Chapter Seven

Trust-Region Methods

The plain Newton method discussed in Chapter 6 was shown to be locally convergent to any critical point of the cost function. The method does not distinguish among local minima, saddle points, and local maxima: all (nondegenerate) critical points are asymptotically stable fixed points of the Newton iteration. Moreover, it is possible to construct cost functions and initial conditions for which the Newton sequence does not converge. There even exist examples where the set of nonconverging initial conditions contains an open subset of search space.

To exploit the desirable superlinear local convergence properties of the Newton algorithm in the context of global optimization, it is necessary to embed the Newton update in some form of descent method. In Chapter 6 we briefly outlined how the Newton equation can be used to generate a descent direction that is used in a line-search algorithm. Such an approach requires modification of the Newton equation to ensure that the resulting sequence of search directions is gradient-related and an implementation of a standard line-search iteration. The resulting algorithm will converge to critical points of the cost function for *all* initial points. Moreover, saddle points and local maxima are rendered unstable, thus favoring convergence to local minimizers.

Trust-region methods form an alternative class of algorithms that combine desirable global convergence properties with a local superlinear rate of convergence. In addition to providing good global convergence, trust-region methods also provide a framework to relax the computational burden of the plain Newton method when the iterates are too far away from the solution for fast local convergence to set in. This is particularly important in the development of optimization algorithms on matrix manifolds, where the inverse Hessian computation can involve solving relatively complex matrix equations.

Trust-region methods can be understood as an enhancement of Newton's method. To this end, however, we need to consider this method from another viewpoint: instead of looking for an update vector along which the derivative of grad f is equal to $-\text{grad}\, f(x_k)$, it is equivalent to think of Newton's method (in $\mathbb{R}^n$) as the algorithm that selects the new iterate x_{k+1} to be the critical point of the quadratic Taylor expansion of the cost function f about x_k.

To this end, the chapter begins with a discussion of generalized quadratic models on manifolds (Section 7.1). Here again, a key role is played by the concept of retraction, which provides a way to pull back the cost function on

the manifold to a cost function on the tangent space. It is therefore sufficient to define quadratic models on abstract vector spaces and to understand how these models correspond to the real-valued function on the manifold $\mathcal{M}$.

Once the notion of a quadratic model is established, a trust-region algorithm can be defined on a manifold (Section 7.2). It is less straightforward to show that all the desirable convergence properties of classical trust-region methods in $\mathbb{R}^n$ still hold, *mutatis mutandis*, for their manifold generalizations. The difficulty comes from the fact that trust-region methods on manifolds do not work with a single cost function but rather with a succession of cost functions whose domains are different tangent spaces. The issue of computing an (approximate but sufficiently accurate) solution of the trust-region subproblems is discussed in Section 7.3. The convergence analysis is carried out in Section 7.4. The chapter is concluded in Section 7.5 with a "checklist" of steps one has to go through in order to turn the abstract geometric trust-region schemes into practical numerical algorithms on a given manifold for a given cost function; this checklist is illustrated for several examples related to Rayleigh quotient minimization.

7.1 MODELS

Several classical optimization schemes rely on successive local minimization of quadratic models of the cost function. In this section, we review the notion of quadratic models in $\mathbb{R}^n$ and in general vector spaces. Then, making use of retractions, we extend the concept to Riemannian manifolds.

7.1.1 Models in $\mathbb{R}^n$

The fundamental mathematical tool that justifies the use of local models is Taylor's theorem (see Appendix A.6). In particular, we have the following results.

Proposition 7.1.1 *Let f be a smooth real-valued function on $\mathbb{R}^n$, $x \in \mathbb{R}^n$, $\mathcal{U}$ a bounded neighborhood of x, and H any symmetric matrix. Then there exists $c > 0$ such that, for all $(x + h) \in \mathcal{U}$,*

$$\big|f(x+h) - \big(f(x) + \partial f(x)h + \tfrac{1}{2}h^T H h\big)\big| \leq c\|h\|^2,$$

where $\partial f(x) := (\partial_1 f(x), \ldots, \partial_n f(x))$. If, moreover, $H_{i,j} = \partial_i \partial_j f(x)$, then there exists $c > 0$ such that, for all $(x + h) \in \mathcal{U}$,

$$\big|f(x+h) - \big(f(x) + \partial f(x)h + \tfrac{1}{2}h^T H h\big)\big| \leq c\|h\|^3.$$

7.1.2 Models in general Euclidean spaces

The first step towards defining quadratic models on Riemannian manifolds is to generalize the above results to (abstract) Euclidean spaces, i.e., finite-dimensional vector spaces endowed with an inner product $\langle \cdot, \cdot \rangle$. This is readily done using the results in Appendix A.6. (Note that Euclidean spaces

are naturally finite-dimensional normed vector spaces for the induced norm $\|\xi\| := \sqrt{\langle \xi, \xi \rangle}$.)

Proposition 7.1.2 *Let f be a smooth real-valued function on a Euclidean space $\mathcal{E}$, $x \in \mathcal{E}$, $\mathcal{U}$ a bounded neighborhood of x, and $H : \mathcal{E} \to \mathcal{E}$ any symmetric operator. Then there exists $c > 0$ such that, for all $(x + h) \in \mathcal{U}$,*

$$\big| f(x+h) - \big(f(x) + \langle \operatorname{grad} f(x), h \rangle + \tfrac{1}{2} \langle H[h], h \rangle \big) \big| \leq c \|h\|^2.$$

If, moreover, $H = \operatorname{Hess} f(x) : h \mapsto \mathrm{D}(\operatorname{grad} f)(x)[h]$, then there exists $c > 0$ such that, for all $(x + h) \in \mathcal{U}$,

$$\big| f(x+h) - \big(f(x) + \langle \operatorname{grad} f(x), h \rangle + \tfrac{1}{2} \langle H[h], h \rangle \big) \big| \leq c \|h\|^3.$$

7.1.3 Models on Riemannian manifolds

Let f be a real-valued function on a Riemannian manifold $\mathcal{M}$ and let $x \in \mathcal{M}$. A model m_x of f around x is a real-valued function defined on a neighborhood of x such that (i) m_x is a "sufficiently good" approximation of f and (ii) m_x has a "simple" form that makes it easy to tackle with optimization algorithms.

The quality of the model m_x is assessed by evaluating how the discrepancy between $m_x(y)$ and $f(y)$ evolves as a function of the Riemannian distance $\operatorname{dist}(x, y)$ between x and y. The model m_x is an *order-q model*, $q > 0$, if there exists a neighborhood $\mathcal{U}$ of x in $\mathcal{M}$ and a constant $c > 0$ such that

$$|f(y) - m_x(y)| \leq c \, (\operatorname{dist}(x, y))^{q+1} \quad \text{for all } y \in \mathcal{U}.$$

Note that an order-q model automatically satisfies $m_x(x) = f(x)$. The following result shows that the order of a model can be assessed using any retraction R on $\mathcal{M}$.

Proposition 7.1.3 *Let f be a real-valued function on a Riemannian manifold $\mathcal{M}$ and let $x \in \mathcal{M}$. A model m_x of f is order-q if and only if there exists a neighborhood $\mathcal{U}$ of x and a constant $c > 0$ such that*

$$|f(y) - m_x(y)| \leq c \, \|R_x^{-1}(y)\|^{q+1} \quad \textit{for all } y \in \mathcal{U}.$$

Proof. In view of the local rigidity property of retractions, it follows that $\mathrm{D}(\operatorname{Exp}^{-1} \circ R_x)(0_x) = \mathrm{id}_{T_x\mathcal{M}}$, hence $\|R_x^{-1}(y)\| = \|(R_x^{-1} \circ \operatorname{Exp}_x^{-1})(\operatorname{Exp}_x y)\| = \Omega(\|\operatorname{Exp}_x y\|) = \Omega(\operatorname{dist}(x, y))$; see Section A.4 for the definition of the asymptotic notation Ω. In other words, there is a neighborhood $\mathcal{U}$ of x and constants c_1 and c_2 such that

$$c_1 \operatorname{dist}(x, y) \leq \|R_x^{-1}(y)\| \leq c_2 \operatorname{dist}(x, y) \quad \text{for all } y \in \mathcal{U}.$$

□

Proposition 7.1.3 yields a conceptually simple way to build an order-q model of f around x. Pick a retraction on $\mathcal{M}$. Consider

$$\widehat{f}_x := f \circ R_x,$$

the pullback of f to $T_x\mathcal{M}$ through R_x. Let $\widehat{m}_x$ be the order-q Taylor expansion of $\widehat{f}_x$ around the origin 0_x of $T_x\mathcal{M}$. Finally, push $\widehat{m}_x$ forward through R_x to obtain $m_x = \widehat{m}_x \circ R_x^{-1}$. This model is an order-q model of f around x. In particular, the obvious order-2 model to choose is

$$\begin{aligned}\widehat{m}_x(\xi) &= \hat{f}_x(0_x) + \mathrm{D}\hat{f}_x(0_x)[\xi] + \tfrac{1}{2}\mathrm{D}^2\hat{f}_x(0_x)[\xi,\xi] \\ &= f(x) + \langle \operatorname{grad} f(x), \xi\rangle + \tfrac{1}{2}\langle \operatorname{Hess} \widehat{f}_x(0_x)[\xi], \xi\rangle.\end{aligned}$$

Note that, since $\mathrm{D}R_x(0_x) = \mathrm{id}_{T_x\mathcal{M}}$ (see Definition 4.1.1), it follows that $\mathrm{D}f(x) = \mathrm{D}\widehat{f}_x(0_x)$, hence $\operatorname{grad} \widehat{f}_{x_k}(0_x) = \operatorname{grad} f(x)$, where $\operatorname{grad} f(x)$ denotes the (Riemannian) gradient of f (see Section 3.6).

In practice, the second-order differentials of the function $\widehat{f}_x$ may be difficult to compute. (The first-order differential is always straightforward since the rigidity condition of the retraction ensures $\mathrm{D}\widehat{f}_x(0_x) = \mathrm{D}f(x)$.) On the other hand, the Riemannian connection ∇ admits nice formulations on Riemannian submanifolds and Riemannian quotient manifolds (see Section 5.3). This suggests the model

$$m_x = \widehat{m}_x \circ R_x^{-1}$$

with

$$\widehat{m}_x(\xi) = f(x) + \langle \operatorname{grad} f(x), \xi\rangle + \tfrac{1}{2}\langle \operatorname{Hess} f(x)[\xi], \xi\rangle, \quad \xi \in T_x\mathcal{M}, \tag{7.1}$$

where the quadratic term is given by the Riemannian Hessian

$$\operatorname{Hess} f(x)[\xi] = \nabla_\xi \operatorname{grad} f(x). \tag{7.2}$$

In general, this model m_x is only order 1 because $\operatorname{Hess} f(x) \neq \operatorname{Hess} \widehat{f}_x(0_x)$. However, if R is a second-order retraction, then $\operatorname{Hess} f(x) = \operatorname{Hess} \widehat{f}_x(0_x)$ (Proposition 5.5.5) and m_x is order 2. More importantly, for any retraction, $\operatorname{Hess} f(x_*) = \operatorname{Hess} \widehat{f}_{x_*}(0_{x_*})$ when x_* is a critical point of f (Proposition 5.5.6).

The quadratic model (7.1) has a close connection to the Newton algorithm. Assuming that the Hessian is nonsingular at x, the critical point ξ_* of $\widehat{m}_x$ satisfies the Newton equation

$$\operatorname{Hess} f(x)[\eta_*] + \operatorname{grad} f(x) = 0.$$

It follows that the geometric Newton method (Algorithm 5), with retraction R and affine connection ∇, defines its next iterate as the critical point of the quadratic model (7.1).

We point out that all the models we have considered so far assume a quadratic form on $T_x\mathcal{M}$, i.e., there is a symmetric operator $H : T_x\mathcal{M} \to T_x\mathcal{M}$ such that

$$\widehat{m}_x(\xi) = f(x) + \langle \operatorname{grad} f(x), \xi\rangle + \tfrac{1}{2}\langle H[\xi], \xi\rangle.$$

Quadratic models are particularly interesting because the problem of minimizing a quadratic function under trust-region constraints is well understood and several algorithms are available (see Section 7.3).

7.2 TRUST-REGION METHODS

We first briefly review the principles of trust-region methods in $\mathbb{R}^n$. Extending the concept to manifolds is straightforward given the material developed in the preceding section.

7.2.1 Trust-region methods in $\mathbb{R}^n$

The basic trust-region method in $\mathbb{R}^n$ for a cost function f consists of adding to the current iterate $x \in \mathbb{R}^n$ the update vector $\eta \in \mathbb{R}^n$, solving (up to some approximation) the *trust-region subproblem*

$$\min_{\eta \in \mathbb{R}^n} m(\eta) = f(x) + \partial f(x)\eta + \tfrac{1}{2}\eta^T H \eta, \qquad \|\eta\| \leq \Delta, \tag{7.3}$$

where H is some symmetric matrix and Δ is the trust-region radius. Clearly, a possible choice for H is the Hessian matrix $H_{i,j} = \partial_i \partial_j f(x)$; classical convergence results guarantee superlinear convergence if the chosen H is "sufficiently close" to the Hessian matrix. The algorithm used to compute an approximate minimizer η of the model within the trust region is termed the *inner iteration*. Once an η has been returned by the inner iteration, the quality of the model m is assessed by forming the quotient

$$\rho = \frac{f(x) - f(x+\eta)}{m(0) - m(\eta)}. \tag{7.4}$$

Depending on the value of ρ, the new iterate $x + \eta$ is accepted or discarded and the trust-region radius Δ is updated. A specific procedure (in the Riemannian setting) is given in Algorithm 10, or see the textbooks mentioned in Notes and References.

7.2.2 Trust-region methods on Riemannian manifolds

We can now lay out the structure of a trust-region method on a Riemannian manifold $(\mathcal{M}, g)$ with retraction R. Given a cost function $f : \mathcal{M} \to \mathbb{R}$ and a current iterate $x_k \in \mathcal{M}$, we use R_{x_k} to locally map the minimization problem for f on $\mathcal{M}$ into a minimization problem for the *pullback* of f under R_{x_k},

$$\widehat{f}_{x_k} : T_{x_k}\mathcal{M} \to \mathbb{R} : \xi \mapsto f(R_{x_k}\xi). \tag{7.5}$$

The Riemannian metric g turns $T_{x_k}\mathcal{M}$ into a Euclidean space endowed with the inner product $g_{x_k}(\cdot,\cdot)$—which we usually denote by $\langle \cdot, \cdot \rangle_{x_k}$—and the trust-region subproblem on $T_{x_k}\mathcal{M}$ reads

$$\begin{aligned} &\min_{\eta \in T_{x_k}\mathcal{M}} \widehat{m}_{x_k}(\eta) = f(x_k) + \langle \operatorname{grad} f(x_k), \eta \rangle + \tfrac{1}{2}\langle H_k[\eta], \eta \rangle, \\ &\text{subject to } \langle \eta, \eta \rangle_{x_k} \leq \Delta_k^2, \end{aligned} \tag{7.6}$$

where H_k is some symmetric operator on $T_{x_k}\mathcal{M}$. (A possible choice is $H_k := \operatorname{Hess} f(x_k)$, the Riemannian Hessian (7.2).) This is called the *trust-region subproblem*.

Next, an (approximate) solution η_k of the Euclidean trust-region subproblem (7.6) is computed using any available method (inner iteration). The candidate for the new iterate is then given by $R_{x_k}(\eta_k)$. Notice that the inner iteration has to operate on the Euclidean space $T_x\mathcal{M}$ which, in the case of embedded submanifolds and quotient manifolds of a matrix space $\mathbb{R}^{n\times p}$, is represented as a linear subspace of $\mathbb{R}^{n\times p}$. Fortunately, most (or even all?) classical algorithms for the trust-region subproblem are readily adapted to Euclidean matrix spaces (see Section 7.3 for details).

The decisions on accepting or rejecting the candidate $R_{x_k}(\eta_k)$ and on selecting the new trust-region radius Δ_{k+1} are based on the quotient

$$\rho_k = \frac{f(x_k) - f(R_{x_k}(\eta_k))}{\widehat{m}_{x_k}(0_{x_k}) - \widehat{m}_{x_k}(\eta_k)} = \frac{\widehat{f}_{x_k}(0_{x_k}) - \widehat{f}_{x_k}(\eta_k)}{\widehat{m}_{x_k}(0_{x_k}) - \widehat{m}_{x_k}(\eta_k)}. \tag{7.7}$$

If ρ_k is exceedingly small, then the model is very inaccurate: the step must be rejected, and the trust-region radius must be reduced. If ρ_k is small but less dramatically so, then the step is accepted but the trust-region radius is reduced. If ρ_k is close to 1, then there is a good agreement between the model and the function over the step, and the trust-region radius can be expanded. If $\rho_k \gg 1$, then the model is inaccurate, but the overall optimization iteration is producing a significant decrease in the cost. In this situation a possible strategy is to increase the trust region in the hope that your luck will hold and that bigger steps will result in a further decrease in the cost, regardless of the quality of the model approximation. This procedure is formalized in Algorithm 10.

Later in the chapter we sometimes drop the subscript k and denote x_{k+1} by x_+.

In general, there is no assumption on the operator H_k in (7.6) other than being a symmetric linear operator. Consequently, the choice of the retraction R does not impose any constraint on $\widehat{m}_{x_k}$. In order to achieve superlinear convergence, however, H_k must approximate the Hessian (Theorem 7.4.11). The issue of obtaining an approximate Hessian in practice is addressed in Section 7.5.1.

7.3 COMPUTING A TRUST-REGION STEP

Step 2 in Algorithm 10 computes an (approximate) solution of the trust-region subproblem (7.6),

$$\min_{\eta\in T_{x_k}\mathcal{M}} \widehat{m}_{x_k}(\eta) = f(x_k) + \langle \operatorname{grad} f(x_k), \eta\rangle + \frac{1}{2}\langle H_k[\eta], \eta\rangle,$$
$$\text{subject to } \langle \eta, \eta\rangle_{x_k} \leq \Delta_k^2.$$

Methods for solving trust-region subproblems in the $\mathbb{R}^n$ case can be roughly classified into two broad classes: (i) methods (based on the work of Moré and Sorensen) that compute nearly exact solutions of the subproblem; (ii) methods that compute an approximate solution to the trust-region subproblem

Algorithm 10 Riemannian trust-region (RTR) meta-algorithm

Require: Riemannian manifold $(\mathcal{M}, g)$; scalar field f on $\mathcal{M}$; retraction R from $T\mathcal{M}$ to $\mathcal{M}$ as in Definition 4.1.1.
Parameters: $\bar{\Delta} > 0$, $\Delta_0 \in (0, \bar{\Delta})$, and $\rho' \in [0, \frac{1}{4})$.
Input: Initial iterate $x_0 \in \mathcal{M}$.
Output: Sequence of iterates $\{x_k\}$.

for $k = 0, 1, 2, \ldots$ **do**
  Obtain η_k by (approximately) solving (7.6);
  Evaluate ρ_k from (7.7);
  if $\rho_k < \frac{1}{4}$ **then**
    $\Delta_{k+1} = \frac{1}{4}\Delta_k$;
  else if $\rho_k > \frac{3}{4}$ and $\|\eta_k\| = \Delta_k$ **then**
    $\Delta_{k+1} = \min(2\Delta_k, \bar{\Delta})$;
  else
    $\Delta_{k+1} = \Delta_k$;
  end if
  if $\rho_k > \rho'$ **then**
    $x_{k+1} = R_x \eta_k$;
  else
    $x_{k+1} = x_k$;
  end if
end for

using a computationally simple iteration that achieves at least the decrease in cost obtained for the Cauchy point. For the trust-region subproblem (7.6), assuming that $\operatorname{grad} f(x_k) \neq 0$, we define the *Cauchy point* as the solution η_k^C of the one-dimensional problem

$$\eta_k^C = \arg\min_{\eta} \{\widehat{m}_{x_k}(\eta) : \eta = -\tau \operatorname{grad} f(x_k),\ \tau > 0,\ \|\eta\| \leq \Delta_k\}, \tag{7.8}$$

which reduces to the classical definition of the Cauchy point when $\mathcal{M} = \mathbb{R}^n$. The *Cauchy decrease* is given by $\widehat{m}_{x_k}(0) - \widehat{m}_{x_k}(\eta_k^C)$. We present a brief outline of the first method before providing a more detailed development of a conjugate gradient algorithm for the second approach that we prefer for the later developments.

7.3.1 Computing a nearly exact solution

The following statement is a straightforward adaptation of a result of Moré and Sorensen to the case of the trust-region subproblem on $T_{x_k}\mathcal{M}$ as expressed in (7.6).

Proposition 7.3.1 *The vector η^* is a global solution of the trust-region subproblem* (7.6) *if and only if there exists a scalar $\mu \geq 0$ such that the*

following conditions are satisfied:

$$(H_k + \mu \,\mathrm{id})\eta^* = -\mathrm{grad}\, f(x_k), \tag{7.9a}$$

$$\mu(\Delta_k - \|\eta^*\|) = 0, \tag{7.9b}$$

$$(H_k + \mu \,\mathrm{id}) \text{ is positive-semidefinite}, \tag{7.9c}$$

$$\|\eta^*\| \leq \Delta_k. \tag{7.9d}$$

This result suggests a strategy for computing the solution of the subproblem (7.6). Either solve (7.9) with $\mu = 0$ or define

$$\eta(\mu) := -(H_k + \mu \,\mathrm{id})^{-1} \,\mathrm{grad}\, f(x_k)$$

and adjust μ to achieve $\|\eta(\mu)\| = \Delta_k$. Several algorithms have been proposed to perform this task; see Notes and References.

7.3.2 Improving on the Cauchy point

In many applications, the dimension d of the manifold $\mathcal{M}$ is extremely large (see, for example, computation of an invariant subspace of a large matrix A in Section 7.5). In such cases, solving the linear system (7.9a) of size d or checking the positive-definiteness of a $d \times d$ matrix (7.9c) is unfeasible. Many algorithms exist that scale down the precision of the solution of the trust-region subproblem (7.6) and lighten the numerical burden.

A number of these methods start by computing the Cauchy point (7.8), and then attempt to improve on it. The improvement strategy is often designed so that, when H_k is positive-definite, and given sufficient iterations, the estimate eventually reaches the minimizer $\eta_k^N = (H_k)^{-1} \,\mathrm{grad}\, f(x_k)$ provided that the minimizer lies within the trust region. Among these strategies, the *truncated conjugate-gradient method* is one of the most popular. Algorithm 11 is a straightforward adaptation of the truncated CG method in $\mathbb{R}^n$ to the trust-region subproblem (7.6) in $T_{x_k}\mathcal{M}$. Note that we use superscripts to denote the evolution of η within the inner iteration, while subscripts are used in the outer iteration.

Several comments about Algorithm 11 are in order.

The following result will be useful in the convergence analysis of trust-region methods.

Proposition 7.3.2 *Let η^i, $i = 0, \ldots, j$, be the iterates generated by Algorithm 11 (truncated CG method). Then $\widehat{m}_{x_k}(\eta^i)$ is strictly decreasing and $\widehat{m}_{x_k}(\eta_k) \leq \widehat{m}_{x_k}(\eta^i)$, $i = 0, \ldots, j$. Further, $\|\eta^i\|$ is strictly increasing and $\|\eta_k\| > \|\eta^i\|$, $i = 0, \ldots, j$.*

The simplest stopping criterion to use in Step 14 of Algorithm 11 is to truncate after a fixed number of iterations. In order to achieve superlinear convergence (see Section 7.4.2), one may take the stopping criterion

$$\|r_{j+1}\| \leq \|r_0\| \min(\|r_0\|^\theta, \kappa), \tag{7.10}$$

where $\theta > 0$ is a real parameter chosen in advance.

Algorithm 11 Truncated CG (tCG) method for the trust-region subproblem

Goal: This algorithm handles Step 2 of Algorithm 10.

1: Set $\eta^0 = 0$, $r_0 = \operatorname{grad} f(x_k)$, $\delta_0 = -r_0$; $j = 0$;
2: **loop**
3: **if** $\langle \delta_j, H_k \delta_j \rangle_{x_k} \leq 0$ **then**
4: Compute τ such that $\eta = \eta^j + \tau \delta_j$ minimizes $\widehat{m}_{x_k}(\eta)$ in (7.6) and satisfies $\|\eta\|_{g_x} = \Delta$;
5: **return** $\eta_k := \eta$;
6: **end if**
7: Set $\alpha_j = \langle r_j, r_j \rangle_{x_k} / \langle \delta_j, H_k \delta_j \rangle_{x_k}$;
8: Set $\eta^{j+1} = \eta^j + \alpha_j \delta_j$;
9: **if** $\|\eta^{j+1}\|_{g_x} \geq \Delta$ **then**
10: Compute $\tau \geq 0$ such that $\eta = \eta^j + \tau \delta_j$ satisfies $\|\eta\|_{g_x} = \Delta$;
11: **return** $\eta_k := \eta$;
12: **end if**
13: Set $r_{j+1} = r_j + \alpha_j H_k \delta_j$;
14: **if** a stopping criterion is satisfied **then**
15: **return** $\eta_k := \eta^{j+1}$;
16: **end if**
17: Set $\beta_{j+1} = \langle r_{j+1}, r_{j+1} \rangle_{x_k} / \langle r_j, r_j \rangle_{x_k}$;
18: Set $\delta_{j+1} = -r_{j+1} + \beta_{j+1} \delta_j$;
19: Set $j = j + 1$;
20: **end loop**

In Steps 4 and 10 of Algorithm 11, τ is found by computing the positive root of the quadratic equation

$$\tau^2 \langle \delta_j, \delta_j \rangle_{x_k} + 2\tau \langle \eta^j, \delta_j \rangle_{x_k} = \Delta_k^2 - \langle \eta^j, \eta^j \rangle_{x_k}.$$

Notice that the truncated CG algorithm is "inverse-free", as it uses H_k only in the computation of $H_k[\delta_j]$.

Practical implementations of Algorithm 11 usually include several additional features to reduce the numerical burden and improve the robustness to numerical errors. For example, the value of $\langle r_{j+1}, r_{j+1} \rangle_{x_k}$ can be stored since it will be needed at the next iteration. Because the Hessian operator H_k is an operator on a vector space of dimension d, where d may be very large, it is important to implement an efficient routine for computing $H_k \delta$. In many practical cases, the tangent space $T_x \mathcal{M}$ to which the quantities η, r, and δ belong will be represented as a linear subspace of a higher-dimensional Euclidean space; to prevent numerical errors it may be useful from time to time to reproject the above quantities onto the linear subspace.

7.4 CONVERGENCE ANALYSIS

In this section, we study the global convergence properties of the RTR scheme (Algorithm 10) without any assumption on the way the trust-region subproblems are solved (Step 2), except that the approximate solution η_k must produce a decrease in the model that is at least a fixed fraction of the Cauchy decrease. Under mild additional assumptions on the retraction and on the cost function, it is shown that the sequences $\{x_k\}$ produced by Algorithm 10 converge to the set of critical points of the cost function. This result is well known in the $\mathbb{R}^n$ case; in the case of manifolds, the convergence analysis has to address the fact that a different lifted cost function $\widehat{f}_{x_k}$ is considered at each iterate x_k.

In the second part of the section we analyze the local convergence of Algorithm 10-11 around nondegenerate local minima. Algorithm 10-11 refers to the RTR framework where the trust-region subproblems are approximately solved using the truncated CG algorithm with stopping criterion (7.10). It is shown that the iterates of the algorithm converge to nondegenerate critical points with an order of convergence of at least $\min\{\theta + 1, 2\}$, where θ is the parameter chosen for the stopping condition (7.10).

7.4.1 Global convergence

The objective of this section is to show that, under appropriate assumptions, the sequence $\{x_k\}$ generated by Algorithm 10 converges to the critical set of the cost function; this generalizes a classical convergence property of trust-region methods in $\mathbb{R}^n$. In what follows, $(\mathcal{M}, g)$ is a Riemannian manifold of dimension d and R is a retraction on $\mathcal{M}$ (Definition 4.1.1). We define the pullback cost

$$\widehat{f} : T\mathcal{M} \mapsto \mathbb{R} : \xi \mapsto f(R\xi) \tag{7.11}$$

and, in accordance with (7.5), $\widehat{f}_x$ denotes the restriction of $\widehat{f}$ to $T_x\mathcal{M}$. We denote by $B_\delta(0_x) = \{\xi \in T_x\mathcal{M} : \|\xi\| < \delta\}$ the open ball in $T_x\mathcal{M}$ of radius δ centered at 0_x, and $B_\delta(x)$ stands for the set $\{y \in M : \mathrm{dist}(x, y) < \delta\}$, where dist denotes the Riemannian distance (i.e., the distance defined in terms of the Riemannian metric; see Section 3.6). We denote by $P_\gamma^{t \leftarrow t_0} v$ the vector of $T_{\gamma(t)}\mathcal{M}$ obtained by parallel translation (with respect to the Riemannian connection) of the vector $v \in T_{\gamma(t_0)}\mathcal{M}$ along a curve γ.

As in the classical $\mathbb{R}^n$ proof, we first show that at least one accumulation point of $\{x_k\}$ is a critical point of f. The convergence result requires that $\widehat{m}_{x_k}(\eta_k)$ be a sufficiently good approximation of $\widehat{f}_{x_k}(\eta_k)$. In classical proofs, this is often guaranteed by the assumption that the Hessian of the cost function is bounded. It is, however, possible to weaken this assumption, which leads us to consider the following definition.

Definition 7.4.1 (radially L-C^1 function) *Let $\hat{f} : T\mathcal{M} \to \mathbb{R}$ be defined as in* (7.11). *We say that $\widehat{f}$ is* radially Lipschitz continuously differentiable

if there exist reals $\beta_{RL} > 0$ *and* $\delta_{RL} > 0$ *such that, for all* $x \in \mathcal{M}$*, for all* $\xi \in T_x\mathcal{M}$ *with* $\|\xi\| = 1$*, and for all* $t < \delta_{RL}$*, it holds that*

$$\left| \frac{\mathrm{d}}{\mathrm{d}\tau} \widehat{f}_x(\tau\xi)|_{\tau=t} - \frac{\mathrm{d}}{\mathrm{d}\tau} \widehat{f}_x(\tau\xi)|_{\tau=0} \right| \leq \beta_{RL} t. \tag{7.12}$$

For the purpose of Algorithm 10, which is a descent algorithm, this condition needs only to be imposed for all x in the level set

$$\{x \in M : f(x) \leq f(x_0)\}. \tag{7.13}$$

A key assumption in the classical global convergence result in $\mathbb{R}^n$ is that the approximate solution η_k of the trust-region subproblem (7.6) produces at least as much decrease in the model function as a fixed fraction of the Cauchy decrease. The definition (7.8) of the Cauchy point is equivalent to the closed-form definition $\eta_k^C = -\tau_k \operatorname{grad} f(x_k)$ with

$$\tau_k = \begin{cases} \frac{\Delta_k}{\|\operatorname{grad} f(x_k)\|} & \text{if } \langle H_k[\operatorname{grad} f(x_k)], \operatorname{grad} f(x_k)\rangle_{x_k} \leq 0, \\ \frac{\Delta_k}{\|\operatorname{grad} f(x_k)\|} \min\left(\frac{\|\operatorname{grad} f(x_k)\|^3}{\Delta_k \langle H_k[\operatorname{grad} f(x_k)], \operatorname{grad} f(x_k)\rangle_{x_k}}, 1 \right) & \text{otherwise.} \end{cases}$$

(Note that the definition of the Cauchy point excludes the case $\operatorname{grad} f(x_k) = 0$, for which convergence to a critical point becomes trivial.) The assumption on the decrease in f then becomes

$$\widehat{m}_{x_k}(0) - \widehat{m}_{x_k}(\eta_k) \geq c_1 \|\operatorname{grad} f(x_k)\| \min\left(\Delta_k, \frac{\|\operatorname{grad} f(x_k)\|}{\|H_k\|} \right), \tag{7.14}$$

for some constant $c_1 > 0$, where $\|H_k\|$ is defined as

$$\|H_k\| := \sup\{\|H_k\zeta\| : \zeta \in T_{x_k}\mathcal{M},\ \|\zeta\| = 1\}. \tag{7.15}$$

In particular, the Cauchy point satisfies (7.14) with $c_1 = \frac{1}{2}$. Hence the tangent vector η_k returned by the truncated CG method (Algorithm 11) satisfies (7.14) with $c_1 = \frac{1}{2}$ since the truncated CG method first computes the Cauchy point and then attempts to improve the model decrease.

With these ideas in place, we can state and prove the first global convergence result. Note that this theorem is presented under weak assumptions; stronger but arguably easier to check assumptions are given in Proposition 7.4.5.

Theorem 7.4.2 *Let* $\{x_k\}$ *be a sequence of iterates generated by Algorithm 10 with* $\rho' \in [0, \frac{1}{4})$*. Suppose that* f *is* C^1 *and bounded below on the level set* (7.13)*, that* $\widehat{f}$ *is radially L-*C^1 *(Definition 7.4.1), and that there is a constant* β *such that* $\|H_k\| \leq \beta$ *for all* k*. Further suppose that all* η_k*'s obtained in Step 2 of Algorithm 10 satisfy the Cauchy decrease inequality* (7.14) *for some positive constant* c_1*. We then have*

$$\liminf_{k\to\infty} \|\operatorname{grad} f(x_k)\| = 0.$$

Proof. From the definition of the ratio ρ_k in (7.7), we have

$$|\rho_k - 1| = \left| \frac{\widehat{m}_{x_k}(\eta_k) - \hat{f}_{x_k}(\eta_k)}{\widehat{m}_{x_k}(0) - \widehat{m}_{x_k}(\eta_k)} \right|. \tag{7.16}$$

Proposition A.6.1 (Taylor) applied to the function $t \mapsto \widehat{f}_{x_k}(t\frac{\eta_k}{\|\eta_k\|})$ yields

$$\begin{aligned} \widehat{f}_{x_k}(\eta_k) &= \widehat{f}_{x_k}(0_{x_k}) + \|\eta_k\| \, \frac{\mathrm{d}}{\mathrm{d}\tau} \widehat{f}_{x_k}(\tau \frac{\eta_k}{\|\eta_k\|}) \Big|_{\tau=0} + \epsilon' \\ &= f(x_k) + \langle \operatorname{grad} f(x_k), \eta_k \rangle_{x_k} + \epsilon', \end{aligned}$$

where $|\epsilon'| = \frac{1}{2}\beta_{RL}\|\eta_k\|^2$ whenever $\|\eta_k\| < \delta_{RL}$ and β_{RL} and δ_{RL} are the constants in the radially L-C^1 property (7.12). Therefore, it follows from the definition (7.6) of $\widehat{m}_{x_k}$ that

$$\begin{aligned} |\widehat{m}_{x_k}(\eta_k) - \hat{f}_{x_k}(\eta_k)| &= \left| \tfrac{1}{2}\langle H_k\eta_k, \eta_k\rangle_{x_k} - \epsilon' \right| \\ &\leq \tfrac{1}{2}\beta\|\eta_k\|^2 + \tfrac{1}{2}\beta_{RL}\|\eta_k\|^2 \leq \beta'\|\eta_k\|^2 \end{aligned} \tag{7.17}$$

whenever $\|\eta_k\| < \delta_{RL}$, where $\beta' = \max(\beta, \beta_{RL})$.

Assume for contradiction that the claim does not hold; i.e., assume there exist $\epsilon > 0$ and a positive index K such that

$$\|\operatorname{grad} f(x_k)\| \geq \epsilon \quad \text{for all } k \geq K. \tag{7.18}$$

From (7.14), for $k \geq K$, we have

$$\begin{aligned} \widehat{m}_{x_k}(0) - \widehat{m}_{x_k}(\eta_k) &\geq c_1 \|\operatorname{grad} f(x_k)\| \min\left(\Delta_k, \frac{\|\operatorname{grad} f(x_k)\|}{\|H_k\|}\right) \\ &\geq c_1\epsilon \, \min\left(\Delta_k, \frac{\epsilon}{\beta'}\right). \end{aligned} \tag{7.19}$$

Substituting (7.17) and (7.19) into (7.16), we have that

$$|\rho_k - 1| \leq \frac{\beta'\|\eta_k\|^2}{c_1\epsilon \min\left(\Delta_k, \frac{\epsilon}{\beta'}\right)} \leq \frac{\beta'\Delta_k^2}{c_1\epsilon \min\left(\Delta_k, \frac{\epsilon}{\beta'}\right)} \tag{7.20}$$

whenever $\|\eta_k\| < \delta_{RL}$. Let $\hat{\Delta}$ be defined as $\hat{\Delta} = \min\left(\frac{c_1\epsilon}{2\beta'}, \frac{\epsilon}{\beta'}, \delta_{RL}\right)$. If $\Delta_k \leq \hat{\Delta}$, then $\min\left(\Delta_k, \frac{\epsilon}{\beta'}\right) = \Delta_k$ and (7.20) becomes

$$|\rho_k - 1| \leq \frac{\beta'\hat{\Delta}\Delta_k}{c_1\epsilon \, \min\left(\Delta_k, \frac{\epsilon}{\beta'}\right)} \leq \frac{\Delta_k}{2 \, \min\left(\Delta_k, \frac{\epsilon}{\beta'}\right)} = \frac{1}{2}.$$

Therefore, $\rho_k \geq \frac{1}{2} > \frac{1}{4}$ whenever $\Delta_k \leq \hat{\Delta}$, so that by the workings of Algorithm 10, it follows (from the argument above) that $\Delta_{k+1} \geq \Delta_k$ whenever $\Delta_k \leq \hat{\Delta}$. It follows that a reduction in Δ_k (by a factor of $\frac{1}{4}$) can occur in Algorithm 10 only when $\Delta_k > \hat{\Delta}$. Therefore, we conclude that

$$\Delta_k \geq \min\left(\Delta_K, \hat{\Delta}/4\right) \quad \text{for all } k \geq K. \tag{7.21}$$

Suppose now that there is an infinite subsequence $\mathcal{K}$ such that $\rho_k \geq \frac{1}{4} > \rho'$ for $k \in \mathcal{K}$. If $k \in \mathcal{K}$ and $k \geq K$, we have from (7.19) that

$$f(x_k) - f(x_{k+1}) = f_{x_k} - \hat{f}_{x_k}(\eta_k)$$
$$\geq \frac{1}{4}(\widehat{m}_{x_k}(0) - \widehat{m}_{x_k}(\eta_k)) \geq \frac{1}{4} c_1 \epsilon \min\left(\Delta_k, \frac{\epsilon}{\beta'}\right). \quad (7.22)$$

Since f is bounded below on the level set containing these iterates, it follows from this inequality that $\lim_{k \in \mathcal{K}, k \to \infty} \Delta_k = 0$, clearly contradicting (7.21). Then such an infinite subsequence as $\mathcal{K}$ cannot exist. It follows that we must have $\rho_k < \frac{1}{4}$ for all k sufficiently large so that Δ_k will be reduced by a factor of $\frac{1}{4}$ on every iteration. Then we have $\lim_{k \to \infty} \Delta_k = 0$, which again contradicts (7.21). Hence our assumption (7.18) is false, and the proof is complete. □

To further show that all accumulation points of $\{x_k\}$ are critical points, we need to make an additional regularity assumption on the cost function f. The convergence result in $\mathbb{R}^n$ requires that f be Lipschitz continuously differentiable. That is, for any $x, y \in \mathbb{R}^n$,

$$\|\text{grad}\, f(y) - \text{grad}\, f(x)\| \leq \beta_1 \|y - x\|. \quad (7.23)$$

A key to obtaining a Riemannian counterpart of this global convergence result is to adapt the notion of being Lipschitz continuously differentiable to the Riemannian manifold $(\mathcal{M}, g)$. The expression $\|x - y\|$ on the right-hand side of (7.23) naturally becomes the Riemannian distance $\text{dist}(x, y)$. For the left-hand side of (7.23), observe that the operation $\text{grad}\, f(x) - \text{grad}\, f(y)$ is not well defined in general on a Riemannian manifold since $\text{grad}\, f(x)$ and $\text{grad}\, f(y)$ belong to two different tangent spaces, namely, $T_x\mathcal{M}$ and T_yM. However, if y belongs to a normal neighborhood of x, then there is a unique geodesic $\alpha(t) = \text{Exp}_x(t \,\text{Exp}_x^{-1}\, y)$ in this neighborhood such that $\alpha(0) = x$ and $\alpha(1) = y$, and we can parallel-translate $\text{grad}\, f(y)$ along α to obtain the vector $P_\alpha^{0 \leftarrow 1}\, \text{grad}\, f(y)$ in $T_x\mathcal{M}$. A lower bound on the size of the normal neighborhoods is given by the *injectivity radius*, defined as

$$i(\mathcal{M}) := \inf_{x \in \mathcal{M}} i_x,$$

where

$$i_x := \sup\{\epsilon > 0 : \text{Exp}_x|_{B_\epsilon(0_x)} \text{ is a diffeomorphism for all } x \in \mathcal{M}\}.$$

This yields the following definition.

Definition 7.4.3 (Lipschitz continuously differentiable) *Assume that $(\mathcal{M}, g)$ has a positive injectivity radius. A real function f on $\mathcal{M}$ is* Lipschitz continuously differentiable *if it is differentiable and if there exists β_1 such that, for all x, y in $\mathcal{M}$ with $\text{dist}(x, y) < i(\mathcal{M})$, it holds that*

$$\|P_\alpha^{0 \leftarrow 1}\, \text{grad}\, f(y) - \text{grad}\, f(x)\| \leq \beta_1 \,\text{dist}(y, x), \quad (7.24)$$

where α is the unique minimizing geodesic with $\alpha(0) = x$ and $\alpha(1) = y$.

Note that (7.24) is symmetric in x and y; indeed, since the parallel transport is an isometry, it follows that

$$\|P_\alpha^{0\leftarrow 1}\operatorname{grad} f(y) - \operatorname{grad} f(x)\| = \|\operatorname{grad} f(y) - P_\alpha^{1\leftarrow 0}\operatorname{grad} f(x)\|.$$

Moreover, we place one additional requirement on the retraction R, that there exist $\mu > 0$ and $\delta_\mu > 0$ such that

$$\|\xi\| \geq \mu \operatorname{dist}(x, R_x\xi) \quad \text{for all } x \in \mathcal{M}, \text{for all } \xi \in T_x\mathcal{M},\ \|\xi\| \leq \delta_\mu. \tag{7.25}$$

Note that for the exponential retraction, (7.25) is satisfied as an equality with $\mu = 1$. The bound is also satisfied when $\mathcal{M}$ is compact (Corollary 7.4.6).

We are now ready to show that under some additional assumptions, the gradient of the cost function converges to zero on the whole sequence of iterates. Here again we refer to Proposition 7.4.5 for a simpler (but slightly stronger) set of assumptions that yield the same result.

Theorem 7.4.4 *Let $\{x_k\}$ be a sequence of iterates generated by Algorithm 10. Suppose that all the assumptions of Theorem 7.4.2 are satisfied. Further suppose that $\rho' \in (0, \frac{1}{4})$, that f is Lipschitz continuously differentiable (Definition 7.4.3), and that* (7.25) *is satisfied for some $\mu > 0$, $\delta_\mu > 0$. It then follows that*

$$\lim_{k\to\infty} \operatorname{grad} f(x_k) = 0.$$

Proof. Consider any index m such that $\operatorname{grad} f(x_m) \neq 0$. Define the scalars

$$\epsilon = \frac{1}{2}\|\operatorname{grad} f(x_m)\|, \quad r = \min\left(\frac{\|\operatorname{grad} f(x_m)\|}{2\beta_1}, i(M)\right) = \min\left(\frac{\epsilon}{\beta_1}, i(M)\right).$$

In view of the Lipschitz property (7.24), we have for all $x \in B_r(x_m)$,

$$\begin{aligned}
\|\operatorname{grad} f(x)\| &= \|P_\alpha^{0\leftarrow 1}\operatorname{grad} f(x)\| \\
&= \|P_\alpha^{0\leftarrow 1}\operatorname{grad} f(x) + \operatorname{grad} f(x_m) - \operatorname{grad} f(x_m)\| \\
&\geq \|\operatorname{grad} f(x_m)\| - \|P_\alpha^{0\leftarrow 1}\operatorname{grad} f(x) - \operatorname{grad} f(x_m)\| \\
&\geq 2\epsilon - \beta_1 \operatorname{dist}(x, x_m) \\
&> 2\epsilon - \beta_1 \min\left(\frac{\|\operatorname{grad} f(x_m)\|}{2\beta_1}, i(M)\right) \\
&\geq 2\epsilon - \frac{1}{2}\|\operatorname{grad} f(x_m)\| \\
&= \epsilon.
\end{aligned}$$

If the entire sequence $\{x_k\}_{k\geq m}$ stays inside the ball $B_r(x_m)$, then we have $\|\operatorname{grad} f(x_k)\| > \epsilon$ for all $k \geq m$, a contradiction to Theorem 7.4.2. Thus the sequence eventually leaves the ball $B_r(x_m)$. Let the index $l \geq m$ be such that x_{l+1} is the first iterate after x_m outside $B_r(x_m)$. Since $\|\operatorname{grad} f(x_k)\| > \epsilon$ for

$k = m, m+1, \ldots, l$, we have, in view of the Cauchy decrease condition (7.14),

$$\begin{aligned} f(x_m) - f(x_{l+1}) &= \sum_{k=m}^{l} f(x_k) - f(x_{k+1}) \\ &\geq \sum_{k=m, x_k \neq x_{k+1}}^{l} \rho' \left(\widehat{m}_{x_k}(0) - \widehat{m}_{x_k}(\eta_k) \right) \\ &\geq \sum_{k=m, x_k \neq x_{k+1}}^{l} \rho' c_1 \|\operatorname{grad} f(x_k)\| \min \left(\Delta_k, \frac{\|\operatorname{grad} f(x_k)\|}{\|B_k\|} \right) \\ &\geq \sum_{k=m, x_k \neq x_{k+1}}^{l} \rho' c_1 \epsilon \min \left(\Delta_k, \frac{\epsilon}{\beta} \right). \end{aligned}$$

We distinguish two cases. If $\Delta_k > \epsilon/\beta$ in at least one of the terms of the sum, then

$$f(x_m) - f(x_{l+1}) \geq \rho' c_1 \epsilon \frac{\epsilon}{\beta}. \tag{7.26}$$

In the other case, we have

$$f(x_m) - f(x_{l+1}) \geq \rho' c_1 \epsilon \sum_{k=m, x_k \neq x_{k+1}}^{l} \Delta_k \geq \rho' c_1 \epsilon \sum_{k=m, x_k \neq x_{k+1}}^{l} \|\eta_k\|. \tag{7.27}$$

If $\|\eta_k\| > \delta_\mu$ in at least one term in the sum, then

$$f(x_m) - f(x_{l+1}) \geq \rho' c_1 \epsilon \delta_\mu. \tag{7.28}$$

Otherwise, (7.27) yields

$$\begin{aligned} f(x_m) - f(x_{l+1}) &\geq \rho' c_1 \epsilon \sum_{k=m, x_k \neq x_{k+1}}^{l} \mu \operatorname{dist}(x_k, R_{x_k} \eta_k) \\ &= \rho' c_1 \epsilon \mu \sum_{k=m, x_k \neq x_{k+1}}^{l} \operatorname{dist}(x_k, x_{k+1}) \\ &\geq \rho' c_1 \epsilon \mu r = \rho' c_1 \epsilon \mu \min \left(\frac{\epsilon}{\beta_1}, i(M) \right). \end{aligned} \tag{7.29}$$

It follows from (7.26), (7.28), and (7.29) that

$$f(x_m) - f(x_{l+1}) \geq \rho' c_1 \epsilon \min \left(\frac{\epsilon}{\beta}, \delta_\mu, \frac{\epsilon \mu}{\beta_1}, i(M)\mu \right). \tag{7.30}$$

Because $\{f(x_k)\}_{k=0}^{\infty}$ is decreasing and bounded below, we have

$$f(x_k) \downarrow f^* \tag{7.31}$$

for some $f^* > -\infty$. It then follows from (7.30) that

$$\begin{aligned} f(x_m) - f^* &\geq f(x_m) - f(x_{l+1}) \\ &\geq \rho' c_1 \epsilon \min\left(\frac{\epsilon}{\beta}, \delta_\mu, \frac{\epsilon\mu}{\beta_1}, i(M)\mu\right) \\ &= \frac{1}{2}\rho' c_1 \|\operatorname{grad} f(x_m)\| \\ &\quad \min\left(\frac{\|\operatorname{grad} f(x_m)\|}{2\beta}, \delta_\mu, \frac{\|\operatorname{grad} f(x_m)\|\mu}{2\beta_1}, i(M)\mu\right). \end{aligned}$$

Taking $m \to \infty$ in the latter expression yields $\lim_{m\to\infty} \|\operatorname{grad} f(x_m)\| = 0$. □

Note that this theorem reduces gracefully to the classical $\mathbb{R}^n$ case, taking $\mathcal{M} = \mathbb{R}^n$ endowed with the classical inner product and $R_x\xi := x + \xi$. Then $i(M) = +\infty > 0$, R satisfies (7.25), and the Lipschitz condition (7.24) reduces to the classical expression, which subsumes the radially L-C^1 condition.

The following proposition shows that the regularity conditions on f and $\widehat{f}$ required in the previous theorems are satisfied under stronger but possibly easier to check conditions. These conditions impose a bound on the Hessian of f and on the "acceleration" along curves $t \mapsto R(t\xi)$. Note also that all these conditions need only be checked on the level set $\{x \in \mathcal{M} : f(x) \leq f(x_0)\}$.

Proposition 7.4.5 *Suppose that $\|\operatorname{grad} f(x)\| \leq \beta_g$ and that $\|\operatorname{Hess} f(x)\| \leq \beta_H$ for some constants β_g, β_H, and all $x \in \mathcal{M}$. Moreover, suppose that*

$$\left\| \frac{\mathrm{D}}{\mathrm{d}t} \frac{\mathrm{d}}{\mathrm{d}t} R(t\xi) \right\| \leq \beta_D \tag{7.32}$$

for some constant β_D, for all $\xi \in T\mathcal{M}$ with $\|\xi\| = 1$ and all $t < \delta_D$, where $\frac{\mathrm{D}}{\mathrm{d}t}$ denotes the covariant derivative along the curve $t \mapsto R(t\xi)$. Then the Lipschitz-C^1 condition on f (Definition 7.4.3) is satisfied with $\beta_L = \beta_H$; the radially Lipschitz-C^1 condition on $\widehat{f}$ (Definition 7.4.1) is satisfied for $\delta_{RL} < \delta_D$ and $\beta_{RL} = \beta_H(1 + \beta_D\delta_D) + \beta_g\beta_D$; and the condition (7.25) on R is satisfied for values of μ and δ_μ satisfying $\delta_\mu < \delta_D$ and $\frac{1}{2}\beta_D\delta_\mu < \frac{1}{\mu} - 1$.

Proof. By a standard Taylor argument (see Lemma 7.4.7), boundedness of the Hessian of f implies the Lipschitz-C^1 property of f.

For (7.25), define $u(t) = R(t\xi)$ and observe that

$$\operatorname{dist}(x, R(t\xi)) \leq \int_0^t \|u'(\tau)\| \, \mathrm{d}\tau$$

where $\int_0^t \|u'(\tau)\| \, \mathrm{d}\tau$ is the length of the curve u between 0 and t. Using the Cauchy-Schwarz inequality and the invariance of the metric by the connection, we have

$$\begin{aligned} \left| \frac{\mathrm{d}}{\mathrm{d}\tau} \|u'(\tau)\| \right| &= \left| \frac{\mathrm{d}}{\mathrm{d}\tau} \sqrt{\langle u'(\tau), u'(\tau)\rangle_{u(\tau)}} \right| = \left| \frac{\langle \frac{\mathrm{D}}{\mathrm{d}t} u'(\tau), u'(\tau)\rangle_{u(\tau)}}{\|u'(\tau)\|} \right| \\ &\leq \frac{\beta_D \|u'(\tau)\|}{\|u'(\tau)\|} \leq \beta_D \end{aligned}$$

for all $t < \delta_D$. Therefore

$$\int_0^t \|u'(\tau)\| \, \mathrm{d}\tau \leq \int_0^t \|u'(0)\| + \beta_D \tau \, \mathrm{d}\tau = \|\xi\| t + \tfrac{1}{2}\beta_D t^2 = t + \tfrac{1}{2}\beta_D t^2,$$

which is smaller than $\frac{t}{\mu}$ if $\frac{1}{2}\beta_D t < \frac{1}{\mu} - 1$.

For the radially Lipschitz-C^1 condition, let $u(t) = R(t\xi)$ and $h(t) = f(u(t)) = \widehat{f}(t\xi)$ with $\xi \in T_x\mathcal{M}$, $\|\xi\| = 1$. Then

$$h'(t) = \langle \operatorname{grad} f(u(t)), u'(t) \rangle_{u(t)}$$

and

$$\begin{aligned} h''(t) &= \frac{\mathrm{D}}{\mathrm{d}t} \langle \operatorname{grad} f(u(t)), u'(t) \rangle_{u(t)} \\ &= \langle \frac{\mathrm{D}}{\mathrm{d}t} \operatorname{grad} f(u(t)), u'(t) \rangle_{u(t)} + \langle \operatorname{grad} f(u(t)), \frac{\mathrm{D}}{\mathrm{d}t} u'(t) \rangle_{u(t)}. \end{aligned}$$

Now, $\frac{\mathrm{D}}{\mathrm{d}t} \operatorname{grad} f(u(t)) = \nabla_{u'(t)} \operatorname{grad} f(u(t)) = \operatorname{Hess} f(u(t))[u'(t)]$. It follows that $|h''(t)|$ is bounded on $t \in [0, \delta_D)$ by the constant $\beta_{RL} = \beta_H(1 + \beta_D \delta_D) + \beta_g \beta_D$. Then

$$|h'(t) - h'(0)| \leq \int_0^t |h''(\tau)| \, \mathrm{d}\tau \leq t\beta_{RL}.$$

□

Corollary 7.4.6 (smoothness and compactness) *If the cost function f is smooth and the Riemannian manifold $\mathcal{M}$ is compact, then all the conditions in Proposition 7.4.5 are satisfied.*

The major manifolds considered in this book (the Grassmann manifold and the Stiefel manifold) are compact, and the cost functions based on the Rayleigh quotient are smooth.

7.4.2 Local convergence

We now state local convergence properties of Algorithm 10-11: local convergence to local minimizers (Theorem 7.4.10) and superlinear convergence (Theorem 7.4.11). We begin with a few preparatory lemmas.

As before, $(\mathcal{M}, g)$ is a Riemannian manifold of dimension d and R is a retraction on $\mathcal{M}$ (Definition 4.1.1). The first lemma is a first-order Taylor formula for tangent vector fields.

Lemma 7.4.7 (Taylor's theorem) *Let $x \in \mathcal{M}$, let $\mathcal{V}$ be a normal neighborhood of x, and let ζ be a C^1 tangent vector field on $\mathcal{M}$. Then, for all $y \in \mathcal{V}$,*

$$P_\gamma^{0 \leftarrow 1} \zeta_y = \zeta_x + \nabla_\xi \zeta + \int_0^1 \left(P_\gamma^{0 \leftarrow \tau} \nabla_{\gamma'(\tau)} \zeta - \nabla_\xi \zeta \right) \mathrm{d}\tau, \tag{7.33}$$

where γ is the unique minimizing geodesic satisfying $\gamma(0) = x$ and $\gamma(1) = y$, and $\xi = \operatorname{Exp}_x^{-1} y = \gamma'(0)$.

Proof. Start from

$$P_\gamma^{0\leftarrow 1}\zeta_y = \zeta_x + \int_0^1 \frac{\mathrm{d}}{\mathrm{d}\tau} P_\gamma^{0\leftarrow\tau}\zeta \,\mathrm{d}\tau = \zeta_x + \nabla_\xi\zeta + \int_0^1 \left(\frac{\mathrm{d}}{\mathrm{d}\tau} P_\gamma^{0\leftarrow\tau}\zeta - \nabla_\xi\zeta\right)\mathrm{d}\tau$$

and use the formula for the connection in terms of the parallel transport (see [dC92, Ch. 2, Ex. 2]), to obtain

$$\frac{\mathrm{d}}{\mathrm{d}\tau} P_\gamma^{0\leftarrow\tau}\zeta = \frac{\mathrm{d}}{\mathrm{d}\epsilon} P_\gamma^{0\leftarrow\tau} P_\gamma^{\tau\leftarrow\tau+\epsilon}\zeta\bigg|_{\epsilon=0} = P_\gamma^{0\leftarrow\tau}\nabla_{\gamma'}\zeta.$$

□

We use this lemma to show that in some neighborhood of a nondegenerate local minimizer v of f, the norm of the gradient of f can be taken as a measure of the Riemannian distance to v.

Lemma 7.4.8 *Let $v \in \mathcal{M}$ and let f be a C^2 cost function such that* $\operatorname{grad} f(v) = 0$ *and* $\operatorname{Hess} f(v)$ *is positive-definite with maximal and minimal eigenvalues $\lambda_{\max}$ and $\lambda_{\min}$. Then, given $c_0 < \lambda_{\min}$ and $c_1 > \lambda_{\max}$, there exists a neighborhood $\mathcal{V}$ of v such that, for all $x \in \mathcal{V}$, it holds that*

$$c_0 \operatorname{dist}(v,x) \leq \|\operatorname{grad} f(x)\| \leq c_1 \operatorname{dist}(v,x). \tag{7.34}$$

Proof. From Taylor (Lemma 7.4.7), it follows that

$$\begin{aligned} P_\gamma^{0\leftarrow 1} \operatorname{grad} f(v) &= \operatorname{Hess} f(v)[\gamma'(0)] \\ &\quad + \int_0^1 \left(P_\gamma^{0\leftarrow\tau} \operatorname{Hess} f(\gamma(\tau))[\gamma'(\tau)] - \operatorname{Hess} f(v)[\gamma'(0)]\right) \mathrm{d}\tau. \end{aligned} \tag{7.35}$$

Since f is C^2 and since $\|\gamma'(\tau)\| = \operatorname{dist}(v,x)$ for all $\tau \in [0,1]$, we have the following bound for the integral in (7.35):

$$\begin{aligned} &\left\| \int_0^1 P_\gamma^{0\leftarrow\tau} \operatorname{Hess} f(\gamma(\tau))[\gamma'(\tau)] - \operatorname{Hess} f(v)[\gamma'(0)] \,\mathrm{d}\tau \right\| \\ &= \left\| \int_0^1 \left(P_\gamma^{0\leftarrow\tau} \circ \operatorname{Hess} f(\gamma(\tau)) \circ P_\gamma^{\tau\leftarrow 0} - \operatorname{Hess} f(v)\right)[\gamma'(0)] \,\mathrm{d}\tau \right\| \\ &\leq \epsilon(\operatorname{dist}(v,x)) \operatorname{dist}(v,x), \end{aligned}$$

where $\lim_{t\to 0} \epsilon(t) = 0$. Since $\operatorname{Hess} f(v)$ is nonsingular, it follows that $|\lambda_{\min}| > 0$. Take $\mathcal{V}$ sufficiently small so that $\lambda_{\min} - \epsilon(\operatorname{dist}(v,x)) > c_0$ and $\lambda_{\max} + \epsilon(\operatorname{dist}(v,x)) < c_1$ for all x in $\mathcal{V}$. Then, using the fact that the parallel translation is an isometry, (7.34) follows from (7.35). □

We need a relation between the gradient of f at $R_x(\xi)$ and the gradient of $\widehat{f}_x$ at ξ.

Lemma 7.4.9 *Let R be a retraction on $\mathcal{M}$ and let f be a C^1 cost function on $\mathcal{M}$. Then, given $v \in M$ and $c_5 > 1$, there exists a neighborhood $\mathcal{V}$ of v and $\delta > 0$ such that*

$$\|\operatorname{grad} f(R\xi)\| \leq c_5 \|\operatorname{grad} \widehat{f}(\xi)\|$$

for all $x \in \mathcal{V}$ and all $\xi \in T_x\mathcal{M}$ with $\|\xi\| \leq \delta$, where $\widehat{f}$ is as in (7.11).

Proof. Consider a parameterization of $\mathcal{M}$ at v and consider the corresponding parameterization of $T\mathcal{M}$ (see Section 3.5.3). We have

$$\partial_i \widehat{f}_x(\xi) = \sum_j \partial_j f(R\xi) A_i^j(\xi),$$

where $A(\xi)$ stands for the differential of R_x at $\xi \in T_x\mathcal{M}$. Then,

$$\begin{aligned}\|\operatorname{grad} \widehat{f}_x(\xi)\|^2 &= \sum_{i,j} \partial_i \widehat{f}_x(\xi) g^{ij}(x)\, \partial_j \widehat{f}_x(\xi) \\ &= \sum_{i,j,k,\ell} \partial_k f(R_x\xi) A_i^k(\xi) g^{ij}(x) A_j^\ell(\xi)\, \partial_\ell f(R_x\xi)\end{aligned}$$

and

$$\|\operatorname{grad} f(R_x\xi)\|^2 = \sum_{i,j} \partial_i f(R_x\xi) g^{ij}(R_x\xi)\, \partial_j f(R_x\xi).$$

The conclusion follows by a real analysis argument, invoking the smoothness properties of R and g, and the compactness of the set $\{(x,\xi) : x \in \mathcal{V},\ \xi \in T_x\mathcal{M},\ \|\xi\| \leq \delta\}$ and using $A(0_x) = \mathrm{id}$. □

Finally, we will make use of Lemma 5.5.6 stating that the Hessians of f and $\widehat{f}$ coincide at critical points.

We now state and prove the local convergence results. The first result states that the nondegenerate local minima are attractors of Algorithm 10-11.

Theorem 7.4.10 (local convergence to local minima) *Consider Algorithm 10-11—i.e., the Riemannian trust-region algorithm where the trust-region subproblems* (7.6) *are solved using the truncated CG algorithm with stopping criterion* (7.10)*—with all the assumptions of Theorem 7.4.2. Let v be a nondegenerate local minimizer of f, i.e.,* $\operatorname{grad} f(v) = 0$ *and* $\operatorname{Hess} f(v)$ *is positive-definite. Assume that $x \mapsto \|H_x^{-1}\|$ is bounded on a neighborhood of v and that* (7.25) *holds for some $\mu > 0$ and $\delta_\mu > 0$. Then there exists a neighborhood $\mathcal{V}$ of v such that, for all $x_0 \in \mathcal{V}$, the sequence $\{x_k\}$ generated by Algorithm 10-11 converges to v.*

Proof. Take $\delta_1 > 0$ with $\delta_1 < \delta_\mu$ such that $\|H_x^{-1}\|$ is bounded on $B_{\delta_1}(v)$, that $B_{\delta_1}(v)$ contains only v as critical point, and that $f(x) > f(v)$ for all $x \in \bar{B}_{\delta_1}(v)$. (In view of the assumptions, such a δ_1 exists.) Take δ_2 small enough that, for all $x \in B_{\delta_2}(v)$, it holds that $\|\eta^*(x)\| \leq \mu(\delta_1 - \delta_2)$, where η^* is the (unique) solution of $H_x\eta^* = -\operatorname{grad} f(x)$; such a δ_2 exists because of Lemma 7.4.8 and the bound on $\|H_x^{-1}\|$. Consider a level set $\mathcal{L}$ of f such that $\mathcal{V} := \mathcal{L} \cap B_{\delta_1}(v)$ is a subset of $B_{\delta_2}(v)$; invoke that $f \in C^1$ to show that such a level set exists. Let $\eta^{tCG}(x,\Delta)$ denote the tangent vector η_k returned by the truncated CG algorithm (Algorithm 11) when $x_k = x$ and $\Delta_k = \Delta$. Then, $\mathcal{V}$ is a neighborhood of v, and for all $x \in \mathcal{V}$ and all $\Delta > 0$, we have

$$\operatorname{dist}(x, x_+) \leq \frac{1}{\mu}\|\eta^{tCG}(x,\Delta)\| \leq \frac{1}{\mu}\|\eta^*\| \leq (\delta_1 - \delta_2),$$

where we used the fact that $\|\eta\|$ is increasing along the truncated CG process (Proposition 7.3.2). It follows from the equation above that x_+ is in $B_{\delta_1}(v)$. Moreover, since $f(x_+) \leq f(x)$, it follows that $x_+ \in \mathcal{V}$. Thus $\mathcal{V}$ is invariant. But the only critical point of f in $\mathcal{V}$ is v, so $\{x_k\}$ goes to v whenever x_0 is in $\mathcal{V}$. □

Now we study the order of convergence of sequences that converge to a nondegenerate local minimizer.

Theorem 7.4.11 (order of convergence) *Consider Algorithm 10-11 with stopping criterion* (7.10)*. Suppose that f is a C^2 cost function on $\mathcal{M}$ and that*

$$\|H_k - \operatorname{Hess} \widehat{f}_{x_k}(0_k)\| \leq \beta_{\mathcal{H}} \|\operatorname{grad} f(x_k)\|, \tag{7.36}$$

i.e., H_k is a sufficiently good approximation of $\operatorname{Hess} \widehat{f}_{x_k}(0_{x_k})$*. Let $v \in \mathcal{M}$ be a nondegenerate local minimizer of f (i.e.,* $\operatorname{grad} f(v) = 0$ *and* $\operatorname{Hess} f(v)$ *is positive-definite). Further assume that* $\operatorname{Hess} \widehat{f}_x$ *is Lipschitz-continuous at 0_x uniformly in x in a neighborhood of v, i.e., there exist $\beta_{L2} > 0$, $\delta_1 > 0$, and $\delta_2 > 0$ such that, for all $x \in B_{\delta_1}(v)$ and all $\xi \in B_{\delta_2}(0_x)$, it holds that*

$$\|\operatorname{Hess} \widehat{f}_x(\xi) - \operatorname{Hess} \widehat{f}_x(0_x)\| \leq \beta_{L2} \|\xi\|, \tag{7.37}$$

where $\|\cdot\|$ on the left-hand side denotes the operator norm in $T_x\mathcal{M}$ defined as in (7.15)*. Then there exists $c > 0$ such that, for all sequences $\{x_k\}$ generated by the algorithm converging to v, there exists $K > 0$ such that for all $k > K$,*

$$\operatorname{dist}(x_{k+1}, v) \leq c \left(\operatorname{dist}(x_k, v)\right)^{\min\{\theta+1, 2\}} \tag{7.38}$$

with $\theta > 0$ as in (7.10)*.*

Proof. The proof relies on a set of bounds which are justified after the main result is proved. Assume that there exist $\tilde{\Delta}, c_0, c_1, c_2, c_3, c_3', c_4, c_5$ such that, for all sequences $\{x_k\}$ satisfying the conditions asserted, all $x \in \mathcal{M}$, all ξ with $\|\xi\| < \tilde{\Delta}$, and all k greater than some K, it holds that

$$c_0 \operatorname{dist}(v, x_k) \leq \|\operatorname{grad} f(x_k)\| \leq c_1 \operatorname{dist}(v, x_k), \tag{7.39}$$

$$\|\eta_k\| \leq c_4 \|\operatorname{grad} \widehat{m}_{x_k}(0)\| \leq \tilde{\Delta}, \tag{7.40}$$

$$\rho_k > \rho', \tag{7.41}$$

$$\|\operatorname{grad} f(R_{x_k}\xi)\| \leq c_5 \|\operatorname{grad} \widehat{f}_{x_k}(\xi)\|, \tag{7.42}$$

$$\|\operatorname{grad} \widehat{m}_{x_k}(\xi) - \operatorname{grad} \widehat{f}_{x_k}(\xi)\| \leq c_3 \|\xi\|^2 + c_3' \|\operatorname{grad} f(x_k)\| \, \|\xi\|, \tag{7.43}$$

$$\|\operatorname{grad} \widehat{m}_{x_k}(\eta_k)\| \leq c_2 \|\operatorname{grad} \widehat{m}_{x_k}(0)\|^{\theta+1}, \tag{7.44}$$

where $\{\eta_k\}$ is the sequence of update vectors corresponding to $\{x_k\}$.

Given the bounds (7.39) to (7.44), the proof proceeds as follows. For all $k > K$, it follows from (7.39) and (7.41) that

$$c_0 \operatorname{dist}(v, x_{k+1}) \leq \|\operatorname{grad} f(x_{k+1})\| = \|\operatorname{grad} f(R_{x_k}\eta_k)\|,$$

from (7.42) and (7.40) that

$$\|\operatorname{grad} f(R_{x_k}\eta_k)\| \leq c_5 \|\operatorname{grad} \widehat{f}_{x_k}(\eta_k)\|,$$

from (7.40) and (7.43) and (7.44) that

$$\begin{aligned}\|\operatorname{grad} \widehat{f}_{x_k}(\eta_k)\| &\leq \|\operatorname{grad} \widehat{m}_{x_k}(\eta_k) - \operatorname{grad} \widehat{f}_{x_k}(\eta_k)\| + \|\operatorname{grad} \widehat{m}_{x_k}(\eta_k)\| \\ &\leq (c_3 c_4^2 + c_3' c_4)\|\operatorname{grad} \widehat{m}_{x_k}(0)\|^2 + c_2 \|\operatorname{grad} \widehat{m}_{x_k}(0)\|^{1+\theta},\end{aligned}$$

and from (7.39) that

$$\|\operatorname{grad} \widehat{m}_{x_k}(0)\| = \|\operatorname{grad} f(x_k)\| \leq c_1 \operatorname{dist}(v, x_k).$$

Consequently, taking K larger if necessary so that $\operatorname{dist}(v, x_k) < 1$ for all $k > K$, it follows that

$$\begin{aligned} c_0 \operatorname{dist}(v, x_{k+1}) \leq & \quad \|\operatorname{grad} f(x_{k+1})\| && (7.45)\\ \leq & \quad c_5(c_3 c_4^2 + c_3' c_4)\|\operatorname{grad} f(x_k)\|^2 + c_5 c_2 \|\operatorname{grad} f(x_k)\|^{\theta+1} && (7.46)\\ \leq & \quad c_5((c_3 c_4^2 + c_3' c_4) c_1^2 (\operatorname{dist}(v, x_k))^2 + c_2 c_1^{1+\theta} (\operatorname{dist}(v, x_k))^{1+\theta}) \\ \leq & \quad c_5((c_3 c_4^2 + c_3' c_4) c_1^2 + c_2 c_1^{1+\theta}) (\operatorname{dist}(v, x_k))^{\min\{2, 1+\theta\}} \end{aligned}$$

for all $k > K$, which is the desired result.

It remains to show that the bounds (7.39)–(7.44) hold under the assumptions or the theorem.

Equation (7.39) comes from Lemma 7.4.8 and is due to the fact that v is a nondegenerate critical point.

We prove (7.40). Since $\{x_k\}$ converges to the nondegenerate local minimizer v where $\operatorname{Hess} \widehat{f}_v(0_v) = \operatorname{Hess} f(v)$ (in view of Lemma 5.5.6), and since $\operatorname{Hess} f(v)$ is positive-definite with $f \in C^2$, it follows from the approximation condition (7.36) and from (7.39) that there exists $c_4 > 0$ such that, for all k greater than some K, H_k is positive-definite and $\|H_k^{-1}\| < c_4$. Given a $k > K$, let η_k^* be the solution of $H_k \eta_k^* = -\operatorname{grad} \widehat{m}_{x_k}(0)$. It follows that $\|\eta_k^*\| \leq c_4 \|\operatorname{grad} \widehat{m}_{x_k}(0)\|$. Since $\{x_k\}$ converges to a critical point of f and since $\operatorname{grad} \widehat{m}_{x_k}(0) = \operatorname{grad} f(x_k)$ in view of (7.6), we obtain that $\|\eta_k^*\| \leq c_4 \|\operatorname{grad} \widehat{m}_{x_k}(0)\| \leq \tilde{\Delta}$ for any given $\tilde{\Delta} > 0$ by choosing K larger if necessary. Then, since the sequence of η_k^j's constructed by the truncated CG inner iteration (Algorithm 11) is strictly increasing in norm (Proposition 7.3.2) and would reach η_k^* at $j = d$ in the absence of the stopping criterion, it follows that (7.40) holds.

We prove (7.41). Let γ_k denote $\|\operatorname{grad} f(x_k)\|$. It follows from the definition of ρ_k that

$$\rho_k - 1 = \frac{\widehat{m}_{x_k}(\eta_k) - \widehat{f}_{x_k}(\eta_k)}{\widehat{m}_{x_k}(0_{x_k}) - \widehat{m}_{x_k}(\eta_k)}. \tag{7.47}$$

From Taylor's theorem, it holds that

$$\begin{aligned}\widehat{f}_{x_k}(\eta_k) = & \widehat{f}_{x_k}(0_{x_k}) + \langle \operatorname{grad} f(x_k), \eta_k \rangle_{x_k} \\ & + \int_0^1 \langle \operatorname{Hess} \widehat{f}_{x_k}(\tau \eta_k)[\eta_k], \eta_k \rangle_{x_k} (1 - \tau) \mathrm{d}\tau.\end{aligned}$$

It follows that

$$\begin{aligned}
&\left|\widehat{m}_{x_k}(\eta_k) - \widehat{f}_{x_k}(\eta_k)\right| \\
&= \left| \int_0^1 \left(\langle H_k[\eta_k], \eta_k \rangle_{x_k} - \langle \operatorname{Hess} \widehat{f}_{x_k}(\tau\eta_k)[\eta_k], \eta_k \rangle_{x_k} \right) (1-\tau)\, \mathrm{d}\tau \right| \\
&\leq \int_0^1 \left| \langle (H_k - \operatorname{Hess} \widehat{f}_{x_k}(0_{x_k}))[\eta_k], \eta_k \rangle_{x_k} \right| (1-\tau)\, \mathrm{d}\tau \\
&\quad + \int_0^1 \left| \langle (\operatorname{Hess} \widehat{f}_{x_k}(0_{x_k}) - \operatorname{Hess} \widehat{f}(\tau\eta_k))[\eta_k], \eta_k \rangle_{x_k} \right| (1-\tau)\, \mathrm{d}\tau \\
&\leq \frac{1}{2}\beta_{\mathcal{H}}\gamma_k \|\eta_k\|^2 + \frac{1}{6}\beta_{L2}\|\eta_k\|^3.
\end{aligned}$$

It then follows from (7.47), using the Cauchy bound (7.14), that

$$|\rho_k - 1| \leq \frac{(3\beta_{\mathcal{H}}\gamma_k + \beta_{L2}\|\eta_k\|)\, \|\eta_k\|^2}{6\gamma_k \min\{\Delta_k, \gamma_k/\beta\}},$$

where β is an upper bound on the norm of H_k. Since $\|\eta_k\| \leq \Delta_k$ and $\|\eta_k\| \leq c_4\gamma_k$, it follows that

$$|\rho_k - 1| \leq \frac{(3\beta_{\mathcal{H}} + \beta_{L2}c_4)\, (\min\{\Delta_k, c_4\gamma_k\})^2}{6 \min\{\Delta_k, \gamma_k/\beta\}}. \tag{7.48}$$

Either Δ_k is active in the denominator of (7.48), in which case we have

$$|\rho_k - 1| \leq \frac{(3\beta_{\mathcal{H}} + \beta_{L2}c_4)\, \Delta_k c_4 \gamma_k}{6\Delta_k} = \frac{(3\beta_{\mathcal{H}} + \beta_{L2}c_4)\, c_4}{6}\gamma_k,$$

or γ_k/β is active in the denominator of (7.48), in which case we have

$$|\rho_k - 1| \leq \frac{(3\beta_{\mathcal{H}} + \beta_{L2}c_4)\, (c_4\gamma_k)^2}{6\gamma_k/\beta} = \frac{(3\beta_{\mathcal{H}} + \beta_{L2}c_4)\, c_4^2\beta}{6}\gamma_k.$$

In both cases, $\lim_{k\to\infty} \rho_k = 1$ since, in view of (7.39), $\lim_{k\to\infty} \gamma_k = 0$.

Equation (7.42) comes from Lemma 7.4.9.

We prove (7.43). It follows from Taylor's formula (Lemma 7.4.7, where the parallel translation becomes the identity since the domain of $\widehat{f}_{x_k}$ is the Euclidean space $T_{x_k}\mathcal{M}$) that

$$\begin{aligned}
\operatorname{grad} \widehat{f}_{x_k}(\xi) = \operatorname{grad} \widehat{f}_{x_k}(0_{x_k}) + \operatorname{Hess} \widehat{f}_{x_k}(0_{x_k})[\xi] \\
+ \int_0^1 \left(\operatorname{Hess} \widehat{f}_{x_k}(\tau\xi) - \operatorname{Hess} \widehat{f}_{x_k}(0_{x_k}) \right)[\xi]\, \mathrm{d}\tau.
\end{aligned}$$

The conclusion comes by the Lipschitz condition (7.37) and the approximation condition (7.36).

Finally, equation (7.44) comes from the stopping criterion (7.10) of the inner iteration. More precisely, the truncated CG loop (Algorithm 11) terminates if $\langle \delta_j, H_k\delta_j \rangle \leq 0$ or $\|\eta_{j+1}\| \geq \Delta$ or the criterion (7.10) is satisfied. Since $\{x_k\}$ converges to v and $\operatorname{Hess} f(v)$ is positive-definite, it follows that H_k is positive-definite for all k greater than a certain K. Therefore, for all

$k > K$, the criterion $\langle \delta_j, H_k \delta_j \rangle \leq 0$ is never satisfied. In view of (7.40) and (7.41), it can be shown that the trust region is eventually inactive. Therefore, increasing K if necessary, the criterion $\|\eta_{j+1}\| \geq \Delta$ is never satisfied for all $k > K$. In conclusion, for all $k > K$, the stopping criterion (7.10) is satisfied each time a computed η_k is returned by the truncated CG loop. Therefore, the truncated CG loop behaves as a classical linear CG method. Consequently, $\operatorname{grad} \widehat{m}_{x_k}(\eta_j) = r_j$ for all j. Choose K such that for all $k > K$, $\|\operatorname{grad} f(x_k)\| = \|\operatorname{grad} \widehat{m}_{x_k}(0)\|$ is so small—it converges to zero in view of (7.39)—that the stopping criterion (7.10) yields

$$\|\operatorname{grad} \widehat{m}_{x_k}(\eta_j)\| = \|r_j\| \leq \|r_0\|^{1+\theta} = \|\operatorname{grad} \widehat{m}_{x_k}(0)\|^{1+\theta}. \tag{7.49}$$

This is (7.44) with $c_2 = 1$. □

The constants in the proof of Theorem 7.4.11 can be chosen as $c_0 < \lambda_{min}$, $c_1 > \lambda_{\max}$, $c_4 > 1/\lambda_{\min}$, $c_5 > 1$, $c_3 \geq \beta_{L2}$, $c_3' \geq \beta_{\mathcal{H}}$, $c_2 \geq 1$, where $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and largest eigenvalue of $\operatorname{Hess} f(v)$, respectively. Consequently, the constant c in the convergence bound (7.38) can be chosen as

$$c > \frac{1}{\lambda_{\min}} \left(\left(\beta_{L2} / \lambda_{\min}^2 + \beta_{\mathcal{H}} / \lambda_{\min} \right) \lambda_{\max}^2 + \lambda_{\max}^{1+\theta} \right). \tag{7.50}$$

A nicer-looking bound holds when convergence is evaluated in terms of the norm of the gradient, as expressed in the theorem below which is a direct consequence of (7.45) and (7.46).

Theorem 7.4.12 *Under the assumptions of Theorem 7.4.11, if $\theta + 1 < 2$, then given $c_g > 1$ and $\{x_k\}$ generated by the algorithm, there exists $K > 0$ such that*

$$\|\operatorname{grad} f(x_{k+1})\| \leq c_g \|\operatorname{grad} f(x_k)\|^{\theta+1}$$

for all $k > K$.

Nevertheless, (7.50) suggests that the algorithm may not perform well when the relative gap $\lambda_{\max}/\lambda_{\min}$ is large. In spite of this, numerical experiments on eigenvalue problems have shown that the method tends to behave as well as, or even better than, other methods in the presence of a small relative gap.

7.4.3 Discussion

The main global convergence result (Theorem 7.4.4) shows that RTR-tCG method (Algorithm 10-11) converges to a set of critical points of the cost function for *all* initial conditions. This is an improvement on the pure Newton method (Algorithm 5), for which only local convergence results exist. However, the convergence theory falls short of showing that the algorithm always converges to a local minimizer. This is not surprising: since we have ruled out the possibility of checking the positive-definiteness of the Hessian of the cost function, we have no way of testing whether a critical point is

a local minimizer or not. (Note as an aside that even checking the positive-definiteness of the Hessian is not always sufficient for determining if a critical point is a local minimizer or not: if the Hessian is singular and nonnegative-definite, then no conclusion can be drawn.) In fact, for the vast majority of optimization methods, only convergence to critical points can be secured unless some specific assumptions (like convexity) are made. Nevertheless, it is observed in numerical experiments with random initial conditions that the algorithm systematically converges to a local minimizer; convergence to a saddle point is observed only on specifically crafted problems, e.g., when the iteration is started on a point that is a saddle point in computer arithmetic. This is due to the fact that the algorithm is a descent method, i.e., $f(x_{k+1}) < f(x_k)$ whenever $x_{k+1} \neq x_k$. Therefore, saddle points or local maxima are unstable fixed points of the algorithm.

There are cases where the bound (7.38) holds with order $\min\{\theta + 1, 3\}$; i.e., by choosing $\theta \geq 2$, one obtains cubic convergence. This situation occurs when the Taylor expansion of the cost function around the limit point has no third-order contribution. Thus, the second-order approximation used in the algorithm becomes an effective third-order approximation, and the order of convergence benefits as expected. In practice, this condition holds more often than one might guess since any cost function that is symmetric around the local minimizer v, i.e., $f(\mathrm{Exp}_x(\xi)) = f(\mathrm{Exp}_x(-\xi))$, will have only even contributions to its Taylor expansion.

7.5 APPLICATIONS

In this section, we briefly review the essential "ingredients" necessary for applying the RTR-tCG method (Algorithm 10-11), and we present two examples in detail as an illustration. One of these examples is optimization of the Rayleigh quotient, leading to an algorithm for computing extreme invariant subpaces of symmetric matrices. Since trust-region algorithms with an exact Hessian can be thought of as enhanced Newton methods, and since the Newton equation for Rayleigh quotient optimization is equivalent to the Jacobi equation (see Notes and References in Chapter 6), it is not surprising that this algorithm has close links with the celebrated Jacobi-Davidson approach to the eigenproblem. The Riemannian trust-region approach sheds new light on the Jacobi-Davidson method. In particular, it yields new ways to deal with the Jacobi equation so as to reduce the computational burden while preserving the superlinear convergence inherited from the Newton approach and obtaining strong global convergence results supported by a detailed analysis.

7.5.1 Checklist

The following elements are required for applying the RTR method to optimizing a cost function f on a Riemannian manifold $(\mathcal{M}, g)$: (i) a tractable

numerical representation for points x on $\mathcal{M}$, for tangent spaces $T_x\mathcal{M}$, and for the inner products $\langle \cdot, \cdot \rangle_x$ on $T_x\mathcal{M}$; (ii) a retraction $R_x : T_x\mathcal{M} \to \mathcal{M}$ (Definition 4.1.1); (iii) formulas for $f(x)$, $\operatorname{grad} f(x)$ and an approximate Hessian H_x that satisfies the properties required for the convergence results in Section 7.4.

Formulas for $\operatorname{grad} f(x)$ and $\operatorname{Hess} \widehat{f}_x(0_x)$ can be obtained by identification in a Taylor expansion of the lifted cost function $\widehat{f}_x$, namely

$$\widehat{f}_x(\eta) = f(x) + \langle \operatorname{grad} f(x), \eta \rangle_x + \tfrac{1}{2} \langle \operatorname{Hess} \widehat{f}_x(0_x)[\eta], \eta \rangle_x + O(\|\eta\|^3),$$

where $\operatorname{grad} f(x) \in T_x\mathcal{M}$ and $\operatorname{Hess} \widehat{f}_x(0_x)$ is a linear transformation of $T_x\mathcal{M}$. A formula for $\operatorname{Hess} \widehat{f}_x(0_x)$ is not needed, though; the convergence theory requires only an "approximate Hessian" H_x that satisfies the approximation condition (7.36). To obtain such an H_x, one can pick $H_x := \operatorname{Hess}(f \circ \widetilde{R}_x)(0_x)$, where $\widetilde{R}$ is any retraction. Then, assuming sufficient smoothness of f, the bound (7.36) follows from Lemmas 7.4.8 and 5.5.6. In particular, the choice $\widetilde{R}_x = \operatorname{Exp}_x$ yields

$$H_x := \nabla \operatorname{grad} f(x) \quad (= \operatorname{Hess} f(x)), \tag{7.51}$$

where ∇ denotes the Riemannian connection, and the model $\widehat{m}_x$ takes the form (7.1). If $\mathcal{M}$ is a Riemannian submanifold or a Riemannian quotient of a Euclidean space, then $\nabla \operatorname{grad} f(x)$ admits a simple formula; see Section 5.3.

7.5.2 Symmetric eigenvalue decomposition

Let $\mathcal{M}$ be the orthogonal group

$$\mathcal{M} = O_n = \{Q \in \mathbb{R}^{n \times n} : Q^T Q = I_n\}.$$

This manifold is an embedded submanifold of $\mathbb{R}^{n \times n}$ (Section 3.3). The tangent spaces are given by $T_Q O_n = \{Q\Omega : \Omega = -\Omega^T\}$ (Section 3.5.7). The canonical Euclidean metric $g(A, B) = \operatorname{tr}(A^T B)$ on $\mathbb{R}^{n \times n}$ induces on O_n the metric

$$\langle Q\Omega_1, Q\Omega_2 \rangle_Q = \operatorname{tr}(\Omega_1^T \Omega_2). \tag{7.52}$$

A retraction $R_Q : T_Q O_n \to O_n$ must be chosen that satisfies the properties stated in Section 7.2. Several possibilities are mentioned in Section 4.1.1.

Consider the cost function

$$f(Q) = \operatorname{tr}(Q^T A Q N),$$

where A and N are $n \times n$ symmetric matrices. For $N = \operatorname{diag}(\mu_1, \ldots, \mu_n)$, $\mu_1 < \cdots < \mu_n$, the minimum of f is realized by the orthonormal matrices of eigenvectors of A sorted in decreasing order of corresponding eigenvalue (this is a consequence of the critical points analysis in Section 4.8). Assume that the retraction R approximates the exponential at least to order 2. With the metric g defined as in (7.52), we obtain

$$\begin{aligned}
\widehat{f}_Q(Q\Omega) &:= f(R_Q(Q\Omega)) \\
&= \operatorname{tr}((I + \Omega + \tfrac{1}{2}\Omega^2 + O(\Omega^3))^T Q^T A Q (I + \Omega + \tfrac{1}{2}\Omega^2 + O(\Omega^3)) N) \\
&= f(Q) + 2\operatorname{tr}(\Omega^T Q^T A Q N) \\
&\quad + \operatorname{tr}(\Omega^T Q^T A Q \Omega N - \Omega^T \Omega Q^T A Q N) + O(\Omega^3),
\end{aligned}$$

from which it follows that

$$\mathrm{D}\widehat{f}_Q(0)[Q\Omega] = 2\,\mathrm{tr}(Q^TAQ\Omega N)$$

$$\tfrac{1}{2}\,\mathrm{D}^2\widehat{f}_Q(0)[Q\Omega_1, Q\Omega_2] = \mathrm{tr}(\Omega_1^TQ^TAQ\Omega_2N - \tfrac{1}{2}(\Omega_1^T\Omega_2 + \Omega_2^T\Omega_1)Q^TAQN)$$

$$\mathrm{grad}\,\widehat{f}_Q(0) = \mathrm{grad}\,f(Q) = Q[Q^TAQ, N]$$

$$\mathrm{Hess}\,\widehat{f}_Q(0)[Q\Omega] = \mathrm{Hess}\,f(Q)[Q\Omega] = \tfrac{1}{2}Q[[Q^TAQ, \Omega], N] + \tfrac{1}{2}Q[[N, \Omega], Q^TAQ],$$

where $[A, B] := AB - BA$. It is now straightforward to replace these expressions in the general formulation of Algorithm 10-11 and obtain a practical matrix algorithm.

An alternative way to obtain $\mathrm{Hess}\,\widehat{f}_Q(0)$ is to exploit Proposition 5.5.5, which yields $\mathrm{Hess}\,\widehat{f}_Q(0) = \nabla\,\mathrm{grad}\,f(Q)$. Since the manifold $\mathcal{M}$ is an embedded Riemannian submanifold of $\mathbb{R}^{n\times p}$, the covariant derivative ∇ is obtained by projecting the derivative in $\mathbb{R}^{n\times p}$ onto the tangent space to $\mathcal{M}$; see Section 5.3.3. We obtain $\mathrm{Hess}\,f(Q)[Q\Omega] = Q\,\mathrm{skew}(\Omega[Q^TAQ, N] + [\Omega^TQ^TAQ + Q^TAQ\Omega, N]$, which yields the same result as above.

All these ingredients can now be used in Algorithm 10-11 to obtain an iteration that satisfies the convergence properties proven in Section 7.4. Convergence to the critical points of the cost function means convergence to the matrices whose column vectors are the eigenvectors of A. Only the matrices containing eigenvectors in decreasing order of eigenvalue can be stable fixed points for the algorithm. They are asymptotically stable, with superlinear convergence, when all the eigenvalues are simple.

7.5.3 Computing an extreme eigenspace

Following up on the geometric Newton method for the generalized eigenvalue problem obtained in Section 6.4.3, we assume again that A and B are $n \times n$ symmetric matrices with B positive-definite, and we consider the generalized eigenvalue problem

$$Av = \lambda Bv.$$

We want to compute the leftmost p-dimensional invariant subspace of the pencil (A, B).

We consider the Rayleigh quotient function f defined by

$$f(\mathrm{span}(Y)) = \mathrm{tr}((Y^TAY)(Y^TBY)^{-1}), \tag{7.53}$$

where Y belongs to the set of full-rank $n \times p$ matrices and $\mathrm{span}(Y)$ denotes the column space of Y. The critical points of f are the invariant subspaces of the pencil (A, B), and the minimizers of f correspond to the leftmost invariant subspaces of (A, B).

In Section 6.4.3, we chose a noncanonical Riemannian metric (6.29) that yields a relatively short formula (6.34) for the Riemannian Hessian of f, obtained using the theory of Riemannian submersions (Proposition 5.3.4). In this section, we again use the definition

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

for the horizontal spaces, and we take

$$R_{\mathcal{Y}}(\xi) = \operatorname{span}(Y + \overline{\xi}_Y), \tag{7.54}$$

where $\mathcal{Y} = \operatorname{span}(Y)$ and $\overline{\xi}_Y$ stands for the horizontal lift of the tangent vector $\xi \in T_{\mathcal{Y}} \operatorname{Grass}(p, n)$. But now we define the Riemannian metric as

$$\langle \xi, \zeta \rangle_{\mathcal{Y}} = \operatorname{tr}\left((Y^T B Y)^{-1} \overline{\xi}_Y{}^T \overline{\zeta}_Y \right). \tag{7.55}$$

Notice that the horizontal space is not orthogonal to the vertical space with respect to the new Riemannian metric (7.55), so we have to renounce using the Riemannian submersion theory. However, we will see that with these choices for the horizontal space, the retraction, and the Riemannian metric, the second-order Taylor development of $\widehat{f}_x := f \circ R_x$ admits quite a simple form.

For the Rayleigh cost function (7.53), using the notation

$$P_{U,V} = I - U(V^T U)^{-1} V^T \tag{7.56}$$

for the projector parallel to the span of U onto the orthogonal complement of the span of V, we obtain

$$\begin{aligned}
&\widehat{f}_{\mathcal{Y}}(\xi) \\
&= f(R_{\mathcal{Y}}(\xi)) = \operatorname{tr}\left(\left((Y + \overline{\xi}_Y)^T B (Y + \overline{\xi}_Y) \right)^{-1} \left((Y + \overline{\xi}_Y)^T A (Y + \overline{\xi}_Y) \right) \right) \\
&= \operatorname{tr}\left((Y^T B Y)^{-1} Y^T A Y \right) + 2 \operatorname{tr}\left((Y^T B Y)^{-1} \overline{\xi}_Y{}^T A Y \right) \\
&\quad + \operatorname{tr}\left((Y^T B Y)^{-1} \overline{\xi}_Y{}^T \left(A \overline{\xi}_Y - B \overline{\xi}_Y (Y^T B Y)^{-1} (Y^T A Y) \right) \right) + O(\|\xi\|^3) \\
&= \operatorname{tr}\left((Y^T B Y)^{-1} Y^T A Y \right) + 2 \operatorname{tr}\left((Y^T B Y)^{-1} \overline{\xi}_Y{}^T P_{BY,BY} A Y \right) \\
&\quad + \operatorname{tr}\left((Y^T B Y)^{-1} \overline{\xi}_Y{}^T P_{BY,BY} \left(A \overline{\xi}_Y - B \overline{\xi}_Y (Y^T B Y)^{-1} (Y^T A Y) \right) \right) \\
&\quad + O(\|\xi\|^3),
\end{aligned} \tag{7.57}$$

where the introduction of the projectors does not modify the expression since $P_{BY,BY} \overline{\xi}_Y = \overline{\xi}_Y$. By identification, using the noncanonical metric (7.55), we obtain

$$\overline{\operatorname{grad} f(\mathcal{Y})}_Y = \overline{\operatorname{grad} \widehat{f}_{\mathcal{Y}}(0)}_Y = 2 P_{BY,BY} A Y \tag{7.58}$$

and

$$\overline{\operatorname{Hess} \widehat{f}_{\mathcal{Y}}(0_{\mathcal{Y}})[\xi]}_Y = 2 P_{BY,BY} \left(A \overline{\xi}_Y - B \overline{\xi}_Y (Y^T B Y)^{-1} (Y^T A Y) \right). \tag{7.59}$$

Notice that $\operatorname{Hess} \widehat{f}_{\mathcal{Y}}(0_{\mathcal{Y}})$ is symmetric with respect to the metric, as required.

We choose to take

$$H_{\mathcal{Y}} := \operatorname{Hess} \widehat{f}_{\mathcal{Y}}(0_{\mathcal{Y}}). \tag{7.60}$$

Consequently, the approximation condition (7.36) is trivially satisfied. The model (7.6) is thus

$$\begin{aligned}
\widehat{m}_{\mathcal{Y}}(\xi) &= f(\mathcal{Y}) + \langle \operatorname{grad} f(\mathcal{Y}), \xi \rangle_{\mathcal{Y}} + \tfrac{1}{2} \langle H_{\mathcal{Y}} \xi, \xi \rangle_{\mathcal{Y}} \\
&= \operatorname{tr}\left((Y^T B Y)^{-1} Y^T A Y \right) + 2 \operatorname{tr}\left((Y^T B Y)^{-1} \overline{\xi}_Y{}^T A Y \right) \\
&\quad + \operatorname{tr}\left((Y^T B Y)^{-1} \overline{\xi}_Y{}^T \left(A \overline{\xi}_Y - B \overline{\xi}_Y (Y^T B Y)^{-1} Y^T A Y \right) \right).
\end{aligned} \tag{7.61}$$

(Observe that for $B = I$ and $p = 1$, and choosing $y = Y$ of unit norm, the model (7.61) becomes $\widehat{m}_Y(\xi) = \mathrm{tr}(y^T A y) + 2\overline{\xi}_y{}^T A y + \overline{\xi}_y{}^T (A - y^T A y I)\overline{\xi}_y$, $y^T \overline{\xi}_y = 0$. Assuming that the Hessian is nonsingular, this model has a unique critical point that is the solution of the Newton equation (6.17) for the Rayleigh quotient on the sphere.)

Since the Rayleigh cost function (7.53) is smooth on $\mathrm{Grass}(p, n)$ and since $\mathrm{Grass}(p, n)$ is compact, it follows that all the assumptions involved in the convergence analysis of the general RTR-tCG algorithm (Section 7.4) are satisfied. The only complication is that we do not have a closed-form expression for the distance involved in the superlinear convergence result (7.38). But since B is fixed and positive-definite, the distances induced by the noncanonical metric (7.55) and by the canonical metric—(7.55) with $B := I$—are locally equivalent, and therefore for a given sequence both distances yield the same rate of convergence. (Saying that two distances dist_1 and dist_2 are *locally equivalent* means that given $x \in \mathcal{M}$, there is a neighborhood $\mathcal{U}$ of x and constants $c_1, c_2 > 0$ such that, for all y in $\mathcal{U}$, we have $c_1 \,\mathrm{dist}_1(x, y) \leq \mathrm{dist}_2(x, y) \leq c_2 \,\mathrm{dist}_1(x, y)$.

We have now all the required information to use the RTR-tCG method (Algorithm 10-11) for minimizing the Rayleigh cost function (7.53) on the Grassmann manifold $\mathrm{Grass}(p, n)$ endowed with the noncanonical metric (7.55). A matrix version of the inner iteration is displayed in Algorithm 12, where we omit the horizontal lift notation for conciseness and define

$$H_Y[Z] := P_{BY,BY}(AZ - BZ(Y^T BY)^{-1} Y^T AY). \tag{7.62}$$

Note that omission of the factor 2 in both the gradient and the Hessian does not affect the sequence $\{\eta\}$ generated by the truncated CG algorithm.

According to the retraction formula (7.54), the returned η_k yields a candidate new iterate

$$Y_{k+1} = (Y_k + \eta_k) M_k,$$

where M_k is chosen such that $Y_{k+1}^T B Y_{k+1} = I$. The candidate is accepted or rejected, and the trust-region radius is updated as prescribed in the outer RTR method (Algorithm 10), where ρ is computed using $\widehat{m}$ as in (7.61) and $\widehat{f}$ as in (7.57).

The resulting algorithm converges to the set of invariant subspaces of (A, B)—which are the critical points of the cost function (7.53)—and convergence to the leftmost invariant subspace $\mathcal{V}$ is expected to occur in practice since the other invariant subspaces are numerically unstable. Moreover, since $\mathcal{V}$ is a nondegenerate local minimum (under our assumption that $\lambda_p < \lambda_{p+1}$), it follows that the rate of convergence is $\min\{\theta + 1, 2\}$, where θ is the parameter appearing in the stopping criterion (7.10) of the inner (truncated CG) iteration.

Numerical experiments illustrating the convergence of the Riemannian trust-region algorithm for extreme invariant subspace computation are presented in Figures 7.1 and 7.2. The θ parameter of the inner stopping criterion (7.10) was chosen equal to 1 to obtain quadratic convergence (see Theorem 7.4.11). The right-hand plot shows a case with a small eigenvalue gap,

Algorithm 12 Truncated CG method for the generalized eigenvalue problem

Require: Symmetric $n \times n$ matrices A and B with B positive-definite; a B-orthonormal full-rank $n \times p$ matrix Y (i.e., $Y^T BY = I$); operator H_Y defined in (7.62); $\Delta > 0$.

Set $\eta^0 = 0 \in \mathbb{R}^{n \times p}$, $r_0 = P_{BY,BY} AY$, $\delta_0 = -r_0$; $j = 0$;
loop
  if $\mathrm{tr}\left(\delta_j^T H_Y[\delta_j]\right) \leq 0$ **then**
    Compute $\tau > 0$ such that $\eta = \eta^j + \tau\delta_j$ satisfies $\mathrm{tr}\left(\eta^T \eta\right) = \Delta$;
    return $\eta_k := \eta$;
  end if
  Set $\alpha_j = \mathrm{tr}\left(r_j^T r_j\right) / \mathrm{tr}\left(\delta_j^T H_Y[\delta_j]\right)$;
  Set $\eta^{j+1} = \eta^j + \alpha_j \delta_j$;
  if $\mathrm{tr}\left(\left(\eta^{j+1}\right)^T \eta^{j+1}\right) \geq \Delta$ **then**
    Compute $\tau \geq 0$ such that $\eta = \eta^j + \tau\delta_j$ satisfies $\mathrm{tr}\left(\eta^T \eta\right) = \Delta$;
    return $\eta_k := \eta$;
  end if
  Set $r_{j+1} = r_j + \alpha H_Y[\delta_j]$;
  Set $\beta_{j+1} = \mathrm{tr}\left(r_{j+1}^T r_{j+1}\right) / \mathrm{tr}\left(r_j^T r_j\right)$;
  Set $\delta_{j+1} = -r_{j+1} + \beta_{j+1}\delta_j$;
  if a stopping criterion is satisfied **then**
    return $\eta_k := \eta^j$;
  end if
end loop

which implies that the smallest eigenvalue of the Hessian of the cost function at the solution is much smaller than its largest eigenvalue. This suggests that the multiplicative constant in the superlinear convergence bound (7.38) is large, which explains why superlinear convergence sets in less clearly than on the left-hand plot featuring a large eigenvalue gap. An experiment with a smaller machine epsilon would reveal a quadratic convergence pattern; i.e., the number of digits of accuracy eventually approximately doubles at each new iterate. (All the experiments described in this book were performed with a machine epsilon of approximately $2 \cdot 10^{-16}$.)

The eigenvalue algorithm resulting from the proposed approach is surprisingly competitive with other eigenvalue methods in spite of the fact that it is just a brute-force application of a general optimization scheme that does not make any attempt to exploit the specific form of the Rayleigh quotient cost function. The efficiency of the algorithm is supported by its similarity to the Jacobi-Davidson eigenvalue method, in particular to the JDCG method of Notay. Nevertheless, the algorithm admits several enhancements, including subspace acceleration techniques and the possible monitoring of the ρ ratio within the inner iteration at a low computational cost. These enhancements, as well as comparisons with state-of-the-art eigenvalue methods, are

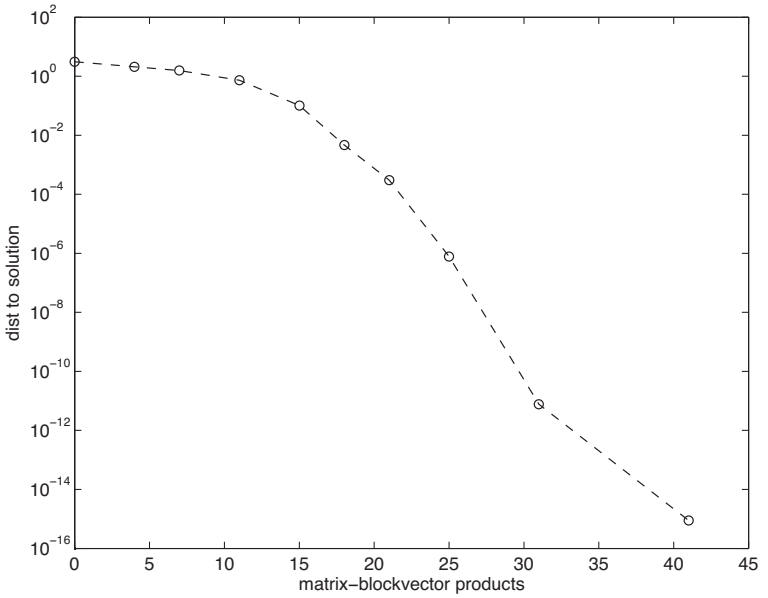

Figure 7.1 Numerical experiments on a trust-region algorithm for minimizing the Rayleigh cost function (7.53) on the Grassmann manifold $\mathrm{Grass}(p, n)$, with $n = 100$ and $p = 5$. $B = I$, and A is chosen with p eigenvalues evenly spaced on the interval $[1, 2]$ and the other $(n - p)$ eigenvalues evenly spaced on the interval $[10, 11]$; this is a problem with a large eigenvalue gap. The horizontal axis gives the number of multiplications of A times a block of p vectors. The vertical axis gives the distance to the solution, defined as the square root of the sum of the canonical angles between the current subspace and the leftmost p-dimensional invariant subspace of A. (This distance corresponds to the geodesic distance on the Grassmann manifold endowed with its canonical metric (3.44).)

presented in articles mentioned in Notes and References.

7.6 NOTES AND REFERENCES

The Riemannian trust-region approach was first proposed in [ABG04]. Most of the material in this section comes from [ABG07].

For more information on trust-region methods in $\mathbb{R}^n$, we refer the reader to Conn *et al.* [CGT00]. Trust-region methods are also discussed in textbooks on numerical optimization such as Nocedal and Wright [NW99]; see also Hei [Hei03], Gould *et al.* [GOST05], and Walmag and Delhez [WD05] for recent developments. Algorithm 10 reduces to [NW99, Alg. 4.1] in the classical $\mathbb{R}^n$ case; variants can be found in Conn *et al.* [CGT00, Ch. 10].

The method for computing an accurate solution of the trust-region sub-

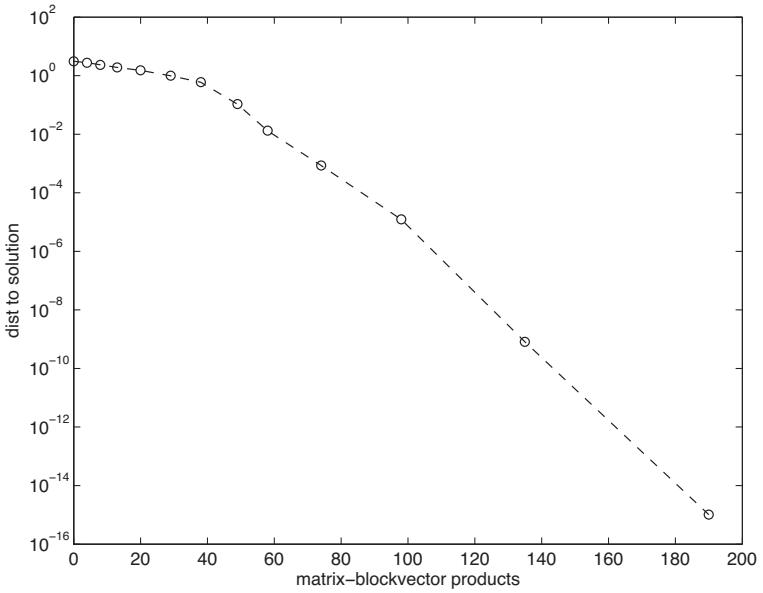

Figure 7.2 Same situation as in Figure 7.1 but now with $B = I$ and $A = \mathrm{diag}(1, \ldots, n)$.

problem is due to Moré and Sorensen [MS83]. Proposition 7.3.1 is a straightforward transcription of [CGT00, Th. 7.4.1], which itself generalizes results from [MS83] (or see [NW99, Th. 4.3]) to general norms.

The truncated CG method presented in Algorithm 11 closely follows the algorithm proposed by Steihaug [Ste83]; see also the work of Toint [Toi81]. Proposition 7.3.2 is due to Steihaug [Ste83, Th. 2.1]. The reader interested in the underlying principles of the Steihaug-Toint truncated CG method should refer to [Ste83], [NW99], or [CGT00].

Besides the truncated CG method, available algorithms for (approximately) solving trust-region subproblems include the dogleg method of Powell [Pow70], the double-dogleg method of Dennis and Mei [DM79], the method of Moré and Sorensen [MS83], the two-dimensional subspace minimization strategy of Byrd *et al.* [BSS88], the method based on the difference of convex functions proposed by Pham Dinh Tao and Le Thi Hoai An [TA98], the truncated Lanczos approach of Gould *et al.* [GLRT99], the matrix-free eigenproblem-based algorithm of Rojas *et al.* [RSS00], and the sequential subspace method of Hager [Hag01, HP05]. These and other methods are discussed in Conn *et al.* [CGT00, §7.5.4].

The classical global convergence results for trust-region methods in $\mathbb{R}^n$ can be found in Nocedal and Wright [NW99] (see in particular Theorem 4.8) and Conn *et al.* [CGT00]. A Taylor development similar to Lemma 7.4.7 can be found in Smith [Smi94]. The principle of the argument of Theorem 7.4.10 is closely related to the capture theorem, see Bertsekas [Ber95, Th 1.2.5]. A

discussion on cubic versus quadratic convergence can be found in Dehaene and Vandewalle [DV00]. A proof of Corollary 7.4.6 is given in [ABG06a].

Experiments, comparisons, and further developments are presented in [ABG06b, ABGS05, BAG06] for the Riemannian trust-region approach to extreme invariant subspace computation (Section 7.5.3). Reference [ABG06b] works out a brute-force application of the Riemannian trust-region method to the optimization of the Rayleigh quotient cost function on the sphere: comparisons are made with other eigenvalue methods, in particular the JDCG algorithm of Notay [Not02] and the Tracemin algorithm of Sameh, Wisniewski, and Tong [SW82, ST00], and numerical results are presented. Reference [ABGS05] proposes a two-phase method that combines the advantages of the (unshifted) Tracemin method and of the RTR-tCG method with and order-2 model; it allows one to make efficient use of a preconditioner in the first iterations by relaxing the trust-region constraint. The implicit RTR method proposed in Baker *et al.* [BAG06] makes use of the particular structure of the eigenvalue problem to monitor the value of the ratio ρ in the course of the inner iteration with little computational overhead, thereby avoiding the rejection of iterates because of poor model quality; for some problems, this technique considerably speeds up the iteration while the iterates are still far away from the solution, especially when a good preconditioner is available.