

# Trust-region methods on Riemannian manifolds

P.-A. Absil<sup>\*†</sup>      C. G. Baker<sup>\*</sup>      K. A. Gallivan<sup>\*†</sup>

## Abstract

A general scheme for trust-region methods on Riemannian manifolds is proposed and analyzed. Among the various approaches available to (approximately) solve the trust-region subproblems, particular attention is paid to the truncated conjugate-gradient technique. The method is illustrated on problems from numerical linear algebra.

**Key words.** Numerical optimization on manifolds, trust-region, truncated conjugate-gradient, Steihaug-Toint, global convergence, local convergence, superlinear, symmetric eigenvalue problem.

## 1 Introduction

Several problems related to numerical linear algebra can be expressed as optimizing a smooth function whose domain is a differentiable manifold. Applications include model reduction, principal component analysis, electronic structure computation and signal processing; see e.g. [LE00] and [HM94] for details. The simplest algorithms for solving optimization problems on manifolds are arguably those based on the idea of steepest descent; see e.g. [HM94, Udr94] and references therein. These algorithms have good global convergence properties but slow (linear) local convergence.

Other methods achieve superlinear convergence by using second-order information on the cost function. Among these methods, Newton's method is conceptually the simplest. The history of Newton's method on manifolds can

---

<sup>\*</sup>School of Computational Science, Florida State University, Tallahassee, FL 32306-4120, USA (<http://www.csit.fsu.edu/{~absil,~cbaker,~gallivan}>).

<sup>†</sup>These authors' work was supported by the USA National Science Foundation under Grant ACI0324944 and by the School of Computational Science of Florida State University through a postdoctoral fellowship. This work was initiated while the first author was a Research Fellow with the Belgian National Fund for Scientific Research (FNRS) at the University of Liège.

be traced back to Luenberger [Lue72], if not earlier. Gabay [Gab82] proposed a Newton method on embedded submanifolds of  $\mathbb{R}^n$ . Smith [Smi93, Smi94] formulated and analyzed the method on general Riemannian manifolds; an equivalent formulation was proposed by Udriște [Udr94]. Related work includes [Shu86, Mah96, EAS98, OW00, Man02, ADM<sup>+</sup>02, DPM03, HT04].

Newton’s method, without modifications, has major drawbacks as a numerical optimization method. The computational cost is often prohibitive, as a linear system has to be solved at each iteration. Moreover, the method is locally attracted to any stationary point, be it a local minimum, local maximum or saddle point. Finally, the method may not even converge to stationary points, unless some strong conditions are satisfied (such as convexity of the cost function).

In the case of cost functions on  $\mathbb{R}^n$ , several techniques exist to improve the convergence properties of Newton’s method. Most of these techniques fall into two categories: line-search methods and trust-region methods; see e.g. [MS84, NW99]. Line-search techniques have been considered on Riemannian manifolds by Udriște [Udr94] and Yang [Yan99]. The main purpose of this paper is to provide a theoretical and algorithmic framework for Riemannian trust-region methods, applicable to multiple problems.

The Riemannian trust-region approach we propose works along the following lines. First, much as in the work of Shub [Shu86, ADM<sup>+</sup>02], a *retraction*  $R$  is chosen on the Riemannian manifold  $M$  that defines for any point  $x \in M$  a one-to-one correspondence  $R_x$  between a neighborhood of  $x$  in  $M$  and a neighborhood of  $0_x$  in the tangent space  $T_x M$  (see Figure 1). Using this retraction, the cost function  $f$  on  $M$  is lifted to a cost function  $\hat{f}_x = f \circ R_x$  on  $T_x M$ . Since  $T_x M$  is an Euclidean space, it is possible to define a quadratic model of  $\hat{f}_x$  and adapt classical methods in  $\mathbb{R}^n$  to compute (in general, approximately) a minimizer of  $\hat{f}_x$  within a trust-region around  $0_x \in T_x M$ . This minimizer is then retracted back from  $T_x M$  to  $M$  using the retraction  $R_x$ . This point is a candidate for the new iterate, which will be accepted or rejected depending on the quality of the agreement between the quadratic model and the function  $f$  itself.

The advantages of a trust-region method over the pure Newton method are multiple. First, under mild conditions, trust-region schemes are provably convergent to a set of stationary points of the cost functions for all initial conditions. Moreover, the cost function is nonincreasing at each iterate which favors convergence to a local minimum. Finally, the presence of a trust-region gives an additional guideline to stop the inner iteration early, hence reducing the computational cost, while preserving the fast local convergence of the exact scheme.

The freedom in the choice of the retraction offers another advantage to the proposed Riemannian trust-region scheme. As in most other optimization algorithms on Riemannian manifolds, the proposed method first computes an update vector in the form of a tangent vector to the manifold at the current iterate. The common practice is then to use the Riemannian exponential mapping to select the next iterate from the update vector; see [Smi94, Udr94, EAS98, Yan99]. However, as pointed out by Manton [Man02, Section IX], the exponential may not be the most appropriate or computationally efficient way of performing the update. Our convergence analysis shows that the good properties of the algorithms hold for all suitably defined retractions (Definition 2.1) and not only for the exponential mapping.

We assume throughout that it is computationally impractical to determine whether the Hessian of the cost function is positive definite; trust-region subproblems are thus solved using inner iterations, such as the truncated conjugate-gradient method, that improve on the so-called Cauchy point by only using the Hessian of the model through its application to a vector. As a consequence, convergence of the trust-region algorithm to stationary points that are not local minima (i.e., saddle points and local maxima) cannot be ruled out. However, because trust-region methods are descent method (the value of the cost function never increases), the situation is fundamentally different from the pure Newton case: convergence to saddle points and local minima of the cost function is numerically unstable and is thus not expected to occur in practice; and indeed, convergence to saddle points and local minima is only observed on very specifically crafted numerical experiments. Moreover, we present a simple randomization technique that explicitly guarantees convergence to local minima with probability one.

The theory and algorithms can be adapted to exploit the properties of specific manifolds and problems in several disciplines. Numerical linear algebra considers several problems that can be analyzed and solved using this approach. A particularly illustrative and computationally efficient application is the computation of the rightmost or leftmost eigenvalue and associated eigenvector of a symmetric/positive-definite matrix pencil  $(A, B)$ . In this case, the manifold can be chosen as the projective space and a possible choice for the cost function is the Rayleigh quotient. The resulting trust-region algorithm can be interpreted as an inexact Rayleigh quotient iteration; we refer to [ABG04a] for details. In Section 5.3, we derive in detail a block generalization of the algorithm, which evolves on the Grassmann manifold. This algorithm is further developed in [ABGS04].

This paper makes use of basic notions of Riemannian geometry and nu-

merical optimization; background can be found in [dC92] and [NW99]. The general concept of trust-region methods on Riemannian manifolds is presented in Section 2. Methods for (approximately) solving the trust-region subproblems are considered in Section 3. Convergence properties are investigated in Section 4. The theory is illustrated on practical examples in Section 5. Conclusions are presented in Section 6.

A summary of the general Riemannian theory appeared in [ABG04c]. The present paper is an abridged and revised version of the technical report [ABG04b].

## 2 The general algorithm

We follow the usual conventions of matrix computations and view  $\mathbb{R}^n$  as the set of column vectors with  $n$  real components. 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 *the trust-region subproblem*

$$\min_{\eta \in \mathbb{R}^n} m(\eta) = f(x) + \partial f(x)\eta + \frac{1}{2}\eta^T \partial^2 f(x)\eta \quad \|\eta\| \leq \Delta \quad (1)$$

where  $\partial f = (\partial_1 f, \dots, \partial_n f)$  is the differential of  $f$ ,  $(\partial^2 f)_{ij} = \partial_{ij}^2 f$  is the Hessian matrix—some convergence results allow for  $\partial^2 f(x)$  in (1) to be replaced by any symmetric matrix, but we postpone this relaxation until later in the development—and  $\Delta$  is the trust-region radius. The quality of the model  $m$  is assessed by forming the quotient

$$\rho = \frac{f(x) - f(x + \eta)}{m(0) - m(\eta)}. \quad (2)$$

Depending on the value of  $\rho$ , the new iterate will be accepted or discarded and the trust-region radius  $\Delta$  will be updated. More details will be given later in this paper; or see e.g. [NW99, CGT00].

With a view towards extending the concept of trust-region subproblem to manifolds, we first consider the case of an abstract *Euclidean space*, i.e., a vector space endowed with an inner product (that is, a symmetric, bilinear, positive-definite form). This generalization to an Euclidean space  $E$  of dimension  $d$  requires little effort since  $E$  may be identified with  $\mathbb{R}^d$  once a basis of  $E$  is chosen (we refer to [Boo75, Section I.2] for a discussion on the distinction between  $\mathbb{R}^n$  and abstract Euclidean spaces). Let  $g(\cdot, \cdot)$  denote the inner product on  $E$ . Given a function  $f : E \rightarrow \mathbb{R}$  and a current iterate  $x \in E$ , one can choose a basis  $(e_i)_{i=1, \dots, d}$  of  $E$  (not necessarily orthonormal

with respect to the inner product) and write a classical  $G$ -norm trust-region subproblem (see e.g. [GLRT99, Section 2])

$$\min_{\bar{\eta} \in \mathbb{R}^d} m(\bar{\eta}) := \bar{f}(\bar{x}) + \partial \bar{f}(\bar{x}) \bar{\eta} + \frac{1}{2} \bar{\eta}^T \partial^2 \bar{f}(\bar{x}) \bar{\eta}, \quad \bar{\eta}^T G \bar{\eta} \leq \Delta_x^2 \quad (3)$$

where  $x = \sum_i \bar{x}_i e_i$ ,  $\eta = \sum_i \bar{\eta}_i e_i$ ,  $\bar{f}(\bar{x}) = f(\sum_i \bar{x}_i e_i)$  and  $G_{ij} = g(e_i, e_j)$ . It can be shown that  $m(\eta)$  does not depend on the choice of basis  $(e_i)_{i=1, \dots, d}$ ; therefore (3) can be written as a coordinate-free expression

$$\begin{aligned} \min_{\eta \in E} m(\eta) &= f(x) + Df(x)[\eta] + \frac{1}{2} D^2 f(x)[\eta, \eta] \\ &= f(x) + g(\text{grad } f(x), \eta) + \frac{1}{2} g(\text{Hess } f[\eta], \eta) \quad \text{s.t. } g(\eta, \eta) \leq \Delta_x^2 \end{aligned} \quad (4)$$

for the trust-region subproblem in the Euclidean space  $E$ .

Now let  $M$  be a *manifold* of dimension  $d$ . Intuitively, this means that  $M$  looks locally like  $\mathbb{R}^d$ . Local correspondences between  $M$  and  $\mathbb{R}^d$  are given by coordinate charts  $\phi_\alpha : \Omega_\alpha \subset M \rightarrow \mathbb{R}^n$ ; see e.g. [dC92] for details. How can we define a trust-region method for a cost function  $f$  on  $M$ ? Given a current iterate  $x$ , it is tempting to choose a coordinate neighborhood  $\Omega_\alpha$  containing  $x$ , translate the problem to  $\mathbb{R}^d$  through the chart  $\phi_\alpha$ , build a quadratic model  $m$ , solve the trust-region problem in  $\mathbb{R}^d$  and bring back the solution to  $M$  through  $\phi_\alpha^{-1}$ . The difficulty is that there are in general infinitely many  $\alpha$ 's such that  $x \in \Omega_\alpha$ . Each choice will yield a different model function  $m \circ \phi_\alpha$  and a different the trust region  $\{y \in M : \|\phi_\alpha(y)\| \leq \Delta\}$ , hence a different next iterate  $x_+$ . This kind of situation is pervasive in numerics on manifolds; it is usually addressed, assuming that  $M$  is a Riemannian manifold, by working in so-called *normal coordinates*.

In order to explain the concept of normal coordinates, we now present a condensed overview of Riemannian geometric concepts; we refer to [dC92, Sak96] for details. In what follows,  $M$  will be a  $(C^\infty)$  Riemannian manifold, i.e.,  $M$  is endowed with a correspondence, called Riemannian metric, which associates to each point  $x$  of  $M$  an inner product  $g_x(\cdot, \cdot)$  on the tangent space  $T_x M$  and which varies differentiably. The Riemannian metric induces a norm  $\|\xi\| = \sqrt{g_x(\xi, \xi)}$  on the tangent spaces  $T_x M$ . Also associated with a Riemannian manifold are the notions of Levi-Civita (or Riemannian) connection  $\nabla$ , parallel transport, geodesic (which, intuitively, generalizes the notion of straight line) and associated exponential map defined by  $\text{Exp}_x \xi = \gamma(1)$  where  $\gamma$  is the geodesic satisfying  $\gamma(0) = x$  and  $\dot{\gamma}(0) = \xi$ . Given a point

$x$  in  $M$ , there is a ball  $B_\epsilon(0_x)$  in  $T_x M$  of radius  $\epsilon$  around the origin  $0_x$  of  $T_x M$  such that  $\text{Exp}_x$  is a diffeomorphism of  $B_\epsilon(0_x)$  onto an open subset of  $M$ . Then  $\text{Exp}_x(B_\epsilon(0_x)) = U$  is called a *normal neighborhood* of  $x$ , and the  $\text{Exp}_x$  defines a diffeomorphism between the Euclidean space  $T_x M$  and  $U$ . The supremum of these  $\epsilon$ 's is the *injectivity radius*  $i_x(M)$  at  $x$ , and  $i(M) := \inf_{x \in M} i_x$  is the *injectivity radius* of  $M$ . Finally, normal coordinates are defined in a normal neighborhood  $U$  by considering an orthonormal basis  $\{e_i\}$  of  $T_x M$  and taking  $(u_1, \dots, u_d)$ ,  $y = \text{Exp}_x(\sum_{i=1}^n u_i e_i)$  as coordinates of  $y$ .

For the purpose of defining a trust-region method, the choice of a basis  $\{e_i\}$  in  $T_x M$  is indifferent, since trust-region subproblems on a Euclidean space like  $T_x M$  admit a coordinate-free expression (4). Therefore, the exponential mapping makes it possible to uniquely define trust-region subproblems on Riemannian manifolds by locally mapping the manifold to the Euclidean space  $T_x M$ .

However, as pointed out in [Man02], the systematic use of the exponential mapping is questionable: other local mappings to  $T_x M$  may reduce the computational cost while preserving the useful convergence properties of the considered method. Therefore, in this paper, we relax the exponential to a class of mappings called *retractions*, a concept that we borrow from [Shu86, ADM<sup>+</sup>02] with some modifications (see also the illustration on Figure 1).

**Definition 2.1 (retraction)** *A retraction on a manifold  $M$  is a mapping  $R$  on the tangent bundle  $TM$  into  $M$  with the following properties. Let  $R_x$  denote the restriction of  $R$  to  $T_x M$ .*

1.  *$R$  is continuously differentiable.*
2.  *$R_x(\xi) = x$  if and only if  $\xi = 0_x$ , the zero element of  $T_x M$ .*
3.  *$\text{DR}_x(0_x) = \text{id}_{T_x M}$ , the identity mapping on  $T_x M$ , with the canonical identification  $T_{0_x} T_x M \simeq T_x M$ .*

It follows from the inverse function theorem (see [dC92, Ch. 0, Theorem 2.10]) that  $R_x$  is a local diffeomorphism at  $0_x$ , namely,  $R_x$  is not only  $C^1$  but also bijective with differentiable inverse on a neighborhood  $V$  of  $0_x$  in  $T_x M$ . In particular, the exponential mapping is a retraction (see Proposition 2.9 in [dC92, Ch. 3] and the proof thereof), and any other retraction can be thought of as a first-order approximation of the exponential mapping. Practical examples of retractions on specific Riemannian manifolds, that may be more tractable computationally than the exponential, are given in Section 5. We point out that the requirements in Definition 2.1 are stronger than needed to obtain the convergence results; in particular, we could allow  $R$  to

be defined only in a certain subset of  $TM$ . However, weaker assumptions would make the forthcoming developments more complicated, and there is no evidence that they would be more relevant in practical applications. For the same reason, we assume throughout that the manifold  $M$  is complete, i.e.,  $\text{Exp } \xi$  exists for all  $\xi$  in  $TM$ .

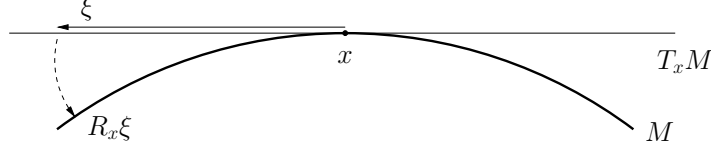


Figure 1: Illustration of retractions.

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

$$\hat{f}_{x_k} : T_{x_k} M \rightarrow \mathbb{R} : \xi \mapsto f(R_{x_k} \xi) \quad (5)$$

The Riemannian metric  $g$  turns  $T_{x_k} M$  into a Euclidean space endowed with the inner product  $g_{x_k}(\cdot, \cdot)$ , and, following (4), the TR subproblem on  $T_{x_k} M$  reads

$$\begin{aligned} \min_{\eta \in T_{x_k} M} m_{x_k}(\eta) &= \hat{f}_{x_k}(0_{x_k}) + D\hat{f}_{x_k}(0_{x_k})[\eta] + \frac{1}{2} D^2 \hat{f}_{x_k}(0_{x_k})[\eta, \eta] \\ &= \hat{f}_{x_k}(0_{x_k}) + g_{x_k}(\text{grad } \hat{f}_{x_k}(0_{x_k}), \eta) + \frac{1}{2} g_{x_k}(\text{Hess } \hat{f}_{x_k}(0_{x_k})[\eta], \eta) \quad \text{s.t. } g_{x_k}(\eta, \eta) \leq \Delta_k^2. \end{aligned} \quad (6)$$

For the global convergence theory it is only required that the second-order term in the model be some symmetric form. Therefore, instead of (6), we will consider the following more general formulation

$$\min_{\eta \in T_{x_k} M} m_{x_k}(\eta) = f(x_k) + g_{x_k}(\text{grad } f(x_k), \eta) + \frac{1}{2} g_{x_k}(\mathcal{H}_{x_k} \eta, \eta) \quad \text{s.t. } g_{x_k}(\eta, \eta) \leq \Delta_k^2, \quad (7)$$

where  $\mathcal{H}_{x_k} : T_{x_k} M \rightarrow T_{x_k} M$  is some symmetric linear operator, i.e.,  $g_{x_k}(\mathcal{H}_{x_k} \xi, \chi) = g_{x_k}(\xi, \mathcal{H}_{x_k} \chi)$ ,  $\xi, \chi \in T_{x_k} M$ . This is called the *trust-region subproblem*.

Next, an (approximate) solution  $\eta_k$  of the Euclidean trust-region subproblem (7) is computed using any available method, referred to as *inner*

iteration (see Section 3). The candidate for the new iterate is then given by

$$x_+ = R_{x_k}(\eta_k). \quad (8)$$

The decision to accept or not the candidate and to update the trust-region radius is based on the quotient

$$\rho_k = \frac{f(x_k) - f(R_{x_k}(\eta_k))}{m_{x_k}(0_{x_k}) - m_{x_k}(\eta_k)} = \frac{\hat{f}_{x_k}(0_{x_k}) - \hat{f}_{x_k}(\eta_k)}{m_{x_k}(0_{x_k}) - m_{x_k}(\eta_k)}. \quad (9)$$

If  $\rho_k$  is exceedingly small, then the model is very inaccurate: the step must be rejected and the trust-region radius must be reduced. If  $\rho_k$  is small but less dramatically so, then the step is accepted but the trust-region radius is reduced. If  $\rho_k$  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.

This procedure can be formalized as the following algorithm; it reduces to [NW99, Alg. 4.1] in the classical  $\mathbb{R}^n$  case (see [CGT00, Ch. 10] for variants).

**Algorithm 1 (RTR – basic Riemannian Trust-Region algorithm)** *Data:*

*Complete Riemannian manifold  $(M, g)$ ; scalar field  $f$  on  $M$ ; retraction  $R$  from  $TM$  to  $M$  as in Definition 2.1.*

*Parameters:  $\bar{\Delta} > 0$ ,  $\Delta_0 \in (0, \bar{\Delta})$ , and  $\rho' \in [0, \frac{1}{4})$ .*

*Input: initial iterate  $x_0 \in M$ .*

*Output: sequence of iterates  $\{x_k\}$ .*

**for**  $k = 0, 1, 2, \dots$

*Obtain  $\eta_k$  by (approximately) solving (7);*

*Evaluate  $\rho_k$  from (9);*

**if**  $\rho_k < \frac{1}{4}$

$\Delta_{k+1} = \frac{1}{4}\Delta_k$

**else if**  $\rho_k > \frac{3}{4}$  and  $\|\eta_k\| = \Delta_k$

$\Delta_{k+1} = \min(2\Delta_k, \bar{\Delta})$

**else**

$\Delta_{k+1} = \Delta_k$ ;

**if**  $\rho_k > \rho'$

$x_{k+1} = R_{x_k}\eta_k$

**else**

$x_{k+1} = x_k$ ;

**end (for).**

In the sequel we will sometimes drop the subscript “ $k$ ” and denote  $x_{k+1}$  by  $x_+$ .



## 2.1 Discussion

The concept of retraction (Definition 2.1) is closely related to the local parameterizations around points introduced in [HT04]. Let  $e_i$ ,  $i = 1, \dots, d$  be smooth vector fields such that  $\{e_i(x)\}_{i=1, \dots, d}$  is an orthogonal basis of  $T_x M$ , and denote by  $\psi_x$  the mapping that sends  $\xi \in T_x M$  to  $(u_1, \dots, u_d)$  such that  $\xi = \sum_i u_i e_i$ . If  $R$  is a smooth retraction, then the mappings  $(u_1, \dots, u_d) \mapsto R_x(\sum u_i e_i)$  define a smooth family of parameterizations. Conversely, if  $\{\mu_x\}_{x \in M}$  is a smooth family of parameterizations, then the mapping  $R_x = \mu_x \circ \psi_x \circ (D(\mu_x \circ \psi_x)(0_x))^{-1}$  is a retraction.

Since  $DR_x(0_x) = \text{id}_{T_x M}$ , it follows that  $\text{grad } \hat{f}_{x_k}(0_x) = \text{grad } f(x)$ , where  $\text{grad } f(x)$ , the gradient of  $f$  at  $x$ , is defined by  $g_x(\text{grad } f(x), \xi) = df_x(\xi)$ ,  $\xi \in T_x M$  (see [dC92, Ch. 3, Ex. 8]). Moreover, we point out that if  $R$  satisfies some second order condition given in Lemma 4.11 page 23, then it holds that  $\text{Hess } \hat{f}_{x_k}(0_x) = \text{Hess } f(x)$ , where  $\text{Hess } f(x) : T_x M \rightarrow T_x M$ , the Hessian (linear) operator, is defined by

$$\text{Hess } f(x)\xi = \nabla_\xi \text{grad } f(x), \quad \xi \in T_x M, \quad (10)$$

see [dC92, Ch. 6, Ex. 11]. The relation  $\text{Hess } \hat{f}_x(0_x) = \text{Hess } f(x)$  is useful in the case of embedded submanifolds; then the Levi-Civita connection  $\nabla$  reduces to a directional derivative in the embedding space followed by a projection onto the tangent space to the manifold, which facilitates the derivation of a formula for  $\text{Hess } f(x)\xi$ .

The Hessian operator  $\xi \in T_x M \mapsto \nabla_\xi \text{grad } f(x) \in T_x M$  defined in (10) is related to the second tensorial derivative  $D^2 f(x)$  by  $D^2 f(\xi, \chi) = g_x(\text{Hess } f(x)\xi, \chi) = \xi\chi f - (\nabla_\xi \chi)f$ , where  $\xi f$  is the usual notation for the directional derivative of the function  $f$  in the direction of the tangent vector  $\xi$ ; see, e.g., [Sak96]. Note that in [Sak96] and in some other references, the word ‘‘Hessian’’ refers to  $D^2 f$ , not to the linear operator (10).

In general, there is no assumption on the operator  $\mathcal{H}_x$  in (7) other than being a symmetric linear operator. Consequently, even though  $m_x$  was initially presented as a model of  $f \circ R_x$ , the choice of the retraction  $R_x$  does not impose any constraint on  $m_x$ . In order to achieve superlinear convergence, however,  $\mathcal{H}_{x_k}$  will be required to be an ‘‘approximate’’ Hessian (Theorem 4.13). Obtaining an appropriate approximate Hessian in practice is addressed in Section 5.1. A possible way of choosing  $\mathcal{H}_x$  is to define  $m_x$  as the quadratic model of  $f \circ \tilde{R}_x$ , where  $\tilde{R}_x$  is a retraction, not necessarily equal to  $R_x$ ; a similar point of view was adopted in [HT04] in the framework of Newton’s method.

We conclude this section by pointing out more explicitly the link between Algorithm 1 and the Riemannian Newton method. Assume that  $\mathcal{H}_{x_k}$  in (7) is the exact Hessian of  $f$  at  $x_k$ , and assume that the exact solution  $\eta^*$  of the trust-region subproblem (7) lies in the interior of the trust region. Then  $\eta^*$  satisfies

$$\text{grad } f + \nabla_{\eta^*} \text{grad } f = 0,$$

which is the Riemannian Newton equation of Smith [Smi93, Smi94] and Udriște [Udr94, Ch. 7, §5]. Note that both authors propose to apply the update vector  $\eta^*$  using the Riemannian exponential retraction; namely, the new iterate is defined as  $x_+ = \text{Exp}_x \eta^*$ . As shown by Smith [Smi93, Smi94], the Riemannian Newton algorithm converges locally quadratically to the nondegenerate stationary points of  $f$ . A cubic rate of convergence is even observed in frequently encountered cases where some symmetry condition holds [AMS04]. We will see in Section 4 that the superlinear convergence property of Newton’s method is preserved by the trust-region modification, while the global convergence properties are improved: the accumulation points are guaranteed to be stationary points regardless of the initial conditions, and among the stationary points only the local minima can be local attractors.

### 3 Computing a trust-region step

We have seen in Section 2 that the use of retractions yields trust-region subproblems expressed in Euclidean spaces  $T_x M$ . Therefore, all the classical methods for solving the trust-region subproblem can be applied.

As mentioned in the introduction, it is assumed here that for some reason, usually related to the large size of the problem under consideration or to the computational efficiency required to outperform alternative methods, it is impractical to check positive-definiteness of  $\mathcal{H}_{x_k}$ ; rather,  $\mathcal{H}_{x_k}$  is only available via its application to a vector. The *truncated conjugate-gradient method* of Steihaug [Ste83] and Toint [Toi81] is particularly appropriate in these circumstances. The following algorithm is a straightforward adaptation of the method of [Ste83] to the trust-region subproblem (7). This algorithm is an *inner iteration* as it is used within the RTR framework (Algorithm 1) to compute an approximate solution of the trust-region subproblems. Note that we use indices in superscript to denote the evolution of  $\eta$  within the inner iteration, while subscripts are used in the outer iteration.

**Algorithm 2 (tCG – truncated CG for the TR subproblem)** Set  $\eta^0 = 0$ ,  $r_0 = \text{grad } f(x_k)$ ,  $\delta_0 = -r_0$ ;  
**for**  $j = 0, 1, 2, \dots$  *until a stopping criterion is satisfied, perform the iteration:*  
    **if**  $g_{x_k}(\delta_j, \mathcal{H}_{x_k} \delta_j) \leq 0$   
        Compute  $\tau$  such that  $\eta = \eta^j + \tau \delta_j$  minimizes  $m_{x_k}(\eta)$  in (7)  
        and satisfies  $\|\eta\|_{g_x} = \Delta$ ;  
        **return**  $\eta$ ;  
    Set  $\alpha_j = g_{x_k}(r_j, r_j) / g_{x_k}(\delta_j, \mathcal{H}_{x_k} \delta_j)$ ;  
    Set  $\eta^{j+1} = \eta^j + \alpha_j \delta_j$ ;  
    **if**  $\|\eta^{j+1}\|_{g_x} \geq \Delta$   
        Compute  $\tau \geq 0$  such that  $\eta = \eta^j + \tau \delta_j$  satisfies  $\|\eta\|_{g_x} = \Delta$ ;  
        **return**  $\eta$ ;  
    Set  $r_{j+1} = r_j + \alpha_j \mathcal{H}_{x_k} \delta_j$ ;  
    Set  $\beta_{j+1} = g_{x_k}(r_{j+1}, r_{j+1}) / g_{x_k}(r_j, r_j)$ ;  
    Set  $\delta_{j+1} = -r_{j+1} + \beta_{j+1} \delta_j$ ;  
**end (for).**

Several comments about Algorithm 2 are in order.

The simplest stopping criterion for Algorithm 2 is to truncate after a fixed number of iterations. In order to improve the convergence rate, a possibility is to stop as soon as an iteration  $j$  is reached for which

$$\|r_j\| \leq \|r_0\| \min(\|r_0\|^\theta, \kappa). \quad (11)$$

Concerning the computation of  $\tau$ , it can be shown that when  $g(\delta_j, \mathcal{H}_{x_k} \delta_j) \leq 0$ ,  $\arg \min_{\tau \in \mathbb{R}} m_k(\eta^j + \tau \delta_j)$  is equal to the positive root of  $\|\eta^j + \tau \delta_k\|_{g_x} = \Delta$ , which is explicitly given by

$$\frac{-g_x(\eta^j, \delta_j) + \sqrt{g_x(\eta^j, \delta_j)^2 - (\Delta^2 - g_x(\eta^j, \eta^j))g_x(\delta_j, \delta_j)}}{g_x(\delta_j, \delta_j)}.$$

Notice that tCG algorithm only requires the following:

- An evaluation of  $\text{grad } f(x)$ .
- A routine that performs line minimizations for the model  $m$ .
- A routine that returns  $\mathcal{H}_{x_k} \delta$  given  $\delta \in T_x M$ .

The algorithm can thus be considered as “inverse-free”. The reader interested in the underlying principles of the Steihaug-Toint truncated CG method should refer to [Ste83], [NW99] or [CGT00].

Alternatives to tCG for approximately solving trust-region subproblems are mentioned in [CGT00, Section 7.5.4]; see also [Hag01, HP04].

## 4 Convergence analysis

In this section, we first study the global convergence properties of the RTR scheme (Algorithm 1), without any assumption on the way the trust-region subproblems (7) are solved, except that the approximate solution  $\eta_k$  must produce a decrease of the model that is at least a fixed fraction of the so-called Cauchy decrease. Under mild additional assumptions on the retraction and the cost function, it is shown that the sequences  $\{x_k\}$  produced by Algorithm 1 converge to the set of stationary 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  $\hat{f}_{x_k}$  is considered at each iterate  $x_k$ .

We then analyze the local convergence of Algorithm 1-2 around nondegenerate local minima. Algorithm 1-2 refers to the RTR framework where the trust-region subproblems are approximately solved using the tCG algorithm with stopping criterion (11). It is shown that the iterates of the algorithm converge to nondegenerate stationary points with an order of convergence  $\min\{\theta + 1, 2\}$  (at least).

### 4.1 Global convergence

The objective of this section is to show that, under appropriate assumptions, the sequence  $\{x_k\}$  generated by Algorithm 1 satisfies  $\lim_{k \rightarrow \infty} \|\text{grad } f(x_k)\| = 0$ ; this generalizes a classical convergence property of trust-region methods in  $\mathbb{R}^n$ , see [NW99, Theorem 4.8].

In what follows,  $(M, g)$  is a complete Riemannian manifold of dimension  $d$ , and  $R$  is a retraction on  $M$  (Definition 2.1). We define

$$\hat{f} : TM \mapsto \mathbb{R} : \xi \mapsto f(R\xi) \quad (12)$$

and, in accordance with (5),  $\hat{f}_x$  denotes the restriction of  $\hat{f}$  to  $T_x M$ . We denote by  $B_\delta(0_x) = \{\xi \in T_x M : \|\xi\| < \delta\}$  the open ball in  $T_x M$  of radius  $\delta$  centered at  $0_x$ , and  $B_\delta(x)$  stands for the set  $\{y \in M : \text{dist}(x, y) < \delta\}$  where  $\text{dist}$  denotes the Riemannian distance. We denote by  $P_\gamma^{t \leftarrow t_0} v$  the vector of  $T_{\gamma(t)} M$  obtained by parallel transporting the vector  $v \in T_{\gamma(t_0)} M$  along a curve  $\gamma$ .

As in the classical  $\mathbb{R}^n$  case (see [NW99, Thm 4.7] or [CGT00, Thm 6.4.5]), we first show that at least one accumulation point of  $\{x_k\}$  is stationary. The convergence result requires that  $m_{x_k}(\eta_k)$  be a sufficiently good approximation of  $\hat{f}_{x_k}(\eta_k)$ . In [CGT00, Thm 6.4.5] this is guaranteed by the assumption

that the Hessian of the cost function is bounded. It is however possible to weaken this assumption<sup>1</sup>, which leads us to consider the following definition.

**Definition 4.1 (radially L- $C^1$  function)** *Let  $\hat{f} : TM \rightarrow \mathbb{R}$  be as in (12). We say that  $\hat{f}$  is radially Lipschitz continuously differentiable if there exist reals  $\beta_{RL} > 0$  and  $\delta_{RL} > 0$  such that, for all  $x \in M$ , for all  $\xi \in TM$  with  $\|\xi\| = 1$ , and for all  $t < \delta_{RL}$ , it holds*

$$\left| \frac{d}{d\tau} \hat{f}_x(\tau\xi)|_{\tau=t} - \frac{d}{d\tau} \hat{f}_x(\tau\xi)|_{\tau=0} \right| \leq \beta_{RL} t. \quad (13)$$

For the purpose of Algorithm 1, which is a descent algorithm, this condition needs only to be imposed for all  $x, y$  in the level set

$$\{x \in M : f(x) \leq f(x_0)\}. \quad (14)$$

A key assumption in the classical global convergence result in  $\mathbb{R}^n$  is that the approximate solution  $\eta_k$  of the trust-region subproblem (7) produces at least as much decrease in the model function as a fixed fraction of the Cauchy decrease; see [NW99, Section 4.3]. Since the trust-region subproblem (7) is expressed on a Euclidean space, the definition of the Cauchy point is adapted from  $\mathbb{R}^n$  without difficulty, and the bound

$$m_k(0) - m_k(\eta_k) \geq c_1 \|\text{grad} f(x_k)\| \min \left( \Delta_k, \frac{\|\text{grad} f(x_k)\|}{\|\mathcal{H}_k\|} \right), \quad (15)$$

for some constant  $c_1 > 0$ , is readily obtained from the  $\mathbb{R}^n$  case, where  $\|\mathcal{H}_k\|$  is defined as

$$\|\mathcal{H}_k\| := \sup\{\|\mathcal{H}_k \zeta\| : \zeta \in T_{x_k} M, \|\zeta\| = 1\}. \quad (16)$$

In particular, the truncated CG method (Algorithm 2) satisfies this bound (with  $c_1 = \frac{1}{2}$ , see [NW99, Lemma 4.5]) since it first computes the Cauchy point and then attempts to improve the model decrease.

With these things 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 4.5.

---

<sup>1</sup>It seems that  $f \in C^1$  is not enough: there is apparently a gap in the proof of [NW99, Thm 4.7].

**Theorem 4.2** *Let  $\{x_k\}$  be a sequence of iterates generated by Algorithm 1 with  $\rho' \in [0, \frac{1}{4})$ . Suppose that  $f$  is  $C^1$  and bounded below on the level set (14), that  $\hat{f}$  is radially  $L$ - $C^1$  (Definition 4.1), and that  $\|\mathcal{H}_k\| \leq \beta$  for some constant  $\beta$ . Further suppose that all approximate solutions  $\eta_k$  of (7) satisfy the Cauchy decrease inequality (15) for some positive constant  $c_1$ . We then have*

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

*Proof.*

First, we perform some manipulation of  $\rho_k$  from (9). Notice that

$$\begin{aligned} |\rho_k - 1| &= \left| \frac{(f(x_k) - \hat{f}_{x_k}(\eta_k)) - (m_k(0) - m_k(\eta_k))}{m_k(0) - m_k(\eta_k)} \right| \\ &= \left| \frac{m_k(\eta_k) - \hat{f}_{x_k}(\eta_k)}{m_k(0) - m_k(\eta_k)} \right|. \end{aligned} \quad (17)$$

Direct manipulations on the function  $t \mapsto \hat{f}_{x_k}(t \frac{\eta_k}{\|\eta_k\|})$  yield

$$\begin{aligned} \hat{f}_{x_k}(\eta_k) &= \hat{f}_{x_k}(0_{x_k}) + \|\eta_k\| \frac{d}{d\tau} \hat{f}_{x_k}(\tau \frac{\eta_k}{\|\eta_k\|})|_{\tau=0} \\ &\quad + \int_0^{\|\eta_k\|} \left( \frac{d}{d\tau} \hat{f}_{x_k}(\tau \frac{\eta_k}{\|\eta_k\|})|_{\tau=t} - \frac{d}{d\tau} \hat{f}_{x_k}(\tau \frac{\eta_k}{\|\eta_k\|})|_{\tau=0} \right) dt \\ &= f(x_k) + g_{x_k}(\text{grad } f(x_k), \eta_k) + \epsilon' \end{aligned}$$

where  $|\epsilon'| < \int_0^{\|\eta_k\|} \beta_{RL} t \, dt = \frac{1}{2} \beta_{RL} \|\eta_k\|^2$  whenever  $\|\eta_k\| < \delta_{RL}$ , and  $\beta_{RL}$  and  $\delta_{RL}$  are the constants in the radially  $L$ - $C^1$  property (13).

Therefore, it follows from the definition (7) of  $m_k$  that

$$\begin{aligned} |m_k(\eta_k) - \hat{f}_{x_k}(\eta_k)| &= \left| \frac{1}{2} g_{x_k}(\mathcal{H}_{x_k} \eta_k, \eta_k) - \epsilon' \right| \\ &\leq \frac{1}{2} \beta \|\eta_k\|^2 + \frac{1}{2} \beta_{RL} \|\eta_k\|^2 \leq \beta' \|\eta_k\|^2 \end{aligned} \quad (18)$$

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

Assume for purpose of contradiction that the theorem does not hold; that is, assume there exist  $\epsilon > 0$  and a positive index  $K$  such that

$$\|\text{grad } f(x_k)\| \geq \epsilon, \quad \forall k \geq K. \quad (19)$$

From (15), for  $k \geq K$ , we have

$$m_k(0) - m_k(\eta_k) \geq c_1 \|\text{grad} f(x_k)\| \min \left( \Delta_k, \frac{\|\text{grad} f(x_k)\|}{\|\mathcal{H}_k\|} \right) \geq c_1 \epsilon \min \left( \Delta_k, \frac{\epsilon}{\beta'} \right). \quad (20)$$

Substituting (18), and (20) into (17), we have that

$$|\rho_k - 1| \leq \frac{\beta' \|\eta_k\|^2}{c_1 \epsilon \min \left( \Delta_k, \frac{\epsilon}{\beta'} \right)} \leq \frac{\beta' \Delta_k^2}{c_1 \epsilon \min \left( \Delta_k, \frac{\epsilon}{\beta'} \right)} \quad (21)$$

whenever  $\|\eta_k\| < \delta_{RL}$ .

We can choose a value of  $\hat{\Delta}$  that allows us to bound the right-hand-side of the inequality (21), when  $\Delta_k \leq \hat{\Delta}$ . Choose  $\hat{\Delta}$  as follows:

$$\hat{\Delta} \leq \min \left( \frac{c_1 \epsilon}{2\beta'}, \frac{\epsilon}{\beta'}, \delta_{RL} \right).$$

This gives us  $\min \left( \Delta_k, \frac{\epsilon}{\beta'} \right) = \Delta_k$ . We can now write (21) as follows:

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

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

Therefore, we conclude that

$$\Delta_k \geq \min \left( \Delta_K, \hat{\Delta}/4 \right), \quad \forall k \geq K. \quad (22)$$

Suppose now that there is an infinite subsequence  $\mathcal{K}$  such that  $\rho_k \geq \frac{1}{4} > \rho'$  for  $k \in \mathcal{K}$  and  $k \geq K$ , we have from (20) that

$$\begin{aligned} f(x_k) - f(x_{k+1}) &= f_{x_k} - \hat{f}_{x_k}(\eta_k) \\ &\geq \frac{1}{4} (m_k(0) - m_k(\eta_k)) \\ &\geq \frac{1}{4} c_1 \epsilon \min \left( \Delta_k, \frac{\epsilon}{\beta'} \right). \end{aligned}$$

Since  $f$  is bounded below on the level set containing these iterates, it follows from this inequality that

$$\lim_{k \in \mathcal{K}, k \rightarrow \infty} \Delta_k = 0,$$

clearly contradicting (22). 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  $\Delta_k$  will be reduced by a factor of  $\frac{1}{4}$  on every iteration. Then we have,  $\lim_{k \rightarrow \infty} \Delta_k = 0$ , which again contradicts (22). Then our original assumption (19) must be false, giving us the desired result.  $\square$

To show that all accumulation points of  $\{x_k\}$  are stationary points, we need to make an additional regularity assumption on the cost function  $f$ . The global convergence result in  $\mathbb{R}^n$ , as stated in [NW99, Theorem 4.8], requires that  $f$  be Lipschitz continuously differentiable. That is to say, for any  $x, y \in \mathbb{R}^n$ ,

$$\|\text{grad}f(y) - \text{grad}f(x)\| \leq \beta_1 \|y - x\|. \quad (23)$$

A key to obtaining a Riemannian counterpart of this global convergence result is to adapt the notion of Lipschitz continuous differentiability to the Riemannian manifold  $(M, g)$ . The expression  $\|x - y\|$  in the right-hand side of (23) naturally becomes the Riemannian distance  $\text{dist}(x, y)$ . For the left-hand side of (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  $T_x M$  and  $T_y M$ . However, if  $y$  belongs to a normal neighborhood of  $x$ , then there is a unique geodesic  $\alpha(t) = \text{Exp}_x(t \text{Exp}_x^{-1} y)$  such that  $\alpha(0) = x$  and  $\alpha(1) = y$ , and we can parallel transport  $\text{grad}f(y)$  along  $\alpha$  to obtain the vector  $P_\alpha^{0 \leftarrow 1} \text{grad}f(y)$  in  $T_x M$ , to yield the following definition.

**Definition 4.3 (Lipschitz continuous differentiability)** *Assume that  $(M, g)$  has an injectivity radius  $i(M) > 0$ . A real function  $f$  on  $M$  is Lipschitz continuous differentiable if it is differentiable and if, for all  $x, y$  in  $M$  such that  $\text{dist}(x, y) < i(M)$ , it holds that*

$$\|P_\alpha^{0 \leftarrow 1} \text{grad}f(y) - \text{grad}f(x)\| \leq \beta_1 \text{dist}(y, x), \quad (24)$$

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

Note that (24) is symmetric in  $x$  and  $y$ ; indeed, since the parallel transport is an isometry, it follows that

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



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

$$\|\xi\| \geq \mu d(x, R_x \xi), \quad \forall x \in M, \forall \xi \in T_x M, \|\xi\| \leq \delta_\mu \quad (25)$$

Note that for the exponential retraction discussed in this paper, (25) is satisfied as an equality, with  $\mu = 1$ . The bound is also satisfied when  $R$  is smooth and  $M$  is compact.

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 4.5 for a simpler (but slightly stronger) set of assumptions that yield the same result.

**Theorem 4.4** *Let  $\{x_k\}$  be a sequence of iterates generated by Algorithm 1. Suppose that all the assumptions of Theorem 4.2 are satisfied. Further suppose that  $\rho' \in (0, \frac{1}{4})$ , that  $f$  is Lipschitz continuously differentiable (Definition 4.3), and that (25) is satisfied for some  $\mu > 0$ . It then follows that*

$$\lim_{k \rightarrow \infty} \text{grad } f(x_k) = 0.$$

*Proof.*

Consider any index  $m$  such that  $\text{grad } f(x_m) \neq 0$ . The satisfaction of the Lipschitz property (24) on the level set (14) gives us

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

for any  $x$  in the level set. Define scalars

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

Define the ball  $B_r(x_m) := \{x : \text{dist}(x, x_m) < r\}$ .

Then for any  $x \in B_r(x_m)$ , we have

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

If the entire sequence  $\{x_k\}_{k \geq m}$  stays inside of the ball  $B_r(x_m)$ , then we would have  $\|\text{grad}f(x_k)\| > \epsilon$  for all  $k \geq m$ , which contradicts the results of Theorem 4.2. Then 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 of  $B_r(x_m)$ . Since  $\|\text{grad}f(x_k)\| > \epsilon$  for  $k = m, m+1, \dots, l$ , we have

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

If  $\Delta_k \leq \epsilon/\beta$  for all  $k = m, m+1, \dots, l$ , then

$$\begin{aligned}
f(x_m) - f(x_{l+1}) &\geq \rho' c_1 \epsilon \sum_{k=m, x_k \neq x_{k+1}}^l \Delta_k \geq \rho' c_1 \epsilon \sum_{k=m, x_k \neq x_{k+1}}^l \|\eta_k\| \\
&\geq \rho' c_1 \epsilon \sum_{k=m, x_k \neq x_{k+1}}^l \mu d(x_k, R_{x_k} \eta_k) \\
&= \rho' c_1 \epsilon \mu \sum_{k=m, x_k \neq x_{k+1}}^l d(x_k, x_{k+1}) \\
&\geq \rho' c_1 \epsilon \mu r = \rho' c_1 \epsilon \mu \min \left( \frac{\epsilon}{\beta_1}, i(M) \right). \tag{26}
\end{aligned}$$

If not, then  $\Delta_k > \epsilon/\beta$  for some  $k \in \{m, \dots, l\}$ , so that

$$f(x_m) - f(x_{l+1}) \geq \rho' c_1 \epsilon \frac{\epsilon}{\beta}. \tag{27}$$

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

$$f(x_k) \downarrow f^*, \tag{28}$$

for some  $f^* > -\infty$ . Then using (26) and (27), we get

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

Assume for the purpose of contradiction that it is not the case that  $\lim_{m \rightarrow \infty} \|\text{grad} f(x_m)\| = 0$ . Then there exists  $\omega > 0$  and an infinite sequence  $\mathcal{K}$  such that

$$\|\text{grad} f(x_k)\| > \omega, \quad \forall k \in \mathcal{K}.$$

Then for  $k \in \mathcal{K}, k \geq m$ , we have

$$\begin{aligned}
f(x_k) - f^* &\geq \frac{1}{2} \rho' c_1 \|\text{grad} f(x_k)\| \min \left( \frac{\|\text{grad} f(x_k)\|}{2\beta}, \frac{\|\text{grad} f(x_k)\| \mu}{2\beta_1}, i(M) \mu \right) \\
&> \frac{1}{2} \rho' c_1 \omega \min \left( \frac{\omega}{2\beta}, \frac{\omega \mu}{2\beta_1}, i(M) \mu \right) \\
&> 0.
\end{aligned}$$

However, this contradicts  $\lim_{k \rightarrow \infty} (f(x_k) - f^*) = 0$ , so that our hypothetical assumption must be false, and

$$\lim_{m \rightarrow \infty} \|\text{grad} f(x_m)\| = 0.$$

□

Note that this theorem reduces gracefully to the classical  $\mathbb{R}^n$  case, taking  $M = \mathbb{R}^n$  endowed with the classical inner product and  $R_x \xi := x + \xi$ . Then  $i(M) = +\infty > 0$ ,  $R$  satisfies (25), the Lipschitz condition (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  $\hat{f}$  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  $t \mapsto Rt\xi$ . Note also that all these conditions need only be checked on the level set  $\{x \in M : f(x) \leq f(x_0)\}$ .

**Proposition 4.5** *Suppose that  $\|\text{grad} f(x)\| \leq \beta_g$  and  $\|\text{Hess} f(x)\| \leq \beta_H$  for some constants  $\beta_g, \beta_H$ , and all  $x \in M$ . Moreover suppose that*

$$\left\| \frac{D}{dt} \frac{d}{dt} Rt\xi \right\| \leq \beta_D \tag{29}$$

for some constant  $\beta_D$ , for all  $\xi \in TM$  with  $\|\xi\| = 1$  and all  $t < \delta_D$ , where  $\frac{D}{dt}$  denotes the covariant derivative along the curve  $t \mapsto Rt\xi$  (see [dC92, Ch. 2, Prop. 2.2]).

Then the Lipschitz- $C^1$  condition on  $f$  (Definition 4.3) is satisfied with  $\beta_L = \beta_H$ ; the radially Lipschitz- $C^1$  condition on  $\hat{f}$  (Definition 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 (25) on  $R$  is satisfied for values of  $\mu$  and  $\delta_\mu$  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 4.7), boundedness of the Hessian of  $f$  implies the Lipschitz- $C^1$  property of  $f$ .

For (25), define  $u(t) = Rt\xi$  and observe that

$$\text{dist}(x, Rt\xi) \leq \int_0^t \|\dot{u}(\tau)\| d\tau$$

where  $\int_0^t \|\dot{u}(\tau)\| d\tau$  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

$$\left| \frac{d}{d\tau} \|\dot{u}(\tau)\| \right| = \left| \frac{d}{d\tau} \sqrt{g_{u(\tau)}(\dot{u}(\tau), \dot{u}(\tau))} \right| = \left| \frac{g_{u(\tau)}(\frac{D}{dt}\dot{u}(\tau), \dot{u}(\tau))}{\|\dot{u}(\tau)\|} \right| \leq \frac{\beta_D \|\dot{u}(\tau)\|}{\|\dot{u}(\tau)\|} \leq \beta_D$$

for all  $t < \delta_D$ . Therefore

$$\int_0^t \|\dot{u}(\tau)\| d\tau \leq \int_0^t \|\dot{u}(0)\| + \beta_D \tau d\tau = \|\xi\|t + \frac{1}{2}\beta_D t^2 = t + \frac{1}{2}\beta_D t^2,$$

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  $u(t) = Rt\xi$  and  $h(t) = f(u(t)) = \hat{f}(t\xi)$  with  $\xi \in T_x M$ ,  $\|\xi\| = 1$ . Then

$$\dot{h}(t) = g_{u(t)}(\text{grad } f(u(t)), \dot{u}(t))$$

and

$$\ddot{h}(t) = \frac{D}{dt} g_{u(t)}(\text{grad } f(u(t)), \dot{u}(t)) = g_{u(t)}(\frac{D}{dt} \text{grad } f(u(t)), \dot{u}(t)) + g_{u(t)}(\text{grad } f(u(t)), \frac{D}{dt} \dot{u}(t)).$$

Now,  $\frac{D}{dt} \text{grad } f(u(t)) = \nabla_{\dot{u}(t)} \text{grad } f(u(t)) = \text{Hess } f(u(t))[\dot{u}(t)]$ . It follows that  $|\ddot{h}(t)|$  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

$$|\dot{h}(t) - \dot{h}(0)| \leq \int_0^t |\ddot{h}(\tau)| d\tau \leq t\beta_{RL}.$$

□

**Remark 4.6 (smoothness and compactness)** *All the above-mentioned conditions on the cost function  $f$  and the retraction  $R$  are satisfied in the frequently encountered case where  $f$  and  $R$  are smooth and the manifold is compact.*

## 4.2 Local convergence

We now state local convergence properties of Algorithm 1-2 (i.e., Algorithm 1 where the trust-region subproblem (7) is solved approximately with Algorithm 2). We first state a few preparation lemmas.

As before,  $(M, g)$  is a complete Riemannian manifold of dimension  $d$ , and  $R$  is a retraction on  $M$  (Definition 2.1). The first lemma is a first-order Taylor formula for tangent vector fields (similar Taylor developments on manifolds can be found in [Smi94]).

**Lemma 4.7 (Taylor)** *Let  $x \in M$ , let  $V$  be a normal neighborhood of  $x$ , and let  $\zeta$  be a  $C^1$  tangent vector field on  $M$ . Then, for all  $y \in V$ ,*

$$P_\gamma^{0 \leftarrow 1} \zeta_y = \zeta_x + \nabla_\xi \zeta + \int_0^1 P_\gamma^{0 \leftarrow \tau} \nabla_{\gamma'(\tau)} \zeta - \nabla_\xi \zeta \, d\tau, \quad (30)$$

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

*Proof.* Start from

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

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

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

□

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

**Lemma 4.8** *Let  $v \in M$  and let  $f$  be a  $C^2$  cost function such that  $\text{grad } f(v) = 0$  and  $\text{Hess } f(v)$  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  $V$  of  $v$  such that, for all  $x \in V$ , it holds that*

$$c_0 \text{dist}(v, x) \leq \|\text{grad } f(x)\| \leq c_1 \text{dist}(v, x). \quad (31)$$

*Proof.* From Taylor (Lemma 4.7), it follows that

$$P_\gamma^{0 \leftarrow 1} \text{grad } f(v) = \text{Hess } f(v)[\gamma'(0)] + \int_0^1 P_\gamma^{0 \leftarrow \tau} \text{Hess } f(\gamma(\tau))[\gamma'(\tau)] - \text{Hess } f(v)[\gamma'(0)] d\tau. \quad (32)$$

Since  $f$  is  $C^2$  and since  $\|\gamma'(\tau)\| = \text{dist}(v, x)$  for all  $\tau \in [0, 1]$ , we have the following bound for the integral in (32):

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

where  $\lim_{t \rightarrow 0} \epsilon(t) = 0$ . Since  $\text{Hess } f(v)$  is nonsingular, it follows that  $|\lambda_{\min}| > 0$ . Take  $V$  sufficiently small so that  $\lambda_{\min} - \epsilon(\text{dist}(v, x)) > c_0$  and  $\lambda_{\max} + \epsilon(\text{dist}(v, x)) < c_1$  for all  $x$  in  $V$ . Then, using the fact that the parallel translation is an isometry, (31) follows from (32).  $\square$

We need a relation between the gradient of  $f$  at  $R_x \xi$  and the gradient of  $\hat{f}_x$  at  $\xi$ .

**Lemma 4.9** *Let  $R$  be a retraction on  $M$  and let  $f$  be a  $C^1$  cost function on  $M$ . Then, given  $v \in M$  and  $c_5 > 1$ , there exists a neighborhood  $V$  of  $v$  and a real  $\delta > 0$  such that*

$$\|\text{grad } f(R\xi)\| \leq c_5 \|\text{grad } \hat{f}(\xi)\|$$

for all  $x \in V$  and all  $\xi \in T_x M$  with  $\|\xi\| \leq \delta$ , where  $\hat{f}$  is as in (12).

*Proof.* Let  $A(\xi)$  denote the differential of  $R_x$  at  $\xi \in T_x M$ . Consider a parameterization of  $M$  at  $v$ , and consider the corresponding parameterization of  $TM$  (see [dC92, Ch. 0, Example 4.1]). Using Einstein's convention (see, e.g., [Sak96]), and denoting  $\partial_i f$  by  $f_{,i}$ , we have

$$\hat{f}_{x,i}(\xi) = f_{,j}(R\xi) A_i^j(\xi),$$

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

$$\|\text{grad } \hat{f}_x(\xi)\|^2 = \hat{f}_{x,i}(\xi) g^{ij}(x) \hat{f}_{x,j}(\xi) = f_{,k}(R_x \xi) A_i^k(\xi) g^{ij}(x) A_j^\ell(\xi) f_{,\ell}(R_x \xi)$$

and

$$\|\text{grad } f(R_x \xi)\|^2 = f_{,j}(R_x \xi) g^{ij}(R_x \xi) f_{,j}(R_x \xi).$$

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

Finally, we need the following result concerning the Hessian at stationary points.

**Lemma 4.10** *Let  $R$  be a  $C^2$  retraction, let  $f$  be a  $C^2$  cost function, and let  $v$  be a stationary point of  $f$  (i.e.,  $\text{grad } f(v) = 0$ ). Then  $\text{Hess } \hat{f}_v(0_v) = \text{Hess } f(v)$ .*

*Proof.* Let  $A$  denote the differential of  $R_v$  at  $0_v$ . Working in a parameterization of  $M$  around  $v$  and using Einstein's convention, one obtains (see [Sak96] for the notation)

$$\begin{aligned} \left( \text{Hess } \hat{f}_v \right)_j^i &= g^{ik} \partial_j \partial_k \hat{f}_v = g^{ik} \partial_j \partial_k (f \circ R_v) = g^{ik} \partial_j \left( \partial_\ell f A_k^\ell \right) \\ &= g^{ik} \partial_\ell f \partial_j A_k^\ell + g^{ik} \partial_\ell \partial_j f A_k^\ell \end{aligned}$$

and

$$(\text{Hess } f)_j^i = g^{ik} \nabla_k \partial_j f = g^{ik} \partial_k \partial_j f - g^{ik} \Gamma_{kj}^\ell \partial_\ell f$$

where  $\Gamma_{kj}^\ell$  are the Christoffel symbols. At  $v$ , one has  $\partial_\ell f = 0$  and  $A_k^\ell$  is the identity, hence  $\text{Hess } \hat{f}_v(0_v) = \text{Hess } f(v)$ .  $\square$

Away from the stationary points, the Hessians  $\text{Hess } f(x)$  and  $\text{Hess } \hat{f}_x(0_x)$  do not coincide. They do coincide if a “zero acceleration” condition (33) is imposed on the retraction. This result will not be used in the convergence analysis but it can be useful in applications, as explained after (10).

**Lemma 4.11** *Suppose that*

$$\frac{D}{dt} \left( \frac{d}{dt} R t \xi \right) \Big|_{t=0} = 0, \quad \text{for all } \xi \in TM, \quad (33)$$

where  $\frac{D}{dt}$  denotes the covariant derivative along the curve  $t \mapsto R t \xi$  (see [dC92, Ch. 2, Prop. 2.2]). Then  $\text{Hess } f(x) = \text{Hess } \hat{f}(0_x)$ .

*Proof.* Observe that  $D^2 f(x)[\xi, \xi] = \frac{d^2}{dt^2} f(\text{Exp}_x t \xi) \Big|_{t=0}$  and  $D^2 \hat{f}(0_x)[\xi, \xi] = \frac{d^2}{dt^2} f(R_x t \xi) \Big|_{t=0} = \frac{d}{dt} (df \frac{d}{dt} R_x t \xi) \Big|_{t=0} = \nabla_\xi df \xi + df \frac{D}{dt} \left( \frac{d}{dt} R_x t \xi \right) \Big|_{t=0}$ . The result follows from the definitions of the Hessians and the one-to-one correspondence between symmetric bilinear forms and quadratic forms.  $\square$

We now state and prove the local convergence results. We first show that the nondegenerate local minima are attractors of Algorithm 1-2. The principle of the argument is closely related to the Capture Theorem, see [Ber95, Theorem 1.2.5].

**Theorem 4.12 (local convergence to local minima)** *Consider Algorithm 1-2—i.e., the Riemannian trust-region algorithm where the trust-region sub-problems (7) are solved using the truncated CG algorithm with stopping criterion (11)—with all the assumptions of Theorem 4.2. Let  $v$  be a nondegenerate local minimum of  $f$ , i.e.,  $\text{grad } f(v) = 0$  and  $\text{Hess } f(v)$  is positive definite. Assume that  $\|\mathcal{H}_k^{-1}\|$  is bounded and that (25) holds for some  $\mu > 0$  and  $\delta_\mu > 0$ . Then there exists a neighborhood  $V$  of  $v$  such that, for all  $x_0 \in V$ , the sequence  $\{x_k\}$  generated by Algorithm 1-2 converges to  $v$ .*

*Proof.* Take  $\delta_1 > 0$  with  $\delta_1 < \delta_\mu$  such that  $B_{\delta_1}(v)$  is a neighborhood of  $v$ , which contains only  $v$  as stationary point, and such that  $f(x) > f(v)$  for all  $x \in \bar{B}_{\delta_1}(v)$ . Take  $\delta_2$  small enough that for all  $x \in B_{\delta_2}(v)$ , it holds that  $\|\eta^*(x)\| \leq \mu(\delta_1 - \delta_2)$ , where  $\eta^*$  is the (unique) solution of  $\mathcal{H}\eta^* = -\text{grad } f(x)$ ; such a  $\delta_2$  exists because of Lemma 4.8 and the bound on  $\|\mathcal{H}_k^{-1}\|$ . Consider a level set  $\mathcal{L}$  of  $f$  such that  $V := \mathcal{L} \cap B_{\delta_1}(v)$  is a subset of  $B_{\delta_2}(v)$ ; invoke that  $f \in C^1$  to show that such a level set exists. Then,  $V$  is a neighborhood of  $v$  and for all  $x \in V$ , we have

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

where we used the fact that  $\|\eta\|$  is increasing along the truncated CG process [Ste83, Thm 2.1]. It follows from the equation above that  $x_+$  is in  $B_{\delta_1}(v)$ . Moreover, since  $f(x_+) \leq f(x)$ , it follows that  $x_+ \in V$ . Thus  $V$  is invariant. But the only stationary point of  $f$  in  $V$  is  $v$ , so  $\{x_k\}$  goes to  $v$  whenever  $x_0$  is in  $V$ .  $\square$

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

**Theorem 4.13 (order of convergence)** *Consider Algorithm 1-2 with stopping criterion (11). Suppose that  $R$  is a  $C^2$  retraction, that  $f$  is a  $C^2$  cost function on  $M$ , and that*

$$\|\mathcal{H}_k - \text{Hess } \hat{f}_{x_k}(0_{x_k})\| \leq \beta_{\mathcal{H}} \|\text{grad } f(x_k)\|, \quad (34)$$

*that is,  $\mathcal{H}_k$  is a sufficiently good approximation of  $\text{Hess } \hat{f}_{x_k}(0_{x_k})$ . Let  $v \in M$  be a nondegenerate local minimum of  $f$ , (i.e.,  $\text{grad } f(v) = 0$  and  $\text{Hess } f(v)$  is positive definite). Further assume that  $\text{Hess } \hat{f}_{x_k}$  is Lipschitz-continuous at  $0_x$  uniformly in  $x$  in a neighborhood of  $v$ , i.e., there exist  $\beta_1 > 0$ ,  $\delta_1 > 0$  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*

$$\|\text{Hess } \hat{f}_{x_k}(\xi) - \text{Hess } \hat{f}_{x_k}(0_{x_k})\| \leq \beta_{L2} \|\xi\|, \quad (35)$$



where  $\|\cdot\|$  in the left-hand side denotes the operator norm in  $T_x M$  defined as in (16).

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  $k > K$ ,

$$\text{dist}(x_{k+1}, v) \leq c (\text{dist}(x_k, v))^{\min\{\theta+1, 2\}} \quad (36)$$

with  $\theta > 0$  as in (11).

*Proof.*

We will show below 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  $x \in M$ , all  $\xi$  with  $\|\xi\| < \tilde{\Delta}$ , and all  $k$  greater than some  $K$ , it holds

$$c_0 \text{dist}(v, x_k) \leq \|\text{grad } f(x_k)\| \leq c_1 \text{dist}(v, x_k), \quad (37)$$

$$\|\eta_k\| \leq c_4 \|\text{grad } m_{x_k}(0)\| \leq \tilde{\Delta}, \quad (38)$$

$$\rho_k > \rho' \quad (39)$$

$$\|\text{grad } f(R_{x_k} \xi)\| \leq c_5 \|\text{grad } \hat{f}_{x_k}(\xi)\|, \quad (40)$$

$$\|\text{grad } m_{x_k}(\xi) - \text{grad } \hat{f}_{x_k}(\xi)\| \leq c_3 \|\xi\|^2 + c'_3 \|\text{grad } f(x_k)\| \|\xi\|, \quad (41)$$

$$\|\text{grad } m_{x_k}(\eta_k)\| \leq c_2 \|\text{grad } m_{x_k}(0)\|^{\theta+1}, \quad (42)$$

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

With these results at hand the proof is concluded as follows. For all  $k > K$ , it follows from (37) and (39) that

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

from (40) and (38) that

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

from (38) and (41) and (42) that

$$\begin{aligned} \|\text{grad } \hat{f}_{x_k}(\eta_k)\| &\leq \|\text{grad } m_{x_k}(\eta_k) - \text{grad } \hat{f}_{x_k}(\eta_k)\| + \|\text{grad } m_{x_k}(\eta_k)\| \\ &\leq (c_3 c_4^2 + c'_3 c_4) \|\text{grad } m_{x_k}(0)\|^2 + c_2 \|\text{grad } m_{x_k}(0)\|^{1+\theta}, \end{aligned}$$

and from (37) that

$$\|\text{grad } m_{x_k}(0)\| = \|\text{grad } f(x_k)\| \leq c_1 \text{dist}(v, x_k).$$

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

$$\begin{aligned} & c_0 \text{dist}(v, x_{k+1}) \\ & \leq \|\text{grad } f(x_{k+1})\| \end{aligned} \quad (43)$$

$$\begin{aligned} & \leq c_5(c_3c_4^2 + c_3'c_4)\|\text{grad } f(x_k)\|^2 + c_5c_2\|\text{grad } f(x_k)\|^{\theta+1} \\ & \leq c_5((c_3c_4^2 + c_3'c_4)c_1^2(\text{dist}(v, x_k))^2 + c_2c_1^{1+\theta}(\text{dist}(v, x_k))^{1+\theta}) \\ & \leq c_5((c_3c_4^2 + c_3'c_4)c_1^2 + c_2c_1^{1+\theta})(\text{dist}(v, x_k))^{\min\{2, 1+\theta\}} \end{aligned} \quad (44)$$

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

It remains to prove the bounds (37)-(42).

Equation (37) comes from Lemma 4.8 and is due to the fact that  $v$  is a nondegenerate critical point.

We prove (38). Since  $\{x_k\}$  converges to the nondegenerate local minimum  $v$  where  $\text{Hess } \hat{f}_v(0_v) = \text{Hess } f(v)$  (see Lemma 4.10) and since  $\text{Hess } f(v)$  is positive definite with  $f \in C^2$ , it follows from the approximation condition (34) and from (37) that there exist  $c_4 > 0$  such that  $\|\mathcal{H}_k^{-1}\| < c_4$  for all  $k$  greater than some  $K$ . Given a  $k > K$ , let  $\eta^*$  be the solution of  $\mathcal{H}_{x_k}\eta^* = -\text{grad } m_{x_k}(0)$ . It follows that  $\|\eta^*\| \leq c_4\|\text{grad } m_{x_k}(0)\|$ . Then, since the sequence of  $\eta_k^j$ 's constructed by the tCG inner iteration (Algorithm 2) is strictly increasing in norm (see [Ste83, Theorem 2.1]) and would eventually reach  $\eta^*$  at  $j = d$ , it follows that (38) holds. The second inequality in (38) comes for any given  $\tilde{\Delta}$  by choosing  $K$  larger if necessary.

We prove (39). Let  $\gamma_k$  denote  $\|\text{grad } f(x_k)\|$ . From the definition (9) of  $\rho_k$ , from the assumption (34) that  $\|\mathcal{H}_k - \text{Hess } \hat{f}_{x_k}\| \leq \beta_{\mathcal{H}}\gamma_k$ , and from the Lipschitz assumption (35) on the Hessian of  $\hat{f}$ , it follows by a classical Taylor argument in the Euclidean space  $T_x M$  that

$$\rho_k = \frac{m_k(0_k) - m_k(\eta_k) + \varepsilon(\|\eta_k\|^3)}{m_k(0_k) - m_k(\eta_k)} = 1 + \frac{\varepsilon(\|\eta_k\|^3)}{m_k(0_k) - m_k(\eta_k)},$$

where  $0 \leq \varepsilon(t) \leq \frac{\beta_{L2} + \beta_{\mathcal{H}}\gamma_k}{6}t$  for all  $t < \delta_2$ . It then follows from  $\|\eta_k\| \leq \Delta_k$ , from the bound (38) and from the Cauchy decrease hypothesis (15), that

$$|\rho_k - 1| \leq \frac{(\beta_{L2} + \beta_{\mathcal{H}}\gamma_k)(\min\{\Delta_k, c_4\gamma_k\})^3}{6\gamma_k \min\{\Delta_k, \gamma_k/\beta\}} \quad (45)$$

where  $\beta$  is an upper bound on the norm of  $\mathcal{H}_k$ . Either,  $\Delta_k$  is active in the denominator of (45), in which case we have

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

Or,  $\gamma_k/\beta$  is active in the denominator of (45), in which case we have

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

In both cases, since  $\lim_{k \rightarrow \infty} \gamma_k = 0$  in view of (37), it follows that  $\lim_{k \rightarrow \infty} \rho_k = 1$ .

Equation (40) comes from Lemma 4.9.

We prove (41). It follows from Taylor's formula (Lemma 4.7, where the parallel translation becomes the identity since the domain of  $\hat{f}_{x_k}$  is the Euclidean space  $T_{x_k}M$ ) that

$$\text{grad } \hat{f}_{x_k}(\xi) = \text{grad } \hat{f}_{x_k}(0_{x_k}) + \text{Hess } \hat{f}_{x_k}(0_{x_k})[\xi] + \int_0^1 \left( \text{Hess } \hat{f}_{x_k}(\tau\xi) - \text{Hess } \hat{f}_{x_k}(0_{x_k}) \right) [\xi] d\tau.$$

The conclusion comes by the Lipschitz condition (35) and the approximation condition (34).

Finally, equation (42) comes from the stopping criterion (11) of the inner iteration. More precisely, the truncated CG loop (Algorithm 2) terminates if either  $g(\delta_j, \mathcal{H}_{x_k}\delta_j) \leq 0$ , or  $\|\eta_{j+1}\| \geq \Delta$ , or the criterion (11) is satisfied. Since  $\{x_k\}$  converges to  $v$  and  $\text{Hess } f(v)$  is positive-definite, it follows that  $\mathcal{H}_{x_k}$  is positive-definite for all  $k$  greater than a certain  $K$ . Therefore, for all  $k > K$ , the criterion  $g(\delta_j, \mathcal{H}_{x_k}\delta_j) \leq 0$  is never satisfied. In view of (38) and (39), 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 (11) is satisfied each time a computed  $\eta_k$  is returned by the tCG loop. Therefore, the tCG loop behaves as a classical linear CG method; see e.g. [NW99, Section 5.1]. Consequently,  $\text{grad } m_{x_k}(\eta_j) = r_j$  for all  $j$ . Choose  $K$  such that for all  $k > K$ ,  $\|\text{grad } f(x_k)\| = \|\text{grad } m_{x_k}(0)\|$  is so small—it converges to zero in view of (37)—that the stopping criterion (11) yields

$$\|\text{grad } m_{x_k}(\eta_j)\| = \|r_j\| \leq \|r_0\|^{1+\theta} = \|\text{grad } m_{x_k}(0)\|^{1+\theta} \text{ or } k \geq d. \quad (46)$$

If the second condition in (46) is active, then it means that the linear CG process has been completed, so  $\text{grad } m_{x_k}(\eta_k^j) = 0$ , and (42) trivially holds. On the other hand, if the first condition in (46) is active, then we obtain (42) with  $c_2 = 1$ .  $\square$

The constants in the proof of Theorem 4.13 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  $\text{Hess } f(v)$  respectively.

Consequently, the constant  $c$  in the convergence bound (36) can be chosen as

$$c > \frac{1}{\lambda_{\min}} \left( (\beta_{L2}/\lambda_{\min}^2 + \beta_{\mathcal{H}}/\lambda_{\min}) \lambda_{\max}^2 + \lambda_{\max}^{1+\theta} \right). \quad (47)$$

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 (43)-(44).

**Theorem 4.14** *Under the assumptions of Theorem 4.13, if  $\theta + 1 < 2$ , then given  $c_g > 1$  and  $\{x_k\}$  generated by the algorithm, there exists  $K > 0$  such that*

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

for all  $k > K$ .

Nevertheless, (43)-(44) 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, or even better than other methods in the presence of a small relative gap [ABG04a].

### 4.3 Discussion

The main global convergence result (Theorem 4.4) shows that RTR-tCG (Algorithm 1-2) converges to a set of stationary points of the cost function for *all* initial conditions. This is an improvement on the pure Newton method, for which only local convergence results exist. However, the convergence theory falls short of showing that the algorithm always converges to a local minimum. This is not surprising: since we have ruled out the possibility of checking positive-definiteness the Hessian of the cost function, we have no way of testing whether a stationary point is a local minimum or not (note as an aside that even checking positive-definiteness of the Hessian is not always sufficient for determining if a stationary point is a local minimum 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 stationary points can be secured unless some specific assumptions (like convexity) are made; see e.g. [Pol97, Ch. 1]. Nevertheless, it is observed in numerical experiments with random initial conditions that the algorithm systematically converges to a local minimum; convergence to a saddle point is only observed on specifically crafted problems, for example 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, convergence to saddle points or local minima is unstable under perturbations.

Concerning the order of convergence to local minima, we point out that there are cases where the bound (36) also holds with “ $\min\{\theta+1, 2\}$ ” replaced by “ $\min\{\theta+1, 3\}$ ”, i.e., cubic convergence can be achieved. This is related to the cubic convergence of the Riemannian Newton method when the cost function is symmetric around the local minimum  $v$ , that is,  $f(\text{Exp}_x(\xi)) = f(\text{Exp}_x(-\xi))$ . This issue is of theoretical importance in applications where state-of-the-art methods converge cubically. Notice however that a cubic method may be less efficient than a quadratic method, even as  $k$  goes to infinity (as pointed out in [DV00], concatenating two steps of a quadratic method yields a quartic method).

## 5 Applications

In this section, we briefly review the essential “ingredients” necessary for applying the RTR-tCG method (Algorithm 1-2) and we present two examples in detail. These examples are presented as illustrations: comparing the resulting algorithms with existing methods and conducting numerical experiments is beyond the scope of this paper. For the problem of computing extreme eigenspaces of matrices, numerical experiments show that the RTR-tCG algorithm can match and sometimes dramatically outperform existing algorithms; experiments, comparisons and further developments are presented in [ABG04b, ABG04a, ABGS04].

### 5.1 Checklist

The following elements are required for applying the RTR method to optimizing a cost function  $f$  on a Riemannian manifold  $(M, g)$ : (i) a tractable numerical representation for points  $x$  on  $M$ , for tangent spaces  $T_x M$ , and for the inner products  $g_x(\cdot, \cdot)$  on  $T_x M$ ; (ii) choice of a retraction  $R_x : T_x M \rightarrow M$  (Definition 2.1); (iii) formulas for  $f(x)$ ,  $\text{grad } f(x)$  and the approximate Hessian  $\mathcal{H}_x$  that satisfies the properties required for the convergence results in Section 4.

Choosing a good retraction amounts to finding an approximation of the exponential mapping that can be computed with low computational cost. Guidelines can be found in [CI01, DN04]. This is an important open research topic.

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

$$\hat{f}_x(\eta) = f(x) + g_x(\text{grad } f(x), \eta) + \frac{1}{2}g_x(\text{Hess } \hat{f}_x(0_x)[\eta], \eta) + O(\|\eta\|^3),$$

where  $\text{grad } f(x) \in T_x M$  and  $\text{Hess } \hat{f}_x(0_x)$  is a linear transformation of  $T_x M$ . In order to obtain an “approximate Hessian”  $\mathcal{H}_x$  that satisfies the approximation condition (34), one can pick  $\mathcal{H}_x := \text{Hess}(f \circ \tilde{R}_x)(0_x)$  where  $\tilde{R}_x$  is any retraction. Then, assuming sufficient smoothness of  $f$ ,  $R$  and  $\tilde{R}$ , the bound (34) follows from Lemmas 4.8 and 4.10. In particular, the choice  $\tilde{R}_x = \text{Exp}_x$  yields  $\mathcal{H}_x = \nabla \text{grad } f(x)$ . If  $M$  is an embedded submanifold of a Euclidean space, then  $\nabla_\eta \text{grad } f(x) = \pi D \text{grad } f(x)[\eta]$  where  $\pi$  denotes the orthogonal projector onto  $T_x M$ .

## 5.2 Symmetric eigenvalue decomposition

Let  $M$  be the orthogonal group,

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

This manifold is an embedded submanifold of  $\mathbb{R}^{n \times n}$ . It can be shown that  $T_Q O_n = \{Q\Omega : \Omega = -\Omega^T\}$ ; see e.g. [HM94]. The canonical Euclidean metric  $g(A, B) = \text{trace}(A^T B)$  on  $\mathbb{R}^{n \times n}$  induces on  $O_n$  the metric

$$g_Q(Q\Omega_1, Q\Omega_2) = \text{trace}(\Omega_1^T \Omega_2). \quad (48)$$

A retraction  $R_Q : T_Q O_n \rightarrow O_n$  must be chosen that satisfies the properties stated in Section 2. The Riemannian geodesic-based choice is

$$R_Q Q\Omega = \text{Exp}_Q Q\Omega = Q \exp(Q(Q^T \Omega)) = Q \exp(\Omega)$$

where  $\exp$  denotes the matrix exponential. However, the matrix exponential is numerically expensive to compute (the computational cost is comparable to solving an  $n \times n$  eigenvalue problem!), which makes it essential to use computationally cheaper retractions. Given a Lie group  $G$  (here the orthogonal group) and its Lie algebra  $\mathfrak{g}$  (here the set of skew-symmetric matrices), there exist several ways of approximating  $\exp(\Omega)$ ,  $\Omega \in \mathfrak{g}$ , by an  $R(\Omega)$  such that  $R(\Omega) \in G$  if  $\Omega \in \mathfrak{g}$ ; these techniques are well-known in geometric integration (see e.g. [CI01] and references therein) and can be applied to our case where  $G$  is the orthogonal group  $O_n$ . For example,  $\exp(\Omega)$  can be approximated by a product of plane (or Givens) rotations [GV96] in such a way that  $R$  is

a second order approximation of the exponential; see [CI01]. This approach has the advantage of being very efficient computationally.

Consider the cost function

$$f(Q) = \text{trace}(Q^T A Q N)$$

where  $A$  and  $N$  are given  $n \times n$  symmetric matrices. For  $N = \text{diag}(\mu_1, \dots, \mu_n)$ ,  $\mu_1 < \dots < \mu_n$ , the minimum of  $f$  is realized by the orthonormal matrices of eigenvectors of  $A$  sorted in increasing order of corresponding eigenvalue; see e.g. [HM94, Section 2.1]. Assume that the retraction  $R$  approximates the exponential at least to order 2. With the metric  $g$  defined as in (48), we obtain

$$\begin{aligned} \hat{f}_Q(Q\Omega) &:= f(R_Q(Q\Omega)) = \text{trace}((I + \Omega + \frac{1}{2}\Omega^2 + O(\Omega^3))^T Q^T A Q (I + \Omega + \frac{1}{2}\Omega^2 + O(\Omega^3)) N) \\ &= f(Q) + 2\text{trace}(\Omega^T Q^T A Q N) + \text{trace}(\Omega^T Q^T A Q \Omega N - \Omega^T \Omega Q^T A Q N) + O(\Omega^3) \end{aligned}$$

from which it follows

$$\begin{aligned} D\hat{f}_Q(0)[Q\Omega] &= 2\text{trace}(Q^T A Q \Omega N) \\ \frac{1}{2}D^2\hat{f}_Q(0)[Q\Omega_1, Q\Omega_2] &= \text{trace}(\Omega_1^T Q^T A Q \Omega_2 N - \frac{1}{2}(\Omega_1^T \Omega_2 + \Omega_2^T \Omega_1) Q^T A Q N) \\ \text{grad } \hat{f}_Q(0) &= \text{grad } f(Q) = Q[Q^T A Q, N] \\ \text{Hess } \hat{f}_Q(0)[Q\Omega] &= \text{Hess } f(Q)[Q\Omega] = \frac{1}{2}Q[[Q^T A Q, \Omega], N] + \frac{1}{2}Q[[N, \Omega], Q^T A Q] \end{aligned}$$

where  $[A, B] := AB - BA$ . It is now straightforward to replace these expressions in the general formulation of Algorithm 1-2 and obtain a practical matrix algorithm. Numerical results are presented in [ABG04b].

An alternative way to obtain  $\text{Hess } \hat{f}_Q(0)$  is to exploit Lemma 4.11 which yields  $\text{Hess } \hat{f}_Q(0) = \nabla \text{grad } f(Q)$ . Since the manifold  $M$  is an embedded Riemannian submanifold of  $\mathbb{R}^{n \times p}$ , the covariant derivative  $\nabla$  is obtained by projecting the derivative in  $\mathbb{R}^{n \times p}$  onto the tangent space to  $M$ ; see [dC92, Ch. 2, sec. 1] or [Boo75, VII.2]. We obtain  $\text{Hess } f(Q)[Q\Omega] = Q \text{skew}(\Omega[Q^T Q Q, N] + [\Omega^T Q^T A Q + Q^T A Q \Omega, N])$ , which yields the same result as above.

### 5.3 Computing an extreme eigenspace of a symmetric definite matrix pencil

We assume that  $A$  and  $B$  are  $n \times n$  symmetric matrices and that  $B$  is positive definite. An eigenspace  $\mathcal{Y}$  of  $(A, B)$  satisfies  $B^{-1}Ay \in \mathcal{Y}$  for all  $y \in \mathcal{Y}$ , which

can also be written  $B^{-1}A\mathcal{Y} \subseteq \mathcal{Y}$  or  $A\mathcal{Y} \subseteq B\mathcal{Y}$ . The simplest example is when  $\mathcal{Y}$  is spanned by a single eigenvector of  $(A, B)$ , i.e., a nonvanishing vector  $y$  such that  $Ay = \lambda By$  for some eigenvalue  $\lambda$ . More generally, an eigenspace can be spanned by a subset of eigenvectors of  $(A, B)$ . For more details we refer to the review of the generalized eigenvalue problem in [Ste01].

Let  $\lambda_1 \leq \dots \leq \lambda_p < \lambda_{p+1} \leq \dots \leq \lambda_n$  be the eigenvalues of the pencil  $(A, B)$ . We consider the problem of computing the (unique) eigenspace  $\mathcal{V}$  of  $A$  associated to the  $p$  leftmost eigenvalues (in other words,  $\mathcal{V}$  is characterized by  $\mathcal{V} = \text{colsp}(V)$  where  $AV = V\text{diag}(\lambda_1, \dots, \lambda_p)$  and  $V^T V = I$ ). We will call  $\mathcal{V}$  the *leftmost*  $p$ -dimensional eigenspace of the pencil  $(A, B)$ . Note that the algorithms we are about to present work equally well for computing the *rightmost* eigenspace: replace  $A$  by  $-A$  throughout and notice that the leftmost eigenspace of  $-A$  is the rightmost eigenspace of  $A$ .

It is well known (see e.g. [SW82, ST00]) that the leftmost eigenspace  $\mathcal{V}$  of  $(A, B)$  is the minimizer of the Rayleigh cost function

$$f(\text{colsp}(Y)) = \text{trace}((Y^T A Y)(Y^T B Y)^{-1}) \quad (49)$$

where  $Y$  is full-rank  $n \times p$  and  $\text{colsp}(Y)$  denotes the column space of  $Y$ . It is readily checked that the right-hand side only depends on  $\text{colsp}(Y)$ .

The domain  $M$  of the cost function  $f$  is the set of  $p$ -dimensional subspaces of  $\mathbb{R}^n$ , called the *Grassmann manifold* and denoted by  $\text{Grass}(p, n)$ . A difficulty with the Grassmann manifold is that it is not directly defined as a submanifold of a Euclidean space (in contrast to the orthogonal group considered in Section 5.2). The first action to take is thus to devise a matrix representation of the elements of  $\text{Grass}(p, n)$  and its tangent vectors. This can be done in several ways.

A possibility is to rely on the one-to-one correspondence between subspaces and projectors; this idea is detailed in [MS85]. Another possibility is to rely on the definition of  $\text{Grass}(p, n)$  as a quotient of Lie groups; see [EAS98] and references therein. Yet another possibility is to rely on coordinate charts on Grassmann (see, e.g., [HM94, Section C4]); this approach is appealing because it uses a minimal set of variables, but it has the drawback of relying on arbitrarily fixed reference points.

A fourth way, which we will follow here, is to consider  $\text{Grass}(p, n)$  as the quotient  $\mathbb{R}_*^{n \times p} / \text{GL}_p$  of the locally Euclidean space  $\mathbb{R}_*^{n \times p}$  (the set of full-rank  $n \times p$  matrices) by the set of transformations that preserve the column space. This approach was developed in [AMS04]. The principle is to allow a subspace to be represented by any  $n \times p$  matrix whose columns span the subspace; that is, the subspaces are represented by bases (which are



allowed to be nonorthonormal, although in practical computations it is often desirable to require some form of orthonormalization). This representation is particularly appropriate in the scope of numerical computations. The set of matrices that represent the same subspace as a matrix  $Y \in \mathbb{R}_*^{n \times p}$  is the *fiber*  $Y\text{GL}_p = \{YM : \det(M) \neq 0\}$ . The *vertical space* at  $Y$  is  $V_Y = \{YM : M \in \mathbb{R}^{p \times p}\}$ . A real function  $h$  on  $\text{Grass}(p, n)$  is represented by its lift  $h_{\uparrow Y} = h(\text{colsp}(Y))$ . To represent a tangent vector  $\xi$  to  $\text{Grass}(p, n)$  at a point  $\mathcal{Y} = \text{colsp}(Y)$ , first define a *horizontal space*  $H_Y$  whose direct sum with  $V_Y$  is the whole  $\mathbb{R}^{n \times p}$ ; then  $\xi$  is uniquely represented by its *horizontal lift*  $\xi_{\uparrow Y}$  defined by the following two conditions: (i)  $\xi_{\uparrow Y} \in H_Y$  and (ii)  $Dh(\mathcal{Y})[\xi] = Dh_{\uparrow}(Y)[\xi_{\uparrow Y}]$  for all real functions  $h$  on  $\text{Grass}(p, n)$ . Therefore, the horizontal space  $H_Y$  represents the tangent space  $T_{\mathcal{Y}}\text{Grass}(p, n)$ .

In this section, with a view to simplifying the derivation of the gradient and Hessian of the Rayleigh cost function (49), we define the horizontal space as

$$H_Y = \{Z \in \mathbb{R}^{n \times p} : Y^T B Z = 0\},$$

which reduces to the definition in [AMS04] when  $B$  is the identity. We then define a noncanonical metric on  $\text{Grass}(p, n)$  as

$$g_{\mathcal{Y}}(\xi, \zeta) = \text{trace} \left( (Y^T B Y)^{-1} \xi_{\uparrow Y}^T \zeta_{\uparrow Y} \right). \quad (50)$$

From now on, the definitions of the gradient, Hessian and Riemannian connection will be with respect to the metric (50). We will use the retraction

$$R_{\mathcal{Y}}(\xi) = \text{colsp}(Y + \xi_{\uparrow Y}) \quad (51)$$

where  $\mathcal{Y} = \text{colsp}(Y)$ .

For the Rayleigh cost function (49), using the notation

$$P_{U,V} = I - U(V^T U)^{-1} V^T \quad (52)$$

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

$$\begin{aligned} \hat{f}_{\mathcal{Y}}(\xi) &= f(R_{\mathcal{Y}}(\xi)) = \text{trace} \left( \left( (Y + \xi_{\uparrow Y})^T B (Y + \xi_{\uparrow Y}) \right)^{-1} \left( (Y + \xi_{\uparrow Y})^T A (Y + \xi_{\uparrow Y}) \right) \right) \\ &= \text{trace} \left( (Y^T B Y)^{-1} Y^T A Y \right) + 2\text{trace} \left( (Y^T B Y)^{-1} \xi_{\uparrow Y}^T A Y \right) \\ &\quad + \text{trace} \left( (Y^T B Y)^{-1} \xi_{\uparrow Y}^T (A \xi_{\uparrow Y} - B \xi_{\uparrow Y} (Y^T B Y)^{-1} (Y^T A Y)) \right) + \text{HOT} \\ &= \text{trace} \left( (Y^T B Y)^{-1} Y^T A Y \right) + 2\text{trace} \left( (Y^T B Y)^{-1} \xi_{\uparrow Y}^T P_{BY, BY} A Y \right) \\ &\quad + \text{trace} \left( (Y^T B Y)^{-1} \xi_{\uparrow Y}^T P_{BY, BY} (A \xi_{\uparrow Y} - B \xi_{\uparrow Y} (Y^T B Y)^{-1} (Y^T A Y)) \right) + \text{HOT}, \end{aligned} \quad (53)$$

where the introduction of the projectors do not modify the expression since  $P_{BY,BY}\xi_{\uparrow Y} = \xi_{\uparrow Y}$ . By identification, using the noncanonical metric (50), we obtain

$$(\text{grad } f(\mathcal{Y}))_{\uparrow Y} = \left( \text{grad } \hat{f}_{\mathcal{Y}}(0) \right)_{\uparrow Y} = 2P_{BY,BY}AY \quad (54)$$

and

$$\left( \text{Hess } \hat{f}_{\mathcal{Y}}(0\mathcal{Y})[\xi] \right)_{\uparrow Y} = 2P_{BY,BY} \left( A\xi_{\uparrow Y} - B\xi_{\uparrow Y}(Y^T BY)^{-1}(Y^T AY) \right). \quad (55)$$

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

We choose to take

$$\mathcal{H}_{\mathcal{Y}} = \text{Hess } \hat{f}_{\mathcal{Y}}(0\mathcal{Y}). \quad (56)$$

Therefore, the approximation condition (34) is trivially satisfied. The model (7) is thus

$$\begin{aligned} m_{\mathcal{Y}}(\xi) &= f(\mathcal{Y}) + g_{\mathcal{Y}}(\text{grad } f(\mathcal{Y}), \xi) + \frac{1}{2}g_{\mathcal{Y}}(\mathcal{H}_{\mathcal{Y}}\xi, \xi) \\ &= \text{trace} \left( (Y^T BY)^{-1}Y^T AY \right) + 2\text{trace} \left( (Y^T BY)^{-1}\xi_{\uparrow Y}^T AY \right) \\ &\quad + \text{trace} \left( (Y^T BY)^{-1}\xi_{\uparrow Y}^T (A\xi_{\uparrow Y} - B\xi_{\uparrow Y}(Y^T BY)^{-1}Y^T AY) \right). \end{aligned} \quad (57)$$

Since the Rayleigh cost function (49) is smooth on  $\text{Grass}(p, n)$ —recall that  $B$  is positive definite—and since  $\text{Grass}(p, n)$  is compact, it follows that all the assumptions involved in the convergence analysis of the general RTR-tCG algorithm (Section 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 (36). (Since the metric (50) is different from the canonical metric, the formulas given in [AMS04] do not apply.) But since  $B$  is fixed and positive definite, the distances induced by the noncanonical metric (50) and by the canonical metric—(50) with  $B := I$ —are locally equivalent, and therefore for a given sequence both distances yield the same rate of convergence.

We have now all the required information to use the RTR-tCG method (Algorithm 1-2) for minimizing the Rayleigh cost function (49) on the Grassmann manifold  $\text{Grass}(p, n)$  endowed with the noncanonical metric (50). This yields the following matrix version of the inner iteration. (We omit the horizontal lift notation for conciseness.) We use the notation

$$\overline{\mathcal{H}}_Y[Z] = P_{BY,BY}(AZ - BZ(Y^T BY)^{-1}Y^T AY). \quad (58)$$

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

**Algorithm 3 (tCG for  $(A, B)$ )** Given two symmetric  $n \times n$  matrices  $A$  and  $B$  with  $B$  positive definite, and a  $B$ -orthonormal full-rank  $n \times p$  matrix  $Y$  (i.e.,  $Y^T B Y = I$ ).

Set  $\eta^0 = 0 \in \mathbb{R}^{n \times p}$ ,  $r_0 = P_{B Y, B Y} A Y$ ,  $\delta_0 = -r_0$ ;

**for**  $j = 0, 1, 2, \dots$  until a stopping criterion is satisfied, perform the iteration:

**if**  $\text{trace} \left( \delta_j^T \overline{\mathcal{H}}_Y [\delta_j] \right) \leq 0$

        Compute  $\tau > 0$  such that  $\eta = \eta^j + \tau \delta_j$   
        satisfies  $\text{trace} (\eta^T \eta) = \Delta$ ;

**return**  $\eta$ ;

    Set  $\alpha_j = \text{trace} \left( r_j^T r_j \right) / \text{trace} \left( \delta_j^T \overline{\mathcal{H}}_Y [\delta_j] \right)$ ;

    Set  $\eta^{j+1} = \eta^j + \alpha_j \delta_j$ ;

**if**  $\text{trace} \left( (\eta^{j+1})^T \eta^{j+1} \right) \geq \Delta$

        Compute  $\tau \geq 0$  such that  $\eta = \eta^j + \tau \delta_j$  satisfies  $\text{trace} (\eta^T \eta) = \Delta$ ;

**return**  $\eta$ ;

    Set  $r_{j+1} = r_j + \alpha \overline{\mathcal{H}}_Y [\delta_j]$ ;

    Set  $\beta_{j+1} = \text{trace} \left( r_{j+1}^T r_{j+1} \right) / \text{trace} \left( r_j^T r_j \right)$ ;

    Set  $\delta_{j+1} = -r_{j+1} + \beta_{j+1} \delta_j$ ;

**end (for).**

According to the retraction formula (51), the returned  $\eta$  yields a candidate new iterate

$$Y_+ = (Y + \eta)M$$

where  $M$  is chosen such that  $Y_+^T B Y_+ = I$ . The candidate is accepted or rejected and the trust-region radius is updated as prescribed in the outer RTR method (Algorithm 1), where  $\rho$  is computed using  $m$  as in (57) and  $\hat{f}$  as in (53).

The resulting algorithm converges to eigenspaces of  $(A, B)$ —which are the stationary points of the cost function (49)—, and convergence to the left-most eigenspace  $\mathcal{V}$  is expected to occur in practice since the other eigenspaces 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  $\theta$  is the parameter appearing in the stopping criterion (11) of the inner (tCG) iteration.

This algorithm is further developed in [ABGS04]. Relations with other methods are investigated in [ABG04a, ABG04b].

## 5.4 Other examples

The Riemannian trust-region algorithm can be applied in general to minimize smooth functions on smooth manifolds where a retraction, the gradient and the Hessian have tractable formulations. Other applications include reduced-rank approximation to matrices, the Procrustes problem, nearest-Jordan structure, trace minimization with a nonlinear term, simultaneous Schur decomposition, and simultaneous diagonalization; see, e.g., [HM94, LE00].

## 6 Conclusion

We have proposed a trust-region approach for optimizing a smooth function on a Riemannian manifold. The method improves on the well-known Riemannian Newton method of Smith and Udriște in three ways. First, the exponential mapping is relaxed to general retractions with a view to reducing computational complexity. Second, a trust-region safeguard is applied for global convergence. Third, early stopping of the inner iteration (yielding inexact solutions of the trust-region subproblems) is allowed under criteria that preserve the convergence properties of the overall algorithm. Taken independently, none of these concepts is new; the novelty is their combination in a general algorithm for optimization on manifolds, aimed at numerical efficiency with reliable global behavior, and supported by a detailed convergence analysis.

## Acknowledgements

The authors wish to thank A. Edelman, U. Helmke, R. Mahony, A. Sameh, R. Sepulchre, S. T. Smith, and P. Van Dooren for useful discussions.

## References

- [ABG04a] P.-A. Absil, C. G. Baker, and K. A. Gallivan, *A truncated-CG style method for symmetric generalized eigenvalue problems*, submitted, 2004.
- [ABG04b] ———, *Trust-region methods on Riemannian manifolds*, Tech. Report FSU-CSIT-04-13, School of Computational Science, Florida State University, July 2004.
- [ABG04c] ———, *Trust-region methods on Riemannian manifolds with applications in numerical linear algebra*, Proceedings of the 16th International Symposium on Mathematical Theory of Networks and Systems (MTNS2004), Leuven, Belgium, 5–9 July 2004, 2004.

- [ABGS04] P.-A. Absil, C. G. Baker, K. A. Gallivan, and A. Sameh, *Adaptive model trust region methods for generalized eigenvalue problems*, ICCS conference paper, to appear in the Lecture Notes in Computer Science, 2004.
- [ADM<sup>+</sup>02] R. L. Adler, J.-P. Dedieu, J. Y. Margulies, M. Martens, and M. Shub, *Newton's method on Riemannian manifolds and a geometric model for the human spine*, IMA J. Numer. Anal. **22** (2002), no. 3, 359–390.
- [AMS04] P.-A. Absil, R. Mahony, and R. Sepulchre, *Riemannian geometry of Grassmann manifolds with a view on algorithmic computation*, Acta Appl. Math. **80** (2004), no. 2, 199–220.
- [Ber95] D. P. Bertsekas, *Nonlinear programming*, Athena Scientific, Belmont, Massachusetts, 1995.
- [Boo75] W. M. Boothby, *An introduction to differentiable manifolds and Riemannian geometry*, Academic Press, 1975.
- [CGT00] A. R. Conn, N. I. M. Gould, and Ph. L. Toint, *Trust-region methods*, MPS/SIAM Series on Optimization, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, and Mathematical Programming Society (MPS), Philadelphia, PA, 2000.
- [CI01] E. Celledoni and A. Iserles, *Methods for the approximation of the matrix exponential in a Lie-algebraic setting*, IMA J. Numer. Anal. **21** (2001), no. 2, 463–488.
- [dC92] M. P. do Carmo, *Riemannian geometry*, Mathematics: Theory & Applications, Birkhäuser Boston Inc., Boston, MA, 1992, Translated from the second Portuguese edition by Francis Flaherty.
- [DN04] J.-P. Dedieu and D. Novitsky, *Symplectic methods for the approximation of the exponential and the Newton sequence on Riemannian submanifolds*, submitted to the Journal of Complexity, 2004.
- [DPM03] Jean-Pierre Dedieu, Pierre Priouret, and Gregorio Malajovich, *Newton's method on Riemannian manifolds: covariant alpha theory*, IMA J. Numer. Anal. **23** (2003), no. 3, 395–419.
- [DV00] J. Dehaene and J. Vandewalle, *New Lyapunov functions for the continuous-time QR algorithm*, Proceedings CD of the 14th International Symposium on the Mathematical Theory of Networks and Systems (MTNS2000), Perpignan, France, July 2000, 2000.
- [EAS98] A. Edelman, T. A. Arias, and S. T. Smith, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix Anal. Appl. **20** (1998), no. 2, 303–353.
- [Gab82] D. Gabay, *Minimizing a differentiable function over a differential manifold*, Journal of Optimization Theory and Applications **37** (1982), no. 2, 177–219.
- [GLRT99] N. I. M. Gould, S. Lucidi, M. Roma, and Ph. L. Toint, *Solving the trust-region subproblem using the Lanczos method*, SIAM J. Optim. **9** (1999), no. 2, 504–525 (electronic).
- [GV96] G. H. Golub and C. F. Van Loan, *Matrix computations, third edition*, Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, 1996.

- [Hag01] W. W. Hager, *Minimizing a quadratic over a sphere*, SIAM J. Optim. **12** (2001), no. 1, 188–208 (electronic).
- [HM94] U. Helmke and J. B. Moore, *Optimization and dynamical systems*, Springer, 1994.
- [HP04] W. W. Hager and S. C. Park, *Global convergence of SSM for minimizing a quadratic over a sphere*, to appear in Math. Comp., 2004.
- [HT04] Knut Hüper and Jochen Trunpf, *Newton-like methods for numerical optimization on manifolds*, Proc. 38th IEEE Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, November 7–10, 2004, 2004.
- [LE00] R. Lippert and A. Edelman, *Nonlinear eigenvalue problems with orthogonal-ity constraints (Section 9.4)*, Templates for the Solution of Algebraic Eigenvalue Problems (Zhaojun Bai, James Demmel, Jack Dongarra, Axel Ruhe, and Henk van der Vorst, eds.), SIAM, Philadelphia, 2000, pp. 290–314.
- [Lue72] David G. Luenberger, *The gradient projection method along geodesics*, Management Sci. **18** (1972), 620–631.
- [Mah96] R. E. Mahony, *The constrained Newton method on a Lie group and the symmetric eigenvalue problem*, Linear Algebra Appl. **248** (1996), 67–89.
- [Man02] J. H. Manton, *Optimization algorithms exploiting unitary constraints*, IEEE Trans. Signal Process. **50** (2002), no. 3, 635–650.
- [MS84] J. J. Moré and D. C. Sorensen, *Newton’s method*, Studies in numerical analysis, MAA Stud. Math., vol. 24, Math. Assoc. America, Washington, DC, 1984, pp. 29–82.
- [MS85] A. Machado and I. Salavessa, *Grassmannian manifolds as subsets of Euclidean spaces*, Res. Notes in Math. **131** (1985), 85–102.
- [NW99] J. Nocedal and S. J. Wright, *Numerical optimization*, Springer Series in Operations Research, Springer-Verlag, New York, 1999.
- [OW00] B. Owren and B. Welfert, *The Newton iteration on Lie groups*, BIT **40** (2000), no. 1, 121–145.
- [Pol97] Elijah Polak, *Optimization*, Applied Mathematical Sciences, vol. 124, Springer-Verlag, New York, 1997, Algorithms and consistent approximations.
- [Sak96] T. Sakai, *Riemannian geometry*, Translations of Mathematical Monographs, no. 149, American Mathematical Society, 1996.
- [Shu86] M. Shub, *Some remarks on dynamical systems and numerical analysis*, Proc. VII ELAM. (L. Lara-Carrero and J. Lewowicz, eds.), Equinoccio, U. Simón Bolívar, Caracas, 1986, pp. 69–92.
- [Smi93] S. T. Smith, *Geometric optimization methods for adaptive filtering*, Ph.D. thesis, Division of Applied Sciences, Harvard University, Cambridge, Massachusetts, 1993.
- [Smi94] Steven T. Smith, *Optimization techniques on Riemannian manifolds*, Hamiltonian and gradient flows, algorithms and control, Fields Inst. Commun., vol. 3, Amer. Math. Soc., Providence, RI, 1994, pp. 113–136.

- [ST00] A. Sameh and Z. Tong, *The trace minimization method for the symmetric generalized eigenvalue problem*, J. Comput. Appl. Math. **123** (2000), 155–175.
- [Ste83] T. Steihaug, *The conjugate gradient method and trust regions in large scale optimization*, SIAM J. Numer. Anal. **20** (1983), 626–637.
- [Ste01] G. W. Stewart, *Matrix algorithms, vol II: Eigensystems*, Society for Industrial and Applied Mathematics, Philadelphia, 2001.
- [SW82] A. H. Sameh and J. A. Wisniewski, *A trace minimization algorithm for the generalized eigenvalue problem*, SIAM J. Numer. Anal. **19** (1982), no. 6, 1243–1259.
- [Toi81] Ph. L. Toint, *Towards an efficient sparsity exploiting Newton method for minimization*, Sparse Matrices and Their Uses (I. S. Duff, ed.), Academic Press, London, 1981, pp. 57–88.
- [Udr94] C. Udriște, *Convex functions and optimization methods on Riemannian manifolds*, Kluwer Academic Publishers, 1994.
- [Yan99] Y. Yang, *Optimization on Riemannian manifold*, Proceedings of the 38th Conference on Decision and Control, 1999.