2011

# Numerical Optimization Methods on Riemannian Manifolds

Chunhong Qi

THE FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

# NUMERICAL OPTIMIZATION METHODS ON RIEMANNIAN MANIFOLDS

By

CHUNHONG QI

A Dissertation submitted to the
Department of Mathematics
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded:
Spring Semester, 2011

The members of the committee approve the dissertation of Chunhong Qi defended on March 22, 2011.

Kyle A. Gallivan
Professor Directing Dissertation

Pierre-Antoine Absil
Professor Co-Directing Dissertation

Dennis Duke
University Representative

Gordon Erlebacher
Committee Member

M. Yousuff Hussaini
Committee Member

Giray Okten
Committee Member

Approved:

Philip L. Bowers, Chair, Department of Mathematics

Joseph Travis, Dean, College of Arts and Sciences

The Graduate School has verified and approved the above-named committee members.

# ACKNOWLEDGMENTS

I would like to give my sincere thanks to the following individuals who, without their help, this dissertation would never have been completed.

First, I would like to thank my advisor, Dr. Kyle Gallivan, not only for his providing careful and patient guidance as my advisor, but also for being a role model as a dedicated professor and researcher. I would like to thank my co-advisor, Dr. Pierre-Antoine Absil, who gave me detailed and delicate instructions in the proving of a series of theorems in this dissertation. I would also like to thank Dr. Dennis Duke, Dr. M. Yousuff Hussaini, Dr. Gordon Erlebacher and Dr. Giray Okten, my committee members, for their contributions and counsel on this dissertation.

Second, I give my great appreciations to my friends Dr.Yanzhao Cao, his wife Hongyu Yu and again Dr. M. Yousuff Hussaini and his wife for giving me lots of help and care in both my study and life in the process pursuing my Ph.D. degree at FSU.

I am grateful to my parents, my parents-in-law who gave me unselfish love and support in my life, especially in taking care of my son. Finally, my deep gratitude is given to my husband, Jianmin Bao, for his help, love, and support of my lifelong pursuit of knowledge.

# TABLE OF CONTENTS

List of Tables . . . . . . . . . . vi

List of Figures . . . . . . . . . . vii

Abstract . . . . . . . . . . viii

**1 INTRODUCTION 1**

1.1 The Problem of Optimization on a Manifold . . . . . . . . . . 1

1.2 Historical Context . . . . . . . . . . 2

1.3 Basic Principles . . . . . . . . . . 4

1.4 Research Overview and Thesis Statement . . . . . . . . . . 7

**2 RIEMANNIAN BFGS ALGORITHM 10**

2.1 History and an Overview . . . . . . . . . . 10

2.2 The General Form of the RBFGS Algorithm . . . . . . . . . . 12

2.3 The Riemannian Dennis-Moré Condition . . . . . . . . . . 14

2.4 Convergence Analysis of RBFGS Algorithm . . . . . . . . . . 20

**3 RIEMANNIAN BFGS ALGORITHM IMPLEMENTATION 34**

3.1 Transformations and Symmetry on a Submanifold of $\mathbb{R}^n$ . . . . . . . . . . 34

3.2 Vector Transport on a Submanifold of $\mathbb{R}^n$ . . . . . . . . . . 44

3.3 Matrix Representations of $\mathcal{B}_k$ on a Submanifold of $\mathbb{R}^n$ . . . . . . . . . . 46

3.4 Transport on the Unit Sphere . . . . . . . . . . 49

3.5 Transport on the compact Stiefel manifold $\mathrm{St}(p, n)$ . . . . . . . . . . 50

3.6 Transport on $\mathrm{OB}(n, N)$ . . . . . . . . . . 51

3.7 Implementation on the Grassmann Manifold Grass $(p, n)$ . . . . . . . . . . 53

**4 RIEMANNIAN ADAPTIVE REGULARIZATION USING CUBICS 55**

4.1 The Algorithm . . . . . . . . . . 55

4.2 Convergence Analysis . . . . . . . . . . 61

**5 EXPERIMENTS 79**

5.1 Problems . . . . . . . . . . 79

5.2 RARC Results . . . . . . . . . . 81

5.3 RBFGS Results . . . . . . . . . . 82

**6 CONCLUSIONS 90**

Bibliography . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 92
Biographical Sketch . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 94

# LIST OF TABLES

5.1 Vector transport Approach 1 vs. Approach 2 for Rayleigh quotient problem 84

5.2 Vector transport Approach 1 vs. Approach 2 for Procrustes problem . . . . 84

5.3 Vector transport Approach 1 vs. Approach 2 for Thomson problem . . . . . 84

5.4 RBFGS Isometric Vector transport vs. Parallel transport . . . . . . . . . . . 86

5.5 Non-isometric vs. Canonical-based isometric vs. Simple implementation . . . 88

5.6 RBFGS vector transports: nonisometric vs. canonical isometric (SVD) vs. isometric(QF) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 88

# LIST OF FIGURES

1.1 Vector transport. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

3.1 Orthogonal and oblique projections relating $t$ and $\tilde{t}$ . . . . . . . . . . . . . . . 39

3.2 Orthogonal and oblique projections relating $t$ and $\tilde{t}$ . . . . . . . . . . . . . . . 40

5.1 RARC convergence Rayleigh Quotient problem: n=100 . . . . . . . . . . . . . 82

5.2 RARC convergence on two Thomson problems . . . . . . . . . . . . . . . . . . 83

5.3 RARC and RGLTR convergence for the Procrustes problem: n=12, p=7 . . 83

5.4 RBFGS Parallel and Vector Transport for Procrustes. n=12, p=7. . . . . . . 86

5.5 RBFGS with 3 transports for Rayleigh quotient. n=100. . . . . . . . . . . . . . 87

5.6 Low-rank (rank 3)approximation problem on Grass$(p, n)$. n=12, p=9 . . . . 89

# ABSTRACT

This dissertation considers the generalization of two well-known unconstrained optimization algorithms for $\mathbb{R}^n$ to solve optimization problems whose constraints can be characterized as a Riemannian manifold. Efficiency and effectiveness are obtained compared to more traditional approaches to Riemannian optimization by applying the concepts of retraction and vector transport. We present a theory of building vector transports on submanifolds of $\mathbb{R}^n$ and use the theory to assess convergence conditions and computational efficiency of the Riemannian optimization algorithms. We generalize the BFGS method which is an highly effective quasi-Newton method for unconstrained optimization on $\mathbb{R}^n$. The Riemannian version, RBFGS, is developed and its convergence and efficiency analyzed. Conditions that ensure superlinear convergence are given.

We also consider the Euclidean Adaptive Regularization using Cubics method (ARC) for unconstrained optimization on $\mathbb{R}^n$. ARC is similar to trust region methods in that it uses a local model to determine the modification to the current estimate of the optimal solution. Rather than a quadratic local model and constraints as in a trust region method, ARC uses a parameterized local cubic model. We present a generalization, the Riemannian Adaptive Regularization using Cubics method (RARC), along with global and local convergence theory.

The efficiency and effectiveness of the RARC and RBFGS methods are investigated and their performance compared to the predictions made by the convergence theory via a series of optimization problems on various manifolds.

# CHAPTER 1

# INTRODUCTION

This dissertation investigates the generalization of algorithms for unconstrained optimization on $\mathbb{R}^n$ to Riemannian manifolds and their analysis, implementation, and evaluation. This is achieved by identifying key components of Riemannian optimization algorithms, analyzing the theoretical properties that influence the convergence of the associated algorithms, and developing novel algorithms and implementations that are significantly more efficient than simple generalizations from $\mathbb{R}^n$ while achieving rigorously guaranteed convergence. The dissertation is organized as follows. In Chapter 1 an overview of the optimization problem on Riemannian manifolds is given followed by a brief history of research on methods for optimization on manifolds and a summary of the basic principles upon which the associated algorithms are built. The chapter ends with an overview of the proposed research and the thesis statement investigated. Chapters 2, 3 and 4 present the details of the two main components of the dissertation: Riemannian manifold versions of the quasi-Newton Broyden-Fletcher-Goldfarb-Shanno algorithm (BFGS) and the Adaptive Regularization using Cubics algorithm (ARC). The discussions include convergence theory as well as algorithmic and implementation issues. Chapter 5 presents the evaluation of the effectiveness of the methods and compares predictions made by the theory to observed performance via numerical experiments.

## 1.1 The Problem of Optimization on a Manifold

Optimization on manifolds (also called **Riemannian optimization**) concerns finding an optimum (global, or more reasonably, local) of a real-valued function $f$ defined over a (smooth) manifold. Roughly speaking, a manifold is a set endowed with coordinate patches that overlap smoothly.

When the function $f$ is read through a coordinate system, it becomes a classical real-valued function, defined on an open subset of $\mathbb{R}^d$ (where $d$ is the dimension of the manifold), to which classical optimization techniques can be applied. There are, however, several reasons not to follow this route. The coordinate patches may not be available explicitly, or they may have an expression that is unwieldy in terms of required floating point operations or memory usage. There is also the issue of switching between the coordinate systems as the algorithm evolves over the manifold. Moreover, resorting to coordinate systems may destroy or hide some useful properties of the manifold $\mathcal{M}$ and the cost function $f$.

Optimization on manifolds is applicable in two broad situations:

1. Classical equality-constrained optimization problems of the form

$$\begin{aligned} &\min f(x) \\ \text{s.t. } & h(x) = 0, \end{aligned}$$

   where $h$ is such that $\{x : h(x) = 0\}$ is a submanifold of $\mathbb{R}^n$. For example, finding the best orientation of a solid object (a problem that appears in pose estimation) is a problem on the special orthogonal group $SO(3)$, which is a submanifold of $\mathbb{R}^{3\times 3}$.

   In view of their formulation, these problems can also be tackled by classical equality-constrained optimization methods. The manifold-based approach offers certain advantages over these methods:

   - All the iterates are feasible (i.e., they satisfy the constraints, $h(x) = 0$); this property is particularly useful when attempting to stop the iteration early.
   - Riemannian optimization algorithms usually enjoy convergence properties akin to unconstrained optimization algorithms. In a sense, these algorithms perform an unconstrained optimization over a constrained set.
   - There is no need to consider Lagrange multipliers or penalty functions. Riemannian optimization is also a way of avoiding the Maratos effect.
   - If $f$ is only defined on $h^{-1}(0)$, then classical infeasible (i.e., the iterates do not satisfy $h(x) = 0$) methods are not applicable.

2. Problems where the objective function has some continuous invariance properties that we want to eliminate for various reasons: efficiency; consistency; applicability of certain convergence results; avoid failure of certain algorithms, e.g., Newton's method, that do not behave satisfactorily in case of degeneracy. For example, a way to impose a low-rank constraint on a symmetric positive-semidefinite matrix $X$ is to factor it as $X = YY^T$, where $Y \in \mathbb{R}^{n\times k}$ with $n > k$, then $YQ$ represents the same matrix $X$ for all orthogonal $Q$.

Optimization on manifolds, therefore, can be thought of as an "informed" way of doing optimization when the cost function has certain invariance properties or when the constraint set possesses a nice smooth geometry.

Applications of Riemannian optimization abound in engineering and the sciences including areas such as: algorithmic questions pertaining to linear algebra, signal processing, data mining, statistical image analysis, financial mathematics, nanostructures, and model reduction of dynamical systems. We refer the reader, e.g., to [4] and the many references therein. Specific problems relevant to multiple applications to be used in our work are also mentioned below.

## 1.2 Historical Context

Even though optimization on manifolds is a relatively new field of research, the concept of optimizing a function over a manifold dates back to the work of Luenberger [19, 20] in the early 1970s, if not earlier. Luenberger mentions the idea of performing line search along geodesics, “which we would use if it were computationally feasible (which it definitely is not)”. This statement is not correct in all generality, as there are important manifolds (such as the sphere and the Grassmann manifold) where geodesics admit closed-form expressions. However, even in this case, Luenberger was correct, in the sense that computing the geodesics is rarely worth the effort: in most optimization algorithms on manifolds, an approximation of the geodesics (in a sense that will be specified later on in this text) is enough to guarantee that the desired convergence properties are achieved. Replacing classical mathematical objects found in Riemannian geometry (such as geodesics, Levi-Civita connections, parallel translation) by approximations of these objects, without losing crucial convergence properties of the algorithms, is one of the cornerstones of our group’s previous work and this dissertation.

Somewhat surprisingly, it is only in around the year 2002 that researchers started to recognize the importance of making room in the collection of optimization methods for a wide class of approximations of geodesics. Indeed, for roughly two decades, the researchers’ concern was chiefly of a theoretical nature: the central research question was to exploit differential-geometric objects in order to formulate optimization strategies on abstract nonlinear manifolds, where the notions of addition and multiplication by a scalar no longer exist. The first research paper to focus on optimization on manifolds was Gabay’s work [15] on minimizing a differentiable function over a differential manifold. Initially, this paper was barely noticed. (According to ISI Web of Knowledge, [15] received only 8 citations before the year 2000.) The area of optimization on manifolds started to gain wider popularity in the 1990s, notably with the seminal works of Helmke and Moore [16] and Edelman et al. [14]. (ISI records a total of 361 citations for [16], including 60 citations over the last two years.)

Gradually, the emphasis shifted towards making the manifold-based approach more practical and flexible, with particular consideration for the efficiency of the resulting numerical algorithms. Optimization on manifolds is now a very active area of research. Many manifold-based algorithms have been proposed or are under development, Ph.D. theses have been presented and are in preparation, and minisymposia and tutorial workshops are being organized. The recent book [4], co-authored by Co-advisor Absil, proposes an introduction to the area, with an emphasis on providing the necessary background in differential geometry instrumental to algorithmic development, and on guiding the reader through the concrete calculations that turn an abstract geometric algorithm into a numerical implementation. It includes a good summary of much of the recent work by the Co-advisors Gallivan and Absil on Riemannian manifolds including that in the dissertation of C. Baker that developed a complete theory, implemented a numerical library and analyzed the performance of a Riemannian trust-region family of methods [6]. Baker’s dissertation also contains a concise introduction to the basic elements of Riemannian optimization methods.

## 1.3 Basic Principles

### 1.3.1 Unconstrained Optimization on a Constrained Space

Roughly speaking, a manifold is a generalization of the Euclidean space $\mathbb{R}^n$ on which the notion of a differentiable scalar field still exists. One can think of a nonlinear manifold as a smooth, curved surface, even though this simple picture does not fully do justice to the generality of the concept. Retaining the notion of differentiability opens the way for preserving concepts such as gradient vector fields and derivatives of vector fields, which are instrumental in many well-known optimization methods in $\mathbb{R}^n$, such as steepest descent, Newton, trust regions or conjugate gradients.

Optimization on manifolds can be intuitively thought of as unconstrained optimization over a constrained search space. As such, optimization algorithms on manifolds are not fundamentally different from classical algorithms for unconstrained optimization in $\mathbb{R}^n$. Indeed, new optimization algorithms on manifolds are often obtained by starting from an algorithm for unconstrained optimization in $\mathbb{R}^n$, extracting the underlying concepts, and rewriting them in such a way that they are well-defined on abstract manifolds. Generally speaking, applying the techniques of optimization on manifolds to a given computational problem involves the following steps. First, one needs to rephrase the problem as an optimization problem on a manifold. Clearly, this is not possible for all problems, but examples abound (several examples are mentioned below, and many other examples can be found, e.g., in [4, 14, 16, 18]). Second, one needs to pick an optimization method, typically from the several classical optimization methods that have been formulated and analyzed for manifold search spaces. The final step is to turn the generic optimization method into a practical numerical algorithm. This entails choosing a representation of the manifold, e.g., encoding via a particular quotient manifold or embedded submanifold, and providing numerical expressions for a handful of differential-geometric objects, such as a Riemannian metric and a retraction. The compartmentalization of the representation of the elements of the manifold, the differential-geometric objects and the algorithm that uses them lends itself to the development of very general and quite powerful software. Generic prototype implementation of algorithms on a Riemannian manifold as well as more specific implementations exploiting the structure of particular problems and manifolds can be obtained, for example, from `http://www.math.fsu.edu/~cbaker/GenRTR`.

### 1.3.2 Analogues of Lines and Planes

Some basic intuition behind the adaptation of algorithms for unconstrained optimization in $\mathbb{R}^n$ can be seen by considering the fact that many have a basic step of $x_{k+1} = x_k + \alpha_k p_k$ where the direction vector $p_k$ may be determined first followed by a one-dimensional search to set the step $\alpha_k$ as in a line search method, or $\alpha_k p_k$ may be set by considering a local constrained optimization of a simplified model of the cost function as in a trust-region method. In either case, the main concern is the ability to generalize the notion of motion for some distance on a line given by a direction vector. Hence, much of the initial manifold work centered around the evaluation of geodesics and was accurately characterized by Luenberger's comment cited earlier.

As an example, consider Newton's method for finding a stationary point of a differentiable function $f$. In $\mathbb{R}^n$, the method reads

$$x_+ = x - (\mathrm{Hess}\, f(x))^{-1}\mathrm{grad}\, f(x),$$

where $x$ is the current iterate, $x_+$ is the new iterate, $\mathrm{grad}\, f(x) = \begin{bmatrix} \partial_1 f(x) & \ldots & \partial_n f(x) \end{bmatrix}^T$ is the gradient of $f$ at $x$ and $\mathrm{Hess}\, f(x)$ is the Hessian matrix of $f$ at $x$ defined by $(\mathrm{Hess}\, f(x))_{ij} = \partial_i \partial_j f(x)$. When $f$ is a function on a nonlinear Riemannian manifold $\mathcal{M}$, most of these operations become undefined. However, the notion of a gradient still exists on an abstract Riemannian manifold, and if one sees the Newton method as iteration that defines $x_+$ as $x+\eta$ where $\eta$ is the vector along which the derivative of the gradient is equal to the negative of the gradient, one is led to the following iteration

$$\begin{aligned} \nabla_\eta \mathrm{grad}\, f &= -\mathrm{grad}\, f(x) \\ x_+ &= \mathrm{Exp}_x(\eta), \end{aligned}$$

where $\nabla$ is the Levi-Civita connection and Exp is the Riemannian exponential. In the 1990s, this was considered “the” Newton iteration on manifolds; see, e.g., [25].

The significant change responsible for the renewed interest in manifold methods came with the work of Shub *et al.* [5]. There the Levi-Civita connection was relaxed to any affine connection, and the Riemannian exponential was relaxed to any **retraction**: a function mapping elements of the tangent space of $x_k$ back to a neighborhood on the manifold. A detailed proof that the resulting algorithm still has local quadratic convergence to the nondegenerate stationary points of $f$ can be found in [4, §6.2]. Several other classical optimization algorithms in $\mathbb{R}^n$ have been generalized to manifolds; those relevant to our proposed research include line-search methods [26], conjugate gradients [25], BFGS [15], various direct-search methods [13], and trust-region methods [2, 6].

The Riemannian Newton method exploiting the idea of a retraction worked in the tangent space to solve the Newton equation for the direction vector and its natural step length of a single step. The work by the Co-advisors Gallivan and Absil and their previous Ph.D. student C. Baker on the theory, implementation and application of the Riemannian trust-region (RTR) method [2, 6] and [4, Chapter 7.0] took this idea to its logical conclusion. Rather than looking for the analogue of a line on the curved space, their approach looked for a series of flat spaces and associated optimization problems to replace the optimization problem on the curved space. Of course, the tangent spaces of the iterates $x_k$ provide a natural series of flat spaces. The retraction is used not only to map tangent vectors back to the manifold, but also to lift the cost function $f(x)$ to the tangent spaces yielding the **lifted cost functions** $\hat{f}_{x_k}(\eta)$ where $\eta \in T_{x_k}\mathcal{M}$. Combining this with Riemannian analogues of the gradient and Hessian produces an algorithm that solves a series of unconstrained optimization problems in $\mathbb{R}^d$ (or at least reduces the lifted cost function sufficiently) via a trust-region method, retracts to the manifold, and decides on step acceptance or rejection and trust-region radius update by considering the relationship between the lifted cost function $\hat{f}_{x_k}(\eta)$, its local model $m_x(\eta)$, and the cost function $f(x)$. The paradigm is sufficient to describe many Riemannian optimization algorithms by simply replacing the notion of using a trust-region method to solve each local problem with other methods. They have shown that, despite using several lifted cost functions $\hat{f}_{x_k}(\eta)$ to define a series of problems, under mild assumptions on the retraction, the cost function and the solution of the local unconstrained optimization problems the method converges globally to the critical points of $f(x)$ and has local superlinear convergence. In practice, convergence is to a local minimizer. Convergence to a saddle point only occurred for carefully constructed malicious situations. This is due to the fact that the method is also shown to be a descent method, and local maxima and saddle points are unstable fixed points of the algorithm. Cubic local convergence was also proven and observed under special circumstances, e.g., when the cost function was symmetric around the local minimizer. This essentially generalizes the behavior of the trust-region method on $\mathbb{R}^n$. Implementations of the RTR method can be obtained from `http://www.math.fsu.edu/~cbaker/GenRTR/`.

### 1.3.3 Transport

When taking a step on a manifold $\mathcal{M}$ from a point $x \in \mathcal{M}$ along a vector $\eta_x \in T_x\mathcal{M}$, it is natural to think about following the geodesic curve $\gamma$ with initial velocity $\eta_x$ and define the new point as $\gamma(1)$. However, whereas geodesics admit closed-form expressions for some specific manifolds, in general, they are the solution of an ordinary differential equation, and are thus costly to compute accurately. Fortunately, as noted earlier, in most optimization algorithms one is content with first-order approximations of the geodesic. This prompted Shub *et al.* [5] to introduce the concept of retraction.

Quite similarly, when one has to subtract two tangent vectors $\xi_x$ and $\xi_y$ at two different points $x$ and $y = \mathrm{Exp}_x(\eta_x)$, it is natural to think about parallel translating one tangent vector to the foot of the other along the curve $t \mapsto \mathrm{Exp}_x(t\eta_x)$. Here again, apart from some specific manifolds where parallel translation admits a closed-form expression, in general, parallel translation requires solving an ordinary differential equation. This prompted the relaxation of the idea and the introduction of the concept of **vector transport**, of which parallel translation is a particular instance [4]. The definition below, illustrated in Figure 1.1, invokes the Whitney sum $TM \oplus TM$, which is defined as the set of all ordered pairs of tangent vectors with same foot.

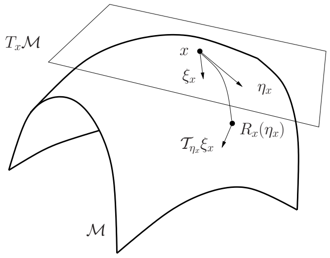

Figure 1.1: Vector transport.

**Definition 1.3.1.** *We define a* **vector transport** *on a manifold $\mathcal{M}$ to be a smooth mapping*

$$T\mathcal{M} \oplus T\mathcal{M} \to T\mathcal{M} : (\eta_x, \xi_x) \mapsto \mathcal{T}_{\eta_x}(\xi_x) \in T\mathcal{M}$$

*satisfying the following properties for all $x \in \mathcal{M}$.*

- *(Associated retraction) There exists a retraction $R$, called the* ***retraction associated with*** $\mathcal{T}$*, such that the following diagram commutes*

$$\begin{tikzcd}
(\eta_x, \xi_x) \arrow[r, "\mathcal{T}"] \arrow[d] & \mathcal{T}_{\eta_x}(\xi_x) \arrow[d, "\pi"] \\
\eta_x \arrow[r, "R"'] & \pi\left(\mathcal{T}_{\eta_x}(\xi_x)\right)
\end{tikzcd}$$

*where $\pi\left(\mathcal{T}_{\eta_x}(\xi_x)\right)$ denotes the foot of the tangent vector $\mathcal{T}_{\eta_x}(\xi_x)$.*

- *(Consistency) $\mathcal{T}_{0_x}\xi_x = \xi_x$ for all $\xi_x \in T_x\mathcal{M}$;*

- *(Linearity) $\mathcal{T}_{\eta_x}(a\xi_x + b\zeta_x) = a\mathcal{T}_{\eta_x}(\xi_x) + b\mathcal{T}_{\eta_x}(\zeta_x)$.*

The first point in Definition 1.3.1 means that $\mathcal{T}_{\eta_x}\xi_x$ is a tangent vector in $T_{R_x(\eta_x)}\mathcal{M}$, where $R$ is the retraction associated with $\mathcal{T}$. When it exists, $(\mathcal{T}_{\eta_x})^{-1}(\xi_{R_x(\eta_x)})$ belongs to $T_x\mathcal{M}$. If $\eta$ and $\xi$ are two vector fields on $\mathcal{M}$, then $(\mathcal{T}_\eta)^{-1}\xi$ is naturally defined as the vector field satisfying

$$\left((\mathcal{T}_\eta)^{-1}\xi\right)_x = \left(\mathcal{T}_{\eta_x}\right)^{-1}\left(\xi_{R_x(\eta_x)}\right).$$

It was shown in [4, §8.2.1] that when any vector transport is used in an approximate Newton method to find zeros of functions defined on a manifold where the Jacobian (or Hessian if in an optimization context) is approximated by finite differences, the resulting algorithm enjoys convergence properties akin to those of approximate Newton method in $\mathbb{R}^n$. As with the introduction of retraction to replace the exponential map, replacing parallel translation by the more general, and sometimes more efficient, concept of vector transport is a key part of our work developing efficient Riemannian optimization algorithms.

## 1.4 Research Overview and Thesis Statement

The development of a complete convergence theory for the RTR method and its successful implementation and application to important problems, by our group, is a significant step in the development of efficient and well-understood optimization algorithms for Riemannian manifolds. However, there is still room for improvement, in several ways. This dissertation concentrates on two: (i) Unconstrained optimization in $\mathbb{R}^n$ is still an active area of research. Several novel algorithms have appeared recently that need to be generalized and analyzed on manifolds. This is the case, for example, of the Adaptive Regularization using Cubics algorithm which, like trust-region methods, advances to a solution using a local model of the cost function [9, 10]. (ii) Some of optimization algorithms have not, or not fully, benefited from the possibility of relaxing certain differential-geometric objects (Riemannian exponential, parallel translation) to wider class of objects that leave leeway for more efficient implementations while preserving convergence properties of the optimization algorithms. We will combine retraction-based ideas with vector transport to generalize and improve important manifold versions of methods on $\mathbb{R}^n$. Of particular interest is the BFGS method, see for example, [12, 22].

### 1.4.1 Riemannian Broyden-Fletcher-Goldfarb-Shanno Algorithm

There are other manifold algorithms, such as conjugate gradients and secant methods, where parallel translation is used to combine two or more tangent vectors from distinct tangent spaces. We will combine retraction-based ideas with vector transport focusing, in particular, on the methods based on a manifold version of the secant condition used in many methods on $\mathbb{R}^n$ such as Riemannian generalizations of the BFGS method. While such generalizations have been proposed in the literature, they are generally based on heuristics and there is currently no convergence analysis that guarantees that the convergence properties associated with the use of parallel translation remain valid when it is replaced by any vector transport. A goal of this dissertation is to fill this gap.

An approximate Jacobian or Hessian at $x \in \mathcal{M}$ is a linear operator in the $d$-dimensional tangent space $T_x\mathcal{M}$. Secant methods in $\mathbb{R}^n$ construct an approximate Jacobian $A_{k+1}$ by imposing the secant equation

$$\xi_{x_{k+1}} - \xi_{x_k} = A_{k+1}\eta_k, \tag{1.1}$$

which can be seen as an underdetermined system of equations with $d^2$ unknowns. The remaining degrees of freedom in $A_{k+1}$ are specified according to some algorithm that uses prior information where possible and also preserves or even improves the convergence properties of the underlying Newton method.

The generalization of the secant condition (1.1) on a manifold $\mathcal{M}$ endowed with a vector transport $\mathcal{T}$ is

$$\xi_{x_{k+1}} - \mathcal{T}_{\eta_k}\xi_{x_k} = A_{k+1}[\mathcal{T}_{\eta_k}\eta_k], \tag{1.2}$$

where $\eta_k$ is the update vector at the iterate $x_k$, i.e., $R_{x_k}(\eta_k) = x_{k+1}$.

In the case where the manifold is Riemannian and $\xi$ is the gradient of a real-valued function $f$ of which a minimizer is sought, it is customary to require the following additional properties. Since the Hessian, $\operatorname{Hess} f(x)$, is symmetric (with respect to the Riemannian metric), one may require that the operator $A_k$ be symmetric for all $k$. Further, in order to guarantee that $\eta_k$ remains a descent direction for $f$, the updating formula may be required to generate a positive-definite operator $A_{k+1}$ whenever $A_k$ is positive-definite. BFGS on $\mathbb{R}^n$ satisfies these properties. This dissertation contains a complete theory addressing when this is possible, its consequences regarding guaranteeing convergence, assessing its necessity, and evaluating its efficiency.

### 1.4.2 Riemannian Adaptive Regularization Using Cubics

Adaptive Regularization using Cubics (ARC) is an unconstrained optimization algorithm recently proposed by Cartis *et al.* [9, 10]. The authors start from an optimization method introduced by Nesterov and Polyak [21] where the objective function is overestimated by a local cubic model, but they modify it in three ways to make it more practical. Cartis *et al.* provide a local and global convergence theory similar to that of trust-region methods, but, remarkably, they also have complexity bounds, and, even more remarkably, the numerical results for small-size problem are overall significantly better than with classical trust-region methods. Our objective is to generalize this method to abstract Riemannian manifolds, analyze its convergence and understand its efficiency tradeoffs. Generalizing the method to abstract manifolds is straightforward, as the operation is similar to the one that performed in [2, 6] for trust-region methods. Extending the convergence analysis and complexity results requires significantly more work. We develop such theory, and the performance of the algorithm is analyzed and compared with state-of-the-art algorithms such as our RTR family of methods.

### 1.4.3 Thesis Statement

To pursue the research goals set out above this dissertation asserts the following thesis:

1. The ARC method on $\mathbb{R}^n$ can be generalized for Riemannian optimization with:
    - convergence theory giving the sufficient conditions for superlinear convergence;
    - efficient implementations based on retractions on embedded submanifolds and quotient manifolds;
    - and demonstrations of its effectiveness on optimization test problems.
2. The BFGS method on $\mathbb{R}^n$ can be generalized for Riemannian optimization with:
    - convergence theory giving the sufficient conditions for superlinear convergence;
    - explanations of the performance effects related to the choice of vector transport, the effect of operator symmetry preservation, and the relationship to true Hessians of the cost function;
    - efficient implementations and analysis of the performance tradeoffs based on vector transport on embedded submanifolds and quotient manifolds;
    - and demonstrations of its effectiveness on optimization test problems.

# CHAPTER 2

# RIEMANNIAN BFGS ALGORITHM

## 2.1 History and an Overview

The BFGS algorithm is one of the most successful methods for unconstrained optimization (see [12, 22]), and it is natural that its generalization would be a topic of interest. However, as mentioned earlier, it is its use of an update to a linear transformation that approximates the evolution of the Hessian or its inverse that makes it particularly challenging on a Riemannian manifold.

The BFGS algorithm on $\mathbb{R}^n$ is given in Algorithm 1 in the form that updates, $B_k$, an approximation to the Hessian at $x_k$. An alternate form that updates an approximation to the inverse of the Hessian at $x_k$ is also used extensively.

**Algorithm 1** The BFGS algorithm on $\mathbb{R}^n$

```
Require: real-valued function $f$ on $\mathbb{R}^n$.
    Goal: Find a local minimizer of $f$ .
    Iutput: Initial iterate $\mathbf{x}_1 \in \mathbb{R}^n$, Hessian approximation $B_1 = I$
    Output: Sequence of iterates $x_k$.
for k =1, 2,... do
    1. Obtain $\eta_k$ by solving: $\eta_k = -B_k^{-1}\nabla f(\mathbf{x}_k)$.
    2. Perform a line search in the direction $\eta_k$ to find an appropriate scale $\alpha$ and update $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha\eta_k$.
    3. Define $s_k = \mathbf{x}_{k+1} - \mathbf{x}_k$ and $\mathbf{y}_k = \nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k)$.
    4. $B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}$
end for
```

The update is a simple rank-2 modification to $B_k$ that preserves symmetry and positive definiteness. It is the generalization of this update in an efficient and effective manner that is a major component in the success of the research discussed in this chapter.

Some work has been done on BFGS for manifolds. Gabay [15, §4.5] discussed a version using parallel transport on submanifolds of $\mathbb{R}^n$. Savas and Lim [23] apply a version on a product of Grassmann manifolds to the problem of best multilinear low-rank approximation of tensors. Brace and Manton [7] have a version on the Grassmann manifold for the problem of weighted low-rank approximations. They made a similar assertion to our thesis concerning the use of transport functions with significantly lower complexity than parallel translation. However, their efficient version was based on heuristics and no rigorous argument was given for the observed performance or performance expectations on other problems.

Gabay's Riemannian BFGS [15, §4.5] differs from the classical BFGS method in $\mathbb{R}^n$ (see, e.g., [22, Alg. 6.1]) in five key aspects: (i) The search space, to which the iterates $x_k$ belong, is a Riemannian submanifold $M$ of $\mathbb{R}^n$ specified by equality constraints; (ii) The search direction at $x_k$ is a tangent vector to $M$ at $x_k$; (iii) The update along the search direction is performed along the geodesic determined by the search direction; (iv) The usual quantities $s_k$ and $y_k$ that appear in the secant equation are tangent vectors to $M$ at $x_{k+1}$, obtained using the Riemannian parallel transport (i.e., the parallel transport induced by the Levi-Civita connection) along the geodesic. (v) The Hessian approximation $\mathcal{B}_k$ is a linear transformation of the tangent space $T_{x_k}M$ that gets updated using a generalized version of the BFGS update formula. This generalized formula specifies recursively how $\mathcal{B}_k$ applies to elements of $T_{x_k}M$.

We propose an algorithm model (or meta-algorithm), dubbed RBFGS, that subsumes Gabay's Riemannian BFGS method. Whereas Gabay's method is fully specified by the Riemannian manifold, the cost function, and the initial iterate, our RBFGS algorithm offers additional freedom in the choice of a retraction and a vector transport. This additional freedom affects points (iii) and (iv) above. For (iii), the curves along which the update is performed are specified by the retraction. For (iv), the Levi-Civita parallel transport is replaced by the more general concept of vector transport. If the retraction is selected as the Riemannian exponential and the vector transport is chosen to be the Levi-Civita parallel transport, then the RBFGS algorithm reduces to Gabay's algorithm (barring variations of minor importance, e.g., in the line-search procedure used).

The impact of the greater freedom offered by the RBFGS algorithm varies according to the manifold of interest. On the sphere, for example, the computational cost of the Riemannian exponential and the Levi-Civita parallel transport is reasonable, and there is not much to be gained by choosing computationally cheaper alternatives. In contrast, as we will show in numerical experiments, when the manifold is the Stiefel manifold, $\mathrm{St}(p, n)$, of orthonormal $p$-frames in $\mathbb{R}^n$, the improvement in computational time can be much more significant.

We also improve on Gabay's work by discussing the practical implementation of the algorithm. When the manifold $M$ is a submanifold of $\mathbb{R}^n$, we offer the alternatives of either representing the tangent vectors and the approximate Hessian using a basis in the tangent spaces, or relying on the canonical inclusion of $M$ in $\mathbb{R}^n$. The latter leads to representations of tangent vectors as $n$-tuples of real numbers and of the approximate Hessian as an $n \times n$ matrix. This approach may offer a strong advantage when the co-dimension of $M$ is sufficiently small.

The proposed RBFGS does not assume that $M$ is a submanifold of a Euclidean space. As such, it can be applied to quotient manifolds as well.

In this chapter we present the general form of the RBFGS algorithm and discuss key aspects of its convergence and implementation on embedded submanifolds and quotient manifolds. Specifically, we present a two-part convergence analysis. In Section 2.3, we propose a general Riemannian line search defined using retraction and vector transport of a local linear operator and develop sufficient conditions to guarantee superlinear convergence. In Section 2.4 we exploit the general line search superlinear result by giving two sets of sufficient conditions for global and superlinear convergence the general form of RBFGS using parallel transport.

In the remainder of the Chapter 2, we discuss the influence of the manifold, the retraction, the transport mechanism, and the implementation details on the performance of RBFGS. Particular attention is paid to a discussion of designing vector transport on an embedded submanifold of $\mathbb{R}^n$.

In Chapter 5, we illustrate performance and check predictions of our theory with a set of Riemannian optimization problems.

## 2.2 The General Form of the RBFGS Algorithm

The general form of the RBFGS algorithm is given in Algorithm 2. This form uses only abstract operations on the manifold, i.e., no specific choices of representation are assumed. Recall that, given a smooth scalar field $f$ on a Riemannian manifold $M$ with Riemannian metric $g$, the gradient of $f$ at $x$, denoted by $\operatorname{grad} f(x)$, is defined as the unique element of $T_xM$ that satisfies:

$$g_x(\operatorname{grad} f(x), \xi) = \mathrm{D}f(x)[\xi], \forall \xi \in T_xM. \tag{2.1}$$

In the rest of the dissertation the subscript indicating the element that defines the relevant tangent space is dropped when it is easily seen from the arguments.

The general form of RBFGS also makes use of the notion of the flat of an element of a tangent space which is also known as the index lowering function, musical isomorphism, and canonical isomorphism, see [1, p. 342]. This allows the update to the approximate Hessian to be written as an operator update with a form similar to that seen in BFGS.

**Definition 2.2.1.** *Let $(M, g)$ be a Riemannian manifold and let $X = X^i\partial_i$ be a vector field on $M$, where $\{\partial_i\}$ is a local frame for the tangent bundle $TM$. The **flat** of $X$ is defined by $X^\flat := g_{ij}X^i dx^j =: X_j dx^j$ where $\{dx^i\}$ is the dual coframe and the metric $g$ is defined locally, using Einstein notation, as $g = g_{ij}dx^i \otimes dx^j$. Equivalently, we have $X^\flat(Y) = g(X, Y)$ for all vectors $X$ and $Y$.*

In order to select a suitable stepsize, a generalization of the Wolfe conditions to a Riemannian manifold is required. The Generalized Wolfe conditions are on $M$ are

$$f(R_{x_k}(\alpha_k\eta_k)) \leq f(x_k) + c_1\alpha_k g(\operatorname{grad} f(x_k), \eta_k) \tag{2.2}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}(f(R_{x_k}(t\eta_k)))|_{t=\alpha_k} \geq c_2\frac{\mathrm{d}}{\mathrm{d}t}(f(R_{x_k}(t\eta_k)))|_{t=0} \tag{2.3}$$

with $0 < c_1 < c_2 < 1$. Condition (2.2) is the Generalized Armijo condition and (2.3) is the Generalized curvature condition.

Other generalizations of the Wolfe conditions are possible. For example, if the vector transport $\mathcal{T}$ is an isometry, then (2.3) can be replaced by:

$$g\Big(\big(\mathcal{T}_{\alpha_k\eta_k}\big)^{-1}\operatorname{grad} f(R_{x_k}(\alpha_k\eta_k)), \eta_k\Big) \geq c_2 g(\operatorname{grad} f(x_k), \eta_k). \tag{2.4}$$

This form transports the tangent vector that is in the tangent space of the potential next iterate $R_{x_k}(\alpha_k \eta_k)$ to $T_{x_k}M$ and applies the Euclidean curvature condition. For parallel transport and the exponential map as the retraction, conditions (2.3) and (2.4) are identical.

**Algorithm 2** General Form of RBFGS

```
1: Given: Riemannian manifold $M$ with Riemannian metric $g$; vector transport $\mathcal{T}$ on
   $M$ with associated retraction $R$; smooth real-valued function $f$ on $M$; initial iterate
   $\mathbf{x}_0 \in M$; initial Hessian approximation $\mathcal{B}_0$.
2: for k = 0, 1, 2, ... do
3:    Obtain $\eta_k \in T_{\mathbf{x}_k}M$ by solving $\mathcal{B}_k \eta_k = -\operatorname{grad} f(\mathbf{x}_k)$.
4:    Perform a line search to find $\alpha_k$ that satisfies conditions (2.2) and (2.3). Set $\mathbf{x}_{k+1} =$
      $R_{\mathbf{x}_k}(\alpha \eta_k)$.
5:    Define $s_k = \mathcal{T}_{\alpha\eta_k}(\alpha\eta_k)$ and $y_k = \operatorname{grad} f(\mathbf{x}_{k+1}) - \mathcal{T}_{\alpha\eta_k}(\operatorname{grad} f(\mathbf{x}_k))$.
6:    Define the linear operator $\mathcal{B}_{k+1} : T_{\mathbf{x}_{k+1}}M \to T_{\mathbf{x}_{k+1}}M$ by

      $\mathcal{B}_{k+1}p = \tilde{\mathcal{B}}_k p - \frac{g(s_k, \tilde{\mathcal{B}}_k p)}{g(s_k, \tilde{\mathcal{B}}_k s_k)} \tilde{\mathcal{B}}_k s_k + \frac{g(y_k, p)}{g(y_k, s_k)} y_k \quad \text{for all } p \in T_{\mathbf{x}_{k+1}}M, \qquad (2.5)$

      or equivalently

      $\mathcal{B}_{k+1} = \tilde{\mathcal{B}}_k - \frac{\tilde{\mathcal{B}}_k s_k (\tilde{\mathcal{B}}_k^* s_k)^\flat}{(\tilde{\mathcal{B}}_k^* s_k)^\flat (s_k)} + \frac{y_k y_k^\flat}{y_k^\flat (s_k)} \qquad (2.6)$

      where $a^\flat$ represents the flat of $a$, $*$ denotes the adjoint with respect to $g$, and

      $\tilde{\mathcal{B}}_k = \mathcal{T}_{\alpha\eta_k} \circ \mathcal{B}_k \circ (\mathcal{T}_{\alpha\eta_k})^{-1}. \qquad (2.7)$

7: end for
```

As with the BFGS algorithm, the RBFGS algorithm can also be reformulated to work with the inverse Hessian approximation $\mathcal{H}_k = \mathcal{B}_k{}^{-1}$ rather than with the Hessian approximation $B_k$. In this case, in Step 6 of RBFGS the following that holds for all $p \in T_{\mathbf{x}_{k+1}}M$ is used

$$\mathcal{H}_{k+1}p = \tilde{\mathcal{H}}_k p - \frac{g(y_k, \tilde{\mathcal{H}}_k p)}{g(y_k, s_k)} s_k - \frac{g(s_k, p_k)}{g(y_k, s_k)} \tilde{\mathcal{H}}_k y_k + \frac{g(s_k, p) g(y_k, \tilde{\mathcal{H}}_k y_k)}{g(y_k, s_k)^2} s_k + \frac{g(s_k, s_k)}{g(y_k, s_k)} p \quad (2.8)$$

or equivalently

$$H_{k+1} = \tilde{H}_k - s_k \, \frac{(\tilde{H}_k^* y_k)^\flat}{(y_k)^\flat (s_k)} - \tilde{H}_k y_k \, \frac{(s_k)^\flat}{(s_k)^\flat (y_k)} + s_k \frac{(y_k)^\flat (\tilde{H}_k^* y_k))(s_k)^\flat}{((y_k)^\flat (s_k))^2} + \frac{(s_k)^\flat (s_k)}{(s_k)^\flat (y_k)} \qquad (2.9)$$

where $a^\flat$ represents the flat of $a$, $*$ denotes the adjoint with respect to $g$, and

$$\tilde{\mathcal{H}}_k = \mathcal{T}_{\eta_k} \circ \mathcal{H}_k \circ (\mathcal{T}_{\eta_k})^{-1}. \qquad (2.10)$$

This yields a mathematically equivalent algorithm. It is useful because it makes it possible to cheaply compute an approximation of the inverse of the Hessian. This may make RBFGS advantageous even in the case where we have a cheap exact formula for the Hessian but not for its inverse or when the cost of solving linear systems is unacceptably high.

## 2.3 The Riemannian Dennis-Moré Condition

In this section, we generalize an important result from [12, Theorem 8.2.4] that guarantees the basic Riemannian line search algorithm $x_{k+1} = R_{x_k}(\eta_k)$, where $\eta_k = -B_k^{-1}F(x_k)$ converges superlinearly. The result, stated in Theorem 2.3.1, is used to prove superlinear convergence of RBFGS in next section.

In the discussions that follow, coordinate expressions in a neighborhood and in tangent spaces are used. For elements of the manifold, $v \in M$, $\hat{v} \in \mathbb{R}^d$ will denote the coordinates defined by a chart $\phi$ over a neighborhood $\mathcal{U}$, i.e., $\hat{v} = \phi(v)$ for $v \in \mathcal{U}$. Coordinate expressions, $\hat{F}(x)$, for elements, $F(x)$, of a vector field $F$ on $M$ are written in terms of the canonical basis of the associated tangent space, $T_xM$ via the coordinate vector fields defined by the chart $\phi$ (see, e.g., [4, §3.5]).

**Lemma 2.3.1.** *Let $\mathcal{U}$ be a compact coordinate neighborhood, and let the hat denote coordinate expressions. Then there is $c_2 > c_1 > 0$ such that, for all $x, y \in \mathcal{U}$, we have*

$$c_1\|\hat{x} - \hat{y}\| \leq dist(x, y) \leq c_2\|\hat{x} - \hat{y}\|,$$

*where $\|\cdot\|$ denotes the Euclidean norm.*

*Proof.* Proof of the first inequality:

Let $\Gamma_{\hat{x},\hat{y}}$ be the set of all smooth curves $\hat{\gamma}$ with $\hat{\gamma}(0) = \hat{x}$ and $\hat{\gamma}(1) = \hat{y}$. We have

$$\begin{aligned} dist(x, y) &= \inf_{\hat{\gamma}\in\Gamma_{\hat{x},\hat{y}}} \int_0^1 \sqrt{\dot{\hat{\gamma}}(t)^T \hat{G}(\hat{\gamma}(t))\dot{\hat{\gamma}}(t)}\mathrm{d}t \\ &\geq \sqrt{\min_{\hat{x}\in\hat{\mathcal{U}}} \lambda_{\min}(\hat{G}(\hat{x}))} \inf_{\hat{\gamma}\in\Gamma_{\hat{x},\hat{y}}} \int_0^1 \sqrt{\dot{\hat{\gamma}}(t)^T\dot{\hat{\gamma}}(t)}\mathrm{d}t \\ &\geq \sqrt{\min_{\hat{x}\in\hat{\mathcal{U}}} \lambda_{\min}(\hat{G}(\hat{x}))} \|\hat{y} - \hat{x}\| \end{aligned}$$

Proof of the second inequality:

Taking $\hat{\gamma}(t) = \hat{x} + t(\hat{y} - \hat{x})$, we have

$$\mathrm{dist}(x, y) \leq \int_0^1 \sqrt{\dot{\hat{\gamma}}(t)^T \hat{G}_{\hat{\gamma}(t)}\dot{\hat{\gamma}}(t)}dt \leq \sqrt{\lambda_{\min}} \int_0^1 \sqrt{\dot{\hat{\gamma}}(t)^T\dot{\hat{\gamma}}(t)}dt = \sqrt{\lambda_{\min}}\|\hat{x} - \hat{y}\|.$$

We have the proof by taking $c_1 = \sqrt{\min_{\hat{x}\in\hat{\mathcal{U}}} \lambda_{\min}(\hat{G}(\hat{x}))}$ and $c_2 = \sqrt{\lambda_{\min}}$.

□

**Lemma 2.3.2.** *Let $M$ be a Riemannian manifold endowed with a vector transport $\mathcal{T}$ and an associated retraction $R$, and let $x_* \in M$. Let $F$ be a smooth vector field on $M$. Then there is a neighborhood $\mathcal{U}$ of $x_*$ and $L > 0$ s.t., $\forall x, y \in \mathcal{U}$:*

$$\left| \|\mathcal{T}^{-1}_{R_y^{-1}x} F(x)\|_y^2 - \|F(x)\|_x^2 \right| \leq L\|F(x)\|_x^2\, dist(x, y).$$

*where $\|F(v)\|_v$ denotes the norm in $T_vM$ defined by the Riemannian metric.*

*Proof.* Let $L(y,x)$ denote $\mathcal{T}^{-1}_{R_y^{-1}x}$. We work in a coordinate chart and let the hat denote the coordinate expressions. We have

$$\begin{aligned} \left| \|\mathcal{T}^{-1}_{R_y^{-1}x} F(x)\|_y^2 - \|F(x)\|_x^2 \right| &= \left| \hat{F}(x)^T \big(\hat{L}(\hat{y},\hat{x})^T \hat{G}(\hat{y}) \hat{L}(\hat{y},\hat{x}) - \hat{G}(\hat{x})\big) \hat{F}(x) \right| && (2.11) \\ &\leq \|\hat{F}(x)\|^2 \|H(\hat{y},\hat{x})\| && (2.12) \\ &\leq c_1 \|\hat{F}(x)\|^2 \|\hat{y} - \hat{x}\| && (2.13) \\ &\leq c_2 \|F(x)\|_x^2 \mathrm{dist}(x,y). && (2.14) \end{aligned}$$

where $H(\hat{y},\hat{x}) = \hat{L}(\hat{y},\hat{x})^T \hat{G}(\hat{y}) \hat{L}(\hat{y},\hat{x}) - \hat{G}(\hat{x})$.

In (2.11), $\hat{G}(\hat{v})$ is the matrix expression of the Riemannian metric on $T_v M$ (see, e.g., [4, (3.29)]). In (2.12), $\|\hat{F}(x)\|$ denotes the classical Euclidean norm of $\hat{F}(x) \in \mathbb{R}^d$, where $d$ is the dimension of $M$, and $\|H(\hat{y},\hat{x})\|$ denotes the induced matrix norm (spectral norm). To get (2.13), take $\mathcal{U}$ bounded and observe that $H$ is smooth and that $H(\hat{x},\hat{x}) = 0$ for all $\hat{x}$. To get (2.14), use Lemma 2.3.1.

□

**Lemma 2.3.3.** *Under the assumption of Lemma 2.3.2, there is a neighborhood $\mathcal{U}$ of $x_*$ and $L' > 0$ s.t., $\forall x, y \in \mathcal{U}$ :*

$$\left| \|\mathcal{T}^{-1}_{R_y^{-1}x} F(x)\|_y - \|F(x)\|_x \right| \leq L' \|F(x)\|_x \, dist(x,y). \tag{2.15}$$

*Proof.* If $\|\mathcal{T}^{-1}_{R_y^{-1}x} F(x)\|_y + \|F(x)\|_x = 0$, then both sides of (2.15) are zero and the claim holds. Otherwise,

$$\begin{aligned} \left| \|\mathcal{T}^{-1}_{R_y^{-1}x} F(x)\|_y - \|F(x)\|_x \right| = \frac{\left| \|\mathcal{T}^{-1}_{R_y^{-1}x} F(x)\|_y^2 - \|F(x)\|_x^2 \right|}{\|\mathcal{T}^{-1}_{R_y^{-1}x} F(x)\|_y + \|F(x)\|_x} &\leq \frac{L \|F(x)\|^2 \mathrm{dist}(x,y)}{c_3 \|F(x)\|} \\ &\leq L' \|F(x)\|_x \mathrm{dist}(x,y) \end{aligned}$$

□

**Definition 2.3.1.** *(Nondegenerate zero) Let $F$ be a smooth vector field on a Riemannian manifold $M$. A point $x_* \in M$ is termed **a nondegenerate zero** of $F$ if $F(x_*) = 0$ and $\nabla_{\xi_{x_*}} F \neq 0$, $\forall \xi_{x_*} \neq 0 \in T_{x_*}M$ for some (and thus all) affine connection $\nabla$ on $M$.*

**Lemma 2.3.4** (Lemma 7.4.7, [4])**.** *Let $x \in M$, let $\mathcal{U}$ be a normal neighborhood of $x$, and let $\zeta$ be a $C^1$ tangent vector field on $M$, then, for all $y \in \mathcal{U}$,*

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

*where $\gamma$ is the unique geodesic in $\mathcal{U}$ satisfying $\gamma(0) = x$ and $\gamma(1) = y$, $P_\gamma^{b \leftarrow a}$ denotes parallel transport along $\gamma(t)$ from $a$ to $b$, and $\xi = \mathrm{Exp}_x^{-1} y = \gamma'(0)$.*

**Lemma 2.3.5.** *Let $F$ be a smooth vector field on a manifold $M$. Let $x_* \in M$ be a nondegenerate zero of $F$, then there exists a neighborhood $\mathcal{U}$ of $x_*$ and $c_0, c_1 > 0$ such that, for all $x \in \mathcal{U}$,*

$$c_0\, dist(x, x_*) \leq \|F(x)\| \leq c_1\, dist(x, x_*). \tag{2.16}$$

*Proof.* Let $\mathbb{D}F(x)$ denote the linear transformation of $T_xM$ defined by $\mathbb{D}F(x)[\xi_x] = \nabla_{\xi_x} F, \forall \xi_x \in T_xM$. Let $\mathcal{U}$ be a normal neighborhood of $x_*$ and, for all $x \in \mathcal{U}$, let $\gamma_x$ denote the unique geodesic in $\mathcal{U}$ satisfying $\gamma_x(0) = x_*$ and $\gamma_x(1) = x$.

From Taylor(Lemma 2.3.4), it follows that

$$P_{\gamma_x}^{0\leftarrow 1} F(x) = \mathbb{D}F(x_*)[\gamma_x'(0)] + \int_0^1 \big(P_{\gamma_x}^{0\leftarrow\tau} \mathbb{D}F(\gamma_x(\tau))[\gamma_x'(\tau)] - \mathbb{D}F(x_*)[\gamma_x'(0)]\big) \mathrm{d}\tau \tag{2.17}$$

Since $F$ is smooth and since $\|\gamma_x'(\tau)\| = \mathrm{dist}(x_*, x), \forall \tau \in [0, 1]$, we have the following bound for the integral:

$$\begin{aligned} & \Big\| \int_0^1 \big(P_{\gamma_x}^{0\leftarrow\tau} \mathbb{D}F(\gamma_x(\tau))[\gamma_x'(\tau)] - \mathbb{D}F(x_*)[\gamma_x'(0)]\big) \mathrm{d}\tau \Big\| \\ = & \Big\| \int_0^1 \big(P_{\gamma_x}^{0\leftarrow\tau} \circ \mathbb{D}F(\gamma_x(\tau)) \circ P_{\gamma_x}^{\tau\leftarrow 0} - \mathbb{D}F(x_*)\big) [\gamma_x'(0)] \mathrm{d}\tau \Big\| \\ \leq & \ \epsilon\big(\mathrm{dist}(x_*, x)\big) \mathrm{dist}(x_*, x), \end{aligned}$$

where $\lim_{t\to 0} \epsilon(t) = 0$.

Since $\mathbb{D}F(x_*)$ is nonsingular, it follows that $\exists c_0, c_1$ such that

$$2c_0\|\xi_{x_*}\| \leq \|\mathbb{D}F(x_*)[\xi_{x_*}]\| \leq \frac{1}{2}c_1\|\xi_{x_*}\|, \forall \xi_{x_*} \in T_{x_*}M \tag{2.18}$$

Take $\mathcal{U}$ sufficiently small such that $\epsilon\big(\mathrm{dist}(x_*, x)\big) < c_0$ and $< \frac{1}{2}c_1$ for all $x \in \mathcal{U}$.

Applying (2.17) yields

$$\begin{aligned} \|F(x)\| = \|P_{\gamma_x}^{0\leftarrow 1} F(x)\| & \leq \frac{1}{2}c_1 \mathrm{dist}(x_*, x) + \frac{1}{2}c_1 \mathrm{dist}(x_*, x) \\ & = c_1 \mathrm{dist}(x_*, x), \text{ for all } x \in \mathcal{U} \end{aligned}$$

and

$$\begin{aligned} \|F(x)\| = \|P_{\gamma_x}^{0\leftarrow 1} F(x)\| & \geq 2c_0 \mathrm{dist}(x_*, x) - c_0 \mathrm{dist}(x_*, x) \\ & = c_0 \mathrm{dist}(x_*, x), \text{ for all } x \in \mathcal{U}. \end{aligned}$$

□

**Lemma 2.3.6.** *Let $F$ be a smooth vector field on a Riemannian manifold $M$ endowed with a vector transport $\mathcal{T}$ and associated retraction $R$. Let $x_* \in M$ be a nondegenerate zero of $F$. Then there exists a neighborhood $\mathcal{V}$ of $0_{x_*} \in T_{x_*}M$ and $c_0, c_1 > 0$ such that, for all $\xi \in \mathcal{V}$,*

$$c_0\|\xi\| \leq \|\mathcal{T}_\xi^{-1}(F\big(R_{x_*}(\xi)\big))\| \leq c_1\|\xi\|. \tag{2.19}$$

*Proof.* Let $G(\xi) = \mathcal{T}_{\xi}^{-1}(F\big(R_{x_*}(\xi))\big)$ and $E(\epsilon) = G(\epsilon\xi)$. We have :

$$\begin{aligned}
\mathcal{T}_{\xi}^{-1}\Big(F\big(R_{x_*}(\xi))\Big) &= E(1) \\
&= E(0) + E'(0) + \int_0^1 E'(\tau) - E'(0) d\tau \qquad (2.20) \\
&= E(0) + \mathrm{D}G(0)\xi + \int_0^1 [\mathrm{D}G(\tau\xi) - \mathrm{D}G(0)]\xi d\tau \qquad (2.21) \\
&= 0 + \widetilde{\mathbb{D}}F(x_*)\xi + \int_0^1 [\mathrm{D}G(\tau\xi) - \mathrm{D}G(0)]\xi d\tau. \qquad (2.22)
\end{aligned}$$

The above (2.20) follows from the fundamental theorem $E(1) - E(0) = \int_0^1 E'(\tau)d\tau$, and (2.21) comes by the chain rule. Observe that $G$ is a function from $T_{x_*}M$ to $T_{x_*}M$, which are vector spaces, thus $\mathrm{D}G$ is the classical derivative of $G$. To get (2.22), observe that $E(0) = \mathcal{T}_{0_{x_*}}^{-1}\big(F(R_{x_*}(0_{x_*}))\big) = F(x_*) = 0$.

Let $\widetilde{\mathbb{D}}F(x)$ denote the derivative at $0_x$ of the function $T_xM \to T_xM : \eta \mapsto \mathcal{T}_{\eta}^{-1}F(R_x(\eta))$. Since $x_*$ is the nondegenerate zero of $F$, $\widetilde{\mathbb{D}}F(x_*)$ is invertible. We have

$$\|\xi\| = \|\widetilde{\mathbb{D}}F(x_*)^{-1}\widetilde{\mathbb{D}}F(x_*)\xi\| \leq \|\widetilde{\mathbb{D}}F(x_*)^{-1}\|\|\widetilde{\mathbb{D}}F(x_*)\xi\|.$$

i.e.

$$\|\widetilde{\mathbb{D}}F(x_*)\xi\| \geq \frac{\|\xi\|}{\|\widetilde{\mathbb{D}}F(x_*)^{-1}\|}. \qquad (2.23)$$

From (2.22), we have

$$\begin{aligned}
\|\mathcal{T}_{\xi}^{-1}(F\big(R_{x_*}(\xi))\big)\| &\geq \|\widetilde{\mathbb{D}}F(x_*)\xi\| - \|\int_0^1 [\mathrm{D}G(\tau\xi) - \mathrm{D}G(0)]\xi \mathrm{d}\tau\| \\
&\geq \frac{1}{\|\widetilde{\mathbb{D}}F(x_*)^{-1}\|}\|\xi\| - \int_0^1 \|\mathrm{D}G(\tau\xi) - \mathrm{D}G(0)\|\|\xi\|\mathrm{d}\tau \\
&\geq \frac{1}{\|\widetilde{\mathbb{D}}F(x_*)^{-1}\|}\|\xi\| - \int_0^1 \alpha\tau\|\xi\|\|\xi\|d\tau, \forall\xi \in \mathcal{V}, \qquad (2.24) \\
&\geq \frac{1}{\|\widetilde{\mathbb{D}}F(x_*)^{-1}\|}\|\xi\| - \frac{1}{2}\alpha\|\xi\|^2, \forall\xi \in \mathcal{V},
\end{aligned}$$

where (2.24) relies on Lipschitz continuity of DG, which holds by taking $\mathcal{V}$ bounded since $G$ is smooth. Taking $\mathcal{V}$ smaller if necessary, we have

$$\|\mathcal{T}_{\xi}^{-1}(F\big(R_{x_*}(\xi))\big)\| \geq \frac{1}{(\widetilde{\mathbb{D}F(x_*)})^{-1}}\|\xi\|, \forall\xi \in \mathcal{V}$$

Let $c_0 = \frac{1}{\|(\widetilde{\mathbb{D}F(x_*)})^{-1}\|}$, this concludes the first inequality in (2.19).

From (2.22), we have

$$\begin{aligned}\|\mathcal{T}_{\xi}^{-1}(F\big(R_{x_*}(\xi)\big))\| &\le \|\widetilde{\mathbb{D}}F(x_*)\xi\| + \|\int_0^1 [\mathrm{DG}(\tau\xi) - \mathrm{DG}(0)]\xi \mathrm{d}\tau\| \\ &\le \|\widetilde{\mathbb{D}}F(x_*)\|\|\xi\| + \int_0^1 \|\mathrm{DG}(\tau\xi) - \mathrm{DG}(0)\|\|\xi\|\mathrm{d}\tau \\ &\le \|\widetilde{\mathbb{D}}F(x_*)\|\|\xi\| + \int_0^1 \alpha\tau\|\xi\|\|\xi\|d\tau, \forall\xi \in \mathcal{V}, \\ &\le \|\widetilde{\mathbb{D}}F(x_*)\|\|\xi\| + \frac{1}{2}\alpha\|\xi\|^2, \forall\xi \in \mathcal{V} \\ &\le \|\widetilde{\mathbb{D}}F(x_*)\|\|\xi\| + \frac{1}{2}\alpha\|\xi\|, \forall\xi \in \mathcal{V}, (\|\xi\| \le 1).\end{aligned}$$

Let $c_1 = \|\widetilde{\mathbb{D}}F(x_*)\| + \frac{1}{2}\alpha$, this concludes the second inequality in (2.19).

□

Finally we note that since $c_1\|\widehat{R_x(\xi)}-\hat{x}\| \le \mathrm{dist}(x, R_x(\xi)) \le c_2\|\widehat{R_x\xi}-\hat{x}\|$, by Lemma 2.3.1, and $\widehat{R_x\xi} = \hat{\xi}+O(\hat{\xi}^2)$, we have that for the retraction $R$ there exist $\mu > 0, \tilde{\mu} > 0$ and $\delta_{\mu,\tilde{\mu}} > 0$ such that for $\forall x$ in a sufficiently small neighborhood of $x^*$ and $\xi \in T_xM, \|\xi\| \le \delta_{\mu,\tilde{\mu}}$

$$\frac{1}{\tilde{\mu}}\|\xi\| \le \mathrm{dist}(x, R_x(\xi)) \le \frac{1}{\mu}\|\xi\|. \tag{2.25}$$

This will be used throughout the remainder of the dissertation.

We are now in a position to state and prove the main result of a necessary and sufficient condition for superlinear convergence of a basic Riemannian line search algorithm.

**Theorem 2.3.1** (Riemannian Dennis-Moré Condition.)**.** *Let $M$ be a manifold endowed with a $C^2$ vector transport $\mathcal{T}$ and an associated retraction $R$. Let $F$ be a $C^2$ tangent vector field on $M$. Also let $M$ be endowed with an affine connection $\nabla$. Let $\mathbb{D}F(x)$ denote the linear transformation of $T_xM$ defined by $\mathbb{D}F(x)[\xi_x] = \nabla_{\xi_x}F$ for all tangent vectors $\xi_x$ to $M$ at $x$. Let $\{\mathcal{B}_k\}$ be a sequence of bounded nonsingular linear transformations of $T_{x_k}M$, where $k = 0, 1, \cdots, x_{k+1} = R_{x_k}(\eta_k)$, and $\eta_k = -\mathcal{B}_k^{-1}F(x_k)$. Assume that $\mathbb{D}F(x^*)$ is nonsingular, $x_k \neq x^*, \forall k$, and $\lim_{k\to\infty} x_k = x^*$. Then $\{x_k\}$ converges superlinearly to $x^*$ and $F(x^*) = 0$ if and only if*

$$\lim_{k\to\infty} \frac{\|[\mathcal{B}_k - \mathcal{T}_{\xi_k}\mathbb{D}F(x^*)\mathcal{T}_{\xi_k}^{-1}]\eta_k\|}{\|\eta_k\|} = 0, \tag{2.26}$$

*where $\xi_k \in T_{x^*}M$ is defined by $\xi_k = R_{x^*}^{-1}(x_k)$, i.e. $R_{x^*}(\xi_k) = x_k$.*

*Proof.* Assume first that (2.26) holds. Since, for $\xi_k \in T_{x^*}M$ and $\eta_k \in T_{x_k}M$ we have

$$\begin{aligned} 0 &= \mathcal{B}_k\eta_k + F(x_k) \\ &= (\mathcal{B}_k - \mathcal{T}_{\xi_k}\mathbb{D}F(x^*)\mathcal{T}_{\xi_k}^{-1})\eta_k + F(x_k) + \mathcal{T}_{\xi_k}\mathbb{D}F(x^*)\mathcal{T}_{\xi_k}^{-1}\eta_k, \end{aligned} \tag{2.27}$$

we have

$$\begin{aligned} -\mathcal{T}_{\eta_k}^{-1}F(x_{k+1}) &= (\mathcal{B}_k - \mathcal{T}_{\xi_k}\mathbb{D}F(x^*)\mathcal{T}_{\xi_k}^{-1})\eta_k + (-\mathcal{T}_{\eta_k}^{-1}F(x_{k+1}) + F(x_k) + \mathcal{T}_{\xi_k}\mathbb{D}F(x^*)\mathcal{T}_{\xi_k}^{-1}\eta_k) \\ &= (\mathcal{B}_k - \mathcal{T}_{\xi_k}\mathbb{D}F(x^*)\mathcal{T}_{\xi_k}^{-1})\eta_k + (-\mathcal{T}_{\eta_k}^{-1}F(x_{k+1}) + F(x_k) + \widetilde{\mathbb{D}}F(x_k)\eta_k) \\ &\quad +(\mathcal{T}_{\xi_k}\widetilde{\mathbb{D}}F(x^*)\mathcal{T}_{\xi_k}^{-1} - \widetilde{\mathbb{D}}F(x_k))\eta_k + \mathcal{T}_{\xi_k}(\mathbb{D}F(x^*) - \widetilde{\mathbb{D}}F(x^*))\mathcal{T}_{\xi_k}^{-1}\eta_k \end{aligned} \tag{2.28}$$

Recall that $\widetilde{\mathbb{D}}F(x)$ denotes the derivative at $0_x$ of the function $T_xM \to T_xM : \eta \mapsto \mathcal{T}_\eta^{-1}F(R_x(\eta))$, we have

$$\lim_{k\to\infty}\frac{\|(-\mathcal{T}_{\eta_k}^{-1}F(x_{k+1})+F(x_k)+\widetilde{\mathbb{D}}F(x_k)\eta_k)\|}{\|\eta_k\|}=0.$$

Since $F$ be $C^2$, we have

$$\lim_{k\to\infty}\frac{\|(\mathcal{T}_{\xi_k}\widetilde{\mathbb{D}}F(x^*)\mathcal{T}_{\xi_k}^{-1}-\widetilde{\mathbb{D}}F(x_k))\eta_k\|}{\|\eta_k\|}=0 \tag{2.29}$$

Since $\lim\limits_{k\to\infty}x_k = x^*$, we have $\lim\limits_{k\to\infty}\|\eta_k\| = 0$ and $\lim\limits_{k\to\infty}\|F(x_k)\| = 0$ if $\mathcal{B}_k$ is bounded. So $F(x^*) = 0$

From Proposition 5.5.6 in [4] which says that $\widetilde{\mathbb{D}}F(0_v) = \mathbb{D}F(x_v)$ if $v$ is a critical point, we have

$$\frac{\|\mathcal{T}_{\xi_k}(\mathbb{D}F(x^*)-\widetilde{\mathbb{D}}F(x^*))\mathcal{T}_{\xi_k}^{-1}\eta_k\|}{\|\eta_k\|}=0,$$

Thus (2.28) yields

$$\lim_{k\to\infty}\frac{\|\mathcal{T}_{\eta_k}^{-1}F(x_{k+1})\|}{\|\eta_k\|}=0. \tag{2.30}$$

From Lemma 2.3.6, we have

$$\|\mathcal{T}_{\xi_{k+1}}^{-1}F(x_{k+1})\| \geq \alpha\|\xi_{k+1}\|, \forall k \geq k_0 \tag{2.31}$$

where $\xi_{k+1} \in T_{x^*}M$ and $R_{x^*}(\xi_{k+1}) = x_{k+1}$.

$$\begin{aligned}
& \|\mathcal{T}_{\eta_k}^{-1}F(x_{k+1})\| && (2.32)\\
= \;& \|\mathcal{T}_{\eta_k}^{-1}F(x_{k+1})\| - \|F(x_k)\| + \|F(x_k)\| - \|\mathcal{T}_{\xi_{k+1}}^{-1}F(x_{k+1})\| + \|\mathcal{T}_{\xi_{k+1}}^{-1}F(x_k)\| \\
\geq \;& \|\mathcal{T}_{\xi_{k+1}}^{-1}F(x_k)\| - \Big|\|\mathcal{T}_{\eta_k}^{-1}F(x_{k+1})\| - \|F(x_k)\|\Big| - \Big|\|F(x_k)\| - \|\mathcal{T}_{\xi_{k+1}}^{-1}F(x_{k+1})\|\Big| \\
\geq \;& \alpha\|\xi_{k+1}\| - L'\|F(x_k)\|\mathrm{dist}(x_k,x_{k+1}) - L'\|F(x_k)\|\mathrm{dist}(x_*,x_{k+1}) && (2.33)\\
\geq \;& \alpha\|\xi_{k+1}\| - c_4\|\xi_{k+1}\|\big(\mathrm{dist}(x_k,x_{k+1}) + \mathrm{dist}(x_*,x_{k+1})\big), && (2.34)
\end{aligned}$$

where (2.33) follows from (2.31) with $k_0$ sufficiently large and Lemma 2.3.3, and (2.34) follows from Lemma 2.3.5 and (2.25).

We have also

$$1/\tilde{\mu}\|\eta_k\| \leq \mathrm{dist}(x_k,x_{k+1}) \leq \mathrm{dist}(x_k,x^*) + \mathrm{dist}(x_{k+1},x^*) \leq 1/\mu\|\xi_k\| + 1/\mu\|\xi_{k+1}\|,$$

that is

$$\|\eta_k\| \leq \tilde{\mu}/\mu(\|\xi_k\| + \|\xi_{k+1}\|).$$

We have

$$\begin{aligned}
0 \;&=\; \lim_{k\to\infty}\frac{\|\mathcal{T}_{\eta_k}^{-1}F(x_{k+1})\|}{\|\eta_k\|} \geq \lim_{k\to\infty}\frac{\alpha\|\xi_{k+1}\|}{\|\eta_k\|}\Big(1-\frac{c_4}{\alpha}\big(\mathrm{dist}(x_k,x_{k+1})+\mathrm{dist}(x_{k+1},x_*)\big)\Big)\\
&=\; \lim_{k\to\infty}\frac{\alpha\|\xi_{k+1}\|}{\|\eta_k\|} \geq \lim_{k\to\infty}\frac{\alpha\|\xi_{k+1}\|}{\tilde{\mu}/\mu(\|\xi_k\|+\|\xi_{k+1}\|)}\\
&=\; \lim_{k\to\infty}\frac{\alpha\|\xi_{k+1}\|/\|\xi_k\|}{\tilde{\mu}/\mu(1+\|\xi_{k+1}\|/\|\xi_k\|)}.
\end{aligned}$$

Hence

$$\lim_{k\to\infty} \frac{\|\xi_{k+1}\|}{\|\xi_k\|} = 0.$$

This is superlinear convergence and this concludes the if portion of the proof.

Conversely, assume that $\{x_k\}$ converges superlinearly to $x^*$ and $F(x^*) = 0$. Since

$$\frac{\|\mathcal{T}_{\eta_k}^{-1} F(x_{k+1})\|}{\operatorname{dist}(x_{k+1}, x_k)} = \frac{\|\mathcal{T}_{\eta_k}^{-1} F(x_{k+1}) - \mathcal{T}_{\xi_k} F(x^*)\|}{\operatorname{dist}(x_k, x^*)} \cdot \frac{\operatorname{dist}(x_k, x^*)}{\operatorname{dist}(x_{k+1}, x_k)}, \tag{2.35}$$

$$\text{from } \left| \frac{\operatorname{dist}(x_{k+1}, x_k)}{\operatorname{dist}(x_k, x^*)} - \frac{\operatorname{dist}(x_k, x^*)}{\operatorname{dist}(x_k, x^*)} \right| \leq \frac{\operatorname{dist}(x_{k+1}, x^*)}{\operatorname{dist}(x_k, x^*)},$$

we have

$$\lim_{k\to\infty} \frac{\operatorname{dist}(x_{k+1}, x_k)}{\operatorname{dist}(x_k, x^*)} = 1. \tag{2.36}$$

(2.36) and the hypothesis on $\mathbb{D}F$ implies (2.30) holds. It then follows from (2.27) that (2.26) is satisfied. This concludes the only if portion of the proof. □

## 2.4 Convergence Analysis of RBFGS Algorithm

In this section, we generalize the two main convergence theorems in the literature for BFGS in $\mathbb{R}^n$ to RBFGS on a Riemannian manifold $M$. These results generalize earlier work by Gabay [15] who gave an outline of a proof of superlinear convergence of RBFGS based on parallel transport on a submanifold of $\mathbb{R}^n$. Specifically, we show under some reasonable assumptions and the requirement that parallel transport is used, the sequence created by Algorithm 2 converges globally to the minimizer of a convex cost function, Theorem 2.4.3, and with a few more assumptions achieves local superlinear convergence, Theorem 2.4.5, for a more general cost function. The work in this section is strongly related to the proofs of the related results given by Dennis and Schnabel [12] and Nocedal and Wright [22]. Our proofs follow their outlines closely when possible with all of the basic objects and properties promoted appropriately to a Riemannian manifold.

A key assumption made for the two main results in their form below is that parallel transport is used. This is more restrictive than the result above in Theorem 2.3.1 where an isometric vector transport was allowed. In fact, our experiments provide substantial evidence that, in practice, both isometric and nonisometric vector transport achieve superlinear convergence with RBFGS. Theorem 2.3.1 is probably a key part of the explanation for this behavior.

### 2.4.1 The global convergence of RBFGS

For BFGS in $\mathbb{R}^n$ given some basic assumptions, preserving the symmetric positive-definiteness when updating the matrix (or its inverse) that defines the basic step is a sufficient condition to achieve global convergence for a convex cost function and local superlinear convergence for a general cost function. In Sections 2.4.1 and 2.4.2, we show similar results for the update of linear transformation $\mathcal{B}_k$ on $T_{x_k}M$ to linear transformation $\mathcal{B}_{k+1}$ on

$T_{x_{k+1}}M$ in the general form of RBFGS. The arguments are expressed in terms of general linear transformations on and between tangent spaces and are not dependent on particular choices of bases. Lemma 2.4.1 is a generalization of the [12, Lemma 9.2.1] to a Riemannian manifold. It is used to justify the update step of the Algorithm 2 and to show that it preserves the positive-definiteness and self-adjointness of all $\mathcal{B}_k$ when the vector transport used is an isometry.

It is straightforward to prove that the linear transformation $\mathcal{B} : T_xM \to T_xM$ is self-adjoint and positive definite with respect to the Riemannian metric, $g$, if and only if there exists some invertible linear transformation $\mathcal{J} : T_xM \to T_xM$ such that

$$\mathcal{B} = \mathcal{J}\mathcal{J}^* \tag{2.37}$$

where $\mathcal{J}^*$ represents the adjoint operator of $\mathcal{J}$ and superposition denotes composition of transformations. In fact, there may be more than one such $\mathcal{J}$. When discussing linear transformations in terms of a matrix on $\mathbb{R}^n$ typically some normalization for the matrix $J$ where $B = JJ^T$ is chosen such as lower triangular with positive diagonal elements, i.e., the Cholesky factorization.

**Lemma 2.4.1.** *Let $s_k, y_k \in T_{x_{k+1}M}$ be as defined in Algorithm 2, $s_k \neq 0$ and assume $\mathcal{T}_\eta$ represents an isometric vector transport in direction $\eta$. Let $\{\mathcal{B}_k\}$ be a sequence of bounded invertible linear transformation of $T_{x_k}M$, where $k = 0, 1, \cdots$. If $\mathcal{B}_k$ on $T_{x_k}M$ is self-adjoint and positive definite with respect to the Riemannian metric, $g$, then there exists an invertible linear transformation, $\mathcal{J}_{k+1}$, on $T_{x_{k+1}}M$ such that*

$$y_k = \mathcal{J}_{k+1}\mathcal{J}_{k+1}^* s_k \tag{2.38}$$

$$\textit{if and only if } g(s_k, y_k) > 0. \tag{2.39}$$

*Proof.* Suppose there is an invertible linear transformation $\mathcal{J}_{k+1}$ on $T_{x_{k+1}}M$ such that

$$\mathcal{J}_{k+1}\mathcal{J}_{k+1}^* s_k = y_k.$$

If $v_k = \mathcal{J}_{k+1}^* s_k$ then

$$0 < g(v_k, v_k) = g(\mathcal{J}_{k+1}^* s_k, \mathcal{J}_{k+1}^* s_k) = g(s_k, \mathcal{J}_{k+1}\mathcal{J}_{k+1}^* s_k) = g(s_k, y_k).$$

which proves the only if portion of the Lemma.

Now assume that $g(s_k, y_k) > 0$. The linear transformation $\mathcal{B}_k$ is assumed self-adjoint and positive definite on $T_{x_k}M$. $\tilde{\mathcal{B}}_k = \mathcal{T}_{\eta_k}\mathcal{B}_k(\mathcal{T}_{\eta_k})^{-1}$ is a self-adjoint positive definite linear transformation on $T_{x_{k+1}}M$ since for any $\zeta_{k+1} \in T_{x_{k+1}}$,

$$\begin{aligned} g(\zeta_{k+1}, \tilde{\mathcal{B}}_k\zeta_{k+1}) &= g(\zeta_{k+1}, \mathcal{T}_{\eta_k}\mathcal{B}_k(\mathcal{T}_{\eta_k})^{-1}\zeta_{k+1}) \\ &= g(\mathcal{T}_{\eta_k}^*\zeta_{k+1}, \mathcal{B}_k(\mathcal{T}_{\eta_k})^{-1}\zeta_{k+1}) = g(\zeta_k, \mathcal{B}_k\zeta_k) > 0, \end{aligned}$$

where $\zeta_k = \mathcal{T}_{\eta_k}^{-1}\zeta_{k+1} = \mathcal{T}_{\eta_k}^*\zeta_{k+1} \in T_{x_k}M$ since $\mathcal{T}_{\eta_k}$ is an isometry. Furthermore, we know that

$$\tilde{\mathcal{B}}_k = \mathcal{T}_{\eta_k}\mathcal{B}_k(\mathcal{T}_{\eta_k})^{-1} = \mathcal{T}_{\eta_k}\mathcal{J}_k(\mathcal{T}_{\eta_k})^{-1}\mathcal{T}_{\eta_k}\mathcal{J}_k^*(\mathcal{T}_{\eta_k})^{-1} = \mathcal{T}_{\eta_k}\mathcal{J}_k(\mathcal{T}_{\eta_k})^{-1}\mathcal{T}_{\eta_k}\mathcal{J}_k^*(\mathcal{T}_{\eta_k})^* = \tilde{\mathcal{J}}_k\tilde{\mathcal{J}}_k^*$$

where $\tilde{\mathcal{J}}_k$ and $\mathcal{J}_k$ are invertible linear transformations on $T_{x_{k+1}}M$ and $T_{x_k}M$ respectively.

We want $\mathcal{J}_{k+1}$ to be a simple update of $\tilde{\mathcal{J}}_k$ such that

$$\mathcal{J}_{k+1}\mathcal{J}_{k+1}^* s_k = y_k \tag{2.40}$$

$$\mathcal{J}_{k+1} v_k = y_k \tag{2.41}$$

$$\mathcal{J}_{k+1}^* s_k = v_k \tag{2.42}$$

Taking $\mathcal{J}_{k+1} = \tilde{\mathcal{J}}_k$ and $v_k = \tilde{\mathcal{J}}_k^{-1} y_k$ satisfies (2.41) but in general not (2.42). However, for almost any $v_k$ there exists a linear transformation $\mathcal{E}(v_k)$ on $T_{x_{k+1}}M$ such that

$$\mathcal{J}_{k+1} v_k = (\tilde{\mathcal{J}}_k + \mathcal{E}(v_k))v_k = y_k$$

One such low complexity transformation is

$$\mathcal{E}(v_k) = \frac{r_k v_k^\flat}{g(v_k, v_k)}, \quad \text{where} \quad r_k = y_k - \tilde{\mathcal{J}}_k v_k$$

To satisfy (2.42), $v_k$ must satisfy

$$v_k = \mathcal{J}_{k+1}^* s_k = (\tilde{\mathcal{J}}_k + \mathcal{E}(v_k))^* s_k = \tilde{\mathcal{J}}_k^* s_k + v_k \frac{r_k^\flat(s_k)}{g(v_k, v_k)} \tag{2.43}$$

the derivation of which uses the following relations:

$$(ab^\flat)^* = ba^\flat, (La)^\flat = a^\flat L^*.$$

This can only be satisfied if

$$v_k = \alpha_k \tilde{\mathcal{J}}_k^* s_k \text{ for some } \alpha_k \in \mathbb{R}. \tag{2.44}$$

Substituting (2.44) into (2.43) and simplifying yields

$$\alpha_k^2 = \frac{g(y_k, s_k)}{g(s_k, \tilde{\mathcal{B}}_k s_k)} \tag{2.45}$$

which has a real solution if and only if $g(y_k, s_k) > 0$, since $g(s_k, \tilde{\mathcal{B}}_k s_k) > 0$. Choosing the positive root yields

$$v_k = +\Big(\frac{g(y_k, s_k)}{g(s_k, \tilde{\mathcal{B}}_k s_k)}\Big)^{1/2} \tilde{\mathcal{J}}_k^* s_k. \tag{2.46}$$

It is easily shown that $\mathcal{J}_{k+1}$ is invertible. Therefore, $\mathcal{B}_{k+1} = \mathcal{J}_{k+1}\mathcal{J}_{k+1}^*$ is self-adjoint and positive definite with respect to the Riemannian metric $g$. □

It remains to show the relationship between the update considered in Lemma 2.4.1 is equivalent to the update given in Algorithm 2.

**Lemma 2.4.2.** *Using the notation and assumptions of Lemma 2.4.1, the sequence of linear transformations $\mathcal{B}_k$ defined by the Lemma is the same as the the sequence defined by Algorithm 2.*

*Proof.* Starting with the definition of the update from Lemma 2.4.1 yields

$$\begin{aligned}
\alpha_k &= +\Big(\frac{y_k^\flat(s_k)}{(\tilde{B}_k^* s_k)^\flat(s_k)}\Big)^{1/2} \\
\mathcal{B}_{k+1} &= \mathcal{J}_{k+1}\mathcal{J}_{k+1}^* \\
&= \Big(\tilde{\mathcal{J}}_k + \frac{(y_k - \tilde{\mathcal{J}}_k v_k)v_k^\flat}{g(v_k, v_k)}\Big)\Big((\tilde{\mathcal{J}}_k)^* + v_k\frac{(y_k - \tilde{\mathcal{J}}_k v_k)^\flat}{g(v_k, v_k)}\Big) \\
&= \tilde{\mathcal{J}}_k(\tilde{\mathcal{J}}_k)^* + \frac{\tilde{\mathcal{J}}_k v_k(y_k^\flat - v_k^\flat(\tilde{\mathcal{J}}_k)^*)}{v_k^\flat v_k} + \frac{(y_k - \tilde{\mathcal{J}}_k v_k)v_k^\flat(\tilde{\mathcal{J}}_k)^*}{v_k^\flat v_k} + \frac{(y_k - \tilde{\mathcal{J}}_k v_k)v_k^\flat v_k(y_k^\flat - v_k^\flat(\tilde{\mathcal{J}}_k)^*)}{(v_k^\flat v_k)^2} \\
&= \tilde{\mathcal{B}}_k + \frac{y_k y_k^\flat - \tilde{\mathcal{J}}_k v_k v_k^\flat(\tilde{\mathcal{J}}_k)^*}{v_k^\flat v_k} \\
&= \tilde{\mathcal{B}}_k + \frac{y_k y_k^\flat - \tilde{\mathcal{J}}_k \alpha_k \tilde{\mathcal{J}}_k^* s_k \alpha_k(\tilde{\mathcal{J}}_k^* s_k)^\flat \tilde{\mathcal{J}}_k^*}{\alpha_k(\tilde{\mathcal{J}}_k^* s_k)^\flat \alpha_k \tilde{\mathcal{J}}_k^* s_k} \\
&= \tilde{\mathcal{B}}_k - \frac{\tilde{\mathcal{B}}_k s_k(\tilde{\mathcal{B}}_k^* s_k)^\flat}{(\tilde{\mathcal{B}}_k^* s_k)^\flat(s_k)} + \frac{y_k y_k^\flat}{y_k^\flat(s_k)},
\end{aligned}$$

which is identical to the update of Algorithm 2. □

We have therefore shown that Algorithm 2 produces a series of linear transformations $\mathcal{B}_k$ on $T_{x_k}M$ that are all self-adjoint and positive definite with respect to the Riemannian metric $g$ if an isometric vector transport is used to define the update. Note that we have not bounded the condition number of $\mathcal{B}_k$.

If we restrict the algorithm considered then global convergence to a set of critical points can be shown. If we restrict the cost function somewhat we can guarantee global convergence to a minimizer.

**Definition 2.4.1** ([4], Definition 7.4.3)**.** ***(Lipschitz continuous differentiability)*** *Assume that $(M, g)$ has a positive injectivity radius $i(M) > 0$. A real function $f$ on $M$ is Lipschitz continuous differentiable if it is differentiable and if there exists $\beta_1$ such that, for all $x, y$ in $M$ with dist$(x, y) < i(M)$, it holds that*

$$\|P_\alpha^{0\leftarrow 1}\mathrm{grad}\, f(y) - \mathrm{grad}\, f(x)\|_x \leq L\; dist(y, x), \tag{2.47}$$

*where $\alpha$ is the unique minimizing geodesic with $\alpha(0) = x$ and $\alpha(1) = y$. Note that* (2.47) *is symmetric in $x$ and $y$. It follows that*

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

Since we enforce the Wolfe conditions, we have the following strong statement about the angles between the direction vectors and gradients at each step that generalizes a result of Zoutendijk as given in [22, Theorem 3.2] to Riemannian manifolds. In this case, we assume that the Riemannian line search algorithm uses the exponential map as the retraction to define the next iterate, $x_{k+1}$, and use the Lipschitz condition in Definition 2.4.1 which is defined in terms of parallel transport. No restriction is placed on the manner in which the direction vector $\eta_k$ is generated beyond the assumptions given in the theorem.

**Theorem 2.4.1** (Riemannian Exponential Map Zoutendijk Condition.)**.** *Consider any iteration of form $x_{k+1} = \mathrm{Exp}_{x_k}(\alpha_k \eta_k)$, where $\eta_k$ is a descent direction and $\alpha_k$ satisfies the Wolfe conditions* (2.2) *and* (2.3)*. Suppose that $f$ is bounded below on $M$ and that $f$ is continuously differentiable in an open set $\mathcal{N}$ containing the level set $\mathcal{L} = \{x : f(x) \leq f(x_0)\}$, where $x_0$ is the starting point of the iteration. Assume also that the gradient* $\mathrm{grad}\, f$ *is Lipschitz continuous on $\mathcal{N}$, then*

$$\sum_{k\geq 0} \cos^2 \theta_k \|\mathrm{grad}\, f(x_k)\|^2 < \infty, \text{ where } \cos\theta_k = \frac{-g(\mathrm{grad}\, f(x_k), \eta_k)}{\|\mathrm{grad}\, f(x_k)\| \|\eta_k\|}. \tag{2.48}$$

*Proof.* Define $\gamma_{\alpha_k \eta_k}(t) = \mathrm{Exp}_{x_k}(t\alpha_k \eta_k)$. Since the retraction is the exponential map, the curvature condition (2.3) is equivalent to

$$g\Big(P^{0\leftarrow 1}_{\gamma_{\alpha_k\eta_k}} \mathrm{grad}\, f(\mathrm{Exp}_{x_k}(\alpha_k \eta_k)) - \mathrm{grad}\, f(x_k), \eta_k\Big) \geq (c_2 - 1) g(\mathrm{grad}\, f(x_k), \eta_k),$$

where $P^{1\leftarrow 0}_{\gamma_{\alpha_k\eta_k}}$ is parallel transport. The righthand inequality of (2.25) holds globally when working with the exponential map and parallel transport along the geodesic. This and the Lipschitz condition (2.47) imply that

$$g\Big(P^{0\leftarrow 1}_{\gamma_{\alpha_k\eta_k}} \mathrm{grad}\, f(\mathrm{Exp}_{x_k}(\alpha_k \eta_k)) - \mathrm{grad}\, f(x_k), \eta_k\Big) \leq \alpha_k (L/\mu) \|\eta_k\|^2.$$

By combining these two inequalities, we have

$$\alpha_k \geq \frac{(c_2 - 1)\mu}{L} \frac{g(\mathrm{grad}\, f(x_k), \eta_k)}{\|\eta_k\|^2}. \tag{2.49}$$

Substituting (2.49) into the first Wolfe condition (2.2), we obtain

$$f(x_{k+1}) \leq f(x_k) - c_1 \frac{(1 - c_2)\mu}{L} \frac{g(\mathrm{grad}\, f(x_k), \eta_k)^2}{\|\eta_k\|^2}$$

This is

$$f(x_{k+1}) \leq f(x_k) - c \cos^2 \theta_k \|\mathrm{grad}\, f(x_k)\|^2, \text{ where } c = c_1(1 - c_2)(\mu/L).$$

By summing this expression over all indices less than or equal to $k$, we have

$$f(x_{k+1}) \leq f(x_0) - c \sum_{j=0}^{k} \cos^2 \theta_j \|\mathrm{grad}\, f(x_j)\|^2. \tag{2.50}$$

since $f$ is bounded below, we have that $f(x_0) - f(x_{k+1})$ is less than some positive constant for all $k$, hence by taking limits in (2.50), we obtain

$$\sum_{k=0}^{\infty} \cos^2 \theta_k \|\mathrm{grad}\, f(x_k)\|^2 < \infty$$

□

It is obvious that $\lim_{k\to\infty} \|\text{grad} f(x_k)\| = 0$ provided that the search directions are never too close to orthogonality with the gradient, i.e. $\cos^2\theta_k$ stay away from 0. This implies that the algorithm would achieve global convergence to a set of stationary points. In practice, given the instability of an iteration at stationary points, such an algorithm is often effective at converging to an isolated minimizer when starting close enough, i.e., the global convergence result is used in a local manner.

We can now prove a generalization of [22, Theorem 8.5] that guarantees global convergence of Algorithm 2 by verifying that the search directions and stepsizes satisfy the conditions of Theorem 2.4.1. We follow a generalized version of the proof of [22, Theorem 8.5] using the notion of an average Riemannian Hessian and a function defined in terms of the trace and determinant of a linear transformation on a tangent space that is self-adjoint positive definite with respect to the Riemannian metric $g$. The proof below depends on the use of parallel transport in the definition of the average Riemannian Hessian and since the Exponential map version of the Zoutendijk condition is used the line search is restricted in the form of its determination of the next iterate.

The assumptions under which we consider the problem are:

**Assumptions 2.4.2.**

1. *The objective function $f$ is twice continuously differentiable .*

2. *The level set $\Omega = \{x \in M : f(x) \leq f(x_0)\}$ is geodesically convex. Let (M, g) be a Riemannian manifold. A subset C of M is said to be a geodesically convex set if, given any two points in C, there is a geodesic arc contained within C that joins those two points.*

3. *There exists positive constants $n$ and $N$ such that*

$$ng(z,z) \leq g(G(x)z,z) \leq Ng(z,z) \text{ for all } z \in T_xM \text{ and } x \in \Omega \tag{2.51}$$

*where $G(x)$ denotes the lifted Hessian $G(x) = \text{Hess}\,\widehat{f}_x(\xi) = \text{Hess}\,f(R_x(\xi))$.*

**Theorem 2.4.3.** *Let $x_0$ be starting point for which Assumptions 2.4.2 is satisfied and let $\mathcal{B}_0$ be any linear transformation on $T_{x_0}M$ that is self-adjoint and positive definite with respect to the Riemannian metric $g$. The sequence $\{x_k\}$ generated by Algorithm 2 using parallel transport and the exponential map as the retraction converges to the minimizer $x^*$ of $f$.*

*Proof.* Define the function $F : [0,1] \to T_{x_{k+1}}M : t \mapsto F(t) \in T_{x_{k+1}}$

$$F(t) := P^{1\leftarrow t}_{\gamma_{\eta_k}} \text{grad}\, f(\gamma_{\eta_k}(t)) \in T_{x_{k+1}}M, \tag{2.52}$$

$$F(1) = \text{grad}\, f(x_{k+1}), F(0) = P^{1\leftarrow 0}_{\gamma_{\eta_k}} \text{grad}\, f(x_k) \tag{2.53}$$

and denote

$$\gamma_{\eta_k} : t \to \text{Exp}(t\eta_k), \dot{\gamma}_{\eta_k}(t) = P^{t\leftarrow 0}_{\gamma_{\eta_k}} \dot{\gamma}_{\eta_k}(0). \tag{2.54}$$

We have $\gamma_{\eta_k}(1) = x_{k+1}$ and

$$F(t-\epsilon) = P^{1\leftarrow t}_{\gamma_{\eta_k}} P^{t\leftarrow t-\epsilon}_{\gamma_{\eta_k}} \text{grad}\, f(\gamma_{\eta_k}(t-\epsilon)). \tag{2.55}$$

$$\begin{aligned} F'(t) &= \lim_{\epsilon \to 0} \frac{F(t) - F(t-\epsilon)}{\epsilon} = -\frac{\mathrm{d}}{d\epsilon} F(t-\epsilon)\Big|_{\epsilon=0} \\ &= -P_{\gamma_{\eta_k}}^{1\leftarrow t} \frac{\mathrm{d}}{d\epsilon} P_{\gamma_{\eta_k}}^{t\leftarrow t-\epsilon} \mathrm{grad}\, f(\gamma_{\eta_k}(t-\epsilon))\Big|_{\epsilon=0} \\ &= -P_{\gamma_{\eta_k}}^{1\leftarrow t} \mathrm{Hess} f(\gamma_{\eta_k}(t)) \big[\frac{\mathrm{d}}{d\epsilon} \gamma_{\eta_k}(t-\epsilon)\big]\Big|_{\epsilon=0} \\ &= P_{\gamma_{\eta_k}}^{1\leftarrow t} \mathrm{Hess} f(\gamma_{\eta_k}(t)) \dot{\gamma}_{\eta_k}(t) \end{aligned}$$

From $F(1) - F(0) = \int_0^1 F'(t)dt$, we have

$$\begin{aligned} &\mathrm{grad}\, f(x_{k+1}) - P_{\gamma_{\eta_k}}^{1\leftarrow 0} \mathrm{grad}\, f(x_k) \\ =& \int_0^1 P_{\gamma_{\eta_k}}^{1\leftarrow t} \mathrm{Hess} f(\gamma_{\eta_k}(t))[\dot{\gamma}_{\eta_k}(t)]dt \\ =& \int_0^1 P_{\gamma_{\eta_k}}^{1\leftarrow t} \mathrm{Hess} f(\gamma_{\eta_k}(t)) P_{\gamma_{\eta_k}}^{t\leftarrow 1} dt \dot{\gamma}_{\eta_k}(1) \\ =& \int_0^1 P_{\gamma_{\eta_k}}^{1\leftarrow t} \mathrm{Hess} f(\gamma_{\eta_k}(t)) P_{\gamma_{\eta_k}}^{t\leftarrow 1} dt s_k, \end{aligned}$$

where $s_k = \dot{\gamma}_{\eta_k}(1)$.

If we define

$$\bar{G}_k = \int_0^1 P_{\gamma_{\eta_k}}^{1\leftarrow t} \mathrm{Hess} f(\gamma_{\eta_k}(t)) P_{\gamma_{\eta_k}}^{t\leftarrow 1} dt, \tag{2.56}$$

it follows that

$$y_k = \mathrm{grad}\, f(x_{k+1}) - P_{\gamma_{\eta_k}}^{1\leftarrow 0} \mathrm{grad}\, f(x_k) = \bar{G}_k \dot{\gamma}_{\eta_k}(1) = \bar{G}_k s_k \tag{2.57}$$

and using (2.51) and (2.57) we obtain

$$\frac{g(y_k, s_k)}{g(s_k, s_k)} = \frac{g(\bar{G}_k s_k, s_k)}{g(s_k, s_k)} \geq n. \tag{2.58}$$

Defining $z_k = \bar{G}_k^{1/2} s_k$ and using the relation (2.57), we have

$$\frac{g(y_k, y_k)}{g(y_k, s_k)} = \frac{g(\bar{G}_k s_k, \bar{G}_k s_k)}{g(\bar{G}_k s_k, s_k)} = \frac{g(z_k, \bar{G}_k z_k)}{g(z_k, z_k)} \leq N. \tag{2.59}$$

and

$$n_k = \frac{g(y_k, s_k)}{g(s_k, s_k)}, N_k = \frac{g(y_k, y_k)}{g(y_k, s_k)}. \tag{2.60}$$

Using (2.58) and (2.59), we have

$$n_k \geq n, N_k \leq N. \tag{2.61}$$

Recall that the values of the Riemannian metric, $g$, and the determinant and trace of a linear transformation on a finite dimensional vector space is independent of the basis (coordinates) used to represent the tangent space and the transformation. So we can work with the abstract operators $\mathcal{B}_k$, $\tilde{\mathcal{B}}_k$ and $\mathcal{B}_{k+1}$ and tangent vectors, $\xi_k$ and $\xi_{k+1}$ and rewrite expressions originally written in terms matrix and Euclidean vectors. Since $P_{\gamma_{\eta_k}}^{1 \leftarrow 0}$ is an isometry, we have

$$\begin{aligned}\operatorname{trace}(\widetilde{\mathcal{B}}_k) = \operatorname{trace}(P_{\gamma_{\eta_k}}^{1 \leftarrow 0} \circ \mathcal{B}_k \circ (P_{\gamma_{\eta_k}}^{1 \leftarrow 0})^{-1}) = \operatorname{trace}(\mathcal{B}_k)\\ \det(\widetilde{\mathcal{B}}_k) = \det(P_{\gamma_{\eta_k}}^{1 \leftarrow 0} \circ \mathcal{B}_k \circ (P_{\gamma_{\eta_k}}^{1 \leftarrow 0})^{-1}) = \det(\mathcal{B}_k).\end{aligned}$$

and $\widetilde{\mathcal{B}}_k$ is self-adjoint and positive definite with respect to the Riemannian metric $g$.

From Step 6 of Algorithm 2, we obtain that

$$\operatorname{trace}(\mathcal{B}_{k+1}) = \operatorname{trace}(\mathcal{B}_k) - \frac{\|\widetilde{\mathcal{B}}_k s_k\|^2}{g(s_k, \widetilde{\mathcal{B}}_k s_k)} + \frac{\|y_k\|^2}{g(y_k, s_k)}. \tag{2.62}$$

Also equation (8.45) in [22] can be converted from Euclidean coordinates for vectors, matrices and inner products to the general form

$$\det(\mathcal{B}_{k+1}) \quad = \quad \det(\mathcal{B}_k) \frac{g(y_k, s_k)}{g(s_k, \widetilde{\mathcal{B}}_k s_k)}. \tag{2.63}$$

If we define

$$\cos\theta_k = \frac{g(s_k, \widetilde{\mathcal{B}}_k s_k)}{\|s_k\| \|\widetilde{\mathcal{B}}_k s_k\|}, \quad q_k = \frac{g(s_k, \widetilde{\mathcal{B}}_k s_k)}{g(s_k, s_k)} \tag{2.64}$$

so that $\theta_k$ is the angle between $s_k$ and $\widetilde{\mathcal{B}}_k s_k$ then we obtain

$$\frac{\|\widetilde{\mathcal{B}}_k s_k\|^2}{g(s_k, \widetilde{\mathcal{B}}_k s_k)} = \frac{\|\widetilde{\mathcal{B}}_k s_k\|^2 \|s_k\|^2}{g(s_k, \widetilde{\mathcal{B}}_k s_k)^2} \frac{g(s_k, \widetilde{\mathcal{B}}_k s_k)}{\|s_k\|^2} = \frac{q_k}{\cos^2\theta_k}. \tag{2.65}$$

In addition, we have from (2.60) that

$$\det(\mathcal{B}_{k+1}) = \det(\mathcal{B}_k) \frac{g(y_k, s_k)}{g(s_k, s_k)} \frac{g(s_k, s_k)}{g(s_k, \widetilde{\mathcal{B}}_k s_k)} = \det(\mathcal{B}_k) \frac{n_k}{q_k}. \tag{2.66}$$

Therefore, since $\mathcal{B}_0$ is self-adjoint and positive definite and parallel transport is an isometry, from Lemma 2.4.1, we know $\mathcal{B}_{k+1}$ is self-adjoint and positive definite if $g(s_k, y_k) > 0$.

Since $y_k = \operatorname{grad} f(x_{k+1}) - P_{\gamma_{\eta_k}}^{1 \leftarrow 0} \operatorname{grad} f(x_k)$, the condition $g(s_k, y_k) > 0$ is equivalent to

$$g(\operatorname{grad} f(x_{k+1}), s_k) \geq g(P_{\gamma_{\eta_k}}^{1 \leftarrow 0} \operatorname{grad} f(x_k), s_k). \tag{2.67}$$

If the line search in Algorithm 2 satisfies the curvature condition (2.3), then (2.67) holds since parallel transport is an isometry. Even without the requirement of (2.3) in Algorithm 2, (2.67) can usually be satisfied if it is in a region without significant negative curvature.

We now combine the trace and determinant with the intent of implicitly bounding the condition number by introducing the following function of a self-adjoint positive definite linear transformation $\mathcal{B}$:

$$\psi(\mathcal{B}) = \operatorname{trace}(\mathcal{B}) - \ln(\det(\mathcal{B})). \tag{2.68}$$

It is not difficult to show that $\psi(\mathcal{B}) > 0$. By using (2.60) and (2.62)- (2.68), we have that

$$\begin{aligned} \psi(\mathcal{B}_{k+1}) &= \psi(\mathcal{B}_k) + N_k - \frac{q_k}{\cos^2\theta_k} - \ln(\det(\mathcal{B}_k)) - \ln n_k + \ln q_k \\ &= \psi(\mathcal{B}_k) + (N_k - \ln n_k - 1) + [1 - \frac{q_k}{\cos^2\theta_k} + \ln\frac{q_k}{\cos^2\theta_k}] + \ln\cos^2\theta_k . \quad (2.69) \end{aligned}$$

Now, since the function $h(t) = 1 - t + \ln t \leq 0$ is nonpositive for all $t > 0$, the term inside the square bracket is nonpositive, and thus from (2.63) and (2.69) we have

$$0 < \psi(\mathcal{B}_{k+1}) \leq \psi(\mathcal{B}_1) + ck + \sum_{j=1}^{k} \ln\cos^2\theta_j, \tag{2.70}$$

where we assume the constant $c = N - \ln n - 1$ to be positive, without loss of generality.

From $\eta_k = -\mathcal{B}_k^{-1}\mathrm{grad}\, f(\mathbf{x}_k)$ and $s_k = P_{\gamma_{\eta_k}}^{1\leftarrow 0}(\alpha\eta_k)$, we know $\cos\theta_k$ is the angle between the steepest decent direction and the search direction. From (2.48) we know that the sequence $\|\mathrm{grad}\, f(\mathbf{x}_k)\|$ generated by the line search algorithm is bounded away from zero only if $\cos\theta_j \to 0$.

Let us now proceed by contradiction and assume that $\cos\theta_j \to 0$. Then there exists $k_1 > 0$ such that for all $j > k_1$, we have

$$\ln\cos^2\theta_j < -2c, \tag{2.71}$$

where $c$ is the constant defined above. using this inequality in (2.70). We find the following relations to be true for all $k > k_1$:

$$\begin{aligned} 0 &< \psi(\mathcal{B}_1) + ck + \sum_{j=1}^{k_1} \ln\cos^2\theta_j + \sum_{j=k_1+1}^{k} (-2c) \\ &= \psi(\mathcal{B}_1) + \sum_{j=1}^{k_1} \ln\cos^2\theta_j + 2ck_1 - ck. \end{aligned}$$

However, the right-hand-side is negative for large $k$, giving a contradiction.Therefore, there exists a subsequence of indices $\{j_k\}$ such that $\{\cos\theta_{j_k}\} \geq \delta > 0$.

By Theorem 2.4.1, this limit implies that liminf $\|\mathrm{grad}\, f(\mathbf{x}_k)\| \to 0$. Since the problem is strongly geodesically convex, the latter limit is enough to prove that $x_k \to x^*$.

□

Note that the convexity of the cost function is only used to guarantee that there is a unique minimizer. One way for this to happen is if $f(x)$ is convex function for the entire domain of interest. However, Theorem 2.4.3 can be used to justify two important conclusions for a more general nonconvex cost function $f(x)$.

**Corollary 2.4.1.** *Suppose $f(x)$ is a nonconvex cost function on $M$ and let $x^* \in M$ be a nondegenerate local minimizer of $f$, i.e.,* $\mathrm{grad}\, f(x^*) = 0$ *and* $\mathrm{Hess}\, f(x^*)$ *is positive definite. Let $x_0$ be starting point that is close enough to $x^*$ so that it is in the neighborhood around $x^*$ where the Hessian is positive definite, i.e., for which Assumptions 2.4.2 are satisfied and* *let $\mathcal{B}_0$ be any linear transformation on $T_{x_0}M$ that is self-adjoint and positive definite with respect to the Riemannian metric $g$.*

*The sequence $\{x_k\}$ generated by Algorithm 2 using parallel transport and the exponential map as the retraction converges to the minimizer $x^*$ of $f$, i.e., it is locally convergent to any nondegenerate minimizer.*

*Additionally, if the convexity assumption is removed from Assumptions 2.4.2 then from any $x_0$ the sequence $\{x_k\}$ generated by Algorithm 2 using parallel transport and the exponential map as the retraction converges to a set of critical points of $f$, i.e., there is global convergence to such a set.*

### 2.4.2 The superlinear convergence of RBFGS

While the theorems above guarantee convergence under certain circumstances, we are interested in achieving acceptably rapid convergence for RBFGS, e.g., superlinear, as is guaranteed with BFGS in $\mathbb{R}^n$. The convergence results for BFGS presented in, for example [22, Theorem 8.6], are given for the general form of RBFGS using parallel transport and the exponential map in Theorem 2.4.5 .

Theorem 2.3.1 identifies a key requirement on the evolution of the action of $\mathcal{B}_k$ in the direction of $\eta_k$ relative to the action of the covariant derivative. Note that this requirement is quite general and only requires the transport be twice continuously differentiable. In order to apply it to proving the superlinear convergence theorem of RBFGS (Theorem 2.4.5), we must identify sufficient conditions on the transport and retraction used in the RBFGS iteration that guarantee the required action of $\mathcal{B}_k$.

The Riemannian manifold version of [22, Theorem 8.6] can be shown by generalizing its proof given the following assumption:

**Assumptions 2.4.4.** *Let $x^* \in M$ be a nondegenerate local minimizer of $f$, i.e.,* $\operatorname{grad} f(x^*) = 0$ *and* $\operatorname{Hess} f(x^*)$ *is positive definite. There is $L > 0$ such that, for all $\xi \in T_{x^*}M$ and all $\eta \in T_{R(\xi)}M$ small enough, we have*

$$\|(P_{\gamma_\eta}^{t\leftarrow 0})^{-1}\mathrm{Hess}f(y)P_{\gamma_\eta}^{t\leftarrow 0} - P_{\gamma_\xi}^{1\leftarrow 0}\mathrm{Hess}f(x^*)(P_{\gamma_\xi}^{1\leftarrow 0})^{-1}\| \le L\max\{dist(y,x^*), dist(x,x^*)\},$$

*for $0 \le t \le 1$ where $x = \mathrm{Exp}_{x^*}(\xi)$, $y = \mathrm{Exp}_x(\eta)$ and $\gamma_\xi(t)$ and $\gamma_\eta(t)$ are the associated geodesics.*

**Theorem 2.4.5.** *Suppose that $f$ is twice continuously differentiable and that the iterates, $x_k$, generated by the RBFGS Algorithm 2 using parallel transport and the exponential map converge to a nondegenerate minimizer $x^* \in M$ at which Assumption 2.4.4 holds. If*

$$\sum_{k=1}^{\infty} dist(x_k, x^*) < \infty \tag{2.72}$$

*holds then $x_k$ converges to $x^*$ superlinearly.*

*Proof.* The algorithm defines $x_{k+1} = \mathrm{Exp}_{x_k}(\eta_k)$, i.e., the stepsize has been included in the definition of $\eta_k$. The tangent vectors $\xi_k, \xi_{k+1} \in T_{x^*}M$ are defined by $\xi_k = \mathrm{Exp}_{x^*}^{-1}(x_k)$ and

$\xi_{k+1} = \mathrm{Exp}_{x^*}^{-1}(x_{k+1})$ and we use the geodesics

$$\begin{gathered}\gamma_{\xi_{k+1}}(t\xi_{k+1}) = \mathrm{Exp}_{x^*}(t\xi_{k+1}), \quad \gamma_{\xi_{k+1}}(0) = x^*, \quad \gamma_{\xi_{k+1}}(1) = x_{k+1} \\ \gamma_{\xi_k}(t\xi_k) = \mathrm{Exp}_{x^*}(t\xi_k), \quad \gamma_{\xi_k}(0) = x^*, \quad \gamma_{\xi_k}(1) = x_k \\ \gamma_{\eta_k}(t\eta_k) = \mathrm{Exp}_{x_k}(t\eta_k), \quad \gamma_{\eta_k}(0) = x_k, \quad \gamma_{\eta_k}(1) = x_{k+1}\end{gathered}$$

We also use the parallel transports

$$\begin{gathered}P^{1\leftarrow 0}_{\gamma_{\xi_{k+1}}} : T_{x^*}M \to T_{x_{k+1}}M \\ P^{1\leftarrow 0}_{\gamma_{\xi_k}} : T_{x^*}M \to T_{x_k}M \\ P^{1\leftarrow 0}_{\gamma_{\eta_k}} : T_{x_k}M \to T_{x_{k+1}}M\end{gathered}$$

Note that

$$\begin{gathered}P^{1\leftarrow 0}_{\gamma_{\xi_{k+1}}} = P^{1\leftarrow 0}_{\gamma_{\eta_k}} \circ P^{1\leftarrow 0}_{\gamma_{\xi_k}} \\ s_k = P^{1\leftarrow 0}_{\gamma_{\eta_k}}(\eta_k) \\ y_k = \mathrm{grad}\, f(x_{k+1}) - P^{1\leftarrow 0}_{\gamma_{\eta_k}}(\mathrm{grad}\, f(x_k)) \\ \tilde{\mathcal{B}}_k = P^{1\leftarrow 0}_{\gamma_{\eta_k}} \mathcal{B}_k (P^{1\leftarrow 0}_{\gamma_{\eta_k}})^{-1}\end{gathered}$$

The 'average' Hessian $\bar{G}$ is as defined in (2.56), Let $G_* = G(x^*) = \mathrm{Hess} f(x^*)$ be invertible, define

$$\tilde{G}_* = P^{1\leftarrow 0}_{\gamma_{\xi_{k+1}}} G_* (P^{1\leftarrow 0}_{\gamma_{\xi_{k+1}}})^{-1}$$

and note that

$$\tilde{G}_*^{1/2} = P^{1\leftarrow 0}_{\gamma_{\xi_{k+1}}} G_*^{1/2} (P^{1\leftarrow 0}_{\gamma_{\xi_{k+1}}})^{-1}.$$

Define the quantities

$$\bar{s}_k = \tilde{G}_*^{1/2} s_k, \quad \bar{y}_k = \tilde{G}_*^{-1/2} y_k, \quad \text{and} \quad \bar{\mathcal{B}}_k = \tilde{G}_*^{-1/2} \tilde{\mathcal{B}}_k \tilde{G}_*^{-1/2}.$$

$$\cos\bar{\theta}_k = \frac{g(\bar{s}_k, \bar{\mathcal{B}}_k \bar{s}_k)}{\|\bar{s}_k\| \|\bar{\mathcal{B}}_k \bar{s}_k\|}, \quad \bar{q}_k = \frac{g(\bar{s}_k, \bar{\mathcal{B}}_k \bar{s}_k)}{\|\bar{s}_k\|^2} \tag{2.73}$$

and

$$\bar{n}_k = \frac{g(\bar{y}_k, \bar{s}_k)}{g(\bar{s}_k, \bar{s}_k)}, \bar{N}_k = \frac{g(\bar{y}_k, \bar{y}_k)}{g(\bar{y}_k, \bar{s}_k)}. \tag{2.74}$$

By pre- and post-multiplying the RBFGS update formula (2.5) by $\tilde{G}_*^{-1/2}$ and grouping terms appropriately, we obtain

$$\bar{\mathcal{B}}_{k+1} = \bar{\mathcal{B}}_k - \frac{\bar{\mathcal{B}}_k s_k (\bar{\mathcal{B}}_k s_k)^\flat}{s_k^\flat (\bar{\mathcal{B}}_k s_k)} + \frac{y_k y_k^\flat}{y_k^\flat (s_k),}$$

This expression has the same form as (2.5), it follows from the steps leading to (2.69) that

$$\psi(\bar{\mathcal{B}}_{k+1}) = \psi(\bar{\mathcal{B}}_k) + (\bar{N}_k - \ln \bar{n}_k - 1) + [1 - \frac{\bar{q}_k}{\cos^2\bar{\theta}_k} + \ln\frac{\bar{q}_k}{\cos^2\bar{\theta}_k}] + \ln\cos^2\bar{\theta}_k \tag{2.75}$$

From Taylor's theorem,

$$\widehat{f}_{x_k}(\eta_k) = \widehat{f}_{x_k}(0_{x_k}) + \langle \mathrm{grad} f(x_k), \eta_k\rangle_{x_k} + \frac{1}{2}\langle \mathrm{Hess}\ \widehat{f}_{x_k}(\tau\eta_k)[\eta_k], \eta_k\rangle_{x_k} \qquad (2.76)$$

holds for some $\tau \in (0,1)$. It follows that $y_k = \mathrm{grad}\, f(\mathbf{x}_{k+1}) - P^{1\leftarrow 0}_{\alpha\eta_k}(\mathrm{grad}\, f(\mathbf{x}_k)) = \bar{G}_k s_k$.

From

$$y_k - \tilde{G}_* s_k = (\bar{G}_k - \tilde{G}_*)s_k$$

we have

$$\bar{y}_k - \bar{s}_k = \tilde{G}_*^{-1/2}(\bar{G}_k - \tilde{G}_*)\tilde{G}_*^{-1/2}\bar{s}_k$$

Using the norms induced by the Riemannian metric $g$ and by assumption (2.4.4) and the isometry of $P^{1\leftarrow 0}_{\gamma_{\xi_{k+1}}}$, we have

$$\begin{aligned} \|\bar{y}_k - \bar{s}_k\| &\leq \|\tilde{G}_*^{-1/2}\|^2\|\bar{G}_k - \tilde{G}_*\|\|\bar{s}_k\| \\ &\leq L\epsilon_k\|\tilde{G}_*^{-1/2}\|^2\|\bar{s}_k\|, \end{aligned}$$

where $\epsilon_k$ is defined by

$$\epsilon_k = \max\{\mathrm{dist}(x_{k+1}, x^*), \mathrm{dist}(x_k, x^*)\}. \qquad (2.77)$$

So

$$\frac{\|\bar{y}_k - \bar{s}_k\|}{\|\bar{s}_k\|} \leq \bar{c}\epsilon_k, \text{ for some positive constant } \bar{c}. \qquad (2.78)$$

From (2.78), we have

$$\|\bar{y}_k\| - \|\bar{s}_k\| \leq \bar{c}\epsilon_k\|\bar{s}_k\| \quad \text{and} \quad \|\bar{s}_k\| - \|\bar{y}_k\| \leq \bar{c}\epsilon_k\|\bar{s}_k\|.$$

and therefore

$$(1 - \bar{c}\epsilon_k)\|\bar{s}_k\| \leq \|\bar{y}_k\| \leq (1 + \bar{c}\epsilon_k)\|\bar{s}_k\|. \qquad (2.79)$$

By squaring (2.78) and using (2.79), we obtain

$$(1 - \bar{c}\epsilon_k)^2\|\bar{s}_k\|^2 - 2g(\bar{y}_k, \bar{s}_k) + \|\bar{s}_k\|^2 \leq \|\bar{y}_k\|^2 - 2g(\bar{y}_k, \bar{s}_k) + \|\bar{s}_k\|^2 \leq \bar{c}^2\epsilon_k^2\|\bar{s}_k\|^2,$$

and therefore

$$2g(\bar{y}_k, \bar{s}_k) \geq (1 - 2\bar{c}\epsilon_k + \bar{c}^2\epsilon_k^2 + 1 - \bar{c}^2\epsilon_k^2)\|\bar{s}_k\|^2 = 2(1 - \bar{c}\epsilon_k)\|\bar{s}_k\|^2.$$

From the definition of $\bar{n}_k$, we have

$$\bar{n}_k = \frac{g(\bar{y}_k, \bar{s}_k)}{\|\bar{s}_k\|^2} \geq 1 - \bar{c}\epsilon_k. \qquad (2.80)$$

By combining (2.79) and (2.80), we obtain

$$\bar{N}_k = \frac{\|\bar{y}_k\|^2}{g(\bar{y}_k, \bar{s}_k)} \leq \frac{1 + \bar{c}\epsilon_k}{1 - \bar{c}\epsilon_k}. \qquad (2.81)$$

Since $x_k \to x^*$, we have that $\epsilon_k \to 0$, and by (2.81) there exists a positive constant $c > \bar{c}$ such that

$$\bar{N}_k \leq 1 + \frac{2\bar{c}}{1 - \bar{c}\epsilon_k}\epsilon_k \leq 1 + c\epsilon_k, \text{ for all sufficiently large } k \tag{2.82}$$

Since $h(t) = 1 - t + \ln t$ is nonpositive, we have

$$\frac{-x}{1-x} - \ln(1-x) = h\Big(\frac{1}{1-x}\Big) \leq 0.$$

For $k$ large enough, we assume that $\bar{c}\epsilon_k < \frac{1}{2}$, and therefore

$$\ln(1 - \bar{c}\epsilon_k) \geq \frac{-\bar{c}\epsilon_k}{1 - \bar{c}\epsilon_k} \geq -2\bar{c}\epsilon_k.$$

From this relation and (2.80), we have

$$\ln \bar{n}_k \geq \ln(1 - \bar{c}\epsilon_k) \geq -2\bar{c}\epsilon_k \geq -2c\epsilon_k, \text{ for all sufficiently large } k. \tag{2.83}$$

From (2.75), (2.82) and (2.83), we can deduce that

$$0 < \psi(\bar{\mathcal{B}}_{k+1}) = \psi(\bar{\mathcal{B}}_k) + 3c\epsilon_k + \big[1 - \frac{\bar{q}_k}{\cos^2\bar{\theta}_k} + \ln\frac{\bar{q}_k}{\cos^2\bar{\theta}_k}\big] + \ln\cos^2\bar{\theta}_k \tag{2.84}$$

By summing (2.84) and using (2.72) we have that

$$\sum_{j=0}^{\infty}\Big(\ln\frac{1}{\cos^2\bar{\theta}_j} - \Big[1 - \frac{\bar{q}_j}{\cos^2\bar{\theta}_j} + \ln\frac{\bar{q}_j}{\cos^2\bar{\theta}_j}\Big]\Big) \leq \psi(\bar{\mathcal{B}}_0) + 3c\sum_{j=0}^{\infty}\epsilon_j < \infty. \tag{2.85}$$

Notice the term in the square brackets is nonpositive, and since $\ln\frac{1}{\cos^2\bar{\theta}_j} \geq 0$ for all $j$, we obtain

$$\lim_{j\to\infty}\ln\frac{1}{\cos^2\bar{\theta}_j} = 0, \lim_{j\to\infty}\Big[1 - \frac{\bar{q}_j}{\cos^2\bar{\theta}_j} + \ln\frac{\bar{q}_j}{\cos^2\bar{\theta}_j}\Big] = 0,$$

which implies that

$$\lim_{j\to\infty}\cos\bar{\theta}_j = 1, \lim_{j\to\infty}\bar{q}_j = 1. \tag{2.86}$$

Recalling (2.65), we have

$$\frac{\|\tilde{G}_*^{-1/2}(\tilde{\mathcal{B}}_k - \tilde{G}_*)s_k\|^2}{\|\tilde{G}_*^{1/2}s_k\|^2} = \frac{\|(\bar{\mathcal{B}}_k - I)\bar{s}_k\|^2}{\|\bar{s}_k\|^2} \tag{2.87}$$

$$= \frac{\|\bar{\mathcal{B}}_k\bar{s}_k\|^2 - 2g(\bar{s}_k, \bar{\mathcal{B}}_k\bar{s}_k) + g(\bar{s}_k, \bar{s}_k)}{g(\bar{s}_k, \bar{s}_k)} \tag{2.88}$$

$$= \frac{\bar{q}_k^2}{\cos^2\bar{\theta}_k} - 2\bar{q}_k + 1. \tag{2.89}$$

By (2.86),

$$\lim_{k\to\infty}\frac{\bar{q}_k^2}{\cos^2\bar{\theta}_k} - 2\bar{q}_k + 1 = 0 \tag{2.90}$$

It is straightforward to show the bound

$$\|\tilde{G}_*^{-1/2}\|^2 \frac{\|\tilde{G}_*^{-1/2}(\tilde{\mathcal{B}}_k - \tilde{G}_*)s_k\|}{\|\tilde{G}_*^{1/2}s_k\|} \geq \frac{(\tilde{\mathcal{B}}_k - \tilde{G}_*)s_k\|}{\|\tilde{G}_*^{1/2}s_k\|}$$

Combining this bound, the limit (2.90) and that $\|\tilde{G}_*^{-1/2}\|$ is constant with respect to the iteration we have the limit

$$\lim_{k\to\infty} \frac{\|(\tilde{\mathcal{B}}_k - P_{\gamma_{\xi_{k+1}}}^{1\leftarrow 0}\,\mathrm{Hess}f(x^*)(P_{\gamma_{\xi_{k+1}}}^{1\leftarrow 0})^{-1})s_k\|}{\|s_k\|} = 0. \tag{2.91}$$

Finally, substituting the appropriate definitions and using the fact that parallel transport is an isometry yields the limit

$$\lim_{k\to\infty} \frac{\|(\mathcal{B}_k - P_{\gamma_{\xi_k}}^{1\leftarrow 0}\mathrm{Hess}f(x^*)(P_{\gamma_{\xi_k}}^{1\leftarrow 0})^{-1})\eta_k\|}{\|\eta_k\|} = 0. \tag{2.92}$$

The desired result that the rate of convergence is superlinear follows from Theorem 2.3.1.

□

# CHAPTER 3

# RIEMANNIAN BFGS ALGORITHM IMPLEMENTATION

A practical implementation of RBFGS requires the following ingredients: (i) an efficient numerical representation for points $x$ on $M$, tangent spaces $T_xM$ and the inner products $g_x(\xi_1, \xi_2)$ on $T_xM$; (ii) an implementation of the chosen retraction $R_x : T_xM \to M$; (iii) efficient formulas for $f(x)$ and $\mathrm{grad}\, f(x)$; (iv) an implementation of the chosen vector transport $\mathcal{T}_{\eta_x}$ and its inverse $(\mathcal{T}_{\eta_x})^{-1}$; (v) a method for solving

$$\mathcal{B}_k \eta_k = -\mathrm{grad}\, f(\mathbf{x}_k), \tag{3.1}$$

where $\mathcal{B}_k$ is defined recursively through (2.5), or alternatively, a method for computing

$$\eta_k = -\mathcal{H}_k \mathrm{grad}\, f(\mathbf{x}_k) \tag{3.2}$$

where $\mathcal{H}_k$ is defined recursively by (2.8). In this section, we summarize our implementation options and their analytical bases.

We consider first the structure of linear transformations on a submanifold of $\mathbb{R}^n$ and the properties of self-adjointness and isometry with respect to the Riemannian metric. A unifying perspective on the design of efficient vector transport/inverse vector transport pairs using the structure is then given in terms of that structure. We then discuss the two main options for the matrix representation of $\mathcal{B}_k$. The choice between these two is the main distinguishing feature in the basic implementations of RBFGS. Refinements are considered for specific submanifolds of $\mathbb{R}^n$ to enhance efficiency. Finally, the case when $M$ is a quotient manifold is discussed using the implementation choices for the Grassmann manifold.

## 3.1 Transformations and Symmetry on a Submanifold of $\mathbb{R}^n$

When implementing RBFGS on a submanifold of $\mathbb{R}^n$, the structure of the $n \times n$ matrix representations of vector transport and inverse vector transport must be considered from an analytical point of view to preserve symmetry when used in the update of $B_k$ to $B_{k+1}$ and from a computational point of view to guarantee efficiency. Since tangent spaces are identified with subspaces of $\mathbb{R}^n$ transformations between subspaces must be considered with respect to symmetry and isometry relative to the Riemannian metric. We have developed a unified point of view of these issues that also lends itself to deriving computationally efficient transport pairs.

Suppose we are given a subspace $\mathcal{S}$ and an inner product $g(x, y)$ for $x,\ y \in \mathcal{S}$. We can then analyze the symmetry of a linear operator $A \in \mathbb{R}^{n\times n}$ restricted to $\mathcal{S}$. We have the usual basis-free characterization of symmetry

**Definition 3.1.1.** *$A \in \mathbb{R}^{n\times n}$ is symmetric with respect to the inner product $g$ on $\mathcal{S}$ if*

$$g(PAPx,\ y) = g(x,\ PAPy)$$

*where $P$ is a projector onto $\mathcal{S}$.*

Symmetry restricted to $\mathcal{S}$ can also be characterized in terms of any basis for $\mathcal{S}$. Suppose the columns of $U_d$, denoted $u_i$, are a basis for $\mathcal{S}$ and for any $x,\ y \in \mathcal{S}$ we write $x = U_d\hat{x}$ and $y = U_d\hat{y}$ for unique $\hat{x},\ \hat{y} \in \mathbb{R}^d$. The inner product $g$ can be written in terms of the basis as

$$g(x, y) = g(U_d\hat{x}, U_d\hat{y}) = \hat{x}^T\hat{G}\hat{y}, \quad \hat{e}_i^T\hat{G}\hat{e}_j = g(u_i, u_j)$$

where $\hat{e}_i \in \mathbb{R}^d$, $1 \le i \le d$, are the canonical basis of $\mathbb{R}^d$. Note $\hat{G} = \hat{G}^T$ since the inner product must be commutative. We therefore have

**Definition 3.1.2.** *Given a basis and an inner product $g$ for $\mathcal{S}$, the linear operator $A \in \mathbb{R}^{n\times n}$ is symmetric on $\mathcal{S}$ with respect to $g$ if*

$$\hat{A}^T\hat{G} = \hat{G}\hat{A}$$

*where $\hat{A} \in \mathbb{R}^{d\times d}$ is $A$ restricted to $\mathcal{S}$ relative to the basis and $\hat{G} \in \mathbb{R}^{d\times d}$ defines $g$ in terms of the basis.*

If we change the basis from $U_d$ to $\tilde{U}_d = U_dM_d$ where $M_d \in \mathbb{R}^{d\times d}$ is nonsingular the inner product and symmetry is invariant but must be expressed in terms of modified matrices.

In the convergence analysis discussion the constraint that the vector transport be an isometry was stated as a sufficient condition for preserving symmetry. This can be formalized as the following theorem.

**Theorem 3.1.1.** *Suppose $(\mathcal{S}_1, g_1)$ and $(\mathcal{S}_2, g_2)$ are inner product spaces with dimension $d$ embedded in $\mathbb{R}^n$ using bases given by the columns of $U_1 \in \mathbb{R}^{n\times d}$ and $U_2 \in \mathbb{R}^{n\times d}$ respectively and the inner products $g_1$ and $g_2$ are defined by $\hat{G}_1 \in \mathbb{R}^{d\times d}$ and $\hat{G}_2 \in \mathbb{R}^{d\times d}$ relative to $U_1$ and $U_2$ respectively Let the linear maps $B_1 : \mathcal{S}_1 \to \mathcal{S}_1$ and $T : \mathcal{S}_1 \to \mathcal{S}_2$ be defined as*

$$B_1 = U_1\hat{B}_1U_1^\dagger \in \mathbb{R}^{n\times n},\ \ T = U_2\hat{T}U_1^\dagger \in \mathbb{R}^{n\times n},\ \ T^\dagger = U_1\hat{T}^{-1}U_2^\dagger \in \mathbb{R}^{n\times n},\ \ \hat{T},\ \hat{B}_1 \in \mathbb{R}^{d\times d}$$

*If $B_1$ is symmetric on $(\mathcal{S}_1, g_1)$ and $\hat{G}_1 = (\hat{T}^T\hat{G}_2\hat{T})$ or equivalently $T$ is an isometry, i.e., $g_1(x_1, y_1) = g_2(Tx_1, Ty_1)$ for all $x_1, y_1 \in \mathcal{S}_1$, then the linear map $B_2$ is symmetric on $(\mathcal{S}_2, g_2)$ where*

$$B_2 = TB_1T^\dagger = U_2(\hat{T}\hat{B}_1\hat{T}^{-1})U_2^\dagger = U_2\hat{B}_2U_2^\dagger \in \mathbb{R}^{n\times n}$$

*Proof.* Note the generalized inverse satisfies

$$TT^{\dagger} = U_2U_2^{\dagger} = I|_{\mathcal{S}_2}$$
$$T^{\dagger}TU_1U_1^{\dagger} = I|_{\mathcal{S}_1}$$

Since $B_1$ is symmetric on $\mathcal{S}_1$ with respect to $g_1$, we have the following equivalences for $B_2$ being symmetric on $\mathcal{S}_2$ with respect to $g_2$:

$$\begin{gathered}\hat{G}_2\hat{B}_2 = \hat{B}_2^T\hat{G}_2 \\ \Leftrightarrow \hat{G}_2\hat{T}\hat{B}_1\hat{T}^{-1} = (\hat{T}\hat{B}_1\hat{T}^{-1})^T\hat{G}_2 \\ \Leftrightarrow \hat{G}_2\hat{T}\hat{B}_1\hat{T}^{-1} = \hat{T}^{-T}\hat{B}_1\hat{T}^T\hat{G}_2 \\ \Leftrightarrow (\hat{T}^T\hat{G}_2\hat{T})\hat{B}_1 = \hat{B}_1(\hat{T}^T\hat{G}_2\hat{T})\end{gathered}$$

Therefore, we have the sufficient condition

$$\hat{G}_1 = (\hat{T}^T\hat{G}_2\hat{T}) \Rightarrow \hat{G}_2\hat{B}_2 = \hat{B}_2^T\hat{G}_2 \tag{3.3}$$

and $B_2$ is symmetric on $\mathcal{S}_2$ with respect to $g_2$.

If $x_1 = U_1\hat{x}_1 \in \mathcal{S}_1$ and $y_1 = U_1\hat{y}_1 \in \mathcal{S}_1$ then

$Tx_1 = U_2\hat{T}\hat{x}_1 = U_2\hat{x}_2 \in \mathcal{S}_2$ and $Ty_1 = U_2\hat{T}\hat{y}_1 = U_2\hat{y}_2 \in \mathcal{S}_2$. It follows that

$$g_1(x_1, y_1) = \hat{x}_1^T\hat{G}_1\hat{y}_1 = \hat{x}_1^T\hat{T}^T\hat{G}_2\hat{T}\hat{y}_1 = \hat{x}_2^T\hat{G}_2\hat{y}_2 = g_2(x_2, y_2) = g_2(Tx_1, Ty_1)$$

and therefore condition (3.3) is equivalent to $T$ being an isometry between $(\mathcal{S}_1, g_1)$ and $(\mathcal{S}_2, g_2)$.

□

If $\mathcal{S}_1$ and $\mathcal{S}_2$ inherit their inner products from the inner product on $\mathbb{R}^n$ defined by $< x, y >= x^TGy$ then we have the following corollary.

**Corollary 3.1.1.** *Using the definitions of Theorem 3.1.1, let $U_1$ and $U_2$ be any pair of orthonormal bases for $\mathcal{S}_1$ and $\mathcal{S}_2$ respectively and assume additionally that the inner products $g_1$ and $g_2$ are defined via the inner product $< x, y >= x^TGy$ on $\mathbb{R}^n$. $T$ is an isometry if and only if $\hat{T}^T\hat{T} = I_d$. In which case, $B_2$ is symmetric on $(\mathcal{S}_2, g_2)$.*

*Proof.* choose $U_1$ and $U_2$ so that they have orthonormal columns relative to the inner product defined by $G$.

$$I_d = \hat{T}^T\hat{T} = \hat{T}^TU_2^TGU_2\hat{T}$$

$$\begin{aligned}U_1^TGU_1 &= \hat{T}^TU_2^TGU_2\hat{T} \\ &= U_1^T(U_1^{\dagger})^T\hat{T}^TU_2^TGU_2\hat{T}U_1^{\dagger}U_1 \\ &= U_1^TT^TGTU_1\end{aligned}$$

We have

$$x_1^TGy_1 = x_1^TT^TGTy_1$$
$$\therefore, g_2(x_2, y_2) = g_2(Tx_1, Ty_1) = g_1(x_1, y_1), \text{ for } \forall x_1,\ y_1 \in \mathcal{S}_1,$$

$B_2$ is symmetric on $(\mathcal{S}_2, g_2)$ follows from Theorem 3.1.1.

□

Under the assumptions of inherited inner products of Corollary 3.1.1 we can characterize a family of useful isometries. We assume $U_1 \in \mathbb{R}^{n\times d}$ and $U_2 \in \mathbb{R}^{n\times d}$ with $\hat{G}_1 = U_1^T G U_1 = U_2^T G U_2 = \hat{G}_2 = I_d$, and $\mathcal{S}_1 = \mathcal{R}(U_1)$ and $\mathcal{S}_2 = \mathcal{R}(U_2)$. Consider the linear map $T : \mathbb{R}^n \to \mathcal{S}_2$ and its restricted inverse defined by projection given by

$$T = U_2 U_2^\dagger U_1 U_1^\dagger = U_2 \hat{T} U_1^\dagger \quad \text{and} \quad T^\dagger = U_1 \hat{T}^{-1} U_2^T G = U_1 \hat{T}^{-1} U_2^\dagger$$

The transformed map is

$$B_2 = T B_1 T^\dagger = U_2 (\hat{T} \hat{B}_1 \hat{T}^{-1}) U_2^\dagger = U_2 \hat{B}_2 U_2^\dagger \in \mathbb{R}^{n\times n}$$

If $\hat{T} = W \Sigma V^T$ is the full SVD of $\hat{T}$ then we have

$$\hat{B}_2 = W \Sigma V^T \hat{B}_1 V \Sigma^{-1} W^T = W \Sigma \tilde{B}_1 \Sigma^{-1} W^T$$

where $\tilde{B_1}^T = \tilde{B}_1$. Since, in general, $\Sigma \neq I_d$ we have $\hat{B}_2 \neq \hat{B}_2^T$ and therefore the sufficient condition of Corollary 3.1.1 is not satisfied and symmetry of $B_2$ on $(\mathcal{S}_2, g_2)$ cannot be guaranteed. Taking $\hat{T} = Q \in \mathbb{R}^{d\times d}$ with $Q^T Q = Q Q^T = I_d$ defines a pair of isometries where

$$\begin{gathered} \hat{B}_2 = Q \hat{B}_1 Q^T \\ \hat{B}_2^T \hat{G}_2 = \hat{B}_2^T = Q \hat{B}_1^T Q^T = Q \hat{B}_1 Q^T = \hat{B}_2 = \hat{B}_2 \hat{G}_2 \end{gathered}$$

and symmetry on $(\mathcal{S}_2, g_2)$ follows.

A particularly useful and easily derived member of this family is an isometry based on canonical bases for $\mathcal{S}_1$ and $\mathcal{S}_2$. We have $\hat{T} = W \Sigma V^T$ which in general is not an orthogonal matrix. Let $\hat{T} = W V^T$ and we have

$$W^T U_2^T G U_1 V = \Sigma = \tilde{U}_2^T G \tilde{U}_1 \quad \text{and} \quad T = U_2 \hat{T} U_1^\dagger = \tilde{U}_2 \tilde{U}_1^\dagger$$

$\tilde{U}_1$ and $\tilde{U}_2$ are the canonical bases with respect to the inner product defined by $G$ and $T$ is an isometry.

Note this family also includes the economical $QR$-based approach with $G = I_n$

$$P_2|_{\mathcal{S}_1} = U_2 U_2^T U_1 U_1^T \quad \text{and} \quad T = qf(U_2 U_2^T U_1) U_1^T = \tilde{U}_2 U_1^T$$

where $qf(A)$ is the rectangular factor with orthonormal columns in the economical $QR$ factorization of $A$.

The particular implementation of the canonical angle/bases isometry, which also happens to be a vector transport/inverse vector transport pair, is not always efficient computationally. However, it can often be made so by looking at equivalent formulations of the transformation. Using the point of view of general projection allows us to characterize isometric and nonisometric transformations between subspaces of $\mathbb{R}^n$ in both an analytical and computationally useful manner. We assume $\mathbb{K} = \mathcal{R}(K)$, $\mathbb{L} = \mathcal{R}(L)$ , $\mathbb{K}^\perp = \mathcal{R}(K_\perp)$,

$\mathbb{L}^\perp = \mathcal{R}(L_\perp)$, and $\mathbb{K} \neq \mathbb{L}$. Projection yields the decomposition of $\mathbb{R}^n$ and the associated split of the identity matrix

$$\mathbb{K} \oplus \mathbb{L}^\perp = \mathbb{R}^n \quad \text{and} \quad I_n = P + P_\perp$$

We also know by definition

$$\forall z \in \mathbb{R}^n, \quad Pz \in \mathbb{K}, \quad z - Pz \in \mathbb{L}^\perp, \quad \therefore \quad P = K(L^T K)^{-1} L^T$$

$$\forall z \in \mathbb{R}^n, \quad P_\perp z \in \mathbb{L}^\perp, \quad z - P_\perp z \in \mathbb{K}, \quad \therefore \quad P_\perp = L_\perp (K_\perp^T L_\perp)^{-1} K_\perp^T$$

For computational purposes, we can use either of the two forms for $P$ and $P_\perp$ and choose the most efficient given the relative sizes of $n$ and the dimension of the manifold $d$:

$$P = K(L^T K)^{-1} L^T \quad \text{and} \quad P = I - L_\perp (K_\perp^T L_\perp)^{-1} K_\perp^T$$

$$P_\perp = L_\perp (K_\perp^T L_\perp)^{-1} K_\perp^T \quad \text{and} \quad P_\perp = I - K(L^T K)^{-1} L^T$$

The effectiveness of this viewpoint is nicely demonstrated by considering an intuitive choice of transformation that is a vector transport but is not, in fact, an isometry. Suppose $\mathcal{M}$, a manifold with dimension $d$ is embedded in $\mathbb{R}^n$. So all elements of the manifold and the tangent bundle are encoded as $n$-vectors. We assume that for each $x \in \mathcal{M}$ we have a matrix $Q_x \in \mathbb{R}^{n \times d}$ such that $T_x = \mathcal{R}(Q_x)$ and $Q_x^T Q_x = I_d$ and a matrix $N_x \in \mathbb{R}^{n \times n-d}$ such that $T_x^\perp = \mathcal{R}(N_x)$ and $N_x^T N_x = I_{n-d}$. The canonical Riemannian metric is

$$g(t_1, t_2) =< t_1, t_2 >= t_1^T t_2$$

for any $(t_1, t_2) \in T_x \times T_x$ and $x \in \mathcal{M}$.

For each $x \in \mathcal{M}$ we need a vector transport, $\mathcal{T} : T_x \to T_{\tilde{x}}$ and inverse vector transport $\mathcal{T}^\dagger : T_{\tilde{x}} \to T_x$ where $\tilde{x} = R_x(d)$ for some direction vector $d \in T_x$ and $R_x(d) : T_x \to \mathcal{M}$ is a retraction. From our discussion above, if these maps are represented as $n \times n$ matrices they have the form

$$\mathcal{T} = Q_{\tilde{x}} \hat{T} Q_x^T \quad \text{and} \quad \mathcal{T}^\dagger = Q_x \hat{T}^{-1} Q_{\tilde{x}}^T.$$

Taking the core mapping $\hat{T}$ such that $\hat{T}^T \hat{T} = I_d$ guarantees the preservation of symmetry of a transported operator. The reconciliation of this form with efficiency, the requirements of vector transport, and the appropriate convergence properties must be considered carefully.

The use of projection to map from an arbitrary $v \in \mathbb{R}^n$ to a subspace can be used to define a transform/inverse transform pair. Intuitively we start with defining the orthogonal projectors

$$\mathbb{K} = \mathbb{L} = T_{\tilde{x}} = \mathcal{R}(Q_{\tilde{x}}), \quad \mathbb{K}^\perp = \mathbb{L}^\perp = T_{\tilde{x}}^\perp = \mathcal{R}(N_{\tilde{x}})$$

$$P : \mathbb{R}^n \to T_{\tilde{x}}, \quad Pv \mapsto Q_{\tilde{x}} Q_{\tilde{x}}^T v \quad \text{and} \quad P_\perp : \mathbb{R}^n \to T_{\tilde{x}}^\perp, \quad P_\perp v \mapsto N_{\tilde{x}} N_{\tilde{x}}^T v$$

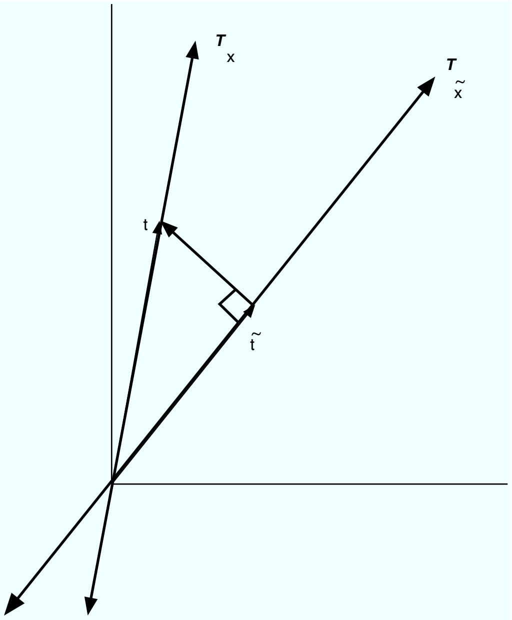

Figure 3.1: Orthogonal and oblique projections relating $t$ and $\tilde{t}$

We can add a projector onto $T_x$ to create the form consistent with our earlier analysis

$$\mathcal{T} = PQ_xQ_x^T = Q_{\tilde{x}}Q_{\tilde{x}}^TQ_xQ_x^T = Q_{\tilde{x}}(Q_{\tilde{x}}^TQ_x)Q_x^T = Q_{\tilde{x}}\hat{T}Q_x^T$$
$$\mathcal{T}^{\dagger} = Q_x\hat{T}^{-1}Q_{\tilde{x}}^T.$$

The two transformation pairs are equivalent when applied to elements of $T_x$. Computationally we do not need to include the $Q_xQ_x^T$ factor if the input is restricted to vectors in $T_x$. So for the transformation we have the straightforward computational choices of

$$\mathcal{T} = Q_{\tilde{x}}Q_{\tilde{x}}^T \quad \text{and} \quad \mathcal{T} = I - \mathcal{T}_{\perp} = I - N_{\tilde{x}}N_{\tilde{x}}^T$$
$$\mathcal{T}_{\perp} = N_{\tilde{x}}N_{\tilde{x}}^T \quad \text{and} \quad \mathcal{T}_{\perp} = I - Q_{\tilde{x}}Q_{\tilde{x}}^T$$

It is easily verified that these are a pair of inverses on $T_x$ and $T_{\tilde{x}}$. The pair is also a vector/inverse vector transport pair but that has not been demonstrated here and characterizing them as such is discussed later. Note, however, that since, in general, $\hat{T}^T\hat{T} \neq I_d$ symmetry is not preserved under this transformation. The geometry of $t \in T_x$, $\tilde{t} = \mathcal{T}t \in T_{\tilde{x}}$, $T_x$ and $T_{\tilde{x}}$ is shown in Figure 3.1. While $\mathcal{T}$ is an orthogonal projector we must have $\mathcal{T}^{\dagger}$ is an oblique projector.

Considering Figure 3.1 yields the intuitive notion that taking two oblique projectors using $T_x$, $T_{\tilde{x}}$ and a third space $\mathbb{L}$ common to both projectors and to which both residuals are orthogonal might yield an orthogonal matrix $\hat{T}$. The proposed situation is shown in Figure 3.2.

Such a space $\mathbb{L}$ always exists under mild assumptions. This yields a pair of isometries and with some care they can be made a vector transport/inverse vector transport pair. This is summarized in the following theorem.

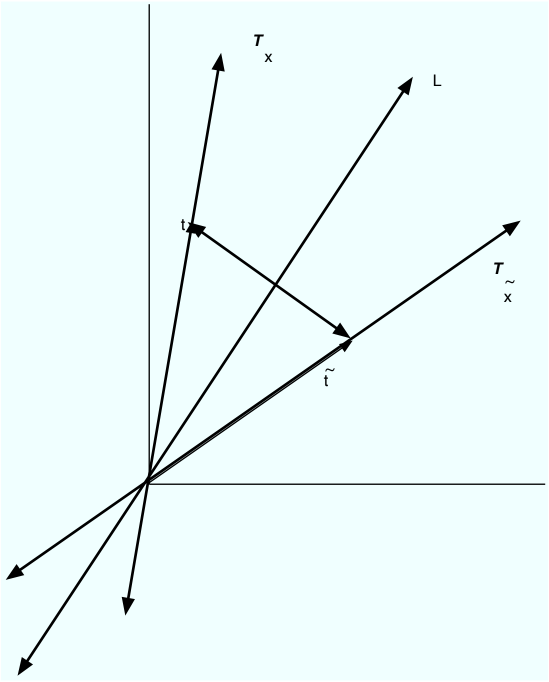

Figure 3.2: Orthogonal and oblique projections relating $t$ and $\tilde{t}$

**Theorem 3.1.2.** *Let $K, \tilde{K} \in \mathbb{R}^{n\times d}$ be such that $K^T K = \tilde{K}^T \tilde{K} = I_d$, $T_x = \mathcal{R}(K)$ and $T_{\tilde{x}} = \mathcal{R}(\tilde{K})$. If $T_x \cap T_{\tilde{x}} = \emptyset$ then for any orthogonal matrix $\hat{T} \in \mathbb{R}^{d\times d}$ there exists $L \in \mathbb{R}^{n\times d}$ with orthonormal columns of the form*

$$L = KM + \tilde{K}\tilde{M}$$

*with nonsingular $M$, $\tilde{M} \in \mathbb{R}^{d\times d}$ that defines a space $\mathbb{L} = \mathcal{R}(L)$ and the associated projectors*

$$\begin{aligned} P &= \tilde{K}\hat{T}K^T, \quad \hat{T} = (L^T\tilde{K})^{-1}(L^T K) \\ \tilde{P} &= K\hat{T}^{-1}\tilde{K}^T, \quad \hat{T}^{-1} = (L^T K)^{-1}(L^T\tilde{K}) \end{aligned}$$

*such that*

$$P\tilde{P} = \tilde{K}\tilde{K}^T \quad \text{and} \quad \tilde{P}P = KK^T.$$

*The projectors define a transform and its inverse between subspaces $T_x$ and $T_{\tilde{x}}$ that are isometries and given the operators*

$$\begin{aligned} A &: T_x \to T_x \\ \tilde{A} = PA\tilde{P} &: T_{\tilde{x}} \to T_{\tilde{x}} \end{aligned}$$

*the symmetry of $A$ on $T_x$ implies the symmetry of of $\tilde{A}$ on $T_{\tilde{x}}$ and vice versa.*

*Proof.* Let

$$L = KM + \tilde{K}\tilde{M}$$

where $M$ and $\tilde{M}$ are to be determined. Recalling $K^TK = \tilde{K}^T\tilde{K} = I_d$ yields

$$(L^TK) = (KM + \tilde{K}\tilde{M})^TK = M^T + \tilde{M}^T\tilde{K}^TK$$
$$(L^T\tilde{K}) = (KM + \tilde{K}\tilde{M})^T\tilde{K} = M^TK^T\tilde{K} + \tilde{M}^T$$

We want

$$(L^TK)Q = (L^T\tilde{K})$$
$$(M^T + \tilde{M}^T\tilde{K}^TK)Q = (M^TK^T\tilde{K} + \tilde{M}^T)$$

Now let $K^T\tilde{K} = U\Sigma V^T$ and define

$$M = U, \quad \tilde{M} = V, \quad Q = UV^T$$

We then have

$$M^TK^T\tilde{K} + \tilde{M}^T = U^TU\Sigma V^T + V^T = \Sigma V^T + V^T$$

$$(M^T + \tilde{M}^T\tilde{K}^TK)Q = U^T + V^TV\Sigma U^T)(UV^T) = \Sigma V^T + V^T$$

as desired. We have

$$(L^T\tilde{K}) = \Sigma V^T + V^T = V^T(V\Sigma V^T + I) = V^T(I - Z\Sigma V^T$$
$$\text{where} \quad Z = -V^T$$

Since $Z$ is orthogonal and $T_x \cap T_{\tilde{x}} = \emptyset$ we have $(L^T\tilde{K})$ is nonsingular and therefore so is $(L^TK)$.

Note that if $L = KM + \tilde{K}\tilde{M}$ does not have orthonormal columns one make them so with $\tilde{L} = LR^{-1}$ and we have

$$Q = (L^TK)^{-1}(L^T\tilde{K}) = (\tilde{L}^TK)^{-1}(\tilde{L}^T\tilde{K})$$

□

The reasoning above can be modified to show that any orthogonal matrix $\hat{T}$ can be placed in the isometric pair when there is no intersection and therefore the construction is universal under those assumptions.

**Theorem 3.1.3.** *Using the assumptions and definitions of Theorem 3.1.2 the orthogonal matrix $\hat{T}$ that defines*

$$P = \tilde{K}\hat{T}K^T$$
$$\hat{T} = (L^T\tilde{K})^{-1}(L^TK)$$
$$\tilde{P} = K\hat{T}^{-1}\tilde{K}^T$$
$$\hat{T}^{-1} = (L^TK)^{-1}(L^T\tilde{K})$$

*can be taken to be any orthogonal matrix and the space $\mathbb{L}$ and an associated orthonormal basis can be determined of the form*

$$L = KM + \tilde{K}\tilde{M}$$

*with nonsingular $M, \;\; \tilde{M} \in \mathbb{R}^{d\times d}$.*

*Proof.* Suppose we want $\hat{T} = Q$ where $Q^TQ = QQ^T = I_d$. We keep $M = U$ and look for $\tilde{M}$ as a function of $Q$. The proof of Theorem 3.1.2 enforces the equality

$$(Q^TU\Sigma V^T - I)\tilde{M} = (V\Sigma U^T - Q^T)U$$

Defining $\hat{V} = Q^TU$ we have

$$\tilde{M} = V(\hat{V}\Sigma - V)^{-1}(V\Sigma - \hat{V})$$

The matrix is well-defined when $\Sigma < I$ which is guaranteed when $T_x \cap T_{\tilde{x}} = \emptyset$. □

Theorem 3.1.2 requires no intersection between the spaces and exploits the result that certain related matrices are nonsingular. If $T_x \cap T_{\tilde{x}} \neq \emptyset$ then the sufficient condition cannot be guaranteed since the canonical bases yield a singular $B$. The following lemma gives the required basic facts.

**Lemma 3.1.1.**

- *If $Q_1 \in \mathbb{R}^{n\times d}$ and $Q_2 \in \mathbb{R}^{n\times d}$ are such that $Q_1^TQ_1 = Q_2^TQ_2 = I_d$ then*

$$Q_1^TQ_2 = U\Gamma V^T, \quad U^TU = V^TV = I_d$$
$$\Gamma = diag(\cos(\theta_1), \ldots, \cos(\theta_d)), \;\; 0 \leq \theta_1 \leq \cdots \leq \theta_d \leq \pi/2$$

- *$\|Q_1^TQ_2\|_2 \leq 1$ since $\|A\|_2 = \max_i \sigma_i$ where $\sigma_i$ are the singular values of $A$*
- *If $\mathcal{R}(Q_1) \cap \mathcal{R}(Q_2) = \emptyset$ then $\cos(\theta_i) < 1$ for $1 \leq i \leq d$ and $\|Q_1^TQ_2\|_2 < 1$.*
- *If $\theta_d < \pi/2$, i.e., $\mathcal{R}(Q_1)$ and $\mathcal{R}(Q_2)$ have no subspaces that are orthogonal to each other, then $Q_1^TQ_2$ is nonsingular.*
- *If $\|A\|_2 < 1$ then*

$$(I - A)^{-1} = \sum_{k=0}^{\infty} A^k$$

We therefore have

**Corollary 3.1.2.**

- *If $\mathcal{R}(Q_1) \cap \mathcal{R}(Q_2) = \emptyset$ then $B = (I_d - {Q_1}^TQ_2)$ is nonsingular.*
- *If $\mathcal{R}(Q_1) \cap \mathcal{R}(Q_2) \neq \emptyset$ then $\exists T_1, \;\; T_2 \in \mathbb{R}^{d\times d}$ such that $T_1^TT_1 = T_2^TT_2 = I_d$ and $B = (I_d - T_1^T{Q_1}^TQ_2T_2)$ is singular, i.e. there are bases for which it is singular.*

*Proof.* The nonsingular result follows directly from the lemma since $\mathcal{R}(Q_1) \cap \mathcal{R}(Q_2) = \emptyset$ implies that $\|{Q_1}^T Q_2\|_2 < 1$.

Singularity for a transformed matrix when there is an intersection of dimension $k$ follows from the fact that there must exist orthogonal transformations, $T_1$ and $T_2$, such that

$$\begin{aligned} \tilde{Q}_1 = Q_1 T_1 = \begin{pmatrix} Q & C_1 \end{pmatrix}, &\quad C_1^T C_1 = I_{d-k} \\ \tilde{Q}_2 = Q_2 T_2 = \begin{pmatrix} Q & C_2 \end{pmatrix}, &\quad C_2^T C_2 = I_{d-k} \\ \mathcal{R}(Q) = T_x \cap T_{\tilde{x}} & \end{aligned}$$

In this coordinate system we have

$$I_d - \tilde{Q}_1^T \tilde{Q}_2 = \begin{pmatrix} 0 & 0 \\ 0 & C_1^T C_2 \end{pmatrix}$$

which is clearly singular. If we are not in that coordinate system then

$$I_d - Q_1^T Q_2$$

may not be singular and

$$T_1^T T_2 - Q_1^T Q_2$$

is singular. □

If $T_x \cap T_{\tilde{x}} \neq \emptyset$ we can still create a projection-based isometric transformation pair in a simple and efficient manner. Let $T_x \cap T_{\tilde{x}}$ have dimension $k < d$ with $\mathcal{R}(Q) = T_x \cap T_{\tilde{x}}$, $Q^T Q = I_k$. Suppose $K, \tilde{K} \in \mathbb{R}^{n \times d}$ be such that $K^T K = \tilde{K}^T \tilde{K} = I_d$, $T_x = \mathcal{R}(K)$ and $T_{\tilde{x}} = \mathcal{R}(\tilde{K})$. We can then choose the bases so that

$$\begin{aligned} K = \begin{pmatrix} Q & K_1 \end{pmatrix}, &\quad K_1^T K_1 = I_{d-k} \\ \tilde{K} = \begin{pmatrix} Q & \tilde{K}_1 \end{pmatrix}, &\quad \tilde{K}_1^T \tilde{K}_1 = I_{d-k} \end{aligned}$$

The projectors that form the pair can be formed by leaving the component of the tangent vectors in $T_x \cap T_{\tilde{x}} \neq \emptyset$ untouched and applying Theorem 3.1.2 to get the pair of transformations between $\mathcal{R}(K_1)$ and $\mathcal{R}(\tilde{K}_1)$. Defining

$$L = \begin{pmatrix} Q & L_1 \end{pmatrix}, \quad L_1^T L_1 = I_{d-k}$$

and noting that

$$\mathcal{R}(K_1) \perp \mathcal{R}(Q) \quad \text{and} \quad \mathcal{R}(K_1) \perp \mathcal{R}(Q) \Rightarrow \mathcal{R}(L_1) \perp \mathcal{R}(Q)$$

we have from Theorem 3.1.2

$$P = \left( QQ^T + \tilde{K}_1 (L_1^T \tilde{K}_1)^{-1} L_1^T \right) KK^T \tag{3.4}$$

$$\tilde{P} = \left( QQ^T + K_1 (L_1^T K_1)^{-1} L_1^T \right) \tilde{K}\tilde{K}^T \tag{3.5}$$

These projectors assume knowledge of the bases associated with all the related spaces. This, of course, could be prohibitive computationally so it is crucial that knowledge of the structure of the tangent spaces and normal spaces be exploited to gain time and space efficiency or to show that efficiency is not possible.

These results give us a mechanism defining and analyzing the isometry and nonisometry properties of transformations between tangent spaces. By virtue of the projection framework it also gives us a set of formulations of the pair of transforms out of which we may select the most computationally efficient.

## 3.2 Vector Transport on a Submanifold of $\mathbb{R}^n$

Since $M$ is assumed to be a submanifold of $\mathbb{R}^n$, tangent spaces $T_xM$ are naturally identified with subspaces of $\mathbb{R}^n$ (see[ [4]§3.5.7] for details). The isometric and nonisometric transformation pairs between such subspaces discussed above are not all vector/inverse vector transport pairs. They must satisfy additional constraints. Using the framework above we have developed the final link in the theory required to analyze and implement vector transport on a submanifold of $\mathbb{R}^n$.

**Definition 3.2.1.** *A **subspace matching function** is a smooth (partial) function*

$$\ell : \mathrm{Gr}(d,n) \times \mathrm{Gr}(d,n) \to L(\mathbb{R}^n, \mathbb{R}^n),$$

*where $L(\mathbb{R}^n, \mathbb{R}^n)$ denotes the set of all linear maps from $\mathbb{R}^n$ into itself, with the following conditions:*

1. *The domain of definition of $\ell$, denoted by $\mathrm{dom}(\ell)$, contains a neighborhood of the diagonal $\Delta_{\mathrm{Gr}(d,n)} = \{(\mathcal{X}, \mathcal{X}) : \mathcal{X} \in \mathrm{Gr}(d,n)\}$.*

2. $$\ell(\mathcal{X}, \mathcal{Y})\mathcal{X} \subseteq \mathcal{Y}. \tag{3.6}$$

3. $$\ell(\mathcal{X}, \mathcal{Y})\mathcal{X}_\perp = \{0\}. \tag{3.7}$$

4. *Consistency:*
$$\ell(\mathcal{X}, \mathcal{X})|_{\mathcal{X}} = \mathrm{id}_{\mathcal{X}}, \quad \text{for all } \mathcal{X} \in \mathrm{Gr}(d,n). \tag{3.8}$$

*If moreover $\ell(\mathcal{X}, \mathcal{Y})|_{\mathcal{X}}$ is an isometry for all $(\mathcal{X}, \mathcal{Y}) \in \mathrm{dom}(\ell)$, where the metric is the one induced from the canonical metric in $\mathbb{R}^n$, then we say that $\ell$ is **isometric**. We say that $\ell$ is **isotropic** if*

$$\ell(U\mathcal{X}, U\mathcal{Y}) = U\ell(\mathcal{X}, \mathcal{Y})U^T$$

*for all $U \in \mathrm{O}_n$; in this case, $\ell$ is fully determined by specifying $\ell(\mathrm{col}(I_{n,d}), \mathcal{Y})$ for all $\mathcal{Y} \in \mathrm{Gr}(d,n)$.*

We will abuse notation and write $\ell(X,Y)$ for $\ell(\mathrm{col}(X), \mathrm{col}(Y))$. From now on, let $\mathcal{M}$ denote a manifold endowed with a retraction $R$.

We have the following characterization of vector transport.

**Theorem 3.2.1.** *If $\ell$ is a subspace matching function (Definition 3.2.1), then $\mathcal{T}$ defined by*

$$\mathcal{T}_{\eta_x}\xi_x = \ell(T_x\mathcal{M}, T_{R(\eta_x)}\mathcal{M})\xi_x \tag{3.9}$$

*is a vector transport.*

*Proof.* From (3.9), we have

$$\mathcal{T}_{\eta_x}\xi_x = \ell(T_x\mathcal{M}, T_{R(\eta_x)}\mathcal{M})\xi_x \in T_{R(\eta_x)}\mathcal{M}$$

so the associated retraction condition in the definition of vector transport is satisfied.

From (3.8), we have

$$\mathcal{T}_{0_x}\xi_x = \ell(T_x\mathcal{M}, T_{R(0_x)}\mathcal{M})\xi_x = \ell(T_x\mathcal{M}, T_x\mathcal{M})\xi_x = \mathrm{id}_{T_x\mathcal{M}}\xi_x = \xi_x.$$

This is the consistency condition of vector transport. We also have

$$\begin{aligned}\mathcal{T}_{\eta_x}(a\xi_x + b\zeta_x) &= \ell(T_x\mathcal{M}, T_{R(\eta_x)}\mathcal{M})(a\xi_x + b\zeta_x)\\ &= \ell(T_x\mathcal{M}, T_{R(\eta_x)}\mathcal{M})a\xi_x + \ell(T_x\mathcal{M}, T_{R(\eta_x)}\mathcal{M})b\zeta_x\\ &= a\mathcal{T}_{\eta_x}(\xi_x) + b\mathcal{T}_{\eta_x}(\zeta_x).\end{aligned}$$

This is the linearity condition of vector transport.

□

In view of (3.6) and (3.7), and restricting from now on to orthonormal $X$ and $Y$, we can write

$$\ell(X, Y) = YQ_{X,Y}X^T, \tag{3.10}$$

where

$$Q_{XM,YN} = N^T Q_{X,Y} M \tag{3.11}$$

to ensure that $\ell$ induces a function on $\mathrm{Gr}(d, n) \times \mathrm{Gr}(d, n)$ through $\ell(\mathrm{col}(X), \mathrm{col}(Y) = \ell(X, Y)$. The smoothness condition imposes that $(X, Y) \mapsto Q_{X,Y}$ is smooth. The consistency condition imposes that

$$Q_{X,X} = I. \tag{3.12}$$

Mapping $\ell$ is isometric if and only if

$$Q_{X,Y} \in \mathrm{O}_d. \tag{3.13}$$

Finally, we have the following result that relates fundamental properties of the mapping $\ell$ defined in terms of a specific form of the core operator $Q_{X,Y}$.

**Theorem 3.2.2.** *If $Q$ is defined by*

$$Q_{X,Y} = W\rho(\Sigma)V^T, \tag{3.14}$$

*where $Y^T X = W\Sigma V^T$ is an SVD and where $\rho$ is such that, for all signed permutation matrix $P$,*

$$P\rho(P^T\Sigma P)P^T = \rho(\Sigma), \tag{3.15}$$

*then isotropy holds for $\ell$ defined through* (3.10)*. Assuming* (3.14) *and* (3.10)*, consistency holds if and only if $\rho(I) = I$, in which case $\ell$ defines a vector transport through* (3.9)*. Still assuming* (3.14) *and* (3.10)*, isometry holds if and only if $\rho(\Sigma) \in \mathrm{O}_d$.*

*Proof.* **Isotropy:**

For any $U \in O_n$, we have $(UY)^T UX = Y^T X$, so

$$Q_{UX,UY} = Q_{X,Y} = W\rho(\Sigma)V^T, \text{ where } Y^T X = W\Sigma V^T, and$$

$$\begin{aligned} \ell(UX, UY) &= (UY)Q_{UX,UY}(UX)^T = UYW\rho(\Sigma)V^T X^T U^T \\ &= U(YW\rho(\Sigma)V^T X^T)U^T = U\ell(X,Y)U^T. \end{aligned}$$

**Consistency:**

If $\rho(I) = I$, then since $X^T X = I_{d,d} = W\Sigma V^T$, we have $W = \Sigma = V = I_{d,d}$ are all identity matrices.

$$\begin{aligned} \ell(X,X)|_{\mathcal{X}} &= XQ_{X,X}X^T|_{\mathcal{X}} = XW\rho(\Sigma)V^T X^T|_{\mathcal{X}} \\ &= X\rho(I_{d,d})X^T|_{\mathcal{X}} = XI_{d,d}X^T|_{\mathcal{X}} = \mathrm{Id}_{\mathcal{X}}, \text{for all } X. \end{aligned}$$

Conversely, if $\ell$ is consistent, then

$$\begin{aligned} I_{n,n}X = \ell(X,X)X &= XQ_{X,X}X^T X = XW\rho(\Sigma)V^T X^T X \\ &= X\rho(I_{d,d})X^T X = X\rho(I_{d,d}), \text{ for all } X. \end{aligned}$$

So $\rho(I_{d,d}) = I$.

**Isometry:**

If $\rho(\Sigma) \in O_d$, then, $W, \rho(\Sigma), V^T$ in (3.14) are all orthogonal. We have that $Q_{X,Y}$ is also orthogonal. From (3.13), isometry of $\ell(X,Y)$ is equivalent to $Q_{X,Y} \in O_d$.

Conversely, if $\ell$ is isometric, we know $Q_{X,Y} \in O_d$.

From (3.14), we have $\rho(\Sigma) = W^T Q_{X,Y} V$. So $Q_{X,Y} \in O_d$ implies that $\rho(\Sigma) \in O_d$ since $W$ and $V$ are all orthogonal.

□

Theorem 3.2.2 characterizes vector transport and isometric vector transport and therefore can be used with the projection framework to analyze and design efficient vector transport/inverse vector transport pairs.

## 3.3 Matrix Representations of $\mathcal{B}_k$ on a Submanifold of $\mathbb{R}^n$

When implementing RBFGS or related methods on a submanifold of $\mathbb{R}^n$ a key consideration is the method in which the linear transformations are represented as matrices. These include $\mathcal{B}_k$, $\mathcal{J}_k$, $\mathcal{H}_k$, and $\mathcal{T}_{\alpha_k \eta_k}$. The choice centers on the efficiency of an $n \times n$ matrix or a $d \times d$ matrix along with a $n \times d$ matrix whose columns form a basis for an appropriate tangent space. In general, we expect a combination of all approaches when considering all of the transformations above.

### 3.3.1 Approach 1: $n \times n$ matrix representations

Approach 1 realizes $\mathcal{B}_k$ as an $n \times n$ matrix $B_k^{(n)}$. When considering $M$ as a submanifold of $\mathbb{R}^n$ and its tangent spaces $T_x M$ naturally identified with subspaces of $\mathbb{R}^n$, it is very common to use the same notation for a tangent vector and its corresponding element of $\mathbb{R}^n$. However, to explain Approach 1, it is useful to distinguish the two objects. To this end, let $\iota_x$ denote the natural inclusion of $T_x M$ in $\mathbb{R}^n$, $\iota_x : \ T_x M \to \mathbb{R}^n, \ \ \xi_x \mapsto \ \iota_x(\xi_x)$.

To represent $\mathcal{B}_k$, we pick $B_k^{(n)} \in \mathbb{R}^{n \times n}$ such that, for all $\xi_{x_k} \in T_{x_k} M$,

$$B_k^{(n)} \iota_{x_k}(\xi_{x_k}) = \iota_{x_k}(\mathcal{B}_k \xi_{x_k}). \tag{3.16}$$

Solving the linear system (3.1) then amounts to finding $\iota_{x_k}(\eta_k)$ in $\iota_{x_k}(T_{x_k} M)$ that satisfies

$$B_k^{(n)} \iota_{x_k}(\eta_k) = -\iota_{x_k}(\operatorname{grad} f(x_k)). \tag{3.17}$$

Note that condition (3.16) does not uniquely specify $B_k^{(n)}$; its action on the normal space is irrelevant or an available degree of freedom depending on the point of view. As a result, $B_k^{(n)}$ may be a singular matrix, i.e., $rank(B_k^{(n)}) = d < n$. In any case, (3.17) will be a consistent set of equations that must be solved on each step of the form of RBFGS that uses (3.1). The matrix $B_k^{(n)}$ could be made nonsingular directly by regularization, but when considering the transport of the operator to get the matrix representations of $\tilde{\mathcal{B}}_k$ and $\mathcal{B}_{k+1}$, the matrix form of the transport and its properties determine the rank of $B_k^{(n)}$. The $n \times n$ matrix representation of transport that naturally arises from the earlier discussion is a rank deficient matrix expressed in terms of an efficient projection form using information about the tangent spaces or the normal spaces involved in the transport. The null space of the transformations would include appropriate normal spaces. In this case, we would have $\mathcal{T}_{\eta_k}$ represented by $T_{\alpha\eta_k}^{(n)} \in \mathbb{R}^{n \times n}$ that satisfies $T_{\alpha\eta_k}^{(n)} \iota_{x_k}(\xi_{x_k}) = \iota_{x_{k+1}}(\mathcal{T}_{\alpha\eta_k} \xi_{x_k})$ for all $\xi_{x_k} \in T_{x_k} M$ and $T_{\alpha\eta_k}^{(n)} \zeta_k = 0$ for all $\zeta_k \perp \iota_{x_k}(T_{x_k} M)$.

The update (2.5) must be expressed using this representation. Since $M$ is an embedded submanifold of $\mathbb{R}^n$, the Riemannian metric is given by $g(\xi_x, \eta_x) = \iota_x(\xi_x)^T \iota_x(\eta_x)$ and the update equation (2.5) is then

$$B_{k+1}^{(n)} = \tilde{B}_k^{(n)} - \frac{\tilde{B}_k^{(n)} \iota_{x_{k+1}}(s_k) \iota_{x_{k+1}}(s_k)^T \tilde{B}_k^{(n)}}{\iota_{x_{k+1}}(s_k)^T \tilde{B}_k^{(n)} \iota_{x_{k+1}}(s_k)} + \frac{\iota_{x_{k+1}}(y_k) \iota_{x_{k+1}}(y_k)^T}{\iota_{x_{k+1}}(y_k)^T \iota_{x_{k+1}}(s_k)},$$

where $\tilde{B}_k^{(n)} = T_{\alpha\eta_k}^{(n)} B_k^{(n)} \big((T_{\alpha\eta_k})^{(n)}\big)^\dagger$ and $\dagger$ denotes the pseudoinverse. The pseudoinverse is used here for the inverse vector transport to emphasize that the $n \times n$ representation of the vector transport/inverse vector transport pair may not be full rank as a transformation on $\mathbb{R}^n$. Clearly, if $rank(T_{\alpha\eta_k}^{(n)}) \neq n$ then $rank(\tilde{B}_k^{(n)}) \neq n$ since the update affects only the action on the tangent spaces. So for this approach, the typical situation is that $T_{\alpha\eta_k}^{(n)}$ is expressed efficiently via some projection-like expression and therefore applying it to a vector or matrix in $\mathbb{R}^n$ or $\mathbb{R}^{n \times n}$ is efficient; and $B_k^{(n)}$, $B_{k+1}^{(n)}$ and $\tilde{B}_k^{(n)}$ will be expressed as dense $n \times n$ matrices that are singular but symmetric and positive definite on the subspaces representing the appropriate tangent space of $M$.

If RBFGS is expressed in terms of (3.2) and propagating $H_k^{(n)} \in \mathbb{R}^{n\times n}$ to represent $\mathcal{H}_k$ similar rank statements can be made. In this case, however, there is no issue of solving consistent singular systems and the update of $H_k^{(n)}$ is essentially propagating $(B_k^{(n)})^\dagger$. As above the typical representation is dense $n \times n$ and symmetric and positive definite on the appropriate subspace.

Solving (3.17) using $B_k^{(n)}$ can be done in three main ways. The first is the most costly and least efficient. A factorization of $B_k^{(n)}$, possibly rank-revealing, could be computed on each step. A very straightforward second way to solve (3.17) is to exploit the CG iterative method for consistent singular systems or ill-conditioned systems. It is straightforward to show that if $B_k^{(n)}$ is symmetric and positive definite on a subspace then given initial conditions in the subspace the CG iteration in $\mathbb{R}^n$ using $B_k^{(n)}$ is equivalent (in exact arithmetic) to CG on problem projected onto the subspace where it is nonsingular. For RBFGS we know that the positive definiteness is preserved but safeguards similar to that used in trust region and line search methods on $\mathbb{R}^n$, e.g., Steighaug CG [22], are easily incorporated. The efficiency concern here centers on the need to perform $O(n^2)$ operations per step of CG on the dense matrix $B_k^{(n)}$.

The third way is to propagate a Cholesky or similar factorization that would be a matrix representation in $\mathbb{R}^{n\times n}$ of the operator $\mathcal{J}_k$ discussed earlier. A key issue in the efficiency of this approach is whether or not the factor is such that its rank is apparent. Recall, that $\mathcal{J}_k$ was transported in our earlier abstract discussion and therefore its matrix representation would be multiplied with the usually singular $T_{\alpha\eta_k}^{(n)}$ and $(T_{\alpha\eta_k}^{(n)})^\dagger$. When, for example, we start with $B_0^{(n)}$ in a factored form (similar to the basis-based Approach 2 below), a strategy that propagates a low-rank factorization based on the Cholesky update idea could for some manifolds result in a very efficient computation. A similar idea of propagating a factorization of $H_k^{(n)}$ could yield for some manifolds a very efficient matrix-vector product for each iteration. These ideas are very much in the spirit of Approach 2 discussed next.

### 3.3.2 Approach 2: $d \times d$ matrix representations

As noted above the representations $T_{\alpha\eta_k}^{(n)}$ and $(T_{\alpha\eta_k}^{(n)})^\dagger$ are expressed in terms of efficient projection-based forms. However, since all of the transformations are only of interest on the $d$-dimensional subspaces corresponding to the appropriate tangent spaces they can all be expressed

$$A_n = U_d \hat{A}_d U_d^\dagger$$

where the columns of $U_d$ form a basis for a tangent space or

$$A_n = U_2 \hat{A}_d U_1^\dagger$$

where the columns of $U_1$ and $U_1$ form bases for the associated tangent spaces. As a result, Approach 2 determines a basis for each $d$-dimensional space of interest and applies the updates and solves systems on the coordinate forms of tangent vectors and linear transformations relative to those bases. All of the formulas have at their core matrix operations using the core transformations. The main efficiency concern is the cost of transport and/or creation of the required bases and the cost of projecting onto the subspaces.

Approach 2 realizes $\mathcal{B}_k$ by a $d\times d$ matrix $B_k^{(d)}$ using bases, where $d$ denotes the dimension of $M$. Given a basis $\underline{E}_k = (E_{k,1}, \ldots, E_{k,d})$ of $T_{x_k}M$, if $\hat{g}_k \in \mathbb{R}^d$ is the vector of coefficients of grad $f(x_k)$ in the basis and $B_k^{(d)}$ is the $d \times d$ matrix representation of $\mathcal{B}_k$ in the basis, then we must solve $B_k^{(d)}\hat{\eta}_k = -\hat{g}_k$ for $\hat{\eta}_k \in \mathbb{R}^d$, and the solution $\eta_k$ of (3.1) is given by $\eta_k = \sum_{i=1}^d E_{k,i}(\hat{\eta}_k)_i$. From (3.16), we have

$$\underline{E}_k^\dagger B_k^{(n)} \underline{E}_k\, \underline{E}_k^\dagger \mathrm{i}_{x_k}(\eta_k) = -\underline{E}_k^\dagger \mathrm{i}_{x_k}(\operatorname{grad} f(x_k))$$

$$B_k^{(d)}\hat{\eta}_k = -\hat{g}_k \tag{3.18}$$

where $\underline{E}_k^\dagger = (\underline{E}_k^T\underline{E}_k)^{-1}\underline{E}_k^T$. The update (2.5) is easily expressed in terms of $\mathbb{R}^d$. Which approach is superior depends on the manifold type and the size of the manifold which are critical in the complexity of creating basis of the tangent space.

Solving (3.18) can be done by computing a factorization on each step or more efficiently the Cholesky factor is easily propagated using the formulas presented abstractly in Lemmas 2.4.1 and 2.4.2 on the core $B_k^{(d)} = L_k^{(d)}(L_k^{(d)})^T$ and transportation of the bases. The update of $L_k^{(d)}$ to $\tilde{L}_k^{(d)}$ and $L_{k+1}^{(d)}$ can be done efficiently using a $QR$ factorization-based strategy presented in [12].

## 3.4 Transport on the Unit Sphere

We view the unit sphere $S^{n-1} = \{x \in \mathbb{R}^n : x^Tx = 1\}$ as a Riemannian submanifold of the Euclidean space $\mathbb{R}^n$ with the inherited inner product on each tangent space. The tangent space at $x$, orthogonal projection onto the tangent space at $x$, and the retraction chosen are given by

$$T_xS^{n-1} = \{\xi \in \mathbb{R}^n \,:\, x^T\xi = 0\}$$
$$\mathrm{P}_x\xi = \xi - xx^T\xi$$
$$R_x(\eta_x) = (x+\eta_x)/\|(x+\eta_x)\|,$$

where $\|\cdot\|$ denotes the Euclidean norm.

$T_x = \mathcal{R}(Q_x)$ has dimension $n-1$ so the projection-based non-isometric vector transport using $P_x = Q_xQ_x^T$ is not useful computationally. If we apply the projection framework and choose the most efficient form of the vector transport/inverse vector transport pair we easily get the orthogonal projector/oblique projector pair on $S^{n-1}$ given by

$$\mathcal{T}_{\eta_x}\xi_x = \left(I - \frac{(x+\eta_x)(x+\eta_x)^T}{\|x+\eta_x\|^2}\right)\xi_x \tag{3.19}$$

$$(\mathcal{T}_{\eta_x})^{-1}(\xi_{R_x(\eta_x)}) = \left(I - \frac{(x+\eta_x)x^T}{x^T(x+\eta_x)}\right)\xi_{R_x(\eta_x)} \tag{3.20}$$

If $\underline{E}_{k+1}^T\underline{E}_k = U\Sigma W^T$ is a full SVD, where $\underline{E}_{k+1}$ and $\underline{E}_k$ are orthonormal basis of $T_{x_{k+1}}$ and $T_{x_k}$ respectively then the projection-based isometric vector transport/inverse vector transport pair based on canonical angles and vectors is

$$\mathcal{T} = \underline{E}_{k+1}U(\underline{E}_kW)^T \quad \text{and} \quad \mathcal{T}^{-1} = \underline{E}_kWU^T\underline{E}_{k+1}^T \tag{3.21}$$

This is also computationally unacceptable. If we apply the projection framework and choose the most efficient form of the vector transport/inverse vector transport pair we easily get the oblique projector pair on $S^{n-1}$ whose actions can be described as follows: for any $t \in T_x$ and $\tilde{t} \in T_{\tilde{x}}$ where $\mathcal{T} t = \tilde{t}$ and $t = \mathcal{T}^\dagger \tilde{t}$ we have the unique decompositions

$$t = t_c + t_\cap \quad \text{and} \quad \tilde{t} = \tilde{t}_c + t_\cap, \quad t_\cap \in T_x \cap T_{\tilde{x}}$$

and the computationally efficient formulas

$$\tilde{r} = (I - \tilde{x}\tilde{x}^T)x, \quad \tilde{q} = \tilde{r}/\|\tilde{r}\|_2, \quad r = (I - xx^T)\tilde{x}, \quad q = r/\|r\|_2$$

$$\begin{aligned} t_\cap &= t - xx^T t - qq^T t = t - \tilde{x}\tilde{x}^T t - \tilde{q}\tilde{q}^T t \\ \tilde{t} &= t_\cap + \hat{\mathcal{T}} t_c = (\tilde{q}q^T)(qq^T)t + t_\cap = \tilde{q}q^T t + t_\cap \\ t &= t_\cap + \hat{\mathcal{T}}^\dagger \tilde{t}_c = (q\tilde{q}^T)\tilde{q}\tilde{q}^T \tilde{t} + t_\cap = q\tilde{q}^T \tilde{t} + t_\cap \end{aligned}$$

For the unit sphere, the Levi-Civita parallel transport of $\xi \in T_x S^{n-1}$ along the geodesic, $\gamma$, from $x$ in direction $\eta \in T_x S^{n-1}$ is [8]

$$P_\gamma^{t \leftarrow 0} \xi = \Big( I_n + (\cos(\|\eta\| t) - 1)\frac{\eta\eta^T}{\|\eta\|^2} - \sin(\|\eta\| t)\frac{x\eta^T}{\|\eta\|} \Big)\xi.$$

This parallel transport and its inverse have computational costs comparable to the efficient forms of the vector transports and their inverses.

### 3.5 Transport on the compact Stiefel manifold $\mathrm{St}(p, n)$

We view the compact Stiefel manifold $\mathrm{St}(p, n) = \{X \in \mathbb{R}^{n \times p} : X^T X = I_p\}$ as a Riemannian submanifold of the Euclidean space $\mathbb{R}^{n \times p}$ endowed with the canonical Riemannian metric $g(\xi, \eta) = \mathrm{tr}(\xi^T \eta)$. The tangent space at $X$ and the associated orthogonal projection are given by

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

We use the retraction given by $R_X(\eta_X) = \mathrm{qf}(X + \eta_X)$, where $\mathrm{qf}(A)$ denotes the $Q$ factor of decomposition of $A \in \mathbb{R}_*^{n \times p}$ as $A = QR$, where $\mathbb{R}_*^{n \times p}$ denotes the set of all nonsingular $n \times p$ matrices, $Q \in \mathrm{St}(p, n)$ and $R$ is an upper triangular $n \times p$ matrix with strictly positive diagonal elements.

Vector transport and its inverse on $\mathrm{St}(p, n)$ are given by

$$\begin{aligned} \mathcal{T}_{\eta_X} \xi_X &= (I - YY^T)\xi_X + Y\mathrm{skew}(Y^T \xi_X) \\ (\mathcal{T}_{\eta_X})^{-1} \xi_Y &= \xi_Y + \zeta, \end{aligned}$$

where $Y := R_X(\eta_X)$, $\zeta$ is in the normal space at $Y$ which implies $\zeta = YS$ where $S$ is a symmetric matrix, and $(\xi_Y + YS) \in T_x\mathrm{St}(p,n)$ which implies $X^T(\xi_Y + YS)$ is skew symmetric. We therefore have

$$X^TYS + SY^TX + X^T\xi_Y + \xi_Y^TX = 0.$$

Therefore, $S$ can be found by solving a Lyapunov equation.

We can also create the projection-based isometric vector transport/inverse vector transport pair in the same way as illustrated in the Unit Sphere ( 3.21).

Or the economical $QR$-based approach

$$T = qf(\underline{E}_{k+1}\underline{E}_{k+1}^T\underline{E}_k)\underline{E}_k^T$$

where $qf(A)$ is the rectangular factor with orthonormal columns in the economical $QR$ factorization of $A$.

For $\mathrm{St}(p,n)$, the parallel transport of $\xi \neq H$ along the geodesic $\gamma(t)$ from $Y$ in direction $H$, denoted by $w(t) = P_\gamma^{t\leftarrow 0}\xi$, satisfies [14, §2.2.3]:

$$w'(t) = -\frac{1}{2}\gamma(t)(\gamma'(t)^Tw(t) + w(t)^T\gamma'(t)), \quad w(0) = \xi. \tag{3.22}$$

In practice, the differential equation is solved numerically and the computational cost of parallel transport may be significantly higher than that of vector transport.

## 3.6 Transport on $\mathrm{OB}(n, N)$

Let $X = [x_1, x_2, \cdots, x_N] \in \mathrm{OB}(n,N)$, where $x_i \in \mathbb{R}^n, x_i^Tx_i = 1,$ for $i = 1$ to $N$. The dimension $d$ of $\mathrm{OB}(n,N))$ is $(n-1)N$.

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

$$\begin{aligned} \ll Z, W \gg_X &= \langle z_1, w_1\rangle_{x_1} + \cdots + \langle z_N, w_N\rangle_{x_N} \\ &= z_1^Tw_1 + \cdots + z_N^Tw_N = \mathrm{tr}(Z^TW), \text{ for } \forall Z, W \in T_XM \end{aligned}$$

The tangent space at $X$

$$T_xM = \{Z = [z_1, \cdots, z_N] \in \mathbb{R}^{n\times N} \Big| x_1^Tz_1 = x_2^Tz_2 = \cdots = x_N^Tz_N = 0\}$$

A choice of retraction is

$$R_X(Z) = \Big[\frac{x_1 + z_1}{\|x_1 + z_1\|}, \cdots, \frac{x_N + z_N}{\|x_N + z_N\|}\Big] \tag{3.23}$$

The orthogonal projection to tangent space is

$$\mathrm{P}_XW = [(I - x_1x_1^T)w_1, \cdots, (I - x_Nx_N^T)w_N] \tag{3.24}$$

Vector transport (and their inverses) of

$$\xi_X = [\xi_1, \xi_2, \cdots, \xi_N] \in T_xM$$

defined by directions

$$\eta_X = [\eta_1, \eta_2, \cdots, \eta_N] \in T_xM$$

simply apply the corresponding transport mechanisms from $S^{n-1}$ componentwise. Denote

$$vt\_i = I - \frac{\big(X(:,i) + \eta(:,i)\big)\big(X(:,i) + \eta(:,i)\big)^T}{\|X(:,i) + \eta_X(:,i)\|^2}$$

$$vt\_inv\_i = I - \frac{\big(X(:,i) + \eta(:,i)\big)X(:,i)^T}{X(:,i)^T\big(X(:,i) + \eta(:,i)\big)}, \text{ for } i = 1, 2, \cdots, N$$

The vector transport is:

$$\text{vec}\{\mathcal{T}_{\eta_X}\xi_X\} = \text{diag}([vt\_1; vt\_2; \cdots; vt\_N])\text{vec}\{\xi_X\} \tag{3.25}$$

The inverse vector transport is:

$$\text{vec}\{(\mathcal{T}_{\eta_X})^{-1}\xi_{R_X(\eta_X)}\} = \text{diag}([vt\_inv\_1; vt\_inv\_2; \cdots; vt\_inv\_N])\text{vec}\{\xi_{R_X(\eta_X)}\} \tag{3.26}$$

Parallel transport and its inverse is computed by simply replacing the componentwise vector transports on $S^{n-1}$ with the parallel transports on $S^{n-1}$. Hence, as with the unit sphere, the parallel and efficient vector transport costs are similar on $\text{OB}(n, N)$.

Let N be the number of $S^{n-1}$, $X = [x_1, x_2, \cdots, x_N] \in S^{n-1} \times \cdots \times S^{n-1}$, $x_i^T x_i = 1$, for $i = 1$ to $N$.

Denote $vt\_i = I - \frac{\big(X(:,i)+\eta(:,i)\big)\big(X(:,i)+\eta(:,i)\big)^T}{\|X(:,i)+\eta_X(:,i)\|^2}$

$vt\_inv\_i = I - \frac{\big(X(:,i)+\eta(:,i)\big)X(:,i)^T}{X(:,i)^T\big(X(:,i)+\eta(:,i)\big)}$, for $i = 1, 2, \cdots, N$

The vector transport is:

$$\text{vec}\{\mathcal{T}_{\eta_X}\xi_X\} = \begin{pmatrix} vt\_1 & & & \\ & vt\_2 & & \\ & & \cdots & \\ & & & vt\_N \end{pmatrix}\text{vec}\{\xi_X\}$$

The inverse vector transport is:

$$\text{vec}\{(\mathcal{T}_{\eta_X})^{-1}\xi_{R_X(\eta_X)}\} = \begin{pmatrix} vt\_inv\_1 & & & \\ & vt\_inv\_2 & & \\ & & \cdots & \\ & & & vt\_inv\_N \end{pmatrix}\text{vec}\{\xi_{R_X(\eta_X)}\}$$

## 3.7 Implementation on the Grassmann Manifold Grass $(p, n)$

All of the manifolds discussed so far were embedded submanifolds. Grass$(p, n)$ has Riemannian quotient manifold structure. Let the structure space $\overline{M}$ be the noncompact Stiefel manifold $\mathbb{R}_*^{n\times p} = \{Y \in \mathbb{R}^{n\times p} : Y \text{ full rank } \}$.

As the Grassmann manifold is not directly defined as a submanifold of a Euclidean space, we must choose a representation for elements of the manifold and their tangent vectors. We choose to represent an element of Grass$(p, n)$, a $p$-dimensional subspace of $\mathbb{R}^n$, by a full-rank $n \times p$ matrix whose columns span that subspace.

The set of matrices that represent the same subspace as a matrix $Y$ is the fiber $Y\mathrm{GL}_p = \{YM : \det(M) \neq 0\}$. The vertical space at $Y$ is the tangent space to the equivalence class.

$$V_y = \{YM : M \in \mathbb{R}^{p\times p}\}.$$

A real function $f$ on Grass$(p, n)$ is represented by its lift $f_{\uparrow Y} = f(\mathrm{colsp}(Y))$. To represent a tangent vector $\xi$ to Grass $(p, n)$ at a point $\mathcal{Y} = \mathrm{colsp}(Y)$, 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) $\mathrm{D}f(\mathcal{Y})[\xi] = \mathrm{D}f_{\uparrow}(Y)[\xi_{\uparrow Y}] = \xi$ for all real functions $f$ on Grass$(p, n)$. Therefore, the horizontal space $H_Y$ represents the tangent space $T_{\mathcal{Y}}$ Grass$(p, n)$. We define the horizontal space as

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

We then define a noncanonical metric on Grass$(p, n)$ as

$$g_Y(\xi, \zeta) = \mathrm{trace}((Y^T Y)^{-1} \xi_{\uparrow Y}^T \zeta_{\uparrow Y})$$

If $Y \in St(n, p)$, then the tangent space of Grassmannian manifold is given by [17]

$$T_{\mathcal{Y}}\mathrm{Grass}(p, n) = \left\{ Z \in \mathbb{R}^{n\times p} : Z = Y^{\perp} K : K \in \mathbb{R}^{(n-p)\times p} \right\}$$

We will use the retraction

$$R_{\mathcal{Y}}(\xi) = \mathrm{span}(Y + \xi_{\uparrow Y}) \tag{3.27}$$

In order to avoid ill-conditioning, we use qf$(Y + \xi_{\uparrow Y}$ instead of span$(Y + \xi_{\uparrow Y}$ as a basis for the subspace $R_{\mathcal{Y}}(\xi)$. where $\mathcal{Y} = \mathrm{colsp}(Y), \mathrm{qf}(A)$ denotes the $Q$ factor of decomposition of $A \in \mathbb{R}_*^{n\times p}$ as $A = QR$, where $Q$ belongs to St$(p, n)$ and $R$ is an upper triangular $n \times p$ matrix with strictly positive diagonal elements.

The non-isometric vector transport/inverse transport pair on the Grassmann manifold can be defined. Let $\mathrm{P}_Y^h : T_Y \bar{N} \to H_Y$, then the vector transport is:

$$(\mathcal{T}_{\eta_{\mathcal{Y}}} \xi_{\mathcal{Y}})_{\uparrow(Y + \eta_{\uparrow Y})} = \mathrm{P}^h_{Y + \eta_{\uparrow Y}} \xi_{\uparrow Y},$$

where $\mathrm{P}_Y^h Z = \big(I - Y(Y^T Y)^{-1} Y^T\big) Z$ and the inverse vector transport is:

$$\left(\mathcal{T}_{\eta_{\mathcal{Y}}}\right)^{-1} \xi_{R_{\mathcal{Y}}(\eta_{\mathcal{Y}})} = \Big(I - (Y + \eta_{\uparrow Y})\big(Y^T(Y + \eta_{\uparrow Y})\big)^{-1} Y^T\Big) \xi_{R_{\mathcal{Y}}(\eta_{\mathcal{Y}})},$$

Finally, the isometric vector transport/inverse vector transport pair is the same as defined by (3.21).

Parallel transport on Grassmannian manifold can be defined as follows. Let $H$ and $\Delta$ be tangent vectors to the Grassmann manifold at $Y$ . The parallel translation of $\Delta$ along the geodesic in the direction $Y(0) = H$ is then

$$\mathcal{T}\Delta(t) = \Big((YV \quad U) - \begin{pmatrix} -\mathrm{sin}\Sigma t \\ \mathrm{cos}\Sigma t \end{pmatrix} U^T + (I - UU^T)\Big)\Delta.$$

where $U\Sigma V^T$ is the compact singular value decomposition of $H$.

# CHAPTER 4

# RIEMANNIAN ADAPTIVE REGULARIZATION USING CUBICS

## 4.1 The Algorithm

The ARC method in $\mathbb{R}^n$, [9, 10], for a cost function $f(x)$ consists of adding to the current iterate $x \in \mathbb{R}^n$ the update vector $\eta \in \mathbb{R}^n$ solving the ARC subproblem

$$\min_{\eta \in \mathbb{R}^n} m(\eta) = f(x) + \partial f(x)\eta + \frac{1}{2}\eta^T B\eta + \frac{1}{3}\sigma\|\eta\|^3$$

where $\partial f = (\partial_1 f, ..., \partial_n f)$ is the differential of $f$, and $B_k$ is, typically, a symmetric approximation to the local Hessian $H(x_k)$. Both parameters in the cubic model, $B_k$ and $\sigma_k > 0$, are dynamic. The quality of the model $m$ is assessed by forming the quotient

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

Depending on the value of $\rho$, the new iterate will be accepted or discarded and the parameter $\sigma$ will be updated.

The analogue with trust region methods on $\mathbb{R}^n$ is clear. As a result, the paradigm developed for the Riemannian trust-region (RTR) method [2, 6] and [4, Chapter 7.0] is sufficient to describe, RARC, the Riemannian optimization form of ARC. RARC uses a series of flat spaces and associated optimization problems to replace the optimization problem on the curved space. The tangent spaces of the iterates $x_k$ provide a natural series of flat spaces. The retraction is used to map tangent vectors back to the manifold and to define the lifted cost function $\hat{f}_{x_k}(\eta)$ where $\eta \in T_{x_k}\mathcal{M}$.

A series of unconstrained optimization problems in $\mathbb{R}^d$ are considered. For each, the lifted cost function is reduced sufficiently, the resulting tangent vector is retracted to the manifold, and a decision on step acceptance or rejection is made. The parameters of the local model are updated by considering the relationship between the lifted cost function $\hat{f}_{x_k}(\eta)$, its local model $m_x(\eta)$, and the cost function $f(x)$. For RTR, despite using several lifted cost functions $\hat{f}_{x_k}(\eta)$ to define a series of problems, the method converges globally to the critical points of $f(x)$ and has local superlinear convergence under mild assumptions on the retraction, the cost function and the solution of the local unconstrained optimization problems. We have observed and proven similar results for RARC. In the remainder of this chapter we follow closely the development and analysis of Cartis et al. [9] for ARC. We generalize each of their results that are required to prove the corresponding convergence results on an arbitrary Riemannian manifold and to generate an efficient computational form of the Riemannian algorithm. In some cases these proofs require significant modification to handle the general Riemannian situation.

The structure of the RARC method on a Riemannian manifold $(M, g\ )$ with retraction $R$ is as follows. Given a cost function $f : M \to \mathbb{R}$ 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

$$\widehat{f}_{x_k} : T_{x_k}M \to \mathbb{R} : \xi \to f(R_{x_k}\xi) \tag{4.1}$$

The Riemannian metric $g$ turns $T_{x_k}M$ into a Euclidean space endowed with the inner product $g_{x_k}(\cdot\,,\cdot)$. We use the following model as the approximation of $\widehat{f}_{x_k}$

$$\widehat{m}_{x_k}(\eta) = f(x_k) + \langle \operatorname{grad} f(x_k), \eta\rangle_{x_k} + \frac{1}{2}\langle B_k[\eta], \eta\rangle_{x_k} + \frac{1}{3}\sigma_k\|\eta\|^3_{x_k} \tag{4.2}$$

where $\|\eta\|_{x_k} = \sqrt{\langle \eta, \eta\rangle_{x_k}} = \sqrt{g(\eta, \eta)}$. Here $\eta \in T_{x_k}M$, $\langle\cdot\,,\cdot\rangle_{x_k} = g_{x_k}(\cdot\,,\cdot)$, where $B_{x_k} : T_{x_k}M \to T_{x_k}M$ is some symmetric linear operator, i.e., $g_{x_k}(B_{x_k}\xi, \chi) = g_{x_k}(\xi, B_{x_k}\chi)$, $\xi, \chi \in T_xM$. The RARC subproblem on $T_{x_k}M$ is

$$\min_{\eta\in T_{x_k}M} \widehat{m}_{x_k}(\eta) \tag{4.3}$$

An approximate solution $\eta_k$ of the RARC subproblem (4.3) is computed using any available method. The candidate for the new iterate is then given by $x_+ = R_{x_k}(\eta_k)$. The decision to accept or not accept the candidate and to update the regularization parameter, $\sigma_k$, is based on the quotient

$$\rho_k = \frac{f(x_k) - f(R_{x_k}(\eta_k))}{\widehat{m}_{x_k}(0_{x_k}) - \widehat{m}_{x_k}(\eta_k)} = \frac{\widehat{f}_{x_k}(0_{x_k}) - \widehat{f}_{x_k}(\eta_k)}{\widehat{m}_{x_k}(0_{x_k}) - \widehat{m}_{x_k}(\eta_k)} \tag{4.4}$$

If $\rho_k$ is exceedingly small, then the model is very inaccurate: the step is rejected and $\sigma$ is increased. If $\rho_k$ is small but less dramatically so, then the step is accepted but $\sigma$ is still increased. If $\rho_k$ is close to 1, then there is a good agreement between the model and the function over the step, and $\sigma$ can be decreased.

We assume in this section that

$$\operatorname{grad} f(x_k) \neq 0, \text{ for all } k \geq 0. \tag{4.5}$$

The following statement is a straightforward adaptation of Theorem 3.1 in [9] to the case of the RARC subproblem on $T_{x_k}M$ as expressed in (4.3).

**Theorem 4.1.1.** *Any $\eta^*_k$ is a global minimizer of $\widehat{m}_{x_k}$ over $T_{x_k}M$ if and only if it satisfies the system of equations*

$$(B_k + \lambda^*_k I)\eta^*_k = -\operatorname{grad} f(x_k), \tag{4.6}$$

*where $\lambda^*_k = \sigma_k\|\eta^*_k\|$ and $B_k + \lambda^*_k I$ is positive semidefinite. If $B_k + \lambda^*_k I$ is positive definite, $\eta^*_k$ is unique.*

Using this theorem, after some manipulations, solving the subproblem is equivalent to finding the root of the secular equation

$$\phi_1(\lambda_k) = \frac{1}{\sqrt{\|\eta_k\|}} - \frac{\sigma_k}{\lambda_k}$$

and the solution of a sequence of linear equations

$$(B_k + \lambda_k I)\eta_k = -\mathrm{grad}\, f(x_k).$$

For acceptable convergence, we need only choose iterates that improve on the associated Cauchy points. An idealized view of this procedure can be formalized as Algorithm 3.

**Algorithm 3** Riemannian Adaptive Regularization using Cubics(RARC) algorithm

```
Require: Complete Riemannian manifold($M$, $g$); real-valued function $f$ on $M$; retraction
  $R_x$ from $T_xM$ to $M$
  Parameters: $\gamma_2 \geq \gamma_1 > 1, 1 > \xi_2 \geq \xi_1 > 0$, and $\sigma_0 > 0$
  Input: Initial iterate $x_0 \in M$
  Output: Sequence of iterates $x_k \in M$
  for k=0, 1, 2 ... until convergence do
    Obtain a step $\eta_k$ in tangent space of $x_k, T_{x_k}M$, for which

        $\widehat{m}_{x_k}(\eta_k) \leq \widehat{m}_{x_k}(\eta_k^C),$        (4.7)

    where we define the Cauchy point $\eta_k^C$, element of $T_{x_k}M$, as the solution of the one-
    dimensional problem

        $\eta_k^C = -\alpha_k^C \mathrm{grad}\, f(x_k)$ and $\alpha_k^C = \underset{\alpha \in \mathbb{R}_+}{\mathrm{argmin}}\, \widehat{m}_{x_k}(-\alpha \mathrm{grad}\, f(x_k))$

    Evaluate $\rho_k$ from (4.4)
    if $\rho_k < \xi_1$ then
      $\sigma_{k+1} \in [\gamma_1\sigma_k, \gamma_2\sigma_k])$   [unsuccessful iteration]
    else if $\rho_k > \xi_2$ then
      $\sigma_k \in (0, \sigma_k]$   [very successful iteration]
    else
      $\sigma_{k+1} \in [\sigma_k, \gamma_1\sigma_k]$   [successful iteration]
    end if
    if $\rho_k \geq \xi_1$ then
      $x_{k+1} = R_{x_k}(\eta_k)$
    else
      $x_{k+1} = x_k$
    end if
  end for
```

In order to turn this into an effective and efficient procedure ARC uses an idea that is applied in various forms in trust region methods. The local cubic model is approximately minimized or more precisely sufficiently reduced in a sequence of nested subspaces related to the current iterate $x_k$. This idea was developed for the Riemannian trust-region (RTR) method by generalizing the truncated CG approach to the trust region method on $\mathbb{R}^n$ [2, 6] and [4, Chapter 7.0].

In the following, we require that $\eta_k$ satisfies

$$\langle \operatorname{grad} f(x_k), \eta_k\rangle_{x_k} + \langle B_k[\eta_k], \eta_k\rangle_{x_k} + \sigma_k \|\eta_k\|^3_{x_k} = 0, k \geq 0 \tag{4.8}$$

$$\langle B_k[\eta_k], \eta_k\rangle_{x_k} + \sigma_k \|\eta_k\|^3_{x_k} \geq 0, k \geq 0 \tag{4.9}$$

Note that since $\operatorname{grad} \widehat{m}_{x_k}(\eta_k) = \operatorname{grad} f(x_k) + B_k[\eta_k] + \sigma_k\|\eta_k\|_{x_k}\eta_k$, (4.8) is equivalent to

$$\langle \operatorname{grad} \widehat{m}_{x_k}(\eta_k), \eta_k\rangle_{x_k} = 0.$$

If $\eta_k$ is a minimizer of $\widehat{m}_{x_k}(\eta_k)$ in a subspace $S_j$, then (4.8) is satisfied due to the following:

$$\eta_k \in \operatorname*{argmin}_{\eta \in S_j} \quad \widehat{m}_{x_k}(\eta_k) \Longrightarrow \langle \operatorname{grad} \widehat{m}_{x_k}(\eta_k), \xi\rangle_{x_k} = 0, \text{ for } \forall \xi \in S_j$$

Lemmas 3.2 and 3.3 of [9] are easily generalized to Lemmas 4.1.1 and 4.1.2.

**Lemma 4.1.1.** *If $\eta_k$ is the global minimizer of $\widehat{m}_{x_k}(\eta)$, for $\eta \in \mathcal{L}_k$, where $\mathcal{L}_k$ is a subspace of $T_{x_k}M$, then $\eta_k$ satisfies* (4.8) *and* (4.9). *Furthermore, letting $Q_k$ denote any orthogonal matrix whose columns form a basis of $\mathcal{L}_k$, we have that*

$$Q_k^T B_k Q_k + \sigma_k \|\eta_k\|_{x_k} I \textit{ is positive semidefinite.}$$

*In particular, if $\eta_k^*$ is the global minimizer of $\widehat{m}_{x_k}(\eta), \eta \in T_{x_k}M$, then $\eta_k^*$ achieves* (4.8) *and* (4.9)*.*

**Lemma 4.1.2.** *Suppose that $\eta_k$ satisfies* (4.8)*. Then*

$$\widehat{f}_{x_k}(0_{x_k}) - \widehat{m}_{x_k}(\eta_k) = \frac{1}{2}\langle B_k[\eta_k], \eta_k\rangle_{x_k} + \frac{2}{3}\sigma_k\|\eta_k\|^3_{x_k} \tag{4.10}$$

*Additionally, if $\eta_k$ also satisfies* (4.9)*, then*

$$\widehat{f}_{x_k}(0_{x_k}) - \widehat{m}_{x_k}(\eta_k) \geq \frac{1}{6}\sigma_k\|\eta_k\|^3_{x_k} \tag{4.11}$$

*Proof.* From (4.2), we have

$$f(x_k) - \widehat{m}_{x_k}(\eta) = -\langle \operatorname{grad} f(x_k), \eta\rangle_{x_k} - \frac{1}{2}\langle B_k[\eta], \eta\rangle_{x_k} - \frac{1}{3}\sigma_k\|\eta\|^3_{x_k} \tag{4.12}$$

From (4.8), we have

$$-\langle \operatorname{grad} f(x_k), \eta_k\rangle_{x_k} = -\langle B_k[\eta_k], \eta_k\rangle_{x_k} - \sigma_k\|\eta_k\|^3_{x_k} = 0, k \geq 0 \tag{4.13}$$

Equation (4.12) becomes (4.10). From (4.9), we get that $\langle B_k[\eta_k], \eta_k\rangle_{x_k} \geq -\sigma_k\|\eta_k\|^3_{x_k}$, which we substitute into (4.10) to get (4.11). □

We assume for the remainder of the discussion that the Cauchy condition (4.7) holds. From (4.5), we have that if $\eta_k$ satisfies (4.8) then

$$\eta_k \neq 0. \tag{4.14}$$

As done in ARC we use the Lanczos method to build an orthogonal basis $\{q_0, \cdots, q_j\}$ for the Krylov space

$$\mathcal{K}(B_k, \operatorname{grad} f(x_k)) = span\{\operatorname{grad} f(x_k), B_k \operatorname{grad} f(x_k), B_k^2 \operatorname{grad} f(x_k) \dots, B_k^j \operatorname{grad} f(x_k)\}.$$

The basic results of Section 6.2 of [9] generalize easily.

Letting $Q_j = (q_0, \cdots, q_j)$ Equation (6.16) of [9] becomes for any two $q_r, q_s \in Q_j$

$$\gamma^0 q_0 = \operatorname{grad} f(x_k), \langle q_r, q_s \rangle = \delta_{r,s}, \langle [B_k] q_r, q_s \rangle = t_{r,s} \tag{4.15}$$

all the $t_{r,s}$ form a symmetric tridiagonal matrix $T_j$. The vector $\eta_j$ solves the following problem:

$$\underset{\eta \in \mathcal{S}_j}{\text{minimize}}\, \widehat{m}_{x_k}(\eta)$$

where

$$\eta \in \mathcal{S}_j = \{\eta \in T_x M \big| \eta = \sum_{i=0}^{j} u_i q_i, u_i \in \mathbb{R}, i = 0, \cdots, j\}$$

It is easily shown that this problem has the form

$$\begin{aligned} \underset{u \in \mathbb{R}^{j+1}}{\text{minimize}}\, \widehat{m}_j(\sum_{i=0}^{j} q_i u_i) &= f(x) + \langle \operatorname{grad} f(x), \eta \rangle_x + \frac{1}{2}\langle B_k[\eta], \eta \rangle_x + \frac{1}{3}\sigma \|\eta\|_x^3 \\ &= f(x) + \gamma_0 e_1^T u + \frac{1}{2} u^T T_j u + \frac{1}{3}\sigma (u^T u)^{3/2} \end{aligned} \tag{4.16}$$

where $u = (u_0, u_1, \cdots, u_j)^T, e_1$ is the first unit vector of approximate length. This is the generalized form of Equation (6.19) of [9].

Equation (4.6) becomes

$$(T_k + \lambda_k^* I) u_k^* = -\operatorname{grad} f(x_k),$$

and the secular equation becomes

$$\phi_1(\lambda_k) = \frac{1}{\|u(\lambda_k)\|_2} - \frac{\sigma_k}{\lambda_k} = 0$$

During each step of RARC, one solves a series of cubic models (4.16) that result from projection to find a point that achieves acceptable reduction in the full cubic model and the cost function. Each reduced, tridiagonal model (4.16) is solved exactly rather than iteratively in the sense its global minimizer is found rather than an approximation to it as one does for each major iteration $k$. This is done using the Newton method described in Algorithm 4. The vector $\eta_k$ is determined by the Generalized Lanczos Adaptive Regularization using Cubics method(GLARC) described in Algorithm 5. Its execution is the dominant computational cost of each moving from $x_k$ to $x_{k+1}$ for RARC.

**Algorithm 4** Newton's method to solve $\phi_1(\lambda)$=0

```
Require: Let $\lambda > \max(0, -\lambda_1)$ be given
  Step1. Factorize $T_j(\lambda) = T_j + \lambda I = LL^T$
  Step2. Solve $LL^T u = -\gamma_0 e_1$ for u
  Step3. Solve Lw=u for w
  Step4. Compute the Newton correction $\Delta\lambda^N = \frac{\lambda\left(\|u\| - \frac{\lambda}{\sigma}\right)}{\|u\| + \frac{\lambda}{\sigma}\left(\frac{\lambda\|w\|^2}{\|u\|^2}\right)}$
  Step5. Replace $\lambda$ by $\lambda + \Delta\lambda^N$
```

**Algorithm 5** The Generalized Lanczos Adaptive Regularization using Cubics (GLARC) method

```
  Let $t_0 = \text{grad}\, f(x_{k-1}), w_{-1} = 0$, and,
  for $j = 0, 1, \ldots, k$ until convergence, perform the iteration:
  $y_j = M^{-1} t_j$;
  $\gamma(j) = \sqrt{\langle t_j, y_j \rangle}$,
  $w_j = t_j / \gamma(j)$,
  $q_j = y_j / \gamma(j)$,
  $\delta_j = \langle q_j, B_k[q_j] \rangle$,
  $t_{j+1} = B_k q_j - \delta_j w_j - \gamma(j) w_{j-1}$,
  Obtain $T_j$ from $T_{j-1}$ using the formula:

      $T(j, j) = \delta(j)$
      $T(j-1, j) = \gamma(j)$
      $T(j, j-1) = T(j-1, j)'$

  Solve the tridiagonal RARC subproblem (4.16) to obtain $u_j$.
  end for: test for convergence using the termination criterion
  Recover $\eta_k = Q_k u_k$ by rerunning the recurrences or obtaining $Q_k$ from backing store.
```

To achieve rapid convergence, we stop as soon as the approximate solution of the RARC algorithm satisfies certain termination criteria. As with ARC, the design of this termination criterion is central to the convergence analysis of RARC. Below we generalize the discussion of Section 3.3 of [9].

The simplest stopping criterion for Algorithm 5 is to stop after a fixed number of iterations. In order to improve the convergence rate, another choice, also used with trust-region methods, is to stop as soon as an iteration $j$ is reached for which

$$\|\mathrm{grad}\, \widehat{m}_{x_k}(\eta_{i,k})\| \leq \theta_{i,k} \|\mathrm{grad}\, f(x_k)\|_{x_k} \tag{4.17}$$

where

$$\theta_{i,k} \overset{\text{def}}{=} \kappa_\theta \min(1, h_{i,k}), \tag{4.18}$$

the $\eta_{i,k}$ are the inner iterates generated by the solver, $\kappa_\theta$ is any constant in (0, 1), and $h_{i,k}$ is a positive parameter. These are the Riemannian forms of Equations (3.23) and (3.24) of [9].

Two choices for $h_{i,k}$ are used

$$h_{i,k} = \|\eta_{i,k}\|, i \geq 0, k \geq 0, \tag{4.19}$$

and

$$h_{i,k} = \|\mathrm{grad}\, f(x_k)\|_{x_k}^{1/2}, i \geq 0, k \geq 0. \tag{4.20}$$

Equations (4.17) and (4.18) allow us to generalize the three termination criteria of [9], which we label in the same manner to facilitate comparison of the results in the remainder of this chapter with the Euclidean results of [9].

$\boxed{\text{TC.h}}$ $\quad \|\mathrm{grad}\, \widehat{m}_{x_k}(\eta_k)\| \leq \theta_k \|\mathrm{grad}\, f(x_k)\|_{x_k},$ where $\theta_k = \kappa_\theta \min(1, h_k), k \geq 0,$

where $h_k \overset{\text{def}}{=} h_{i,k} > 0$ with $i$ being the last inner iteration. For the choice (4.19), we have

$\boxed{\text{TC.s}}$ $\quad \|\mathrm{grad}\, \widehat{m}_{x_k}(\eta_k)\| \leq \theta_k \|\mathrm{grad}\, f(x_k)\|_{x_k},$ where $\theta_k = \kappa_\theta \min(1, \|\eta_k\|_{x_k}), k \geq 0,$

while for the choice (4.20), we obtain

$\boxed{\text{TC.g}}$ $\quad \|\mathrm{grad}\, \widehat{m}_{x_k}(\eta_k)\| \leq \theta_k \|\mathrm{grad}\, f(x_k)\|_{x_k},$ where $\theta_k = \kappa_\theta \min(1, \|\mathrm{grad}\, f(x_k)\|_{x_k}^{1/2}), k \geq 0,$

This inner convergence criterion seeks linear convergence early on, and superlinear convergence after some threshold. It is referred to as the $\kappa/\theta$ convergence criterion.

## 4.2 Convergence Analysis

The convergence analysis for RARC in this section successfully generalizes the very satisfactory results available for ARC on $\mathbb{R}^n$. In particular, we are able to show under a series of increasingly tight assumptions that RARC converges globally to first-order critical points, converges Q-superlinearly or Q-quadratically to nondegenerate local minimizers, and finally converges globally to second-order critical points.

Beginning in Section 4.2.1, there are many theorems and lemmas concerning the convergence property of RARC. The proofs of some are the straight-forward generalizations of those of corresponding results in [9] and are skipped. In general, however, we follow the proof techniques of Cartis, Gould and Toint [9] very closely, generalizing as needed. We use a similar manner of labeling assumptions and notation as close as possible to that in [9] in order to facilitate comparison of the Euclidean and Riemannian results. Specifically, our assumptions labeled *RM.#* are Riemannian versions of the assumptions on the model and correspond to the assumptions labeled *AM.#* in [9]. Our assumptions labeled *RF.#* are Riemannian versions of the assumptions on the cost function. Their correspondence to those labeled *AF.#* in [9] are as follows:

- *RF.*1 corresponds to *AF.*3.
- *RF.*2 corresponds to *AF.*2 but is slightly weaker in the Riemannian version.
- *RF.*2 ′ has no direct corresponding assumption in terms of Riemannian generalization but it plays the role of *AF.*4 in the proofs.
- *RF.*3 corresponds to *AF.*5.
- *RF.*4 corresponds to *AF.*6.

It is also important when comparing to remember that the metric $g$ (and therefore norms of tangent vectors) is defined in each tangent space and is not identical to distance on the manifold, and that our expressions and statements are abstract and are true for any choice of representation of vectors in, or operators on, a particular tangent space.

### 4.2.1 Global Convergence to First-order Critical Points

The results in this section generalize those of Section 2.2 of [9]. The index set of all successful iterations of the RARC algorithm is denoted by

$$\mathcal{S} \overset{\text{def}}{=} \{k \geq 0 : \text{ iteration } k \text{ successful or very successful }\}. \tag{4.21}$$

Lemma 4.2.1 generalizes [9, Lemma 2.1] and shows that the difference between the cost function and the cubic model value is bounded below.

Our first assumption concerns the continuity of the cost function.

$\boxed{\text{RF.1}}$ $f \in C^2(M)$

**Lemma 4.2.1.** *If RF.1 holds and $\eta_k$ satisfies (4.7) then for $k \geq 0$, we have*

$$\begin{aligned}\widehat{f}_{x_k}(0_{x_k}) - \widehat{m}_{x_k}(\eta_k) &\geq \widehat{f}_{x_k}(0_{x_k}) - \widehat{m}_{x_k}(\eta_k^C)\\ &\geq \frac{\|\operatorname{grad} f(x_k)\|_{x_k}^2}{6\sqrt{2}\max[1+\|B_k\|_{x_k},\, 2\sqrt{\sigma_k \|\operatorname{grad} f(x_k)\|_{x_k}}]}\\ &= \frac{\|\operatorname{grad} f(x_k)\|_{x_k}}{6\sqrt{2}} \min\left[\frac{\|\operatorname{grad} f(x_k)\|_{x_k}}{1+\|B_k\|_{x_k}},\, \frac{1}{2}\sqrt{\frac{\|\operatorname{grad} f(x_k)\|_{x_k}}{\sigma_k}}\right].\end{aligned} \tag{4.22}$$

Lemma 4.2.2 generalizes [9, Lemma 2.2] and shows that the norm of the step is bounded above. We assume, as in [9] that

$\boxed{\text{RM.1}}$ $\|B_k\| \leq \kappa_B,$ for all $k \geq 0$ and some $\kappa_B \geq 0$ where the norm is the operator norm induced by the metric $g$ on the tangent space $T_{x_k}M$.

**Lemma 4.2.2.** *If RF.1 holds and $\eta_k$ satisfies (4.7) then*

$$\|\eta_k\|_{x_k} \leq \frac{3}{\sigma_k} \max(\kappa_B,\ \sqrt{\sigma_k \|\operatorname{grad} f(x_k)\|_{x_k}}), k \geq 0. \tag{4.23}$$

As with the Euclidean results we must show when under certain conditions, a step $k$ is very successful. If $\widehat{f}_{x_k}(0_{x_k}) > \widehat{m}_{x_k}(\eta_k)$ and given $\rho_k$ in (4.4) it follows that

$$\rho_k > \xi_2 \Longleftrightarrow r_k = \widehat{f}_{x_k}(\eta_k) - \widehat{f}_{x_k}(0_{x_k}) - \xi_2[\widehat{m}_{x_k}(\eta_k) - \widehat{f}_{x_k}(0_{x_k})] < 0 \tag{4.24}$$

and we have

$$r_k = \widehat{f}_{x_k}(\eta_k) - \widehat{m}_{x_k}(\eta_k) + (1 - \xi_2)[\widehat{m}_{x_k}(\eta_k) - \widehat{f}_{x_k}(0_{x_k})], k \geq 0. \tag{4.25}$$

We will show in Theorem 4.2.1 that at least one accumulation point of $\{x_k\}$ is a critical point of $f$. The following definition is required to generalize the required Euclidean assumptions

**Definition 4.2.1.** ***(radially L-$C^1$ function [[4], Definition 7.4.1])*** *Let $\widehat{f} : TM \to \mathbb{R}$ be defined as in* (4.1)*. We say that $\widehat{f}$ is radially Lipschitz continuously differentiable if there exist reals $\beta_{RL} > 0$ and $\delta_{RL} > 0$ such that, for all $x \in M$, for all $\xi \in T_xM$ with $\|\xi\| = 1$, and for all $t < \delta_{RL}$, it holds that*

$$\left| \frac{d}{d\tau} \widehat{f}_x(\tau\xi)|_{\tau=t} - \frac{d}{d\tau} \widehat{f}_x(\tau\xi)|_{\tau=0} \right| \leq \beta_{RL} t. \tag{4.26}$$

The convergence result requires that $\widehat{m}_{x_k}(\eta_k)$ be a sufficiently good approximation of $\widehat{f}_{x_k}(\eta_k)$. In classical proofs, this is often guaranteed by the assumption that the Hessian of the cost function is bounded. It is however possible to weaken this assumption, which leads us to consider the following assumption.

$\boxed{\text{RF.2}}$ $\widehat{f}$ **is radially** $L - C^1$ (See Definition 4.2.1) with $\beta_{RL} > 1$

The following convergence result for RARC that generalizes [9, Theorem 2.5] in straightforward manner and therefore the proof is omitted. The generalization of [9, Corollary 2.6] in Corollary 4.2.1 below, however, requires more careful consideration for the Riemannian situation.

**Theorem 4.2.1.** *If RF.1, RF.2 and RM.1 hold and $\{f(x_k)\}$ is bounded below then*

$$\liminf_{k\to\infty} \|\operatorname{grad} f(x_k)\|_{x_k} = 0. \tag{4.27}$$

In order to show the next convergence result, we need to make an additional regularity assumption on the cost function $f$, that is, the Lipschitz continuous differentiability given in Definition 2.4.1.

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

$$\|\xi\|_x \geq \mu \mathrm{dist}(x, R_x\xi), \text{ for all } x \in M, \text{ for all } \xi \in T_xM, \|\xi\|_x \leq \delta_\mu. \tag{4.28}$$

In contrast to the Euclidean case, $\|\eta_k\|_{x_k}$ is, in general, different from $\mathrm{dist}(x_k, R_{x_k}(\eta_k))$. We use (4.28) to fall back to a suitable bound. We then have the following generalization of [9, Corollary 2.6].

**Corollary 4.2.1.** *Let RF.1, RF.2 and RM.1 hold and assume $\{f(x_k)\}$ is bounded below. If, additionally, $f$ is Lipschitz continuously differentiable (Definition 2.4.1), and* (4.28) *is satisfied for some $\mu > 0, \delta_\mu > 0$ then*

$$\lim_{k\to\infty} \|\mathrm{grad}\, f(x_k)\|_{x_k} = 0. \tag{4.29}$$

*Proof.* Based on the assumptions we know there are infinitely many successful iterations. Since $\{f(x_k)\}$ is bounded below and there is a subsequence of successful iterates, indexed by $\{t_i\} \subseteq \mathcal{S}$ such that

$$\|\mathrm{grad}\, f(x_{t_i})\|_{x_{t_i}} \geq 2\epsilon \tag{4.30}$$

for some $\epsilon > 0$ and for all $i$. We only consider the successful iterates. For each $t_i$, there is a first successful iteration $l_i > t_i$ such that $\|\mathrm{grad} f(x_{l_i})\| < \epsilon$ . Thus $\{l_i\} \subseteq \mathcal{S}$ and

$$\|\mathrm{grad}\, f(x_k)\|_{x_k} \geq \epsilon, \ \textit{for}\ t_i \leq k < l_i, \text{ and } \|\mathrm{grad}\, f(x_{l_i})\| < \epsilon. \tag{4.31}$$

Let

$$\mathcal{K} \overset{\text{def}}{=} \{k \in \mathcal{S} : t_i \leq k < l_i\}, \tag{4.32}$$

where the subsequences $\{t_i\}$ and $\{l_i\}$ were defined above. Since $\mathcal{K} \subseteq \mathcal{S}$, the construction of the RARC algorithm, RM.1 and Lemma 4.2.1 provide that for each $k \in \mathcal{K}$,

$$\begin{aligned} f(x_k) - f(x_{k+1}) &\geq \xi_1[\widehat{m}_{x_k}(0_{x_k}) - \widehat{m}_{x_k}(\eta_k)] \\ &\geq \frac{\xi_1}{6\sqrt{2}}\|\mathrm{grad}\, f(x_k)\|_{x_k} \cdot \min\left(\frac{1}{2}\sqrt{\frac{\|\mathrm{grad}\, f(x_k)\|_{x_k}}{\sigma_k}}, \frac{\|\mathrm{grad}\, f(x_k)\|_{x_k}}{1+\kappa_B}\right). \end{aligned} \tag{4.33}$$

Using (4.31) gives the equivalent form

$$f(x_k) - f(x_{k+1}) \geq \frac{\xi_1\epsilon}{6\sqrt{2}} \cdot \min\left(\frac{1}{2}\sqrt{\frac{\|\mathrm{grad}\, f(x_k)\|_{x_k}}{\sigma_k}}, \frac{\epsilon}{1+\kappa_B}\right), k \geq \mathcal{K}. \tag{4.34}$$

$\{f(x_k)\}$ is convergent since it is monotonically decreasing and bounded from below, and (4.34) implies

$$\frac{\sigma_k}{\|\mathrm{grad}\, f(x_k)\|_{x_k}} \to \infty, k \in \mathcal{K}, k \to \infty, \tag{4.35}$$

and furthermore, due to (4.31),

$$\sigma_k \to \infty, k \in \mathcal{K}, k \to \infty. \tag{4.36}$$

It follows from (4.35) that

$$\sqrt{\frac{\sigma_k}{\|\mathrm{grad}\, f(x_k)\|_{x_k}}} \geq \frac{1+\kappa_B}{2\epsilon}, \text{ for all } k \in \mathcal{K} \text{ sufficiently large,} \tag{4.37}$$

and from (4.34) we then have

$$\sqrt{\frac{\|\mathrm{grad}\, f(x_k)\|_{x_k}}{\sigma_k}} \leq \frac{12\sqrt{2}}{\xi_1 \epsilon}[f(x_k) - f(x_{k+1})], \text{ for all } k \in \mathcal{K} \text{ sufficiently large.} \tag{4.38}$$

For each $l_i$ and $t_i$, we have

$$\mathrm{dist}(x_{l_i}, x_{t_i}) \leq \sum_{k=t_i, k\in\mathcal{K}}^{l_i-1} \mathrm{dist}(x_k, x_{k+1}) = \sum_{k=t_i, k\in\mathcal{K}}^{l_i-1} \mathrm{dist}(x_k, R_{x_k}(\eta_k)) \leq \sum_{k=t_i, k\in\mathcal{K}}^{l_i-1} \frac{1}{\mu}\|\eta_k\|_{x_k}. \tag{4.39}$$

Recall now the upper bound (4.23) on $\|\eta_k\|_{x_k}, k \geq 0$, in Lemma 2.2. It follows from (4.31) and (4.36) that

$$\sqrt{\sigma_k \|\mathrm{grad}\, f(x_k)\|_{x_k}} \geq \kappa_B, \text{ for all } k \in \mathcal{K} \text{ sufficiently large,}$$

and thus (4.23) becomes

$$\|\eta_k\|_{x_k} \leq 3\sqrt{\frac{\|\mathrm{grad}\, f(x_k)\|_{x_k}}{\sigma_k}}, \text{ for all } k \in \mathcal{K} \text{ sufficiently large.}$$

Now (4.38) and (4.39) provide

$$\mathrm{dist}(x_{l_i}, x_{t_i}) \leq \frac{3}{\mu} \sum_{k=t_i, k\in\mathcal{K}}^{l_i-1} \sqrt{\frac{\|\mathrm{grad}\, f(x_k)\|_{x_k}}{\sigma_k}} \leq \frac{36\sqrt{2}}{\xi_1 \epsilon}[f(x_{t_i}) - f(x_{l_i})]$$

for all $t_i$ and $l_i$ sufficiently large. Observe that $\{f(x_{t_i}) - f(x_{l_i})\}$ converges to zero as $i \to \infty$ since $\{f(x_j)\}$ is convergent. Therefore, $\mathrm{dist}(x_{l_i}, x_{t_i})$ converges to zero as $i \to \infty$, and by Lipschitz continuous differentiability, $\|P_\alpha^{0\leftarrow 1}\mathrm{grad}\, f(x_{l_i}) - \mathrm{grad}\, f(x_{t_i})\|_{x_{t_i}}$ tends to zero. This is a contradiction, since (4.30) and (4.31) imply

$$\|P_\alpha^{0\leftarrow 1}\mathrm{grad}\, f(x_{l_i}) - \mathrm{grad}\, f(x_{t_i})\|_{x_{t_i}} \geq \|\mathrm{grad}\, f(x_{t_i})\|_{x_{t_i}} - \|\mathrm{grad}\, f(x_{l_i})\|_{x_{l_i}} \geq \epsilon.$$

□

### 4.2.2 Fast Convergence

In the following, we analyze the local convergence of Algorithm 3 around nondegenerate local minima. We show asymptotic convergence properties of the RARC in the presence of local convexity. We then prove the RARC algorithm converges at least Q-superlinearly.

We assume that $\operatorname{grad} f(x_k) \neq 0,$ for all $k \geq 0$ and we have the following generalizations of the Euclidean assumptions of [9]

$\boxed{\text{RF.2}\,'}$ $\quad \|\operatorname{Hess} \widehat{f}_x\|_x \leq \kappa_H$ for all $x \in X$ and some $\kappa_H \geq 1$,

where $X$ is some subset of $TM$ containing the line segments in each $TM_{x_k}$ defined by $t\eta_k$ where $0 \leq t \leq 1$, $k \in \mathcal{S}$, and $\mathcal{S}$ is as defined in (4.21).

$\boxed{\text{RF.3}}$ $\quad$ Hess $\widehat{f}_x$ is Lipschitz-continuous at $0_x$ uniformly in a neighborhood of $v, (v \in M$ is a nondegenerate local minimizer of $f$, i.e., $\operatorname{grad} f(v) = 0$ and $\operatorname{Hess} f(v)$ is positive definite) i.e., there exists $L_* > 0, \delta_1 > 0$, and $\delta_2 > 0$ such that, for all $x \in B_{\delta_1}(v)$ and all $\xi \in B_{\delta_2}(0_x)$, it holds that

$$\|\operatorname{Hess} \widehat{f}_x(\xi) - \operatorname{Hess} \widehat{f}_x(0_x)\|_x \leq L_* \|\xi\|_x. \tag{4.40}$$

$\boxed{\text{RM.2}}$ $\quad \frac{\|(B_k - \operatorname{Hess} \widehat{f}_{x_k}(0_{x_k}))\eta_k\|_{x_k}}{\|\eta_k\|_{x_k}} \to 0,$ whenever $\|\operatorname{grad} f(x_k)\|_{x_k} \to 0,$

$\boxed{\text{RM.3}}$ $\quad \|\operatorname{Hess} \widehat{f}_{x_k}(0_{x_k}) - B_k\|_{x_k} \to 0, k \to \infty,$ whenever $\|\operatorname{grad} f(x_k)\|_{x_k} \to 0, k \to \infty,$

$\boxed{\text{RM.4}}$ $\quad \|(\operatorname{Hess} \widehat{f}_{x_k}(0_{x_k}) - B_k)\eta_k\|_{x_k} \leq C\|\eta_k\|^2_{x_k},$ for all $k \geq 0,$ and some constant $C > 0$.

Let

$$R_k(\eta_k) \overset{\text{def}}{=} \frac{\langle B_k[\eta_k], \eta_k\rangle_{x_k}}{\|\eta_k\|^2_{x_k}} \tag{4.41}$$

denote the Rayleigh quotient of $\eta_k$ with respect to $B_k$, representing the curvature of the quadratic part of the model $m_k$ along the step. Lemma 4.2.3 generalizes [9, Lemma 4.1].

**Lemma 4.2.3.** *If RF.1 holds and $\eta_k$ satisfies* (4.8) *then*

$$\widehat{f}_{x_k}(0_{x_k}) - \widehat{m}_{x_k}(\eta_k) \geq \frac{1}{2} R_k(\eta_k) \|\eta_k\|^2_{x_k}, \tag{4.42}$$

*where $R_k(\eta_k)$ is the Rayleigh quotient (*4.41*). In particular,*

$$\widehat{f}_{x_k}(0_{x_k}) - \widehat{m}_{x_k}(\eta_k) \geq \frac{1}{2} \lambda_{\min}(B_k) \|\eta_k\|^2_{x_k},$$

*where $\lambda_{\min}(B_k)$ denotes the leftmost eigenvalue of $B_k$.*

Lemma 4.2.4 generalizes [9, Lemma 4.2] and shows the relationship of the norm of $\eta_k$ and that of $\|\operatorname{grad} f(x_k)\|_{x_k}$ when the Rayleigh quotient (4.41) is positive.

**Lemma 4.2.4.** *Suppose that RF.1 holds and that $\eta_k$ satisfies (*4.8*). If the Rayleigh quotient (*4.41*) is positive, then*

$$\|\eta_k\|_{x_k} \leq \frac{1}{R_k(\eta_k)} \|\operatorname{grad} f(x_k)\|_{x_k} \tag{4.43}$$

*and if $B_k$ is positive definite, then*

$$\|\eta_k\|_{x_k} \leq \frac{1}{\lambda_{\min}(B_k)} \|\operatorname{grad} f(x_k)\|_{x_k}. \tag{4.44}$$

*Proof.* The Cauchy-Schwarz inequality and (4.8) imply

$$R_k(\eta_k)\|\eta_k\|_{x_k}^2 \leq \langle B_k[\eta_k], \eta_k\rangle_{x_k} + \sigma_k\|\eta_k\|_{x_k}^3 = -\langle \operatorname{grad} f(x_k), \eta_k\rangle_{x_k} \leq \|\operatorname{grad} f(x_k)\|_{x_k} \cdot \|\eta_k\|_{x_k}.$$

By definition $R_k(\eta_k) > 0$ and (4.14) implies $\eta_k \neq 0$. Therefore, the first and the last terms above give (4.43) and the bound (4.44) follows from (4.43) and the Rayleigh quotient inequality. □

Theorem 4.2.2 generalizes [9, Theorem 4.3] and shows that under some further assumption all iterations are eventually very successful and $\sigma_k$ is bounded from above.

**Theorem 4.2.2.** *If RF.1,RF.2, RM.1 and RM.2 hold, $\eta_k$ satisfies (4.8), and*

$$x_k \to x_*, \text{ as } k \to \infty, \tag{4.45}$$

*where* $\operatorname{Hess} f(x_*)$ *is positive definite then there exists $R_{\min} > 0$ such that*

$$R_k(\eta_k) \geq R_{\min}, \text{ for all } k \text{ sufficiently large.} \tag{4.46}$$

*We also have*

$$\|\eta_k\|_{x_k} \leq \frac{1}{R_{\min}} \|\operatorname{grad} f(x_k)\|_{x_k}, \text{ for all } k \text{ sufficiently large,} \tag{4.47}$$

*all iterations are eventually very successful, and $\sigma_k$ is bounded from above.*

*Proof.* $\{f(x_k)\}$ is bounded below due to the continuity of $f$ and the limit (4.45).By Corollary 4.2.1 $x_*$ is a first-order critical point and $\|\operatorname{grad} f(x_k)\|_{x_k} \to 0$. RM.2 and $\|\operatorname{grad} f(x_k)\|_{x_k} \to 0$ imply

$$\frac{\|(\operatorname{Hess} \widehat{f}_{x_k}(0_{x_k}) - B_k)\eta_k\|_{x_k}}{\|\eta_k\|_{x_k}} \to 0, k \to \infty \tag{4.48}$$

We therefore have a Riemannian Dennis−Moré condition that holds. Since $\operatorname{Hess} \widehat{f}_{x_*}(0_*) = \operatorname{Hess} f(x_*)$ is positive definite, so is $\operatorname{Hess} \widehat{f}_{x_k}(0_{x_k})$ in a neighborhood of $x_*$, i.e., for all $k$ sufficiently large, and there must exist a constant $R_{\min}$ such that

$$\frac{\langle \eta_k, \operatorname{Hess} \widehat{f}_{x_k}(0_{x_k})\eta_k\rangle_{x_k}}{\|\eta_k\|_{x_k}^2} > 2R_{\min} > 0 \tag{4.49}$$

for all sufficiently large $k$.

It follows from (4.41), (4.48) and (4.49), that for all sufficiently large $k$,

$$\begin{aligned} 2R_{\min}\|\eta_k\|_{x_k}^2 \leq \langle \eta_k, \operatorname{Hess} \widehat{f}_{x_k}(0_{x_k})\eta_k\rangle_{x_k} &= \langle \eta_k, [\operatorname{Hess}\widehat{f}_{x_k}(0_{x_k}) - B_k]\eta_k\rangle_{x_k} + \langle \eta_k, B_k[\eta_k]\rangle_{x_k} \\ &\leq [R_{\min} + R(\eta_k)]\|\eta_k\|_{x_k}^2. \end{aligned}$$

This gives (4.46). The bound (4.47) results from (4.43) and (4.46).

It follows from (4.22) and (4.5) that

$$f(x_k) > \widehat{m}_{x_k}(\eta_k), k \geq 0. \tag{4.50}$$

From (4.50), we know that (4.24) holds. Let $r_k$ be the expression in (4.25), we show it is bounded above and negative for all $k$ sufficiently large.

From Taylor's Theorem, we have

$$\widehat{f}_{x_k}(\eta_k) = \widehat{f}_{x_k}(0_{x_k}) + \langle \text{grad} f(x_k), \eta_k \rangle_{x_k} + \frac{1}{2} \langle \text{Hess } \widehat{f}_{x_k}(\tau\eta_k)[\eta_k], \eta_k \rangle_{x_k} \tag{4.51}$$

for some $\tau \in (0, 1)$. It follows that

$$\begin{aligned} & \widehat{f}_{x_k}(\eta_k) - \widehat{m}_{x_k}(\eta_k) \\ = \; & \frac{1}{2} \langle \text{Hess} \widehat{f}_{x_k}(\tau\eta_k)[\eta_k], \eta_k \rangle_{x_k} - \frac{1}{2} \langle B_k[\eta_k], \eta_k \rangle_{x_k} - \frac{1}{3} \sigma_k \|\eta\|_{x_k}^3 \\ \leq \; & \frac{1}{2} \langle (\text{Hess} \widehat{f}_{x_k}(\tau\eta_k) - B_k)[\eta_k], \eta_k \rangle_{x_k}, \end{aligned}$$

and thus

$$\widehat{f}_{x_k}(\eta_k) - \widehat{m}_{x_k}(\eta_k) \leq \frac{1}{2} \|(\text{Hess} \widehat{f}_{x_k}(\tau\eta_k) - B_k)\eta_k\|_{x_k} \cdot \|\eta_k\|_{x_k}, \tag{4.52}$$

where $\tau\eta_k$ belongs to the line segment between $0_{x_k}$ and $\eta_k$. It follows from (4.42) in Lemma 4.2.3 and (4.46) that for all sufficiently large $k$

$$\widehat{f}_{x_k}(0_{x_k}) - \widehat{m}_{x_k}(\eta_k) \geq \frac{1}{2} R_{\min} \|\eta_k\|_{x_k}^2. \tag{4.53}$$

Using (4.25), (4.52) and (4.53) yields for all sufficiently large $k$

$$r_k \leq \frac{1}{2} \|\eta_k\|^2 \left\{ \frac{\|(\text{Hess } \widehat{f}_{x_k}(\tau\eta_k) - B_k)\eta_k\|}{\|\eta_k\|_{x_k}} - (1 - \xi_2) R_{\min} \right\}. \tag{4.54}$$

For $k \geq 0$, we have

$$\begin{aligned} \frac{\|(\text{Hess } \widehat{f}_{x_k}(\tau\eta_k) - B_k)\eta_k\|_{x_k}}{\|\eta_k\|_{x_k}} \leq \quad & \|\text{Hess } \widehat{f}_{x_k}(0_{x_k}) - \text{Hess } \widehat{f}_{x_k}(\tau\eta_k)\|_{x_k} + \\ & \frac{\|(\text{Hess } \widehat{f}_{x_k}(0_{x_k}) - B_k)\eta_k\|_{x_k}}{\|\eta_k\|_{x_k}}. \end{aligned} \tag{4.55}$$

The bound $\|\tau\eta_k - 0_{x_k}\| \leq \|\eta_k\|_{x_k}$ follows since $\tau\eta_k$ is on the line segment between $0_{x_k}$ and $\eta_k$. This along with (4.47) and $\|\text{grad } f(x_k)\|_{x_k} \to 0$, imply $\|\tau\eta_k - 0_{x_k}\| \to 0$ which combined with (4.45) and Hess $\widehat{f}_x$ continuous implies that $\|\text{Hess } \widehat{f}_{x_k}(0_{x_k}) - \text{Hess } \widehat{f}_{x_k}(\tau\eta_k)\| \to 0$, as $k \to \infty$. We have that as $k \to \infty$

$$\frac{\|(\text{Hess } \widehat{f}_{x_k}(\tau\eta_k) - B_k)\eta_k\|_{x_k}}{\|\eta_k\|_{x_k}} \to 0,$$

from (4.48) and (4.55). Therefore, for all sufficiently large $k$

$$\|(\mathrm{Hess}\,\widehat{f}_{x_k}(\tau\eta_k) - B_k)\eta_k\|_{x_k}/\|\eta_k\|_{x_k} < (1-\xi_2)R_{\min}.$$

This, together with (4.14) and (4.54), imply for all sufficiently large $k$ that $r_k < 0$ and the iteration $k$ is very successful. Finally, $\sigma_k$ is bounded from above since on the very successful steps of the RARC algorithm $\sigma_k$ is cannot increase. □

Theorem 4.2.3 generalizes a combination of [9, Theorems 4.4 and 4.5] and shows that, the sequence of iterates $\{x_k\}$ converges to a local minimizer under certain conditions.

**Theorem 4.2.3.** *Let $x_*$ be a nondegenerate local minimizer of $f$, i.e.,* $\mathrm{grad}\, f(x_*) = 0$ *and* $\mathrm{Hess}\, f(x_*)$ *is positive definite, and assume there exists $\underline{\lambda} > 0$ such that*

$$\lambda_{\min}(B_k) \geq \underline{\lambda} \tag{4.56}$$

*on a neighborhood of $x_*$ and that* (4.28) *holds for some $\mu > 0$ and $\delta_\mu > 0$. If RF.1, RF.2, RM.1 and* (4.8) *hold, $\{f(x_k)\}$ is bounded below, then there exists a neighborhood $\mathcal{V}$ of $x_*$ such that, for all $x_0 \in \mathcal{V}$, the sequence $\{x_k\}$ generated converges to $x_*$.*

*Proof.* Take $\delta_1 > 0$ with $\delta_1 < \delta_\mu$ such that (4.56) holds on $B_{\delta_1}(x_*)$, that $B_{\delta_1}(x_*)$ contains only $x_*$ as critical point, and that $f(x) > f(x_*)$ for all $x \in \bar{B}_{\delta_1}(x_*)$. (In view of the assumptions, such a $\delta_1$ exists.)

From Lemma 4.2.4, we have

$$\|\eta_k\|_{x_k} \leq \frac{1}{\underline{\lambda}}\|\mathrm{grad}\, f(x_k)\|_{x_k}.$$

From [4, Lemma 7.4.8], we have, given $c > \lambda_{\max}$, the maximal eigenvalue of $\mathrm{Hess}\, f(x_*)$, there exists a neighborhood $\mathcal{V}$ of $x_*$ such that, for all $x_k \in \mathcal{V}$, it holds that

$$\|\mathrm{grad}\, f(x_k)\|_{x_k} \leq c \;\mathrm{dist}\;(x_*, x_k).$$

Take $\delta_2$ small enough, such that $\mathrm{dist}(x_*, x_k) \leq \delta_2 \leq \frac{\underline{\lambda}\mu}{c+\underline{\lambda}\mu}\delta_1$, for all $x_k \in B_{\delta_2}(x_*)$, then

$$\|\eta_k\|_{x_k} \leq \mu(\delta_1 - \delta_2).$$

From (4.28), the following inequalities hold

$$\mathrm{dist}(x_k, x_+) \leq \frac{1}{\mu}\|\eta_k\|_{x_k} \leq \delta_1 - \delta_2.$$

It follows from the equation above that $x_+$ is in $B_{\delta_1}(x_*)$. Moreover, since $f(x_+) \leq f(x)$, it follows that $x_+ \in \mathcal{V}$ . Thus $\mathcal{V}$ is invariant. But the only critical point of $f$ in $\mathcal{V}$ is $x_*$, so $\{x_k\}$ goes to $x_*$ whenever $x_0$ is in $\mathcal{V}$ . □

We next show (Corollaries 4.2.2 and 4.2.3) that the RARC algorithm is at least Q-superlinearly convergent under certain conditions. We begin with Lemma 4.2.5 that generalizes [9, Lemma 4.6].

**Lemma 4.2.5.** *If RF.1, RF.2 ′ and TC.h hold then for each $k \in \mathcal{S}$, with $\mathcal{S}$ defined in (4.21), we have*

$$(1-\kappa_\theta)\|\mathrm{grad}\widehat{f}_{x_k}(\eta_k)\|_{x_k} \leq \Big\| \int_0^1 \mathrm{Hess}\widehat{f}_{x_k}(\tau\eta_k)d\tau - \mathrm{Hess}\widehat{f}_{x_k}(0_{x_k}) \Big\|_{x_k} \|\eta_k\|_{x_k} + \\ \|(\mathrm{Hess}\widehat{f}_{x_k}(0_{x_k}) - B_k)\eta_k\|_{x_k} + \kappa_\theta\kappa_H h_k \|\eta_k\|_{x_k} + \sigma_k\|\eta_k\|^2_{x_k}, \tag{4.57}$$

*where $\kappa_\theta \in (0,1)$ occurs in TC.h.*

*Proof.* For $k \in \mathcal{S}$ we have

$$\begin{aligned} \|\mathrm{grad}\widehat{f}_{x_k}(\eta_k)\|_{x_k} &\leq \|\mathrm{grad}\widehat{f}_{x_k}(\eta_k) - \mathrm{grad}\,\widehat{m}_{x_k}(\eta_k)\|_{x_k} + \|\mathrm{grad}\,\widehat{m}_{x_k}(\eta_k)\|_{x_k} \\ &\leq \|\mathrm{grad}\widehat{f}_{x_k}(\eta_k) - \mathrm{grad}\,\widehat{m}_{x_k}(\eta_k)\|_{x_k} + \theta_k\|\mathrm{grad}\,f(x_k)\|_{x_k}, \end{aligned} \tag{4.58}$$

where the last inequality follows from TC.h.

We have

$$\mathrm{grad}\,\widehat{f}_{x_k}(\eta) = \mathrm{grad}\,\widehat{f}_{x_k}(0_{x_k}) + \int_0^1 \mathrm{Hess}\,\widehat{f}_{x_k}(\tau\eta)[\eta]d\tau$$

and

$$\mathrm{grad}\,\widehat{m}_{x_k}(\eta_k) = \mathrm{grad}f(x_k) + B_k\eta_k + \sigma_k\langle\eta,\eta\rangle^{\frac{1}{2}}\eta$$

and it follows that

$$\|\mathrm{grad}\,\widehat{f}_{x_k}(\eta) - \mathrm{grad}\,\widehat{m}_{x_k}(\eta_k)\|_{x_k} \leq \Big\| \int_0^1 (\mathrm{Hess}\,\widehat{f}_{x_k}(\tau\eta_k) - B_k)[\eta_k]d\tau \Big\| + \sigma_k\|\eta_k\|^2_{x_k}. \tag{4.59}$$

Using Taylor's Theorem and RF.2 ′ gives

$$\|\mathrm{grad}\,f(x_k)\|_{x_k} = \Big\|\mathrm{grad}\widehat{f}_{x_k}(\eta_k) - \int_0^1 \mathrm{Hess}\,\widehat{f}_{x_k}(\tau\eta_k)[\eta_k]d\tau\Big\| \leq \|\mathrm{grad}\widehat{f}_{x_k}(\eta_k)\|_{x_k} + \kappa_H\|\eta_k\|_{x_k}. \tag{4.60}$$

Substituting (4.60) and (4.59) into (4.58), yields

$$(1-\theta_k)\|\mathrm{grad}\widehat{f}_{x_k}(\eta_k)\|_{x_k} \leq \Big\| \int_0^1 (\mathrm{Hess}\,\widehat{f}_{x_k}(\tau\eta_k) - B_k)[\eta_k]d\tau \Big\|_{x_k} + \theta_k\kappa_H\|\eta_k\|_{x_k} + \sigma_k\|\eta_k\|^2_{x_k}. \tag{4.61}$$

and, since $\theta_k \leq \kappa_\theta h_k$ and $\theta_k \leq \kappa_\theta$ for TC.h, this is equivalent to

$$(1-\theta_k)\|\mathrm{grad}\widehat{f}_{x_k}(\eta_k)\|_{x_k} \leq \Big\| \int_0^1 (\mathrm{Hess}\,\widehat{f}_{x_k}(\tau\eta_k) - B_k)[\eta_k]d\tau \Big\|_{x_k} + \kappa_\theta\kappa_H h_k\|\eta_k\|_{x_k} + \sigma_k\|\eta_k\|^2_{x_k}. \tag{4.62}$$

Combining (4.62) and the triangle inequality

$$\Big\| \int_0^1 (\mathrm{Hess}\,\widehat{f}_{x_k}(\tau\eta_k) - B_k)[\eta_k]d\tau \Big\|_{x_k} \leq \Big\| \int_0^1 \mathrm{Hess}\,\widehat{f}_{x_k}(\tau\eta_k)d\tau - \mathrm{Hess}\,\widehat{f}_{x_k}(0_{x_k}) \Big\|_{x_k} \cdot \|\eta_k\|_{x_k} + \\ \|(\mathrm{Hess}\,\widehat{f}_{x_k}(0_{x_k}) - B_k)\eta_k\|_{x_k}$$

yields (4.57). □

Lemma 4.2.6 generalizes [9, Lemma 4.7] and shows that $\eta_k$ is bounded below under certain conditions when the TC.h criterion is used.

**Lemma 4.2.6.** *Suppose $k \to \infty$ we have $x_k \to x_*$ and let RF.1, RF.2 ′, RM.2 hold. If TC.h is satisfied with*

$$h_k \to 0, \text{ as } k \to \infty, k \in \mathcal{S}. \tag{4.63}$$

*then $\eta_k$ satisfies*

$$\|\eta_k\|_{x_k}(d_k + \sigma_k \|\eta_k\|_{x_k}) \geq (1 - \kappa_\theta)\|\text{grad}\widehat{f}_{x_k}(\eta_k)\|_{x_k} \text{ for all } k \in \mathcal{S}, \tag{4.64}$$

*where $d_k > 0$ for all $k \geq 0$, and*

$$d_k \to 0, \text{ as } k \to \infty, k \in \mathcal{S}. \tag{4.65}$$

*Proof.* The inequality (4.57) can be expressed as

$$(1-\kappa_\theta)\|\text{grad}\widehat{f}_{x_k}(\eta_k)\|_{x_k} \leq \Big[\Big\| \int_0^1 (\text{Hess}\, \widehat{f}_{x_k}(\tau\eta_k) - \text{Hess}\, \widehat{f}_{x_k}(0_{x_k}))d\tau \Big\|_{x_k} + \frac{\|(\text{Hess}\, \widehat{f}_{x_k}(0_{x_k}) - B_k)[\eta_k]\|_{x_k}}{\|\eta_k\|_{x_k}} + \kappa_\theta \beta_{RL} h_k\Big] \|\eta_k\|_{x_k} + \sigma_k \|\eta_k\|^2_{x_k},$$

where $k \in \mathcal{S}$ . If $d_k$ denotes the expression multiplying $\|\eta_k\|_{x_k}$ then since $h_k > 0$ we have $d_k > 0$ . Since the Hess $\widehat{f}_x$ is continuous and $\tau\eta_k$ is on the line segment between $0_{x_k}$ and $\eta_k$ for all $\tau \in (0,1)$, and $x_k \to x_*$, it follows that as $k \to \infty$

$$\Big\| \int_0^1 (\text{Hess}\, \widehat{f}_{x_k}(\tau\eta_k) - \text{Hess}\, \widehat{f}_{x_k}(0_{x_k}))d\tau \Big\|_{x_k} \to 0. \tag{4.66}$$

Recalling that $\|\text{grad}\, f(x_k)\| \to 0$ due to Corollary 4.2.1, RM.2, (4.63) and (4.66) imply that $d_k \to 0$, as the index $k \in \mathcal{S}$ increases. □

Corollary 4.2.2 generalizes [9, Corollary 4.8] and shows that the RARC algorithm is asymptotically Q-superlinearly convergent.

**Corollary 4.2.2.** *If RF.1, RF.2, RF.2 ′, RM.1 and RM.2 hold, $\eta_k$ satisfies* (4.8)*, and*

$$x_k \to x_*, \text{ as } k \to \infty, \tag{4.67}$$

*where* Hess $f(x_*)$ *is positive definite and if, additionally, TC.h holds with $h_k \to 0, k \to \infty, k \in \mathcal{S}$ then*

$$\frac{\|\text{grad}\, f(x_{k+1})\|_{x_{k+1}}}{\|\text{grad}\, f(x_k)\|_{x_k}} \to 0, \text{ as } k \to \infty \tag{4.68}$$

*and*

$$\frac{dist(x_{k+1}, x_*)}{dist(x_k, x_*)} \to 0, \text{ as } k \to \infty. \tag{4.69}$$

*The limits* (4.68) *and* (4.69) *hold when $h_k = \|\eta_k\|_{x_k}$ or $h_k = \|\text{grad}\, f(x_k)\|^{1/2}_{x_k}$,*
*$k \geq 0$, which are the termination criteria TC.s and TC.g, respectively.*

*Proof.* Since the hypothesis in this corollary also satisfies the conditions of Lemma 4.2.6, considering the conclusion in Theorem 4.2.2, (4.64) gives

$$\|\eta_k\|_{x_k}(d_k + \sigma_{\text{sup}}\|\eta_k\|_{x_k}) \geq (1-\kappa_\theta)\|\text{grad}\widehat{f}_{x_k}(\eta_k)\|_{x_k}, \forall k \text{ sufficiently large,} \tag{4.70}$$

where $\sigma_{\text{sup}}$ is the upper bound of $\sigma_k$, $d_k > 0$ and $\kappa_\theta \in (0, 1)$. From (4.47), (4.70) can be written for all sufficiently large $k$ as

$$\begin{aligned}
&\frac{1}{R_{\min}}\Big(d_k + \frac{\sigma_{\text{sup}}}{R_{\min}}\|\text{grad}\, f(x_k)\|_{x_k}\Big)\|\text{grad}\, f(x_k)\|_{x_k} \\
&\geq \|\eta_k\|_{x_k}(d_k + \sigma_{\text{sup}}\|\eta_k\|_{x_k}) \\
&\geq (1-\kappa_\theta)\|\text{grad}\widehat{f}_{x_k}(\eta_k)\|_{x_k},
\end{aligned}$$

and since $\text{grad}\, f(x_k) \neq 0$ and [[3], Lemma 4.9], we have

$$\begin{aligned}
\frac{\|\text{grad}\, f(x_{k+1})\|_{x_{k+1}}}{\|\text{grad}\, f(x_k)\|_{x_k}} &\leq c_5 \frac{\|\text{grad}\widehat{f}_{x_k}(\eta_k)\|_{x_k}}{\|\text{grad}\, f(x_k)\|} \\
&\leq c_5\left(\frac{R_{\min}d_k + \sigma_{\text{sup}}\|\text{grad}\, f(x_k)\|_{x_k}}{R_{\min}^2(1-\kappa_\theta)}\right), \forall k \text{ sufficiently large.}
\end{aligned} \tag{4.71}$$

From (4.65), the fact that all iterations with sufficiently large $k$ are successful and Corollary 4.2.1, we have the following

$$d_k \to 0 \text{ and } \|\text{grad}\, f(x_k)\|_{x_k} \to 0, \text{ as } k \to \infty. \tag{4.72}$$

It follows that as $k \to \infty$ The right-hand side of (4.71) tends to zero and therefore (4.68) holds.

Using Taylor expansions of $\text{grad}\, f(x_k)$ and $\text{grad}\, f(x_{k+1})$ around $x_*$, and recalling that $\text{grad} f(x_*) = 0$ with positive definite $\text{Hess}\, \widehat{f}_{x_*}(0)$ yields the limit (4.69).

The limit $\|\text{grad}\, f(x_k)\|_{x_k} \to 0$ and (4.47) imply that the choices of $h_k$ in TC.s and TC.g converge to zero, and thus the limits (4.68) and (4.69) hold for these choices of $h_k$ .

□

Lemma 4.2.7 generalizes [9, Lemma 4.9]. It makes a local Lipschitz continuity assumption on $\text{Hess}\, \widehat{f}_x$ in a neighborhood of a nondegenerate local minimizer of $f$, i.e., RF.3.

**Lemma 4.2.7.** *Let RF.1, RF.2, RF.3, RM.4 and TC.s hold. Suppose also that $x_k \to x_*$, as $k \to \infty$. If*

$$\sigma_k \leq \sigma_{\max} \text{ for all } k \geq 0 \tag{4.73}$$

*for some $\sigma_{\max} > 0$, then $\eta_k$ satisfies*

$$\|\eta_k\|_{x_k} \geq \kappa_g^* \sqrt{\|\text{grad}\widehat{f}_{x_k}(\eta_k)\|_{x_k}}, \text{ for all sufficiently large } k \in \mathcal{S}, \tag{4.74}$$

*where $\kappa_g^*$ is the positive constant*

$$\kappa_g^* \overset{def}{=} \sqrt{\frac{1-\kappa_\theta}{\frac{1}{2}L_* + C + \sigma_{\max} + \kappa_\theta\beta_{RL}}}. \tag{4.75}$$

*Proof.* Since the conditions of Lemma 4.2.5 are satisfied with $h_k = \|\eta_k\|_{x_k}$ and given RM.4 and (4.73), it follows that for any sufficiently large $k \in \mathcal{S}$, (4.57) can be written

$$\begin{aligned}(1-\kappa_\theta)\|\mathrm{grad}\widehat{f}_{x_k}(\eta_k)\|_{x_k} \leq \quad & \left\| \int_0^1 (\mathrm{Hess}\,\widehat{f}_{x_k}(\tau\eta_k) - \mathrm{Hess}\,\widehat{f}_{x_k}(0_{x_k}))d\tau \right\|_{x_k} \cdot \|\eta_k\|_{x_k} + \\ & C\|\eta_k\|_{x_k}^2 + (\sigma_{\max} + \kappa_\theta\beta_{RL})\|\eta_k\|_{x_k}^2. \qquad (4.76)\end{aligned}$$

Given $x_k \to x_*$, RF.3 and the fact that $\tau\eta_k$ is on the line segment defined by $0_{x_k}$ and $\eta_k$ for any $\tau \in (0,1)$, imply that for all sufficiently large $k \in \mathcal{S}$ we have

$$\begin{aligned}\left\| \int_0^1 (\mathrm{Hess}\,\widehat{f}_{x_k}(\tau\eta_k) - \mathrm{Hess}\,\widehat{f}_{x_k}(0_{x_k}))d\tau \right\| \quad & \leq \quad \int_0^1 \|\mathrm{Hess}\,\widehat{f}_{x_k}(\tau\eta_k) - \mathrm{Hess}\,\widehat{f}_{x_k}(0_{x_k})\|d\tau \\ & \leq \quad \frac{1}{2}L_*\|\eta_k\|_{x_k}\end{aligned}$$

The result (4.74) follows from (4.75) and writing (4.76)

$$(1-\kappa_\theta)\|\mathrm{grad}\widehat{f}_{x_k}(\eta_k)\|_{x_k} \leq (\frac{1}{2}L_* + C + \sigma_{\max} + \kappa_\theta\beta_{RL})\|\eta_k\|_{x_k}^2. \qquad (4.77)$$

□

Corollary 4.2.3 generalizes [9, Corollary 4.10] and shows that the RARC algorithm is asymptotically Q-quadratic convergent.

**Corollary 4.2.3.** *Assume that RF.1, RF.2, RF.3, RM.1, RM.2, RM.4 and TC.s hold. If $x_k \to x_*$, as $k \to \infty$, where $\mathrm{Hess}\, f(x_*)$ is positive definite, and $\eta_k$ satisfies (4.8) then, as $k \to \infty$, $\mathrm{grad}\, f(x_k)$ converges to zero, and $x_k$ to $x_*$, Q-quadratically.*

*Proof.* The conditions required in Lemma 4.2.7 are assumed to hold, so we have for all sufficiently large $k$

$$\|\eta_k\|_{x_k} \geq \kappa_g^* \sqrt{\|\mathrm{grad}\widehat{f}_{x_k}(\eta_k)\|_{x_k}}, \qquad (4.78)$$

where $\kappa_g^* > 0$. Therefore, given (4.46) it follows for all sufficiently large $k$

$$\frac{1}{R_{\min}}\|\mathrm{grad}\, f(x_k)\|_{x_k} \geq \|\eta_k\|_{x_k} \geq \kappa_g^* \sqrt{\|\mathrm{grad}\widehat{f}_{x_k}(\eta_k)\|_{x_k}}.$$

We have from (4.71) that

$$\frac{\|\mathrm{grad}\, f(x_{k+1})\|_{x_{k+1}}}{\|\mathrm{grad}\, f(x_k)\|_{x_k}^2} \leq c_5 \frac{\|\mathrm{grad}\widehat{f}_{x_k}(\eta_k)\|_{x_k}}{\|\mathrm{grad}\, f(x_k)\|_{x_k}^2} \leq c_5 \frac{1}{R_{\min}^2(\kappa_g^*)^2}, \text{ for all } k \text{ sufficiently large,}$$

and therefore $\mathrm{grad}\, f(x_k)$ converges Q-quadratically. By Taylor's theorem, the iterates convergent Q-quadratically. □

### 4.2.3 Global Convergence to Second-order Critical Points

In this section, we prove that the RARC method converges globally to a second order critical point of the cost function under appropriate assumptions and conditions, including

$$\sigma_k \geq \sigma_{\min}, \text{ for } k \geq 0, \tag{4.79}$$

for some $\sigma_{\min} > 0$.

Denote by $B_\delta(0_x) = \{\xi \in T_xM : \xi < \delta\}$ the open ball in $T_xM$ of radius $\delta$ centered at $0_x$, and $B_\delta(x)$ stands for the set $\{y \in M : \operatorname{dist}(x, y) < \delta\}$.

Lemma 4.2.8 generalizes [9, Lemma 5.1] and shows that the size of the steps $\eta_k$ approaches 0 under certain conditions.

**Lemma 4.2.8.** *Suppose $\{f(x_k)\}$ is bounded below by $f_{low}$. If $\eta_k$ satisfies* (4.8) *and* (4.9)*, $\sigma_k$, satisfies* (4.79) *and RF.1 holds then we have for $k \in \mathcal{S}$*

$$\|\eta_k\|_{x_k} \to 0, \text{ as } k \to \infty. \tag{4.80}$$

Lemma 4.2.9 generalizes [9, Lemma 5.2] and shows that $\sigma_k$ is bounded above given the following assumption on the Hessian of the lifted cost function:

$\boxed{\text{RF.4}}$ Hess $\widehat{f}_x$ is Lipschitz-continuous at $0_x$ uniformly in $x$, *i.e.*, there exists $L > 0$, $\delta_1 > 0$, such that, for all $x \in M$ and all $\xi \in B_{\delta_1}(0_x)$, it holds that

$$\|\text{Hess}\, \widehat{f}_x(\xi) - \text{Hess}\, \widehat{f}_x(0_x)\| \leq L\|\xi\|.$$

**Lemma 4.2.9.** *If RF.1, RF.4 and RM.4 hold then for all $k \geq 0$*

$$\sigma_k \leq \max(\sigma_0, \frac{3}{2}\gamma_2(C + L)) \overset{def}{=} L_0. \tag{4.81}$$

Next, we generalize [9, Theorem 5.3] as Theorem 4.2.4 to show that, at successful steps $\eta_k$, the limit points of the sequence of both Rayleigh quotients of $B_k$ and of the Hessian of the lifted cost function, $\text{Hess}\widehat{f}_{x_k}(0_{x_k})$, are nonnegative.

**Theorem 4.2.4.** *Suppose $\{f(x_k)\}$ be bounded below by $f_{low}$, If $\eta_k$ satisfies* (4.8) *and* (4.9)*, $\sigma_k$, satisfies* (4.79)*, RF.1, RF.4 and RM.4 hold then*

$$\liminf_{\substack{k\to\infty \\ k\in\mathcal{S}}} R_k(\eta_k) \geq 0 \text{ and } \liminf_{\substack{k\to\infty \\ k\in\mathcal{S}}} \frac{\langle \eta_k, \text{Hess}\, \widehat{f}_{x_k}(0_{x_k})[\eta_k]\rangle_{x_k}}{\|\eta_k\|^2_{x_k}} \geq 0. \tag{4.82}$$

*Proof.* For all $k \geq 0$ such that $R_k(\eta_k) < 0$, (4.9), (4.14) and (4.81) imply

$$L_0\|\eta_k\|_{x_k} \geq \sigma_k\|\eta_k\|_{x_k} \geq -R_k(\eta_k) = |R_k(\eta_k)|. \tag{4.83}$$

If $k \in \mathcal{K}$, where $\mathcal{K} = \{k \in \mathcal{S} \big| R_k(\eta_k) < 0\}$, then (4.80) and (4.83) imply $\{R_k(\eta_k)\}_{k\in\mathcal{S}} \to 0$ and the first limit in (4.82) follows. Through some manipulation on $R_k(\eta_k)$ and employing RM.4, we obtain the following inequalities

$$\begin{aligned} R_k(\eta_k) &\leq \frac{\|(\text{Hess}\widehat{f}_{x_k}(0_{x_k}) - B_k)[\eta_k]\|_{x_k}}{\|\eta_k\|_{x_k}} + \frac{\langle \eta_k, \text{Hess}\widehat{f}_{x_k}(0_{x_k})[\eta_k]\rangle_{x_k}}{\|\eta_k\|^2_{x_k}} \\ &\leq C\|\eta_k\|_{x_k} + \frac{\langle \eta_k, \text{Hess}\widehat{f}_{x_k}(0_{x_k})\eta_k\rangle_{x_k}}{\|\eta_k\|^2_{x_k}}, k \geq 0. \end{aligned} \tag{4.84}$$

The second inequality in (4.82) now follows from the first inequality, (4.80) and (4.84). □

A second order retraction is one that either satisfies the zero initial acceleration condition

$$\frac{\mathrm{D}^2}{Dt^2}R(t\xi)\Big|_{t=0} = 0 \text{ for all } \xi \in T_xM.$$

When $R$ is a second order retraction, or $x$ is a critical point,

$$\mathrm{Hess}\, f(x) = \mathrm{Hess}\,(f \circ R_x)(0_x).$$

For a general retraction, the Hessian of the cost function and the lifted cost function do not match in this manner and we need a more general relationship in order to turn (4.92) below, which is written in terms of the Hessian of $\hat{f}(x)$, into a statement written in terms of the Riemannian Hessian Hess $f$.

In Theorem 4.2.5 we show that for general retractions

$$\lim_{\substack{k\to\infty\\ k\in\mathcal{S}}} \inf\{\lambda_{\min}(Q_k^T \mathrm{Hess}\widehat{f}_{x_k}(0_{x_k})Q_k)\} \geq 0 \Longrightarrow \lim_{\substack{k\to\infty\\ k\in\mathcal{S}}} \inf\{\lambda_{\min}(Q_k^T \mathrm{Hess} f(x_k)Q_k)\} \geq 0,$$

where $Q_k$ is any matrix whose columns form an orthonormal basis of $\mathcal{L}_k$, which is a subspace of $T_{x_k}M$,

$$Q_k = (q_1, \cdots, q_j), \text{ for any } q_i, q_k \in Q_k, \langle q_i, q_k\rangle = \delta_{i,k}.$$

We will prove this theorem by contradiction. Letting $(x_k, Q_k)$ represent $(x_k, q_1, \cdots, q_j)$ and $\{v_k\}_{\mathcal{K}}$ to denote a set of objects from a sequence with indices in the set $\mathcal{K}$, the main task is to show that the set $\{(x_k, Q_k)\}_{\mathcal{K}}$ is a subset of a compact subset of a Whitney sum, denoted by $St(j, U')$, when $k$ is sufficiently large. We first propose two lemmas. In Lemma 4.2.10, we construct a bijection and $C^\infty$ mapping $\widetilde{\psi}$ from a Whitney sum to a subset of $\mathbb{R}^{n\times(j+1)}$, and show it equals a particular subset of $\mathbb{R}^{n\times(j+1)}$. In Lemma 4.2.11, we show that $\widetilde{\psi}(\mathrm{St}(j, U'))$ is compact subset of $\mathbb{R}^{n\times(j+1)}$. The compactness of $\mathrm{St}(j, U')$ follows from the fact that

$$\mathrm{St}(j, U') = \widetilde{\psi}^{-1}\widetilde{\psi}(\mathrm{St}(j, U')).$$

Let

$$St(j, M) := \{(x, q_1, \cdots, q_j)\big| x \in M,\, q_1, \cdots, q_j \in T_xM,\, g_x(q_i, q_k) = \langle q_i, q_k\rangle_x = \delta_{i,k}, \forall i, k \in 1, \cdots j\}$$

be an orthonormal $j$ frame. Denote the Whitney sum

$$TM \oplus \cdots \oplus TM := \{(\xi_1, \cdots, \xi_j)\big| \exists x \in M : \xi_i \in T_xM,\, \forall i = 1, \cdots, j\}$$

and observe that $TM \oplus \cdots \oplus TM \supset \mathrm{St}(j, M)$.

**Lemma 4.2.10.** *Assume the sequence $\{x_k\}$ converges to a critical point $x_*$. Let $(U, \psi)$ be a chart of $M$ with $x_* \in U$, and $\widetilde{\psi}$ be a bijection and $C^\infty$ mapping as following:*

$$\begin{array}{rcl} \widetilde{\psi} : U \times (TU \oplus \cdots \oplus TU) & \longmapsto & \psi(U) \times \mathbb{R}^{n\times j} \subseteq \mathbb{R}^{n\times(j+1)} : \\ (x, \xi_1, \cdots, \xi_j) & \longmapsto & (\psi(x), D\psi(x)[\xi_1], \cdots, D\psi(x)[\xi_j]) \end{array}$$

*Let $U'$ be such that $\psi(U')$ is compact, $x_* \in int(U')$ and $\psi(U') \subseteq \psi(U)$, then we have:*

$$\widetilde{\psi}(St(j, U')) = \{(y, V) \in \psi(U') \times \mathbb{R}^{n\times j} : V^T G_y V = I\} \subseteq \mathbb{R}^n \times \mathbb{R}^{n\times j}, \tag{4.85}$$

*where $e_i^T G_y e_k := g_{\psi^{-1}(y)}(D\psi^{-1}(y)[e_i], D\psi^{-1}(y)[e_k])$.*

*Proof.* Denote $S = \{(y, V) \in \psi(U') \times \mathbb{R}^{n\times j} : V^T G_y V = I\}$.
(1) Show $\widetilde{\psi}(\mathrm{St}(j, U')) \subset S$:
Let

$$(x, q_1, \cdots, q_j) \in \mathrm{St}(j, U'), g_x(q_i, q_k) = \langle q_i, q_k\rangle_x = \delta_{ik}.$$

We have

$$\widetilde{\psi}(x, q_1, \cdots, q_j) = (\psi(x), \mathrm{D}\psi(x)[q_1], \cdots, \mathrm{D}\psi(x)[q_j]).$$

$$\text{set } y = \psi(x), V = \mathrm{D}\psi(x)[q_1], \cdots, \mathrm{D}\psi(x)[q_j].$$

From the definition of $G_y$, we have

$$(\mathrm{D}\psi(x)[q_i])^T G_y \mathrm{D}\psi(x)[q_k] = \delta_{ik}.$$

So

$$V^T G_y V = I, \widetilde{\psi}(\mathrm{St}(j, U')) \in S.$$

(2) Show $S \subset \widetilde{\psi}(\mathrm{St}(j, U'))$:
Let

$$(y, V) \in S, \text{ then } y \in \psi(U') \text{ and } V^T G_y V = I.$$

Denote $V = [v_1, v_2, \cdots, v_j]$, and take

$$(x, Q) = (\psi^{-1}(y), \mathrm{D}\psi^{-1}(y)[v_1], \cdots, \mathrm{D}\psi^{-1}(y)[v_j])$$

Therefore,

$$(x, Q) \in \mathrm{St}(j, U'), \widetilde{\psi}(x, Q) = (y, V)$$

$$V^T G_y V = I \Longrightarrow (\mathrm{D}\psi(x)[q_i])^T G_y \mathrm{D}\psi(x)[q_k] = \delta_{ik}$$

We have

$$\langle v_i, v_k\rangle = \delta_{ik}, (y, V) \in \widetilde{\psi}(\mathrm{St}(j, U')).$$

□

**Lemma 4.2.11.** *$\widetilde{\psi}(St(j, U'))$ as described in Lemma 4.2.10 is a compact subset of $\psi(U) \times \mathbb{R}^{n\times j}$.*

*Proof.* Since $\widetilde{\psi}(\mathrm{St}(j, U'))$ is subset $\mathbb{R}^{n\times(j+1)}$, we just need to show it is closed and bounded.

Since $y \in \psi(U') \longmapsto G_y$ is $C^0$ mapping, we know $y \longmapsto \lambda_{\min}(G_y)$ is also a $C^0$ mapping. Also since $\psi(U')$ is compact, we have $\lambda_{\min}(G_y) > 0, \forall y \in \psi(U')$. So $\exists \underline{\lambda} > 0$, such that

$$\lambda_{\min}(G_y) \geq \underline{\lambda}, \forall y \in \psi(U').$$

Define the norm on $\mathbb{R}^n \times \mathbb{R}^{n\times j}$

$$\|(y, V)\|_F^2 = \|y\|^2 + \|V\|_F^2.$$

Since $y$ is in a compact set $\psi(U')$, $\exists b_1 \geq 0$, such that

$$\|y\|^2 \leq b_1, \forall y \in \psi(U'). \tag{4.86}$$

Since $V^T G y V = I$, we have

$$\|V\|_F^2 \leq \frac{1}{\underline{\lambda}}, \forall y \in \psi(U'). \tag{4.87}$$

From (4.86) and (4.87), we have

$$\|(y, V)\|_F^2 \leq b_1 + \frac{1}{\underline{\lambda}}, \forall (y, V) \in \widetilde{\psi}(\mathrm{St}(j, U')). \tag{4.88}$$

Therefore bounded is proven.

Let $\{(y_k, V_k)\}_{k \in \mathbb{N}} \subset \widetilde{\psi}(\mathrm{St}(j, U'))$, such that $\{(y_k, V_k)\} \longrightarrow (y_*, V_*)$. For any $k$, $V_k^T G_{y_k} V_k = I$, we have

$$V_*^T G_{y_*} V_* = \lim_{k \to \infty} V_k^T G_{y_k} V_k = I, \tag{4.89}$$

so $(y_*, V_*) \in \widetilde{\psi}(\mathrm{St}(j, U'))$. That is any limit point of $\{(y_k, V_k)\}$ is still in $\widetilde{\psi}(\mathrm{St}(j, U'))$ and closed is proven.

□

We can now prove the required relationship between the Hessian of the cost function and the Hessian of the lifted cost function for a general retraction.

**Theorem 4.2.5.**

$$\lim_{\substack{k \to \infty \\ k \in \mathcal{S}}} \inf\{\lambda_{\min}(Q_k^T \mathrm{Hess}\widehat{f}_{x_k}(0_{x_k}) Q_k)\} \geq 0 \Longrightarrow \lim_{\substack{k \to \infty \\ k \in \mathcal{S}}} \inf\{\lambda_{\min}(Q_k^T \mathrm{Hess} f(x_k) Q_k\} \geq 0, \tag{4.90}$$

*where $Q_k$ is any matrix whose columns form an orthonormal basis of a subspace of $T_{x_k} M$.*

*Proof.* Denote $a_k = \lambda_{\min}(Q_k^T \mathrm{Hess}\widehat{f}_{x_k}(0_{x_k}) Q_k), b_k = \lambda_{\min}(Q_k^T \mathrm{Hess} f(x_k) Q_k)$. Suppose (4.90) does not hold, then there exists an index set $\mathcal{K} \subseteq \mathcal{S}$, such that $\{b_k\}_{\mathcal{K}} \to \epsilon$, for some $\epsilon < 0$. Further restrict this index set in such a way that $Q_k$ contains a fixed number of columns, say $j$. From Lemma 4.2.11, we know $\{(x_k, Q_k)\}_{\mathcal{K}} = (x_k, q_1, \cdots, q_j)_{\mathcal{K}} \subset St(j, U')$, which is a compact subset of a Whitney sum, for all $k$ sufficiently large, so $\exists \mathcal{K}' \subseteq \mathcal{K}$ and critical point $(x_*, Q_*) \in St(j, U')$, such that $\{(x_k, Q_k)\}_{\mathcal{K}'} \longrightarrow (x_*, Q_*)$.

So

$$\lim_{\substack{k \to \infty \\ k \in \mathcal{K}'}} a_k = \lambda_{\min}(Q_*^T \mathrm{Hess}\widehat{f}_{x_*}(0_{x_*}) Q_*) \geq 0.$$

By continuity, we have $\lim_{\substack{k \to \infty \\ k \in \mathcal{K}'}} b_k = \lambda_{\min}(Q_*^T \mathrm{Hess} f(x_*) Q_*) \geq 0$, which yields the desired contradiction. □

The result above shows that (4.92) can be turned into a statement about the Riemannian Hessian. The proof of Theorem 4.2.6 is straight-forward to generalize from the Euclidean proof of [9, Theorem 5.4]. It is therefore omitted.

**Theorem 4.2.6.** *Assume that $\{f(x_k)\}$ is bounded below by $f_{low}$, that $\sigma_k$ satisfies* (4.79)*, that $\eta_k$ is a global minimizer of $\widehat{m}_{x_k}$ over a subspace $\mathcal{L}_k$, and let $Q_k$ be any orthogonal matrix whose columns form a basis of $\mathcal{L}_k$.*

*If RF.1, RF.4 and RM.4 hold then any subsequence of negative leftmost eigenvalues* $\{\lambda_{\min}(Q_k^T B_k Q_k)\}$ *converges to zero as* $k \to \infty, k \in \mathcal{S}$*, and thus*

$$\lim_{\substack{k \to \infty \\ k \in \mathcal{S}}} \inf\{\lambda_{\min}(Q_k^T B_k Q_k)\} \geq 0. \tag{4.91}$$

*Additionally, if RF.2, RM.1 and RM.3 hold then any subsequence of negative leftmost eigenvalues* $\{\lambda_{\min}(Q_k^T \mathrm{Hess}\, \widehat{f}_{x_k}(0_{x_k}) Q_k)\}$ *converges to zero as* $k \to \infty, k \in \mathcal{S}$ *, and thus*

$$\lim_{\substack{k \to \infty \\ k \in \mathcal{S}}} \inf\{\lambda_{\min}(Q_k^T \mathrm{Hess}\, \widehat{f}_{x_k}(0_{x_k}) Q_k)\} \geq 0. \tag{4.92}$$

*which implies*

$$\lim_{\substack{k \to \infty \\ k \in \mathcal{S}}} \inf\{\lambda_{\min}(Q_k^T \mathrm{Hess} f(x_k) Q_k)\} \geq 0 \tag{4.93}$$

*Furthermore, if* $Q_k$ *becomes a full orthogonal basis of* $T_x M$ *as* $k \to \infty, k \in \mathcal{S}$*, then, provided it exists, any limit point of the sequence of iterates* $\{x_k\}$ *is second-order critical.*

# CHAPTER 5

# EXPERIMENTS

## 5.1 Problems

Four problems are used to illustrate various aspects of the performance of the proposed Riemannian optimization algorithms: Rayleigh quotient minimization, a matrix Procrustes problem, Thomson's problem, and a weighted low-rank matrix approximation problem. In this section, the cost functions and other relevant information are presented for each problem.

In [24], the RBFGS is used in for ICA Learning on the Oblique manifold for image processing problems for unmixing natural images, brain MRI classification for axial slices and land typology estimation from multispectral infrared visible imaging spectrometer (MIVIS) data. The RBFGS method was compared to other ICA methods and a substantial improvement in the solution accuracy and computational efficiency was observed.

**Rayleigh quotient minimization on $S^{n-1}$**

For a symmetric matrix $A$, the unit-norm eigenvector, $v$, corresponding to the smallest eigenvalue, defines the two global minima, $\pm v$, of the Rayleigh quotient $f : S^{n-1} \to \mathbb{R}$, $x \mapsto x^T Ax$. The gradient and Hessian of $f$ are given by

$$\begin{aligned} \operatorname{grad} f(x) &= 2\mathrm{P}_x(Ax) = 2(Ax - xx^T Ax) \\ \mathrm{Hess} f(x) : T_xM \ &\to \ T_xM : \ \eta \to \ \nabla_\eta \operatorname{grad} f(x) \\ \text{where} \quad \nabla_\eta \operatorname{grad} f(x) &= 2\mathrm{P}_x(A\eta - \eta x^T Ax) = 2(\mathrm{P}_x A \mathrm{P}_x \eta - \eta x^T Ax) \end{aligned}$$

**Matrix Procrustes problem on the Stiefel manifold** $\mathrm{St}(p, n)$

On $\mathrm{St}(p, n)$ we consider a matrix Procrustes problem that minimizes the cost function $f : \mathrm{St}(p, n) \to \mathbb{R},\ X \to \|AX - XB\|_F$ given $n \times n$ and $p \times p$ matrices $A$ and $B$ respectively.

Consider the cost function embedded in $\mathbb{R}^{n\times p}$:

$$\bar{f} : \mathbb{R}^{n\times p} \to \mathbb{R} : X \to \|AX - XB\|_F, \quad \text{with} \ \ f = \bar{f}\big|_{St(p,n)}$$

The gradient of $f$ on the submanifold of $\mathbb{R}^{n\times p}$ used to represent $\mathrm{St}(p, n)$ is

$$\begin{gathered} \operatorname{grad} f(X) = \mathrm{P}_X \operatorname{grad} \bar{f}(X) = Q - X \operatorname{sym}(X^T Q), \ \text{where} \\ Q := A^T AX - A^T XB - AXB^T + XBB^T. \end{gathered}$$

The Hessian is given by

$$\begin{aligned}\operatorname{Hess} f(X)[Z] = \mathrm{P}_X \operatorname{Dgrad} f(X)[Z] = \operatorname{Dgrad} f(X)[Z] - X\operatorname{sym}(X^T \operatorname{Dgrad} f(X)[Z])\\ \operatorname{Dgrad} f(X)[Z] = \mathrm{D}Q(X)[Z] - X\operatorname{sym}(Z^T Q) - Z\operatorname{sym}(X^T Q)\\ \mathrm{D}Q(X)[Z] = A^T AZ - A^T ZB - AZB^T + ZBB^T\end{aligned}$$

where sym($Q$) is the symmetric part of the matrix $Q$.

**Thomson problem on** $\mathrm{OB}(n, N) : S^{n-1} \times \cdots \times S^{n-1}$

Let $x_i \in \mathbb{R}^n$, $1 \leq i \leq N$, be such that $x_i^T x_i = 1$ and consider the cost function

$$f : [x_1, x_2, \cdots, x_N] \longmapsto \sum_{\substack{i,j=1\\ i\neq j}}^{N} \frac{1}{\|x_i - x_j\|^2}$$

The physical meaning of the optimization this cost function is to optimally arrange $N$ repulsive particles on a sphere and determine the minimum energy configuration of these particles.

We view this as an optimization problem on $\mathrm{OB}(n, N)$ the elements of which have the form $X = [x_1, x_2, \cdots, x_N] \in M, \quad x_i^T x_i = 1,$ for $i = 1$ to $N$.

To compute the gradient and Hessian of $f$, we first consider the cost function on entire embedding space and compute its gradient

$$\bar{f} : \mathbb{R}^n \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R} : X \to \sum_{\substack{i,j=1\\ i\neq j}}^{N} \frac{1}{\|x_i - x_j\|^2}$$

$$\mathrm{D}\bar{f}(X)[Z] = \Big(\sum_{j=2}^{N} \frac{1}{(1 - x_1^T x_j)^2} x_j^T\Big) z_1 + \cdots + \Big(\sum_{\substack{j=1\\ j\neq i}}^{N} \frac{1}{(1 - x_i^T x_j)^2} x_j^T\Big) z_i + \cdots + \Big(\sum_{j=1}^{N-1} \frac{1}{(1 - x_N^T x_j)^2} x_j^T\Big) z_N$$

It follows that the gradient on a submanifold is the projection of the gradient on the embedding manifold, $\operatorname{grad} f(X) = P_X \operatorname{grad} \bar{f}(X)$. The orthogonal projection of $W \in \mathbb{R}^{n\times N}$ to $T_x M$ is

$$\mathrm{P}_X W = [(I - x_1 x_1^T) w_1, \cdots, (I - x_N x_N^T) w_N]$$

and therefore

$$\begin{aligned}\operatorname{grad} f(X) = \Big[(I - x_1 x_1^T) \sum_{j=2}^{N} \frac{1}{(1 - x_1^T x_j)^2} x_j, \cdots, (I - x_i x_i^T) \sum_{\substack{j=1\\ j\neq i}}^{N} \frac{1}{(1 - x_i^T x_j)^2} x_j,\\ \cdots (I - x_N x_N^T) \sum_{j=1}^{N-1} \frac{1}{(1 - x_N^T x_j)^2} x_j\Big]\end{aligned}$$

Finally the Hessian is given by

$$\operatorname{Hess} f(X)[Z] = \mathrm{P}_X \operatorname{Dgrad} f(X)[Z].$$

**The weighted low-rank matrix approximations on** $\mathrm{Grass}(p, n)$

Given a data matrix $X \in \mathbb{R}^{p\times n}$ and a weighting matrix $Q \in \mathbb{R}^{pn\times pn}$ that defines

$$\|X - R\|_Q^2 = \mathrm{vec}\{X - R\}^T Q \mathrm{vec}\{X - R\},$$

the problem is to find that matrix $R_*$ such that

$$R_* = \arg \min_{\substack{R\in\mathbb{R}^{p\times n} \\ \mathrm{rank}\{R\}\leq r}} \|X - R\|_Q^2 \tag{5.1}$$

An alternative formulation, suggested by Brace and Manton [7], is to rewrite (5.1) as the equivalent double minimization

$$\min_{\substack{N\in\mathbb{R}^{n\times(n-r)} \\ N^T N=1}} \min_{\substack{R\in\mathbb{R}^{p\times n} \\ RN=0}} \|X - R\|_Q^2 \tag{5.2}$$

The inner minimization has a closed form solution, call it $f(N)$, given by:

$$f(N) = \mathrm{vec}\{X\}^T (N \otimes I_p) \Big[(N \otimes I_p)^T Q^{-1}(N \otimes I_p)\Big]^{-1} (N \otimes I_p)^T \mathrm{vec}\{X\} \tag{5.3}$$

This cost function depends only on the range space of $N$, rather than the actual value of $N$. That is, $f(NQ) = f(N)$ for any orthogonal matrix $Q$. If $N$ minimizes $f(N)$ then the solution to the original problem (5.1) is the unique matrix $R$ satisfying

$$\mathrm{vec}\{R\} = \mathrm{vec}\{X\} - Q^{-1}(N \otimes I_p)\Big[(N \otimes I_p)^T Q^{-1}(N \otimes I_p)\Big]^{-1}(N \otimes I_p)^T \mathrm{vec}\{X\}$$

From the retraction formula (3.27), the local cost function (5.3) becomes

$$g(K) = f(N + N^{\perp}K). \tag{5.4}$$

Finally, the gradient of $g(K)$ at $K = 0$ is

$$\mathrm{grad}\, g(0) = 2(N^{\perp})^T (X - C)^T A \tag{5.5}$$

where $A \in \mathbb{R}^{p\times(n-r)}$ and $C \in \mathbb{R}^{p\times n}$ are the unique matrices satisfying

$$\mathrm{vec}\{A\} = \Big[(N \otimes I_p)^T Q^{-1}(N \otimes I_p)\Big]^{-1} \mathrm{vec}\{XN\}$$
$$\mathrm{vec}\{C\} = Q^{-1}\mathrm{vec}\{AN^T\}$$

## 5.2 RARC Results

The RARC convergence analysis predicts rapid convergence and our observations coincide with our analysis. For each problem several initial conditions were tried and similar convergence rates and ultimate cost function values were observed. Given that ARC is similar in spirit to the trust-region method on $\mathbb{R}^n$ it was compared in [11] to trust region method (GLTR). We therefore compare RARC with the Riemannian version of Lanczos-based trust

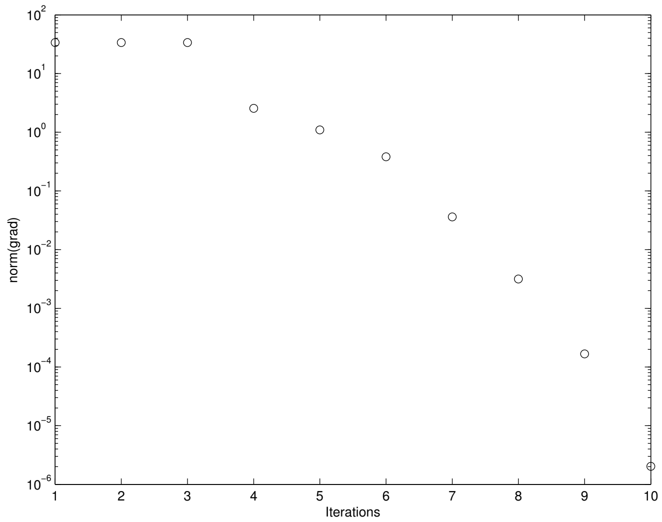

Figure 5.1: RARC convergence Rayleigh Quotient problem: n=100

region method (RGLTR). The code is a modified version of the truncated-CG-based Riemannian trust region method (RTR) of C.Baker's Ph.D. dissertation. The experiments show correspondingly that the RARC is competitive compared with Riemannian GLTR. Figures 5.1 and 5.2 show the rapid convergence of RARC for the Rayleigh quotient problem and for two Thomson problems. For both of these problems the performance of RGLTR is similar to RARC. Figure 5.3 however illustrates a potential benefit of RARC compared to RGLTR. Note that the convergence of RGLTR on the Procrustes problem exhibits the plateau effect so often seen in Lanczos iterations before converging very rapidly and as a result doing significantly more work. RARC, even though Lanczos is also used, benefits from its regularization parameter in its local cubic model and exhibits superior performance.

## 5.3 RBFGS Results

### 5.3.1 Approach 1 and Approach 2 for $H_k$ update form of RBFGS

We have observed that the $H_k$ update form of RBFGS tends to be more computationally efficient than the $B_k$ update forms. So in this section we present data that compares the $H_k$ update form for the three submanifold problems (Rayleigh Quotient, Thomson Problem, and Procrustes Problem) using Approach 1 or Approach 2. The data uses the most efficient isometric vector transport for the manifold associated with each problem. We expect the same convergence rates of course but there is a significant variation in computational cost between the two approaches and the choices of vector transport.

For the unit sphere and the oblique manifold, the simple and efficient isometric vector transport is the most efficient of all the isometric vector transports. Its efficiency arises from careful choice of the projection form used when applying vector transport and its inverse transport. The choice means we can avoid the need for a basis of the tangent space. The

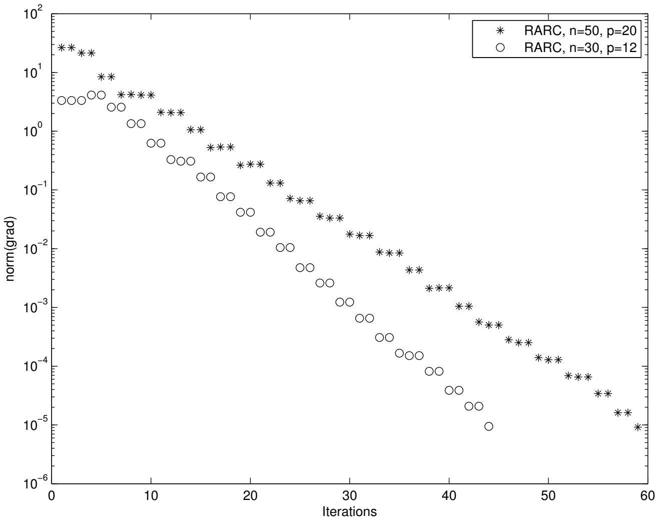

Figure 5.2: RARC convergence on two Thomson problems

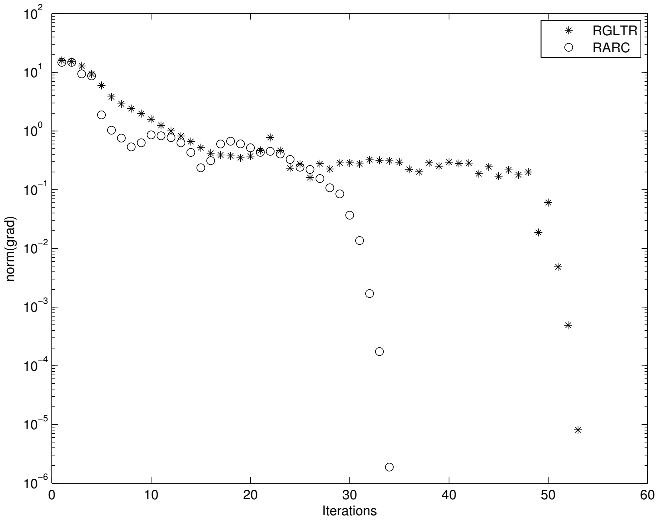

Figure 5.3: RARC and RGLTR convergence for the Procrustes problem: n=12, p=7

other isometric vector transports require generating a basis at each step of the algorithm (or transporting a basis from the previous step). For the compact Stiefel we do not have a form of isometric vector transport that is basis-free, but the data below uses the most efficient available. So, in general, Approach 1 is not completely independent of tangent space bases computationally. Since Approach 2 works in the core coordinates, it uses the basis not only for vector transport but also for projection of the problem before transporting and updating. As a result, we expect it to be more costly computationally.

Table 5.1: Vector transport Approach 1 vs. Approach 2 for Rayleigh quotient problem

| Case | Approach 1 | Approach 1 | Approach 2 | Approach 2 |
|---|---|---|---|---|
| | ( n=100) | (n=300) | (n=100) | (n=300) |
| Time | 0.21 | 4.6 | 0.54 | 11 |
| Iteration | 68 | 92 | 72 | 97 |

Table 5.2: Vector transport Approach 1 vs. Approach 2 for Procrustes problem

| Case | Approach 1 | Approach 1 | Approach 2 | Approach 2 |
|---|---|---|---|---|
| | ( n=7, p=4) | (n=12, p=7) | (n=7, p=4) | (n=12, p=7) |
| Time | 0.24 | 1.4 | 3.17 | 41 |
| Iteration | 47 | 79 | 47 | 79 |

Table 5.3: Vector transport Approach 1 vs. Approach 2 for Thomson problem

| Case | Approach 1 | Approach 1 | Approach 2 | Approach 2 |
|---|---|---|---|---|
| | ( n=30, N=12) | (n=50, N=20) | (n=30, N=12) | (n=50, N=20) |
| Time | 3.12 | 59 | 6.5 | 132 |
| Iteration | 22 | 24 | 24 | 25 |

Since, analytically, the $H_k$ update is identical to the $B_k$ update and Cholesky factor update form, the only source of convergence differences is numerical. In fact, these forms are observed to converge at the same rate. Similarly, Approach 1 and Approach 2 are identical and analytically converge at the same rate. We should, however, expect a potentially large difference in time. Tables 5.1, 5.2, and 5.3 contain Approach 1 and Approach 2 of the $H_k$ update form for the three problems with two problem sizes each. For each problem, the number of iterations are seen to be identical or very close for a given problem size. We also observe the significant difference in computing time between the two approaches. For the larger problem sizes on each manifold we have factors of 2.4, 29.3 and 2.2 in computing time.

We conclude that $H_k$ update form with Approach 1 is preferred. Also, it is crucial for efficiency to analyze the computational form of vector transport, as recommended in our projection framework, to make sure the most efficient and, if possible, a basis-free version is used. Approach 1 could further benefit from propagating rank factorizations of the $n \times n$ matrices $H_k$. (In this case, propagation of a low-rank Cholesky factorization would make the $B_k$ update version competitive as well but that option is not explored in these results.) Approach 2 has potential savings only if there is an efficient manner of projecting the problem to the core coordinates. This of course may simply be due to the tangent space having a relatively small dimension.

### 5.3.2 Analysis of experimental results using parallel transport and vector transport

Since parallel transport and vector transport by projection have similar computational costs on $S^{n-1}$, the corresponding RBFGS versions have a similar computational cost per iteration. Therefore, we would expect any performance difference measured by time to reflect differences in rates of convergence. Based on the discussions above we consider this question in this section using the Approach 1 implementation of the $H_k$ update form of RBFGS. Columns 2 and 3 of Table 5.4 show that vector transport produces a convergence rate very close to parallel transport and the times are close as expected. This is encouraging from the point of view that the more flexible vector transport did not significantly degrade the convergence rate of RBFGS.

Given that vector transport by projection is significantly less expensive computationally than parallel transport on $\mathrm{St}(p, n)$, for the Procrustes problem, we would expect a significant improvement in performance as measured by time if the vector transport version manages to achieve a convergence rate similar to parallel transport. The times in columns 4 and 5 of Table 5.4 show an advantage to the vector transport version larger than the computational complexity predicts. The iteration counts provide an explanation. Encouragingly, the use of vector transport actually improves convergence compared to parallel transport. We note that the parallel transport version performs the required numerical integration of a differential equation with a stepsize sufficiently small so that decreasing it does not improve the convergence rate of RBFGS but no smaller to avoid unnecessary computations. The data here and below provides strong evidence that a careful consideration of the choice of vector transport may have significant beneficial effects on both cost per step and overall convergence.

The vector transports used in Table 5.4 for the Rayleigh Quotient problem and Procrustes problem were both efficient isometric forms. Interestingly, even when they are replaced by a nonisometric vector transport, though they do not guarantee the preservation of symmetry on every step, they both converge very effectively (97 iterations and 83 for the the Rayleigh Quotient and Procrustes respectively). As mentioned earlier we have a conjecture as to how this can be shown to be consistent with our convergence theory and this will be pursued further. Figure 5.4 illustrates in more detail the significant improvement in convergence rate achieved for vector transport in RBFGS on the Procrustes problem. This does beg the question however of what an isometric vector transport can achieve.

Table 5.5 shows the number of iterations and time required for RBFGS on the Rayleigh

Table 5.4: RBFGS Isometric Vector transport vs. Parallel transport

| | Rayleigh<br>$n = 300$ | | Procrustes<br>$(n, p) = (12, 7)$ | |
|---|---|---|---|---|
| | Vector | Parallel | Vector | Parallel |
| Time (sec.) | 4.6 | 4.2 | 1.4 | 259 |
| Iteration | 92 | 95 | 79 | 175 |

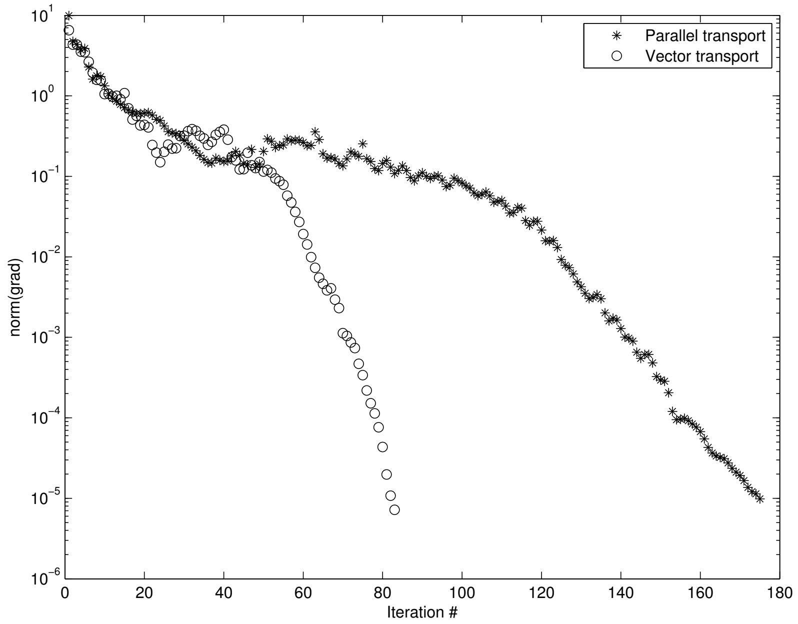

Figure 5.4: RBFGS Parallel and Vector Transport for Procrustes. n=12, p=7.

Quotient using an efficient nonisometric vector transport, the SVD form of the canonical-based isometric vector transport and an equivalent but computationally efficient form of the canonical-based isometric vector transport. As expected, the isometric vector transports converge at the same rate, as does the nonisometric version, while the efficient isometric vector transport derived by using the variety of implementations apparent from the projection framework uses less time than the SVD based implementation and is competitive with the efficient nonisometric version. Figure 5.5 shows the convergence of the nonisometric vector transport and the isometric vector transport in RBFGS on the Rayleigh Quotient compared to transformation between tangent spaces that is not a vector transport but has similarly high computational efficiency. In fact this transformation is the efficient isometric vector transport with a malicious sign change added that violates the requirements of vector transport. The results show that the vector transport property is crucial in achieving the desired results and much more important that the efficiency of each application of a transformation between tangent spaces.

Table 5.6 includes results with three different vector transports: nonisometric, canonical-based isometric and QR-based isometry denote isometric(QF) . They all have similar convergence rate for the Procrustes problem. But the nonisometric on has better efficiency. This is further evidence that the nonisometric vector transport can converge effectively.

The low-rank matrix approximation problem compares our rigorously analyzed RBFGS algorithms to the highly heuristic and computationally inexpensive per step RBFGS algorithm of Brace and Manton [7, Algorithm 2] that essentially completely ignores vector transport. Figure 5.6 shows that the convergence using RBFGS with the canonical-based vector transport is significantly better than the Brace-Manton heuristic form. Even more important is the fact that the RBFGS version required only 7 seconds vs. 21 seconds required by Brace-Manton. Other experiments show that nonisometric vector transport and canonical-based vector transports have similar timing and convergence results for the low-rank approximation problem.

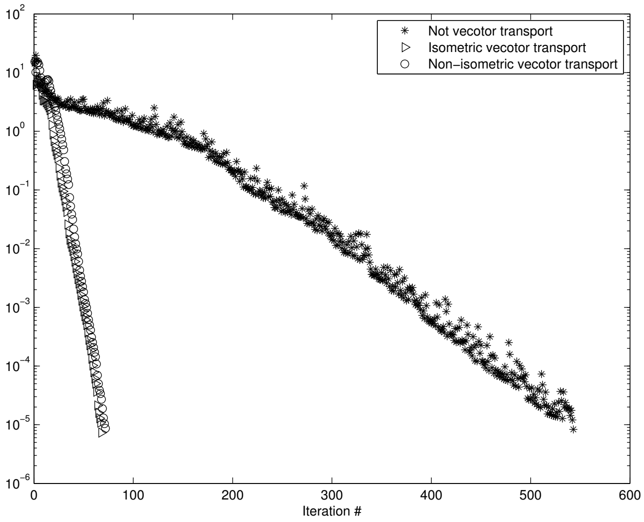

Figure 5.5: RBFGS with 3 transports for Rayleigh quotient. n=100.

Table 5.5: Non-isometric vs. Canonical-based isometric vs. Simple implementation

| | Rayleigh<br>$n = 300$ | | | |
|---|---|---|---|---|
| | Non-isometric | Canonical isometric | Simple imple. | Isometric(QF) |
| Time (sec.) | 4.0 | 20 | 4.6 | 17.5 |
| Iteration | 97 | 92 | 92 | 99 |

Table 5.6: RBFGS vector transports: nonisometric vs. canonical isometric (SVD) vs. isometric(QF)

| | Procrustes<br>$n = 12, p = 7$ | | |
|---|---|---|---|
| | Non-isometric | Canonical isometric | Isometric(QF) |
| Time (sec.) | 4.0 | 1.4 | 1.3 |
| Iteration | 83 | 79 | 80 |

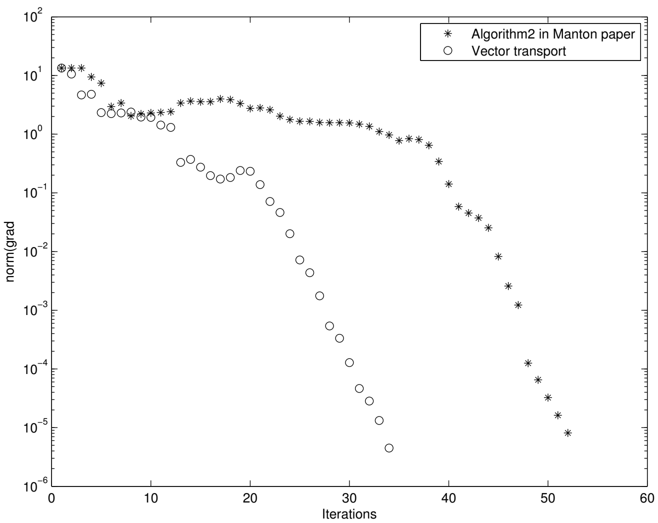

Figure 5.6: Low-rank (rank 3)approximation problem on Grass$(p, n)$. n=12, p=9

# CHAPTER 6

# CONCLUSIONS

We have generalized quasi-Newton algorithms to a Riemannian manifold (with an emphasis on RBFGS) and we have proven several important convergence results. The most general, and perhaps most significant, is the Riemannian Dennis-Moré Condition, Theorem 2.3.1, that gives a necessary and sufficient condition for a Riemannian quasi-Newton algorithm that defines its search direction based on vector transport, its associated retraction and the transport of a linear transformation to achieve superlinear convergence.

We have given a generalization of the Wolfe conditions that is a key aspect of the convergence analysis of Riemannian quasi-Newton algorithms. This discussion includes an alternative curvature condition associated with a general vector transport.

For the general form of RBFGS given as Algorithm 2, if the vector transport is assumed to be an isometry then Lemma 2.4.1 shows the crucial fact that the transport and update of the linear transformation that defines RBFGS preserves the self-adjoint and positive definite properties that are very important in proving convergence of RBFGS.

Our main convergence results for Algorithm 2 currently require the restriction to the use of parallel transport with the exponential map as the associated retraction. The restriction is due to our reliance on an average Hessian in the convergence analysis in a manner analogous to that of [22]. The assumption allows the derivation of the Riemannian Exponential Map Zoutendijk Condition in Theorem 2.4.1. This condition along with the positive definite property provides the basis for proceeding to demonstrate convergence of RBFGS using parallel transport by showing that linear transformation also has bounded condition.

Theorems 2.4.3 and 2.4.5 are the main results specific to RBFGS using parallel transport. Theorem 2.4.3 guarantees

1. global convergence of RBFGS using parallel transport to a unique minimizer when the cost $f(x)$ is convex on the domain of interest.

2. global convergence of RBFGS using parallel transport to a set of stationary points when the cost $f(x)$ is not convex on the domain of interest.

3. local convergence of RBFGS using parallel transport to a nondegenerate minimizer, $x^*$, when the cost $f(x)$ is not convex on the domain of interest but the initial guess $x_0$ is sufficiently close to $x^*$.

Given a converging iteration to a nondegenerate minimizer Theorem 2.4.5 guarantees superlinear convergence for RBFGS using parallel transport by showing that the Riemannian Dennis-Moré condition is satisfied.

We have therefore generalized to a Riemannian manifold all of the key convergence theorems for BFGS, in particular, and transport-based quasi-Newton algorithms, in general without relying on special assumptions such as $M$ being a submanifold of $\mathbb{R}^n$. The key limitation to our results is the fact that the reliance on the use of the average Hessian requires, thus far, the restriction to RBFGS using parallel transport. A generic vector transport/retraction pair does not satisfy all of the properties needed for this approach to the convergence proofs. For any particular choice, the required properties may be satisfied, depending on the method of construction of the vector transport/retraction pair, and can be specifically checked.

The fundamental results Theorem 2.3.1 and Lemma 2.4.1 do not have this restriction and they indicate that a more general result for RBFGS with isometric vector transport should be possible. In our experiments, RBFGS using isometric vector transport produces a bounded condition in addition to the positive definiteness guaranteed by Lemma 2.4.1 and therefore superlinear convergence is expected. In fact, our experiments with RBFGS using both isometric and nonisometric vector transport indicate that Theorem 2.3.1 and Lemma 2.4.1 are satisfied and superlinear convergence has been observed consistently.

As an alternative to earlier work on the Riemannian Trust Region family of methods, we have successfully generalized the Euclidean ARC algorithm and completed the convergence analysis for the resulting algorithm RARC. It successfully generalizes the very satisfactory convergence results available for ARC on $\mathbb{R}^n$. In particular, we have shown under a series of reasonable assumptions that RARC:

- converges globally to first-order critical points,
- converges Q-superlinearly or Q-quadratically to local minimizers,
- and converges globally to local minimizers.

We have provided leading empirical evidence that RARC can converge reasonably quickly for a set of simple test problems. Further work comparing RARC, as we have implemented it, to other more aggressively optimized methods such C. Baker's numerical library of a Riemannian trust-region family of methods [6] based on a Steighaug CG-like approach to solving the local minimization problem is needed.

# BIBLIOGRAPHY

[1] Ralph Abraham, Jerrold E. Marsden, and Tudor S. Ratiu. *Manifolds, Tensor Analysis and Applications*. Springer, New Jersey, second edition.

[2] P.-A. Absil, C. G. Baker, and K. A. Gallivan. Trust-region methods on Riemannian manifolds. *Found. Comput. Math.*, 7(3):303–330, July 2007.

[3] P.-A. Absil, C. G. Baker, and K. A. Gallivan. Trust-region methods on Riemannian manifolds. *Found. Comput. Math.*, 2007.

[4] P.-A. Absil, R. Mahony, and R. Sepulchre. Optimization algorithms on matrix manifolds. *Princeton University Press*, 2008.

[5] Roy L. Adler, Jean-Pierre Dedieu, Joseph Y. Margulies, Marco Martens, and Mike Shub. Newton's method on Riemannian manifolds and a geometric model for the human spine. *IMA J. Numer. Anal.*, 22(3):359–390, July 2002.

[6] Christopher G. Baker. *Riemannian manifold trust-region methods with applications to eigenproblems*. PhD thesis, School of Computational Science, Florida State University, Summer Semester 2008.

[7] Brace and J. H. Manton. An improved bfgs-on-manifold algorithm for computing weighted low-rank approximations. *In Proceedings 2006 International Symposium on the Mathematical Theory of Networks and Systems.*, pages 1735–1738, 2006.

[8] N.Del Buono and C.Elia. Computation of few lyapunov exponents by geodesic based algorithms. *Future Generation Computer systems*, 19:425–430, 2003.

[9] Coralia Cartis, Nicholas I. M. Gould, and Philippe Toint. Adaptive cubic overestimation methods for unconstrained optimization. Part I: motivation, convergence and numerical results. *Math. Program., Ser. A, DOI 10.1007/s10107-009-0286-5*, 2009.

[10] Coralia Cartis, Nicholas I. M. Gould, and Philippe Toint. Adaptive cubic overestimation methods for unconstrained optimization. Part II: worst-case function- and derivative-evaluation complexity. *Math. Program., Ser. A, DOI 10.1007/s10107-009-0337-y*, 2010.

[11] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. Trust-region methods. *SIAM*, 2000.

[12] J. E. Dennis and Robert B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*. Springer, New Jersey, 1983.

[13] David W. Dreisigmeyer. Direct search algorithms over Riemannian manifolds. Optimization Online 2007-08-1742, December 2006.

[14] Alan Edelman, Tomás A.Arias, and Steven T.Smith. The geometry of algorithms with orthogonality constrains. *SIAM J.Matrix Anal.Appl.*, 20(2):303–353, 1998.

[15] D. Gabay. Minimizing a differentiable function over a differentialmanifold. *J. Optim. Theory Appl.*, 37(2):177–219, 1982.

[16] U. Helmke and J. Moore. Optimization and dynamical systems. *Springer-Verlag*, 1994.

[17] J.H.Manton. Optimization algorithms exploiting unitary constrains. *IEEE Transactions on Signal Processing.*, 50:635–650, 2002.

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

[19] David G. Luenberger. The gradient projection method along geodesics. *Management Sci.*, 18:620–631, 1972.

[20] David G. Luenberger. *Introduction to Linear and Nonlinear Programming.* Addison-Wesley, Reading, MA, 1973.

[21] Yurii Nesterov and B. T. Polyak. Cubic regularization of Newton method and its global performance. *Math. Program.*, 108(1, Ser. A):177–205, 2006.

[22] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization.* Springer Series in Operations Research. Springer-Verlag, New York, 1999.

[23] B. Savas and L.-H. Lim. Best multilinear rank approximation of tensors with quasinewton methods on grassmannians. *Technical Report LITH-MAT-R-2008-1-SE*, pages 69–92, 2008.

[24] Easter Selvan, Umberto Amato, Chunhong Qi, Kyle A. Gallivan, Michele Larobina, and Bruno Alfano. Unconstrained optimizers for unconstrained learning on the oblique manifold using parzen density estimation. submitted for publication, 2010.

[25] Steven T. Smith. Optimization techniques on Riemannian manifolds. In Anthony Bloch, editor, *Hamiltonian and gradient flows, algorithms and control*, volume 3 of *Fields Inst. Commun.*, pages 113–136. Amer. Math. Soc., Providence, RI, 1994.

[26] Y. Yang. Globally convergent optimization algorithms on Riemannian manifolds: Uniform framework for unconstrained and constrained optimization. *J. Optim. Theory Appl.*, 132(2):245–265, 2007.

# BIOGRAPHICAL SKETCH

Chunhong Qi completed her Bachelor degree in Applied Mathematics in 1990 and her Master degree in Automatic Control in 1996, both from Heilongjiang University in China. She enrolled in the doctoral program of at The Florida State University in 2005. After obtaining her master degree in Applied and Computational Mathematics in 2009, she is currently under the advisement of Prof. Kyle A Gallivan and Prof. Pierre-Antoine Absil. Chunhong Qi's research interests include optimization methods on manifolds, time-invariant system control and image processing.

She has published two papers and submitted one paper.

- Chunhong Qi, Kyle A. Gallivan, P.-A. Absil. *Riemannian BFGS algorithm with applications*, **Recent Advances in Optimization and its Applications in Engineering**, Springer-Verlag, pp. 183-192, 2010. (full paper refereed)

- Chunhong Qi, Kyle A. Gallivan, P.-A. Absil. *An Efficient Riemannian BFGS Algorithm for Manifold Optimization*, **Proceedings of 2010 Mathematical Theory of Networks and Systems**, July 5-9, 2010, Budapest, Hungary, pp. 2221-2227, 2010. (extended abstract refereed)

- Easter Selvan, Umberto Amato, Chunhong Qi, Kyle A. Gallivan, Michele Larobina, and Bruno Alfano. *Unconstrained Optimizers for Unconstrained Learning on the Oblique Manifold using Parzen Density Estimation,* submitted.