ for all k. Since $\mathcal{L}$ is compact, $\{x_k\}_{k\in\mathcal{K}}$ has an accumulation point x_* in $\mathcal{L}$. By the continuity of $\operatorname{grad} f$, one has $\|\operatorname{grad} f(x_*)\| \geq \epsilon$; i.e., x_* is not critical, a contradiction to Theorem 4.3.1. □

4.4 STABILITY OF FIXED POINTS

Theorem 4.3.1 states that only critical points of the cost function f can be accumulation points of sequences $\{x_k\}$ generated by Algorithm 1. This result gives useful information on the behavior of Algorithm 1. Still, it falls short of what one would expect of an optimization method. Indeed, Theorem 4.3.1 does not specify whether the accumulation points are local minimizers, local maximizers, or *saddle points* (critical points that are neither local minimizers nor local maximizers).

Unfortunately, avoiding saddle points and local maximizers is too much to ask of a method that makes use of only first-order information on the cost function. We illustrate this with a very simple example. Let x_* be any critical point of a cost function f and consider the sequence $\{(x_k, \eta_k)\}$, $x_k = x_*$, $\eta_k = 0$. This sequence satisfies the requirements of Algorithm 1, and $\{x_k\}$ trivially converges to x_* even if x_* is a saddle point or a local minimizer.

Nevertheless, it is observed in practice that unless the initial point x_0 is carefully crafted, methods within the framework of Algorithm 1 do produce sequences whose accumulation points are local minima of the cost function. This observation is supported by the following stability analysis of critical points.

Let F be a mapping from $\mathcal{M}$ to $\mathcal{M}$. A point $x_* \in \mathcal{M}$ is a *fixed point* of F if $F(x_*) = x_*$. Let $F^{(n)}$ denote the result of n applications of F to x, i.e.,

$$F^{(1)}(x) = F(x), \quad F^{(i+1)}(x) = F(F^{(i)}(x)),\ i = 1, 2, \ldots.$$

A fixed point x_* of F is a *stable point* of F if, for every neighborhood $\mathcal{U}$ of x_*, there exists a neighborhood $\mathcal{V}$ of x_* such that, for all $x \in \mathcal{V}$ and all positive integer n, it holds that $F^{(n)}(x) \in \mathcal{U}$. The fixed point x_* is *asymptotically stable* if it is stable, and, moreover, $\lim_{n\to\infty} F^{(n)}(x) = x_*$ for all x sufficiently close to x_*. The fixed point x_* is *unstable* if it is not stable; in other words, there exists a neighborhood $\mathcal{U}$ of x_* such that, for all neighborhood $\mathcal{V}$ of x_*, there is a point $x \in \mathcal{V}$ such that $F^{(n)}(x) \notin \mathcal{U}$ for some n. We say that F is a *descent mapping* for a cost function f if

$$f(F(x)) \leq f(x) \quad \text{for all } x \in \mathcal{M}.$$

Theorem 4.4.1 (unstable fixed points) *Let $F : \mathcal{M} \to \mathcal{M}$ be a descent mapping for a smooth cost function f and assume that for every $x \in \mathcal{M}$, all the accumulation points of $\{F^{(k)}(x)\}_{k=1,2,\ldots}$ are critical points of f. Let x_* be a fixed point of F (thus x_* is a critical point of f). Assume that x_* is not a local minimum of f. Further assume that there is a compact neighborhood $\mathcal{U}$ of x_* where, for every critical point y of f in $\mathcal{U}$, $f(y) = f(x_*)$. Then x_* is an unstable point for F.*

Proof. Since x_* is not a local minimum of f, it follows that every neighborhood $\mathcal{V}$ of x_* contains a point y with $f(y) < f(x_*)$. Consider the sequence $y_k := F^{(k)}(y)$. Suppose for the purpose of establishing a contradiction that $y_k \in \mathcal{U}$ for all k. Then, by compactness, $\{y_k\}$ has an accumulation point z in $\mathcal{U}$. By assumption, z is a critical point of f, hence $f(z) = f(x_*)$. On the other hand, since F is a descent mapping, it follows that $f(z) \leq f(y) < f(x_*)$, a contradiction. □

The assumptions made about f and F in Theorem 4.4.1 may seem complicated, but they are satisfied in many circumstances. The conditions on F are satisfied by any method in the class of Algorithm 1. As for the condition on the critical points of f, it holds for example when f is real analytic. (This property can be recovered from Łojasiewicz's gradient inequality: if f is real analytic around x_*, then there are constants $c > 0$ and $\mu \in [0, 1)$ such that

$$\|\operatorname{grad} f(x)\| \geq c|f(x) - f(x_*)|^{\mu}$$

for all x in some neighborhood of x_*.)

We now give a stability result.

Theorem 4.4.2 (capture theorem) *Let $F : \mathcal{M} \to \mathcal{M}$ be a descent mapping for a smooth cost function f and assume that, for every $x \in \mathcal{M}$, all the accumulation points of $\{F^{(k)}(x)\}_{k=1,2,\dots}$ are critical points of f. Let x_* be a local minimizer and an isolated critical point of f. Assume further that $\operatorname{dist}(F(x), x)$ goes to zero as x goes to x_*. Then x_* is an asymptotically stable point of F.*

Proof. Let $\mathcal{U}$ be a neighborhood of x_*. Since x_* is an isolated local minimizer of f, it follows that there exists a closed ball

$$\overline{B}_\epsilon(x_*) := \{x \in \mathcal{M} : \operatorname{dist}(x, x_*) \le \epsilon\}$$

such that $\overline{B}_\epsilon(x_*) \subset \mathcal{U}$ and $f(x) > f(x_*)$ for all $x \in \overline{B}_\epsilon(x_*) - \{x_*\}$. In view of the condition on $\operatorname{dist}(F(x), x)$, there exists $\delta > 0$ such that, for all $x \in B_\delta(x_*)$, $F(x) \in \overline{B}_\epsilon(x_*)$. Let α be the minimum of f on the compact set $\overline{B}_\epsilon(x_*) - B_\delta(x_*)$. Let

$$\mathcal{V} = \{x \in \overline{B}_\epsilon(x_*) : f(x) < \alpha\}.$$

This set is included in $B_\delta(x_*)$. Hence, for every x in $\mathcal{V}$, it holds that $F(x) \in \overline{B}_\epsilon(x_*)$, and it also holds that $f(F(x)) \le f(x) < \alpha$ since F is a descent mapping. It follows that $F(x) \in \mathcal{V}$ for all $x \in \mathcal{V}$, hence $F^{(n)}(x) \in \mathcal{V} \subset \mathcal{U}$ for all $x \in \mathcal{V}$ and all n. This is stability. Moreover, since by assumption x_* is the only critical point of f in $\mathcal{V}$, it follows that $\lim_{n\to\infty} F^{(n)}(x) = x_*$ for all $x \in \mathcal{V}$, which shows asymptotic stability. □

The additional condition on $\operatorname{dist}(F(x), x)$ in Theorem 4.4.2 is not satisfied by every instance of Algorithm 1 because our accelerated line-search framework does not put any restriction on the step length. The distance condition is satisfied, for example, when η_k is selected such that $\|\eta_k\| \le c\|\operatorname{grad} f(x_k)\|$ for some constant c and x_{k+1} is selected as the Armijo point.

In this section, we have assumed for simplicity that the next iterate depends only on the current iterate: $x_{k+1} = F(x_k)$. It is possible to generalize the above result to the case where x_{k+1} depends on x_k and on some "memory variables": $(x_{k+1}, y_{k+1}) = F(x_k, y_k)$.

4.5 SPEED OF CONVERGENCE

We have seen that, under reasonable assumptions, if the first iterate of Algorithm 1 is sufficiently close to an isolated local minimizer x_* of f, then the generated sequence $\{x_k\}$ converges to x_*. In this section, we address the issue of how fast the sequence converges to x_*.

4.5.1 Order of convergence

A sequence $\{x_k\}_{k=0,1,\dots}$ of points of $\mathbb{R}^n$ is said to converge linearly to a point x_* if there exists a constant $c \in (0, 1)$ and an integer $K \ge 0$ such that, for all $k \geq K$, it holds that $\|x_{k+1} - x_*\| \leq c\|x_k - x_*\|$. In order to generalize this notion to manifolds, it is tempting to fall back to the $\mathbb{R}^n$ definition using charts and state that a sequence $\{x_k\}_{k=0,1,\dots}$ of points of a manifold $\mathcal{M}$ converges linearly to a point $x_* \in \mathcal{M}$ if, given a chart $(\mathcal{U}, \psi)$ with $x \in \mathcal{U}$, the sequence $\{\psi(x_k)\}_{k=0,1,\dots}$ converges linearly to $\psi(x_*)$. Unfortunately, the notion is not independent of the chart used. For example, let $\mathcal{M}$ be the set $\mathbb{R}^n$ with its canonical manifold structure and consider the sequence $\{x_k\}_{k=0,1,\dots}$ defined by $x_k = 2^{-k}e_1$ if k is even and by $x_k = 2^{-k+2}e_2$ if k is odd. In the identity chart, this sequence is not linearly convergent because of the requirement that the constant c be smaller than 1. However, in the chart defined by $\psi(xe_1 + ye_2) = xe_1 + (y/4)e_2$, the sequence converges linearly with constant $c = \frac{1}{2}$.

If $\mathcal{M}$ is a Riemannian manifold, however, then the induced Riemannian distance makes it possible to define linear convergence as follows.

Definition 4.5.1 (linear convergence) *Let $\mathcal{M}$ be a Riemannian manifold and let* dist *denote the Riemannian distance on $\mathcal{M}$. We say that a sequence $\{x_k\}_{k=0,1,\dots}$* converges linearly *to a point $x_* \in \mathcal{M}$ if there exists a constant $c \in (0,1)$ and an integer $K \geq 0$ such that, for all $k \geq K$, it holds that*

$$\operatorname{dist}(x_{k+1}, x_*) \leq c \operatorname{dist}(x_k, x_*). \tag{4.16}$$

The limit

$$\limsup_{k\to\infty} \frac{\operatorname{dist}(x_{k+1}, x_*)}{\operatorname{dist}(x_k, x_*)}$$

is called the linear convergence factor *of the sequence. An iterative algorithm on $\mathcal{M}$ is said to* converge locally linearly *to a point x_* if there exists a neighborhood $\mathcal{U}$ of x_* and a constant $c \in (0,1)$ such that, for every initial point $x_0 \in \mathcal{U}$, the sequence $\{x_k\}$ generated by the algorithm satisfies* (4.16).

A convergent sequence $\{x_k\}$ on a Riemannian manifold $\mathcal{M}$ converges linearly to x_* with constant c if and only if

$$\|R_{x_*}^{-1}(x_{k+1}) - R_{x_*}^{-1}(x_*)\| \leq c\|R_{x_*}^{-1}(x_k) - R_{x_*}^{-1}(x_*)\|$$

for all k sufficiently large, where R is any retraction on $\mathcal{M}$ and $\|\cdot\|$ denotes the norm on $T_{x_*}\mathcal{M}$ defined by the Riemannian metric. (To see this, let Exp_{x_*} denote the exponential mapping introduced in Section 5.4, restricted to a neighborhood $\hat{\mathcal{U}}$ of 0_{x_*} in $T_{x_*}\mathcal{M}$ such that $\mathcal{U} := \mathrm{Exp}_{x_*}(\hat{\mathcal{U}})$ is a normal neighborhood of x_*. We have $\operatorname{dist}(x, x_*) = \|\mathrm{Exp}_{x_*}^{-1}(x) - \mathrm{Exp}_{x_*}^{-1}(x_*)\| = \|\mathrm{Exp}_{x_*}^{-1}(x)\|$ for all $x \in \mathcal{U}$. Moreover, since Exp is a retraction, we have $\mathrm{D}(R_{x_*}^{-1} \circ \mathrm{Exp}_{x_*})(0_{x_*}) = \mathrm{id}$. Hence $\|R_{x_*}^{-1}(x) - R_{x_*}^{-1}(x_*)\| = \|\mathrm{Exp}_{x_*}^{-1}(x) - \mathrm{Exp}_{x_*}^{-1}(x_*)\| + o(\|\mathrm{Exp}_{x_*}^{-1}(x) - \mathrm{Exp}_{x_*}^{-1}(x_*)\|) = \operatorname{dist}(x, x_*) + o(\operatorname{dist}(x, x_*))$.)

In contrast to linear convergence, the notions of superlinear convergence and order of convergence can be defined on a manifold independently of any other structure.

Definition 4.5.2 *Let $\mathcal{M}$ be a manifold and let $\{x_k\}_{k=0,1,\ldots}$ be a sequence on $\mathcal{M}$ converging to x_*. Let $(\mathcal{U}, \psi)$ be a chart of $\mathcal{M}$ with $x \in \mathcal{U}$. If*

$$\lim_{k\to\infty} \frac{\|\psi(x_{k+1}) - \psi(x_*)\|}{\|\psi(x_k) - \psi(x_*)\|} = 0,$$

then $\{x_k\}$ is said to converge superlinearly *to x_*. If there exist constants $p > 0$, $c \geq 0$, and $K \geq 0$ such that, for all $k \geq K$, there holds*

$$\|\psi(x_{k+1}) - \psi(x_*)\| \leq c\|\psi(x_k) - \psi(x_*)\|^p, \tag{4.17}$$

then $\{x_k\}$ is said to converge to x_* with order at least p. An iterative algorithm on a manifold $\mathcal{M}$ is said to converge locally to a point x_* with order at least p if there exists a chart $(\mathcal{U}, \psi)$ at x_* and a constant $c > 0$ such that, for every initial point $x_0 \in \mathcal{U}$, the sequence $\{x_k\}$ generated by the algorithm satisfies (4.17)*. If $p = 2$, the convergence is said to be* quadratic*, and* cubic *if $p = 3$.*

Since by definition charts overlap smoothly, it can be shown that the definitions above do not depend on the choice of the chart $(\mathcal{U}, \psi)$. (The multiplicative constant c depends on the chart, but for any chart, there exists such a constant.)

Theorem 4.5.3 below gives calculus-based local convergence results for iterative methods defined by $x_{k+1} = F(x_k)$, where the iteration mapping $F : \mathcal{M} \to \mathcal{M}$ has certain smoothness properties.

Theorem 4.5.3 *Let $F : \mathcal{M} \to \mathcal{M}$ be a C^1 mapping whose domain and range include a neighborhood of a fixed point x_* of F.*

(i) If $\mathrm{D}F(x_*) = 0$, then the iterative algorithm with iteration mapping F converges locally superlinearly to x_*.
(ii) If $\mathrm{D}F(x_*) = 0$ and F is C^2, then the iterative algorithm with mapping F converges locally quadratically to x_*.

Although Theorem 4.5.3 is very powerful for smooth iteration mappings, it is rarely useful for practical line-search and trust-region methods because of the nondifferentiability of the step selection process.

4.5.2 Rate of convergence of line-search methods*

In this section we give an asymptotic convergence bound for Algorithm 1 when η_k is chosen as $-\mathrm{grad}\, f(x_k)$, without any further assumption on how x_{k+1} is selected.

The result invokes the smallest and largest eigenvalues of the Hessian of f at a critical point x_*. We have not yet given a definition for the Hessian of a cost function on a Riemannian manifold. (This is done in Section 5.5.) Nevertheless, regardless of this definition, it makes sense to talk about the eigenvalues of the Hessian at a critical point because of the following results.

Lemma 4.5.4 *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $x_* \in \mathbb{R}^n$ such that $\mathrm{D}f(x_*) = 0$. Let $F : \mathbb{R}^n \to \mathbb{R}^n$ and $y_* \in \mathbb{R}^n$ such that $F(y_*) = x_*$ and that the* Jacobian matrix *of F at y_*,*

$$\mathrm{J}_F(y_*) := \begin{bmatrix} \partial_1 F^1(y_*) & \cdots & \partial_n F^1(y_*) \\ \vdots & \ddots & \vdots \\ \partial_1 F^n(y_*) & \cdots & \partial_n F^n(y_*) \end{bmatrix},$$

is orthogonal (i.e., $\mathrm{J}_F^T(y_*)\mathrm{J}_F(y_*) = I$). Let H be the Hessian matrix of f at x_*; i.e., $H_{ij} = \partial_i\partial_j f(x_*)$. Let $\hat{H}$ be the Hessian matrix of $f \circ F$ at y_*. Then $\lambda(H) = \lambda(\hat{H})$; i.e., the spectrum of H and the spectrum of $\hat{H}$ are the same.

Proof. Since $\partial_j(f \circ F)(y) = \sum_k \partial_k f(F(y))\, \partial_j F^k(y)$, we have

$$\begin{aligned} \hat{H}_{ij} &= \partial_i\partial_j(f \circ F)(y_*) \\ &= \sum_{k,\ell} \partial_\ell\partial_k f(F(y_*))\, \partial_i F^\ell(y_*)\, \partial_j F^k(y_*) + \sum_k \partial_k f(F(y_*))\, \partial_i\partial_j F^k(y_*). \end{aligned}$$

Since x_* is a critical point of f, it follows that $\partial_k f(F(y_*)) = 0$. Hence we have, in matrix notation,

$$\hat{H} = \mathrm{J}_F^T(y_*) H \mathrm{J}_F(y_*) = \mathrm{J}_F^{-1}(y_*) H \mathrm{J}_F(y_*).$$

This shows that H and $\hat{H}$ have the same spectrum because they are related by a similarity transformation. □

Corollary 4.5.5 *Let f be a cost function on a Riemannian manifold $(\mathcal{M}, g)$ and let $x_* \in \mathcal{M}$ be a critical point of f, i.e.,* $\operatorname{grad} f(x_*) = 0$. *Let $(\mathcal{U}, \psi)$ be any chart such that $x_* \in \mathcal{U}$ and that the representation of g_{x_*} in the chart is the identity, i.e., $g_{ij} = \delta_{ij}$ at x_*. Then the spectrum of the Hessian matrix of $f \circ \psi^{-1}$ at $\psi(x_*)$ does not depend on the choice of ψ.*

We can now state the main result of this section. When reading the theorem below, it is useful to note that $0 < r_* < 1$ since $\beta, \sigma \in (0, 1)$. Also, in common instances of Algorithm 1, the constant c in the descent condition (4.12) is equal to 1, hence (4.18) reduces to $f(x_{k+1}) - f(x_*) \leq r\,(f(x_k) - f(x_*))$.

Theorem 4.5.6 *Let $\{x_k\}$ be an infinite sequence of iterates generated by Algorithm 1 with $\eta_k := -\operatorname{grad} f(x_k)$, converging to a point x_*. (By Theorem 4.3.1, x_* is a critical point of f.) Let $\lambda_{H,\min}$ and $\lambda_{H,max}$ be the smallest and largest eigenvalues of the Hessian of f at x_*. Assume that $\lambda_{H,\min} > 0$ (hence x_* is a local minimizer of f). Then, given r in the interval $(r_*, 1)$ with $r_* = 1 - \min\left(2\sigma\bar{\alpha}\lambda_{H,\min}, 4\sigma(1-\sigma)\beta\frac{\lambda_{H,\min}}{\lambda_{H,\max}}\right)$, there exists an integer $K \geq 0$ such that*

$$f(x_{k+1}) - f(x_*) \leq (r + (1-r)(1-c))\,(f(x_k) - f(x_*)) \tag{4.18}$$

for all $k \geq K$, where c is the parameter in Algorithm 1.

Proof. Let $(\mathcal{U}, \psi)$ be a chart of the manifold $\mathcal{M}$ with $x_* \in \mathcal{U}$. We use the notation $\zeta_x := -\operatorname{grad} f(x)$. Coordinate expressions are denoted with a hat, e.g., $\hat{x} := \psi(x)$, $\hat{\mathcal{U}} = \psi(\mathcal{U})$, $\hat{f}(\hat{x}) := f(x)$, $\hat{\zeta}_{\hat{x}} := \mathrm{D}\psi\,(x)\,[\zeta_x]$, $\hat{R}_{\hat{x}}(\hat{\zeta}) := \psi(R_x(\zeta))$. We also let $y_{\hat{x}}$ denote the Euclidean gradient of $\hat{f}$ at $\hat{x}$, i.e., $y_{\hat{x}} := (\partial_1 \hat{f}(\hat{x}), \ldots, \partial_d \hat{f}(\hat{x}))^T$. We let $G_{\hat{x}}$ denote the matrix representation of the Riemannian metric in the coordinates, and we let $H_{\hat{x}_*}$ denote the Hessian matrix of $\hat{f}$ at $\hat{x}_*$. Without loss of generality, we assume that $\hat{x}_* = 0$ and that $G_{\hat{x}_*} = I$, the identity matrix.

The major work is to obtain, at a current iterate x, a suitable upper bound on $f(R_x(t^A\zeta_x))$, where t^A is the Armijo step (so $t^A\zeta_x$ is the Armijo point). The Armijo condition is

$$\begin{aligned} f(R_x(t^A\zeta_x)) &\leq f(x) - \sigma\langle\zeta_x, t^A\zeta_x\rangle \\ &\leq f(x) - \sigma t^A\langle\zeta_x, \zeta_x\rangle. \end{aligned} \tag{4.19}$$

We first give a lower bound on $\langle\zeta_x, \zeta_x\rangle$ in terms of $f(x)$. Recall from (3.32) that $\hat{\zeta}_{\hat{x}} = G_{\hat{x}}^{-1} y_{\hat{x}}$, from which it follows that

$$\langle\zeta_x, \zeta_x\rangle = \hat{\zeta}_{\hat{x}}^T G_{\hat{x}} \hat{\zeta}_{\hat{x}} = y_{\hat{x}} G_{\hat{x}}^{-1} y_{\hat{x}} = \|y_{\hat{x}}\|^2 (1 + O(\hat{x})) \tag{4.20}$$

since we have assumed that G_0 is the identity. It follows from $y_{\hat{x}} = H_0\hat{x} + O(\hat{x}^2)$ and $\hat{f}(\hat{x}) = \hat{f}(0) + \frac{1}{2}\hat{x}^T H_0 \hat{x} + O(\hat{x}^3)$ that, given $\epsilon \in (0, \lambda_{H,\min})$,

$$\hat{f}(\hat{x}) - \hat{f}(0) = \frac{1}{2} y_{\hat{x}}^T H_0^{-1} y_{\hat{x}} + O(\hat{x}^3) \leq \frac{1}{2}\frac{1}{\lambda_{H,\min} - \epsilon}\|y_{\hat{x}}\|^2 \tag{4.21}$$

holds for all $\hat{x}$ sufficiently close to 0. From (4.20) and (4.21), we conclude that, given $\epsilon \in (0, \lambda_{H,\min})$,

$$f(x) - f(x_*) \leq \frac{1}{2}\frac{1}{\lambda_{H,\min} - \epsilon}\langle\zeta_x, \zeta_x\rangle, \tag{4.22}$$

which is the desired lower bound on $\langle\zeta_x, \zeta_x\rangle$. Using (4.22) in (4.19) yields

$$f(R_x(t^A\zeta_x)) - f(x_*) \leq (1 - 2(\lambda_{H,\min} - \epsilon)\sigma t^A)(f(x) - f(x_*)). \tag{4.23}$$

We now turn to finding a lower bound on the Armijo step t^A. We use the notation

$$\gamma_{\hat{x},u}(t) := \hat{f}(\hat{R}_{\hat{x}}(tu))$$

and

$$h_x(t) = f(R_x(-t\zeta_x)).$$

Notice that $h_x(t) = \gamma_{\hat{x},-\hat{\zeta}_{\hat{x}}}(t)$ and that $\dot{h}_x(0) = -\langle\zeta_x, \zeta_x\rangle = \dot{\gamma}_{\hat{x},-\hat{\zeta}_{\hat{x}}}(0)$, from which it follows that the Armijo condition (4.19) reads

$$h_x(t^A) \leq h_x(0) - \sigma t^A \dot{h}_x(0). \tag{4.24}$$

We want to find a lower bound on t^A. From a Taylor expansion of h_x with the residual in Lagrange form (see Appendix A.6), it follows that the t's at which the left- and right-hand sides of (4.24) are equal satisfy

$$t = \frac{-2(1-\sigma)\dot{h}_x(0)}{\ddot{h}_x(\tau)},$$

where $\tau \in (0, t)$. In view of the definition of the Armijo point, we conclude that

$$t^A \geq \min\left(\bar{\alpha}, \frac{-2\beta(1-\sigma)\dot{h}_x(0)}{\max_{\tau\in(0,\bar{\alpha})} \ddot{h}_x(\tau)}\right). \tag{4.25}$$

Let $B_\delta := \{\hat{x} : \|\hat{x}\| < \delta\}$ and

$$M := \sup_{\hat{x}\in B_\delta, \|u\|=1, t\in(0,\bar{\alpha}\|\hat{\zeta}_{\hat{x}}\|)} \ddot{\gamma}_{\hat{x},u}(t).$$

Then $\max_{\tau\in(0,\bar{\alpha})} \ddot{h}_x(\tau) \leq M\|\hat{\zeta}_{\hat{x}}\|^2$. Notice also that $\ddot{\gamma}_{\hat{x},u}(0) = u^T H_0 u \leq \lambda_{H,\max}\|u\|^2$, so that $M \to \lambda_{H,\max}$ as $\delta \to 0$. Finally, notice that $\dot{h}_x(0) = -\hat{\zeta}_{\hat{x}}^T G_{\hat{x}} \hat{\zeta}_{\hat{x}} = \|\hat{\zeta}_{\hat{x}}\|^2(1 + O(\hat{x}))$. Using these results in (4.25) yields that, given $\epsilon > 0$,

$$t^A \geq \min\left(\bar{\alpha}, \frac{2\beta(1-\sigma)}{\lambda_{H,\max} + \epsilon}\right) \tag{4.26}$$

holds for all x sufficiently close to x_*.

We can now combine (4.26) and (4.23) to obtain a suitable upper bound on $f(R_x(t^A\zeta_x))$:

$$f(R_x(t^A\zeta_x)) - f(x_*) \leq c_1(f(x) - f(x_*)) \tag{4.27}$$

with

$$c_1 = 1 - \sigma \min\left(\bar{\alpha}, \frac{2\beta(1-\sigma)}{\lambda_{H,\max} + \epsilon}\right) 2(\lambda_{H,\min} - \epsilon).$$

Finally, the bound (4.27), along with the bound (4.12) imposed on the value of f at the next iterate, yields

$$\begin{aligned}
f(x_{k+1}) - f(x_*) &= f(x_{k+1}) - f(x_k) + f(x_k) - f(x_*) \\
&\leq -c(f(x_k) - f(R_{x_k}(t_k^A\zeta_{x_k}))) + f(x_k) - f(x_*) \\
&= (1-c)(f(x_k) - f(x_*)) + c(f(R_{x_k}(t_k^A\zeta_{x_k})) - f(x_*)) \\
&\leq (1 - c + c\,c_1)(f(x_k) - f(x_*)) \\
&= (c_1 + (1-c_1)(1-c))(f(x_k) - f(x_*)),
\end{aligned}$$

where $c \in (0, 1)$ is the constant in the bound (4.12). □

4.6 RAYLEIGH QUOTIENT MINIMIZATION ON THE SPHERE

In this section we apply algorithms of the class described by Algorithm 1 to the problem of finding a minimizer of

$$f : S^{n-1} \to \mathbb{R} : x \mapsto x^T A x, \tag{4.28}$$

the Rayleigh quotient on the sphere. The matrix A is assumed to be symmetric ($A = A^T$) but not necessarily positive-definite. We let λ_1 denote the smallest eigenvalue of A and v_1 denote an associated unit-norm eigenvector.

4.6.1 Cost function and gradient calculation

Consider the function

$$\overline{f} : \mathbb{R}^n \to \mathbb{R} : x \mapsto x^T A x,$$

whose restriction to the unit sphere S^{n-1} yields (4.28).

We view S^{n-1} as a Riemannian submanifold of the Euclidean space $\mathbb{R}^n$ endowed with the canonical Riemannian metric

$$\overline{g}(\xi, \zeta) = \xi^T \zeta.$$

Given $x \in S^{n-1}$, we have

$$\mathrm{D}\overline{f}\,(x)\,[\zeta] = \zeta^T A x + x^T A \zeta = 2\zeta^T A x$$

for all $\zeta \in T_x\mathbb{R}^n \simeq \mathbb{R}^n$, from which it follows, recalling the definition (3.31) of the gradient, that

$$\operatorname{grad} \overline{f}(x) = 2Ax.$$

The tangent space to S^{n-1}, viewed as a subspace of $T_x\mathbb{R}^n \simeq \mathbb{R}^n$, is

$$T_x S^{n-1} = \{\xi \in \mathbb{R}^n : x^T \xi = 0\}.$$

The normal space is

$$(T_x S^{n-1})^\perp = \{x\alpha : \alpha \in \mathbb{R}\}.$$

The orthogonal projections onto the tangent and the normal space are

$$\mathrm{P}_x \xi = \xi - x x^T \xi, \qquad \mathrm{P}_x^\perp \xi = x x^T \xi.$$

It follows from the identity (3.37), relating the gradient on a submanifold to the gradient on the embedding manifold, that

$$\operatorname{grad} f(x) = 2\mathrm{P}_x(Ax) = 2(Ax - x x^T A x). \tag{4.29}$$

The formulas above are summarized in Table 4.1.

4.6.2 Critical points of the Rayleigh quotient

To analyze an algorithm based on the Rayleigh quotient cost on the sphere, the first step is to characterize the critical points.

Proposition 4.6.1 *Let $A = A^T$ be an $n \times n$ symmetric matrix. A unit-norm vector $x \in \mathbb{R}^n$ is an eigenvector of A if and only if it is a critical point of the Rayleigh quotient* (4.28)*.*

Proof. Let x be a critical point of (4.28), i.e., $\operatorname{grad} f(x) = 0$ with $x \in S^{n-1}$. From the expression (4.29) of $\operatorname{grad} f(x)$, it follows that x statisfies $Ax = (x^T A x)x$, where $x^T A x$ is a scalar. Conversely, if x is a unit-norm eigenvector of A, i.e., $Ax = \lambda x$ for some scalar λ, then a left multiplication by x^T yields $\lambda = x^T A x$ and thus $Ax = (x^T A x)x$, hence $\operatorname{grad} f(x) = 0$ in view of (4.29). □

We already know from Proposition 2.1.1 that the two points $\pm v_1$ corresponding to the "leftmost" eigendirection are the global minima of the Rayleigh quotient (4.28). Moreover, the other eigenvectors are not local minima:

Table 4.1 Rayleigh quotient on the unit sphere.

	Manifold (S^{n-1})	Embedding space ($\mathbb{R}^n$)
cost	$f(x) = x^T Ax$, $x \in S^{n-1}$	$\overline{f}(x) = x^T Ax$, $x \in \mathbb{R}^n$
metric	induced metric	$\overline{g}(\xi, \zeta) = \xi^T \zeta$
tangent space	$\xi \in \mathbb{R}^n : x^T \xi = 0$	$\mathbb{R}^n$
normal space	$\xi \in \mathbb{R}^n : \xi = \alpha x$	$\emptyset$
projection onto tangent space	$\mathrm{P}_x \xi = (I - xx^T)\xi$	identity
gradient	$\operatorname{grad} f(x) = \mathrm{P}_x \operatorname{grad} \overline{f}(x)$	$\operatorname{grad} \overline{f}(x) = 2Ax$
retraction	$R_x(\xi) = \operatorname{qf}(x + \xi)$	$R_x(\xi) = x + \xi$

Proposition 4.6.2 *Let $A = A^T$ be an $n \times n$ symmetric matrix with eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$ and associated orthonormal eigenvectors $v_1, \ldots, v_n$. Then*

(i) $\pm v_1$ are local and global minimizers of the Rayleigh quotient (4.28)*; if the eigenvalue λ_1 is simple, then they are the only minimizers.*
(ii) $\pm v_n$ are local and global maximizers of (4.28)*; if the eigenvalue λ_n is simple, then they are the only maximizers.*
(iii) $\pm v_q$ corresponding to interior eigenvalues (i.e., strictly larger than λ_1 and strictly smaller than λ_n) are saddle points of (4.28)*.*

Proof. Point (i) follows from Proposition 2.1.1. Point (ii) follows from the same proposition by noticing that replacing A by $-A$ exchanges maxima with minima and leftmost eigenvectors with rightmost eigenvectors. For point (iii), let v_q be an eigenvector corresponding to an interior eigenvalue λ_q and consider the curve $\gamma : t \mapsto (v_q + tv_1)/\|v_q + tv_1\|$. Simple calculus shows that

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}(f(\gamma(t))|_{t=0} = \lambda_1 - \lambda_q < 0.$$

Likewise, for the curve $\gamma : t \mapsto (v_q + tv_n)/\|v_q + tv_n\|$, we have

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}(f(\gamma(t))|_{t=0} = \lambda_n - \lambda_q > 0.$$

It follows that v_q is a saddle point of the Rayleigh quotient f. □

It follows from Proposition 4.6.1 and the global convergence analysis of line-search methods (Proposition 4.3.1) that all methods within the class of Algorithm 1 produce iterates that converge to the set of eigenvectors of A. Furthermore, in view of Proposition 4.6.1, and since we are considering descent methods, it follows that, if λ_1 is simple, convergence is stable to $\pm v_1$ and unstable to all other eigenvectors.

Hereafter we consider the instances of Algorithm 1 where

$$\eta_k := -\operatorname{grad} f(x_k) = 2(Ax_k - x_k x_k^T A x_k).$$

It is clear that this choice of search direction is gradient-related. Next we have to pick a retraction. A reasonable choice is (see Example 4.1.1)

$$R_x(\xi) := \frac{x+\xi}{\|x+\xi\|}, \tag{4.30}$$

where $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^n$, $\|y\| := \sqrt{y^T y}$. Another possibility is

$$R_x(\xi) := x \cos\|\xi\| + \frac{\xi}{\|\xi\|}\sin\|\xi\|, \tag{4.31}$$

for which the curve $t \mapsto R_x(t\xi)$ is a big circle on the sphere. (The second retraction corresponds to the exponential mapping defined in Section 5.4.)

4.6.3 Armijo line search

We now have all the necessary ingredients to apply a simple backtracking instance of Algorithm 1 to the problem of minimizing the Rayleigh quotient on the sphere S^{n-1}. This yields the matrix algorithm displayed in Algorithm 2. Note that with the retraction R defined in (4.30), the function $f(R_{x_k}(t\eta_k))$ is a quadratic rational function in t. Therefore, the Armijo step size is easily computed as an expression of the reals $\eta_k^T\eta_k$, $\eta_k^T A\eta_k$, $x_k^T A\eta_k$, and $x_k^T A x_k$.

Algorithm 2 Armijo line search for the Rayleigh quotient on S^{n-1}

Require: Symmetric matrix A, scalars $\overline{\alpha} > 0$, $\beta, \sigma \in (0,1)$.
Input: Initial iterate x_0, $\|x_0\| = 1$.
Output: Sequence of iterates $\{x_k\}$.

```
for $k = 0, 1, 2, \dots$ do
    Compute $\eta_k = -2(Ax_k - x_k x_k^T A x_k)$.
    Find the smallest integer $m \geq 0$ such that
        $f(R_{x_k}(\overline{\alpha}\beta^m\eta_k)) \leq f(x_k) - \sigma\overline{\alpha}\beta^m\eta_k^T\eta_k$,
    with $f$ defined in (4.28) and $R$ defined in (4.30).
    Set
        $x_{k+1} = R_{x_k}(\overline{\alpha}\beta^m\eta_k)$.
end for
```

Numerical results for Algorithm 2 are presented in Figure 4.3 for the case $A = \operatorname{diag}(1, 2, \dots, 100)$, $\sigma = 0.5$, $\overline{\alpha} = 1$, $\beta = 0.5$. The initial point x_0 is chosen from a uniform distribution on the sphere. (The point x_0 is obtained by normalizing a vector whose entries are selected from a normal distribution).

Let us evaluate the upper bound r_* on the linear convergence factor given by Theorem 4.5.6. The extreme eigenvalues $\lambda_{H,\min}$ and $\lambda_{H,\max}$ of the Hessian at the solution v_1 can be obtained as

$$\lambda_{H,\min} = \min_{v_1^T u=0, u^T u=1} \left.\frac{\mathrm{d}^2(f(\gamma_{v_1,u}(t)))}{\mathrm{d}t^2}\right|_{t=0}$$

$$\lambda_{H,\max} = \max_{v_1^T u=0, u^T u=1} \left.\frac{\mathrm{d}^2(f(\gamma_{v_1,u}(t)))}{\mathrm{d}t^2}\right|_{t=0},$$

where

$$\gamma_{v_1,u}(t) := R_{v_1}(tu) = \frac{v_1 + tu}{\|v_1 + tu\|}.$$

This yields

$$\left.\frac{\mathrm{d}^2(f(\gamma_{v_1,u}(t)))}{\mathrm{d}t^2}\right|_{t=0} = 2(u^T A u - \lambda_1)$$

and thus

$$\lambda_{H,\min} = \lambda_2 - \lambda_1, \quad \lambda_{H,\max} = \lambda_n - \lambda_1.$$

For the considered numerical example, it follows that the upper bound on the linear convergence factor given by Theorem 4.5.6 is $r_* = 0.9949....$ The convergence factor estimated from the experimental result is below 0.97, which is in accordance with Theorem 4.5.6. This poor convergence factor, very close to 1, is due to the small value of the ratio

$$\frac{\lambda_{H,\min}}{\lambda_{H,\max}} = \frac{\lambda_2 - \lambda_1}{\lambda_n - \lambda_1} \approx 0.01.$$

The convergence analysis of Algorithm 2 is summarized as follows.

Theorem 4.6.3 *Let $\{x_k\}$ be an infinite sequence of iterates generated by Algorithm 2. Let $\lambda_1 \leq \cdots \leq \lambda_n$ denote the eigenvalues of A.*

(i) The sequence $\{x_k\}$ converges to the eigenspace of A associated to some eigenvalue.

(ii) The eigenspace related to λ_1 is an attractor of the iteration defined by Algorithm 2. The other eigenspaces are unstable.

(iii) Assuming that the eigenvalue λ_1 is simple, the linear convergence factor to the eigenvector $\pm v_1$ associated with λ_1 is smaller or equal to

$$r_* = 1 - 2\sigma(\lambda_2 - \lambda_1)\min\left(\overline{\alpha}, \frac{2\beta(1-\sigma)}{\lambda_n - \lambda_1}\right).$$

Proof. Points (i) and (iii) follow directly from the convergence analysis of the general Algorithm 1 (Theorems 4.3.1 and 4.5.6). For (ii), let $\mathcal{S}_1 := \{x \in S^{n-1} : Ax = \lambda_1 x\}$ denote the eigenspace related to λ_1. Any neighborhood of $\mathcal{S}_1$ contains a sublevel set $\mathcal{L}$ of f such that the only critical points of f in $\mathcal{L}$ are the points of $\mathcal{S}_1$. Any sequence of Algorithm 2 starting in $\mathcal{L}$ converges to $\mathcal{S}_1$. The second part follows from Theorem 4.4.1. □

4.6.4 Exact line search

In this version of Algorithm 1, x_{k+1} is selected as $R_{x_k}(t_k\eta_k)$, where

$$t_k := \operatorname*{arg\,min}_{t>0} f(R_{x_k}(t\eta_k)).$$

We consider the case of the projected retraction (4.30), and we define again $\eta_k := -\operatorname{grad} f(x_k)$. It is assumed that $\operatorname{grad} f(x_k) \neq 0$, from which it also follows that $\eta_k^T A x_k \neq 0$. An analysis of the function $t \mapsto f(R_{x_k}(t\eta_k))$ reveals that it admits one and only one minimizer $t_k > 0$. This minimizer is the positive solution of a quadratic equation. In view of the particular choice of the retraction, the points $\pm R_{x_k}(t_k\eta_k)$ can also be expressed as

$$\operatorname*{arg\,min}_{x\in S^{n-1},x\in \operatorname{span}\{x_k,\eta_k\}} f(x),$$

which are also equal to

$$\pm Xw,$$

where $X := [x_k, \frac{\eta_k}{\|\eta_k\|}]$ and w is a unit-norm eigenvector associated with the smaller eigenvalue of the interaction matrix $X^T A X$.

Numerical results are presented in Figure 4.3. Note that in this example the distance to the solution as a function of the number of iterates is slightly better with the selected Armijo method than with the exact line-search method. This may seem to be in contradiction to the fact that the exact line-search method chooses the optimal step size. However, the exact minimization only implies that if the two algorithms start at the same point x_0, then the *cost function* will be lower at the *first* iterate of the exact line-search method than at the *first* iterate of the Armijo method. This does not imply that the distance to the solution will be lower with the exact line search. Neither does it mean that the exact line search will achieve a lower cost function at subsequent iterates. (The first step of the Armijo method may well produce an iterate from which a larger decrease can be obtained.)

4.6.5 Accelerated line search: locally optimal conjugate gradient

In this version of Algorithm 1, η_k is selected as $-\operatorname{grad} f(x_k)$ and x_{k+1} is selected as $R_{x_k}(\xi_k)$, where ξ_k is a minimizer over the two-dimensional subspace of $T_{x_k}\mathcal{M}$ spanned by η_k and $R_{x_k}^{-1}(x_{k-1})$, as described in (4.13). When applied to the Rayleigh quotient on the sphere, this method reduces to the locally optimal conjugate-gradient (LOCG) algorithm of A. Knyazev. Its fast convergence (Figure 4.3) can be explained by its link with conjugate-gradient (CG) methods (see Section 8.3).

4.6.6 Links with the power method and inverse iteration

The power method,

$$x_{k+1} = \frac{Ax_k}{\|Ax_k\|},$$

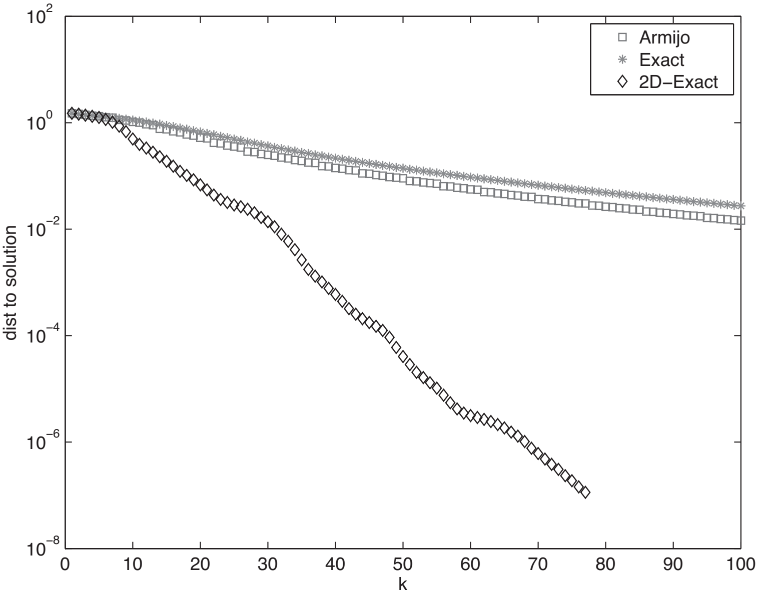

Figure 4.3 Minimization of the Rayleigh quotient of $A = \operatorname{diag}(1, 2, \ldots, n)$ on S^{n-1}, with $n = 100$. The distance to the solution is defined as the angle between the direction of the current iterate and the eigendirection associated with the smallest eigenvalue of A.

is arguably the simplest method for eigenvector computation. Let A be a symmetric matrix, assume that there is an eigenvalue λ that is simple and larger in absolute value than all the other eigenvalues, and let v denote the corresponding eigenvector. Then the power method converges to $\pm v$ for almost all initial points x_0.

We mention, as a curiosity, a relation between the power method and the steepest-descent method for the Rayleigh quotient on the sphere. Using the projective retraction (4.30), the choice $t_k = \frac{1}{2x_k^T A x_k}$ yields

$$R_{x_k}(t_k \operatorname{grad} f(x_k)) = \frac{Ax_k}{\|Ax_k\|},$$

i.e., the power method.

There is no such relation for the inverse iteration

$$x_{k+1} = \frac{A^{-1}x_k}{\|A^{-1}x_k\|}.$$

In fact, inverse iteration is in general much more expensive computationally than the power method since the former requires solving a linear system of size n at each iteration while the latter requires only a matrix-vector multiplication. A comparison between inverse iteration and the previous direct methods in terms of the number of iterations is not informative since an iteration of inverse iteration is expected to be computationally more demanding than an iteration of the other methods.

4.7 REFINING EIGENVECTOR ESTIMATES

All the critical points of the Rayleigh quotient correspond to eigenvectors of A, but only the extreme eigenvectors correspond to extrema of the cost function. For a given cost function f, it is, however, possible to define a new cost function that transforms *all* critical points of f into (local) minimizers. The new cost function is simply defined by

$$\tilde{f}(x) := \|\operatorname{grad} f(x)\|^2.$$

In the particular case of the Rayleigh quotient (4.28), one obtains

$$\tilde{f} : S^{n-1} \to \mathbb{R} : x \mapsto \|\mathrm{P}_x Ax\|^2 = x^T A \mathrm{P}_x Ax = x^T A^2 x - (x^T Ax)^2,$$

where $\mathrm{P}_x = (I - xx^T)$ is the orthogonal projector onto the tangent space $T_x S^{n-1} = \{\xi \in \mathbb{R}^n : x^T \xi = 0\}$. Following again the development in Section 3.6.1, we define the function

$$\overline{f} : \mathbb{R}^n \to \mathbb{R} : x \mapsto x^T A^2 x - (x^T Ax)^2$$

whose restriction to S^{n-1} is $\tilde{f}$. We obtain

$$\operatorname{grad} \overline{f}(x) = 2(A^2 x - 2Axx^T Ax),$$

hence

$$\operatorname{grad} \tilde{f}(x) = \mathrm{P}_x(\operatorname{grad} \overline{f}(x)) = 2\mathrm{P}_x(AAx - 2Axx^T Ax).$$

Applying a line-search method to the cost function $\tilde{f}$ provides a descent algorithm that (locally) converges to any eigenvector of A.

4.8 BROCKETT COST FUNCTION ON THE STIEFEL MANIFOLD

Following up on the study of descent algorithms for the Rayleigh quotient on the sphere, we now consider a cost function defined as a weighted sum $\sum_i \mu_i x_{(i)}^T A x_{(i)}$ of Rayleigh quotients on the sphere under an orthogonality constraint, $x_{(i)}^T x_{(j)} = \delta_{ij}$.

4.8.1 Cost function and search direction

The cost function admits a more friendly expression in matrix form:

$$f : \mathrm{St}(p, n) \to \mathbb{R} : X \mapsto \operatorname{tr}(X^T AXN), \tag{4.32}$$

where $N = \operatorname{diag}(\mu_1, \cdots, \mu_p)$, with $0 \leq \mu_1 \leq \ldots \leq \mu_p$, and $\mathrm{St}(p, n)$ denotes the orthogonal Stiefel manifold

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

As in Section 3.3.2, we view $\mathrm{St}(p, n)$ as an embedded submanifold of the Euclidean space $\mathbb{R}^{n\times p}$. The tangent space is (see Section 3.5.7)

$$\begin{aligned} T_X \,\mathrm{St}(p, n) &= \{Z \in \mathbb{R}^{n\times p} : X^T Z + Z^T X = 0\} \\ &= \{X\Omega + X_\perp K : \ \Omega^T = -\Omega, \ K \in \mathbb{R}^{(n-p)\times p}\}. \end{aligned}$$

We further consider $\mathrm{St}(p,n)$ as a Riemannian submanifold of $\mathbb{R}^{n\times p}$ endowed with the canonical inner product

$$\langle Z_1, Z_2\rangle := \mathrm{tr}\left(Z_1^T Z_2\right).$$

It follows that the normal space to $\mathrm{St}(p,n)$ at a point X is

$$(T_X \,\mathrm{St}(p,n))^\perp = \{XS : \; S^T = S\}.$$

The orthogonal projection P_X onto $T_X \,\mathrm{St}(p,n)$ is given by

$$\mathrm{P}_X Z = Z - X\,\mathrm{sym}(X^T Z) = (I - XX^T)Z + X\,\mathrm{skew}(X^T Z),$$

where

$$\mathrm{sym}(M) := \tfrac{1}{2}(M + M^T), \qquad \mathrm{skew}(M) = \tfrac{1}{2}(M - M^T)$$

denote the symmetric part and the skew-symmetric part of the decomposition of M into a symmetric and a skew-symmetric term.

Consider the function

$$\overline{f} : \mathbb{R}^{n\times p} \to \mathbb{R} : X \mapsto \mathrm{tr}(X^T AXN),$$

so that $f = \overline{f}\big|_{\mathrm{St}(p,n)}$. We have

$$\mathrm{D}\overline{f}\,(X)\,[Z] = 2\;\mathrm{tr}\left(Z^T AXN\right),$$

hence

$$\mathrm{grad}\,\overline{f}(X) = 2AXN$$

and

$$\begin{aligned}\mathrm{grad}\,f(X) &= \mathrm{P}_X\,\mathrm{grad}\,\overline{f}(X)\\ &= 2AXN - 2X\,\mathrm{sym}(X^T AXN)\\ &= 2AXN - XX^T AXN - XNX^T AX.\end{aligned}$$

It remains to select a retraction. Choices are proposed in Section 4.1.1, such as

$$R_X(\xi) := \mathrm{qf}(X + \xi).$$

This is all we need to turn various versions of the general Algorithm 1 into practical matrix algorithms for minimizing the cost fuction (4.32) on the orthogonal Stiefel manifold.

4.8.2 Critical points

We now show that X is a critical point of f if and only if the columns of X are eigenvectors of A.

The gradient of f admits the expression

$$\begin{aligned}\mathrm{grad}\,f(X) &= 2(I - XX^T)AXN + 2X\,\mathrm{skew}(X^T AXN) \qquad (4.33)\\ &= 2(I - XX^T)AXN + X[X^T AX, N],\end{aligned}$$

Table 4.2 Brockett cost function on the Stiefel manifold.

	Manifold ($\mathrm{St}(p,n)$)	Total space ($\mathbb{R}^{n\times p}$)
cost	$\mathrm{tr}(X^TAXN)$, $X^TX = I_p$	$\mathrm{tr}(X^TAXN)$, $X \in \mathbb{R}^{n\times p}$
metric	induced metric	$\langle Z_1, Z_2\rangle = \mathrm{tr}(Z_1^TZ_2)$
tangent space	$Z \in \mathbb{R}^{n\times p} : \mathrm{sym}(X^TZ) = 0$	$\mathbb{R}^{n\times p}$
normal space	$Z \in \mathbb{R}^{n\times p} : Z = XS,\ S^T = S$	$\emptyset$
projection onto tangent space	$\mathrm{P}_XZ = Z - X\,\mathrm{sym}(X^TZ)$	identity
gradient	$\mathrm{grad}\, f(X) = \mathrm{P}_X\,\mathrm{grad}\,\overline{f}(X)$	$\mathrm{grad}\,\overline{f}(X) = 2AXN$
retraction	$R_X(Z) = \mathrm{qf}(X + Z)$	$R_X(Z) = X + Z$

where

$$[A, B] := AB - BA$$

denotes the (matrix) commutator of A and B. Since the columns of the first term in the expression of the gradient belong to the orthogonal complement of $\mathrm{span}(X)$, while the columns of the second term belong to $\mathrm{span}(X)$, it follows that $\mathrm{grad}\, f(X)$ vanishes if and only if

$$(I - XX^T)AXN = 0 \tag{4.34}$$

and

$$[X^TAX, N] = 0. \tag{4.35}$$

Since N is assumed to be invertible, equation (4.34) yields

$$(I - XX^T)AX = 0,$$

which means that

$$AX = XM \tag{4.36}$$

for some M. In other words, $\mathrm{span}(X)$ is an invariant subspace of A. Next, in view of the specific form of N, equation (4.35) implies that X^TAX is diagonal which, used in (4.36), implies that M is diagonal, hence the columns of X are eigenvectors of A. Showing conversely that any such X is a critical point of f is straightfoward.

In the case $p = n$, $\mathrm{St}(n, n) = O_n$, and critical points of the Brockett cost function are orthogonal matrices that diagonalize A. (Note that $I - XX^T = 0$, so the first term in (4.33) trivially vanishes.) This is equivalent to saying that the columns of X are eigenvectors of A.

4.9 RAYLEIGH QUOTIENT MINIMIZATION ON THE GRASSMANN MANIFOLD

Finally, we consider a generalized Rayleigh quotient cost function on the Grassmann manifold. The Grassmann manifold is viewed as a Riemannian quotient manifold of $\mathbb{R}_*^{n\times p}$, which allows us to exploit the machinery for steepest-descent methods on quotient manifolds (see, in particular, Sections 3.4, 3.5.8, 3.6.2, and 4.1.2).

4.9.1 Cost function and gradient calculation

We start with a review of the Riemannian quotient manifold structure of the Grassmann manifold (Section 3.6.2). Let the structure space $\overline{\mathcal{M}}$ be the noncompact Stiefel manifold $\mathbb{R}_*^{n\times p} = \{Y \in \mathbb{R}^{n\times p} : Y \text{ full rank}\}$. We consider on $\overline{\mathcal{M}}$ the equivalence relation

$$X \sim Y \quad \Leftrightarrow \quad \exists M \in \mathbb{R}_*^{n\times p} : Y = XM.$$

In other words, two elements of $\mathbb{R}_*^{n\times p}$ belong to the same equivalence class if and only if they have the same column space. There is thus a one-to-one correspondence between $\mathbb{R}_*^{n\times p}/\sim$ and the set of p-dimensional subspaces of $\mathbb{R}^n$. The set $\mathbb{R}_*^{n\times p}/\sim$ has been shown (Proposition 3.4.6) to admit a unique structure of quotient manifold, called the Grassmann manifold and denoted by $\mathrm{Grass}(p,n)$ or $\mathbb{R}_*^{n\times p}/\mathrm{GL}_p$. Moreover, $\mathbb{R}_*^{n\times p}/\mathrm{GL}_p$ has been shown (Section 3.6.2) to have a structure of Riemannian quotient manifold when $\mathbb{R}_*^{n\times p}$ is endowed with the Riemannian metric

$$\overline{g}_Y(Z_1, Z_2) = \mathrm{tr}\left((Y^TY)^{-1}Z_1^TZ_2\right).$$

The vertical space at Y is by definition the tangent space to the equivalence class of $\pi^{-1}(\pi(Y)) = \{YM : M \in \mathbb{R}_*^{p\times p}\}$, which yields

$$\mathcal{V}_Y = \{YM : M \in \mathbb{R}^{p\times p}\}.$$

The horizontal space at Y is defined as the orthogonal complement of the vertical space with respect to the metric $\overline{g}$, which yields

$$\mathcal{H}_Y = \{Z \in \mathbb{R}^{n\times p} : Y^TZ = 0\}.$$

Given $\xi \in T_{\mathrm{span}(Y)}\,\mathrm{Grass}(p,n)$, there exists a unique horizontal lift $\overline{\xi}_Y \in T_Y\mathbb{R}_*^{n\times p}$ satisfying

$$\mathrm{D}\pi(Y)[\overline{\xi}_Y] = \xi.$$

Since

$$\overline{g}(\overline{\xi}_{YM}, \overline{\zeta}_{YM}) = \overline{g}(\overline{\xi}_Y, \overline{\zeta}_Y)$$

for all $M \in \mathbb{R}_*^{p\times p}$, it follows that $(\mathrm{Grass}(p,n), g)$ is a Riemannian quotient manifold of $(\mathbb{R}_*^{n\times p}, \overline{g})$ with

$$g_{\mathrm{span}(Y)}(\xi, \zeta) := \overline{g}_Y(\overline{\xi}_Y, \overline{\zeta}_Y).$$

In other words, the canonical projection π is a Riemannian submersion from $(\mathbb{R}_*^{n\times p}, \overline{g})$ to $(\mathrm{Grass}(p,n), g)$.

Let A be an $n \times n$ symmetric matrix, not necessarily positive-definite. Consider the cost function on the total space $\mathbb{R}_*^{n\times p}$ defined by

$$\overline{f} : \mathbb{R}_*^{n\times p} \to \mathbb{R} : Y \mapsto \mathrm{tr}\left((Y^T Y)^{-1} Y^T A Y\right). \tag{4.37}$$

Since $\overline{f}(YM) = \overline{f}(Y)$ whenever $M \in \mathbb{R}_*^{p\times p}$, it follows that $\overline{f}$ induces a function f on the quotient $\mathrm{Grass}(p,n)$ such that $\overline{f} = f \circ \pi$. The function f can be described as

$$f : \mathrm{Grass}(p,n) \to \mathbb{R} : \mathrm{span}(Y) \mapsto \mathrm{tr}\left((Y^T Y)^{-1} Y^T A Y\right). \tag{4.38}$$

This function can be thought of as a generalized Rayleigh quotient. Since $\overline{f}$ is smooth on $\mathbb{R}_*^{n\times p}$, it follows from Proposition 3.4.5 that f is a smooth cost function on the quotient $\mathrm{Grass}(p,n)$.

In order to obtain an expression for the gradient of f, we will make use of the trace identities (A.1) and of the formula (A.3) for the derivative of the inverse of a matrix. For all $Z \in \mathbb{R}^{n\times p}$, we have

$$\begin{aligned} \mathrm{D}\overline{f}\,(Y)\,[Z] = \mathrm{tr}&\left(-(Y^T Y)^{-1}(Z^T Y + Y^T Z)(Y^T Y)^{-1} Y^T A Y\right) \\ &+ \mathrm{tr}\left((Y^T Y)^{-1} Z^T A Y\right) + \mathrm{tr}\left((Y^T Y)^{-1} Y^T A Z\right). \end{aligned} \tag{4.39}$$

For the last term, we have, using the two properties (A.1) of the trace,

$$\mathrm{tr}\left((Y^T Y)^{-1} Y^T A Z\right) = \mathrm{tr}\left(Z^T A Y (Y^T Y)^{-1}\right) = \mathrm{tr}\left((Y^T Y)^{-1} Z^T A Y\right).$$

Using the same properties, the first term can be rewritten as

$$-2\ \mathrm{tr}\left((Y^T Y)^{-1} Z^T Y (Y^T Y)^{-1} Y^T A Y\right).$$

Replacing these results in (4.39) yields

$$\begin{aligned} \mathrm{D}\overline{f}\,(Y)\,[Z] &= \mathrm{tr}\left((Y^T Y)^{-1} Z^T\, 2(AY - Y(Y^T Y)^{-1} Y^T A Y)\right) \\ &= \overline{g}_Y(Z, 2(AY - Y(Y^T Y)^{-1} Y^T A Y)). \end{aligned}$$

It follows that

$$\mathrm{grad}\,\overline{f}(Y) = 2\left(AY - Y(Y^T Y)^{-1} Y^T A Y\right) = \mathrm{P}_Y^h(2AY),$$

where

$$\mathrm{P}_Y^h = (I - Y(Y^T Y)^{-1} Y^T)$$

is the orthogonal projection onto the horizontal space. Note that, in accordance with the theory in Section 3.6.2, $\mathrm{grad}\,\overline{f}(Y)$ belongs to the horizontal space. It follows from the material in Section 3.6.2, in particular (3.39), that

$$\overline{\mathrm{grad}\,f}_Y = 2\mathrm{P}_Y^h A Y = 2\left(AY - Y(Y^T Y)^{-1} Y^T A Y\right).$$

Table 4.3 Rayleigh quotient cost function on the Grassmann manifold.

	$\mathrm{Grass}(p, n)$	Total space $\mathbb{R}_*^{n\times p}$
cost	$\mathrm{span}(Y) \mapsto \overline{f}(Y)$	$\overline{f}(Y) = \mathrm{tr}((Y^TY)^{-1}Y^TAY)$
metric	$g_{\mathrm{span}(Y)}(\xi, \zeta) = \overline{g}_Y(\overline{\xi}_Y, \overline{\zeta}_Y)$	$\overline{g}_Y(Z_1, Z_2)$ $= \mathrm{tr}((Y^TY)^{-1}Z_1^TZ_2)$
horizontal space	$Z \in \mathbb{R}^{n\times p} : Y^TZ = 0$	/
projection onto horizontal space	$\mathrm{P}_Y^h Z = Z - Y(Y^TY)^{-1}Y^TZ$	/
gradient	$\overline{\mathrm{grad}\, f}_Y = \mathrm{grad}\, \overline{f}(Y)$	$\mathrm{grad}\, \overline{f}(Y) = \mathrm{P}_Y^h(2AY)$
retraction	$R_{\mathrm{span}(Y)}(\xi) = \mathrm{span}(Y + \overline{\xi}_Y)$	$R_Y(Z) = Y + Z$

4.9.2 Line-search algorithm

In order to obtain a line-search algorithm for the Rayleigh quotient on the Grassmann manifold, it remains to pick a retraction. According to Section 4.1.2, a natural choice is

$$R_{\mathrm{span}(Y)}(\xi) = \mathrm{span}(Y + \overline{\xi}_Y). \tag{4.40}$$

In other words, $(Y + \overline{\xi}_Y)M$ is a matrix representation of $R_{\mathrm{span}(Y)}(\xi)$ for any $M \in \mathbb{R}_*^{p\times p}$. The matrix M can be viewed as a normalization factor that can be used to prevent the iterates from becoming ill-conditioned, the best-conditioned form being orthonormal matrices. We now have all the necessary elements (see the summary in Table 4.3) to write down explicitly a line-search method for the Rayleigh quotient (4.38).

The matrix algorithm obtained by applying the Armijo line-search version of Algorithm 1 to the problem of minimizing the generalized Rayleigh quotient (4.38) is stated in Algorithm 3.

The following convergence results follow from the convergence analysis of the general line-search Algorithm 1 (Theorems 4.3.1 and 4.5.6).

Theorem 4.9.1 *Let $\{Y_k\}$ be an infinite sequence of iterates generated by Algorithm 3. Let $\lambda_1 \leq \cdots \leq \lambda_n$ denote the eigenvalues of A.*

(i) The sequence $\{\mathrm{span}(Y_k)\}$ converges to the set of p-dimensional invariant subspaces of A.

(ii) Assuming that the eigenvalue λ_p is simple, the (unique) invariant subspace associated with $(\lambda_1, \ldots, \lambda_p)$ is asymptotically stable for the iteration defined by Algorithm 3, and the convergence is linear with a factor smaller than or equal to

$$r_* = 1 - 2\sigma(\lambda_{p+1} - \lambda_p)\min\left(\bar{\alpha}, \frac{2\beta(1-\sigma)}{\lambda_n - \lambda_1}\right).$$

Algorithm 3 Armijo line search for the Rayleigh quotient on Grass(p, n)

```
Require: Symmetric matrix $A$, scalars $\overline{\alpha} > 0$, $\beta, \sigma \in (0, 1)$.
Input: Initial iterate $Y_0 \in \mathbb{R}^{n \times p}$, $Y_0$ full rank.
Output: Sequence of iterates $\{Y_k\}$.
for $k = 0, 1, 2, \ldots$ do
    Compute $\eta_k = -2(AY - Y(Y^T Y)^{-1} AY)$.
    Find the smallest integer $m \geq 0$ such that
        $\overline{f}(Y_k + \overline{\alpha}\beta^m \eta_k) \leq \overline{f}(Y_k) - \sigma\overline{\alpha}\beta^m \operatorname{tr}(\eta_k^T \eta_k)$,
    with $\overline{f}$ defined in (4.37).
    Select $Y_{k+1} := (Y_k + \overline{\alpha}\beta^m \eta_k)M$, with some invertible $p \times p$ matrix
    $M$ chosen to preserve good conditioning. (For example, select $Y_{k+1}$ as
    the Q factor of the QR decomposition of $Y_k + \overline{\alpha}\beta^m \eta_k$.)
end for
```

The other invariant subspaces are unstable.

Numerical results are presented in Figure 4.4.

4.10 NOTES AND REFERENCES

Classical references on numerical optimization include Bertsekas [Ber95], Dennis and Schnabel [DS83], Fletcher [Fle01], Luenberger [Lue73], Nash and Sofer [NS96], Polak [Pol71], and Nocedal and Wright [NW99].

The choice of the qualification *complete* for Riemannian manifolds is not accidental: it can be shown that a Riemannian manifold $\mathcal{M}$ is complete (i.e., the domain of the exponential is the whole $T\mathcal{M}$) if and only if $\mathcal{M}$, endowed with the Riemannian distance, is a complete metric space; see, e.g., O'Neill [O'N83].

The idea of using computationally efficient alternatives to the Riemannian exponential was advocated by Manton [Man02, § IX] and was also touched on in earlier works [MMH94, Smi94, EAS98]. Retraction mappings are common in the field of algebraic topology [Hir76]. The definition of retraction used in this book comes from Shub [Shu86]; see also Adler *et al.* [ADM$^+$02]. Most of the material about retractions on the orthogonal group comes from [ADM$^+$02].

Selecting a computationally efficient retraction is a crucial step in developing a competitive algorithm on a manifold. This problem is linked to the question of approximating the exponential in such a way that the approximation resides on the manifold. This is a major research topic in computational mathematics, with important recent contributions; see, e.g., [CI01, OM01, IZ05, DN05] and references therein.

The concept of a locally smooth family of parameterizations was introduced by Hüper and Trumpf [HT04].

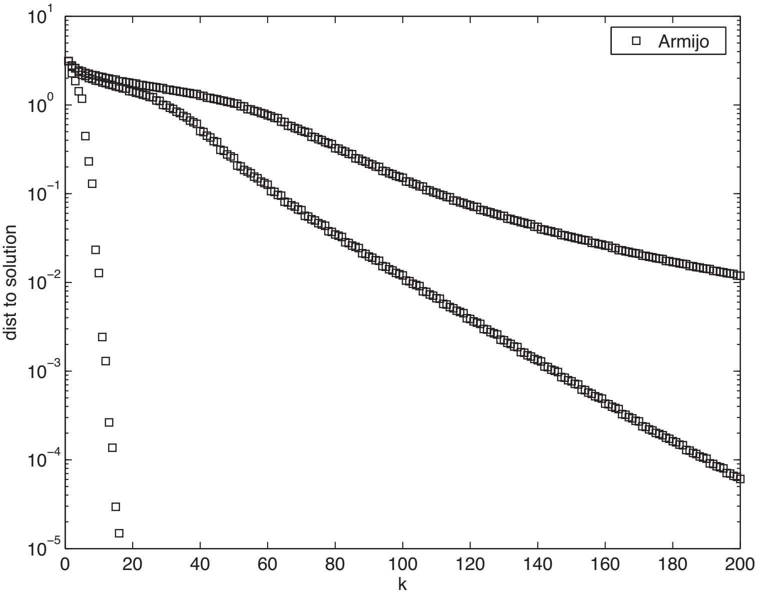

Figure 4.4 Rayleigh quotient minimization on the Grassmann manifold of p-planes in $\mathbb{R}^n$, with $p = 5$ and $n = 100$. Upper curve: $A = \mathrm{diag}(1, 2, \ldots, 100)$. Middle curve: $A = \mathrm{diag}(1, 102, 103, \ldots, 200)$. Lower curve: $A = \mathrm{diag}(1, \ldots, 5, 106, 107, \ldots, 200)$.

Details on the QR and polar decompositions and algorithms to compute them can be found in Golub and Van Loan [GVL96]; the differentiability of the qf mapping is studied in Dehane [Deh95], Dieci and Eirola [DE99], and Chern and Dieci [CD00]. Formulas for the differential of qf and other smooth matrix functions can be found in Dehaene [Deh95].

Definition 4.2.1, on gradient-related sequences, is adapted from [Ber95]. Armijo's backtracking procedure was proposed in [Arm66] (or see [NW99, Ber95] for details).

Several key ideas for line-search methods on manifolds date back to Luenberger [Lue73, Ch. 11]. Luenberger proposed to use a search direction obtained by projecting the gradient in $\mathbb{R}^n$ onto the tangent space of the constraint set and mentioned the idea of performing a line search along the geodesic, "which we would use if it were computationally feasible (which it definitely is not)". He also proposed an alternative to following the geodesic that corresponds to retracting orthogonally to the tangent space. Other early contributions to optimization on manifolds can be found in Gabay [Gab82]. Line-search methods on manifolds are also proposed and analyzed in Udrişte [Udr94]. Recently, Yang [Yan07] proposed an Armijo line-search strategy along geodesics. Exact and approximate line-search methods were proposed for matrix manifolds in a burst of research in the early 1990s [MMH94, Mah94, Bro93, Smi94]. Algorithm 1 comes from [AG05].

Many refinements exist for choosing the step length in line-search methods. For example, the backtracking parameter β can be adapted during the backtracking procedure. We refer to Dennis and Schnabel [DS83, §6.3.2] and Ortega and Rheinboldt [OR70].

The non-Hausdorff example given in Section 4.3.2 was inspired by Brickell and Clark [BC70, Ex. 3.2.1], which refers to Haefliger and Reeb [HR57].

For a local convergence analysis of classical line-search methods, see, e.g., Luenberger [Lue73] or Bertsekas [Ber95]. The proof of Theorem 4.3.1 (the global convergence of line-search methods) is a generalization of the proof of [Ber95, Prop. 1.2.1]. In Section 4.4, it is pointed out that convergence to critical points that are not local minima cannot be ruled out. Another undesirable behavior that cannot be ruled out in general is the existence of several (even infinitely many) accumulation points. Details can be found in Absil *et al.* [AMA05]; see also [GDS05]. Nevertheless, such algorithms do converge to single accumulation points, and the gap between theory and practice should not prevent one from utilizing the most computationally effective algorithm.

The notions of stability of fixed points have counterparts in dynamical systems theory; see, e.g., Vidyasagar [Vid02] or Guckenheimer and Holmes [GH83]. In fact, iterations $x_{k+1} = F(x_k)$ can be thought of as discrete-time dynamical systems.

Further information on Łojasiewicz's gradient inequality can be found in Łojasiewicz [Łoj93]. The concept of Theorem 4.4.2 (the capture theorem) is borrowed from Bertsekas [Ber95]. A coordinate-free proof of our Theorem 4.5.6 (local convergence of line-search methods) is given by Smith [Smi94] in the particular case where the next iterate is obtained via an exact line search minimization along geodesics. Optimization algorithms on the Grassmann manifold can be found in Smith [Smi93], Helmke and Moore [HM94], Edelman *et al.* [EAS98], Lippert and Edelman [LE00], Manton [Man02], Manton *et al.* [MMH03], Absil *et al.* [AMS04], and Liu *et al.* [LSG04].

Gradient-descent algorithms for the Rayleigh quotient were considered as early as 1951 by Hestenes and Karush [HK51]. A detailed account is given in Faddeev and Faddeeva [FF63, §74, p. 430]. There has been limited investigation of line-search descent algorithms as numerical methods for linear algebra problems since it is clear that such algorithms are not competitive with existing numerical linear algebra algorithms. At the end of his paper on the design of gradient systems, Brockett [Bro93] provides a discrete-time analog, with an analytic step-size selection method, for a specific class of problems. In independent work, Moore *et al.* [MMH94] (see also [HM94, p. 68]) consider the symmetric eigenvalue problem directly. Chu [Chu92] proposes numerical methods for the inverse singular value problem. Smith *et al.* [Smi93, Smi94, EAS98] consider line-search and conjugate gradient updates to eigenspace tracking problems. Mahony *et al.* [Mah94, MHM96] proposes gradient flows and considers discrete updates for principal component analysis. A related approach is to consider explicit integration of the gradient flow dynamical system with a numerical integration technique that preserves the underlying matrix constraint. Moser and Veselov [MV91] use this approach directly in building numerical algorithms for matrix factorizations. The literature on structure-preserving integration algorithms is closely linked to work on the integration of Hamiltonian systems. This field is too vast to cover here, but we mention the excellent review by Iserles *et al.* [IMKNZ00] and an earlier review by Sanz-Serna [SS92].

The locally optimal conjugate gradient algorithm for the symmetric eigenvalue problem is described in Knyazev [Kny01]; see Hetmaniuk and Lehoucq [HL06] for recent developments. The connection between the power method and line-search methods for the Rayleigh quotient was studied in Mahony *et al.* [MHM96].

More information on the eigenvalue problem can be found in Golub and van der Vorst [GvdV00], Golub and Van Loan [GVL96], Parlett [Par80], Saad [Saa92], Stewart [Ste01], Sorensen [Sor02], and Bai *et al.* [BDDR00].

Linearly convergent iterative numerical methods for eigenvalue and subspace problems are not competitive with the classical numerical linear algebra techniques for one-off matrix factorization problems. However, a domain in which linear methods are commonly employed is in tracking the principal subspace of a covariance matrix associated with observations of a noisy signal. Let $\{x_1, x_2, \ldots\}$ be a sequence of elements of vectors in $\mathbb{R}^n$ and define

$$E_k^N = \frac{1}{N} \sum_{i=k+1}^{k+N} x_i x_i^T \in \mathbb{R}^{n \times n}, \qquad A_k^N = \begin{bmatrix} x_{k+1} & \cdots & x_N \end{bmatrix} \in \mathbb{R}^{n \times N}.$$

The signal subspace tracking problem is either to track a principal subspace of the covariance matrix E_k^N (a Hermitian eigenspace problem) or to directly track a signal subspace of the signal array A_k^N (a singular value problem). Common and Golub [CG90] studied classical numerical linear algebra techniques for this problem with linear update complexity. More recent review material is provided in DeGroat *et al.* [DDL99]. Most (if not all) high-accuracy linear complexity algorithms belong to a family of power-based algorithms [HXC$^+$99]. This includes the Oja algorithm [Oja89], the PAST algorithm [Yan95], the NIC algorithm [MH98b], and the Bi-SVD algorithm [Str97], as well as gradient-based updates [FD95, EAS98]. Research in this field is extremely active at this time, with the focus on reduced-complexity updates [OH05, BDR05]. We also refer the reader to the Bayesian geometric approach followed in [Sri00, SK04].

In line-search algorithms, the limit case where the step size goes to zero corresponds to a continuous-time dynamical system of the form $\dot{x} = \eta_x$, where $\eta_x \in T_x\mathcal{M}$ denotes the search direction at $x \in \mathcal{M}$. There is a vast literature on continuous-time systems that solve computational problems, spanning several areas of computational science, including, but not limited to, linear programming [BL89a, BL89b, Bro91, Fay91b, Hel93b], continuous nonlinear optimization [Fay91a, LW00], discrete optimization [Hop84, HT85, Vid95, AS04], signal processing [AC98, Dou00, CG03], balanced realization of linear systems [Hel93a, GL93], model reduction [HM94, YL99], and automatic control [HM94, MH98a, GS01]. Applications in linear algebra, and especially in eigenvalue and singular value problems, are particularly abundant. Important advances in the area have come from the work on isospectral flows in the early 1980s. We refer the reader to Helmke and Moore [HM94] as the seminal monograph in this area and the thesis of Dehaene [Deh95] for more information; see also [Chu94, DMV99, CG02, Prz03, MA03, BI04, CDLP05, MHM05] and the many references therein.