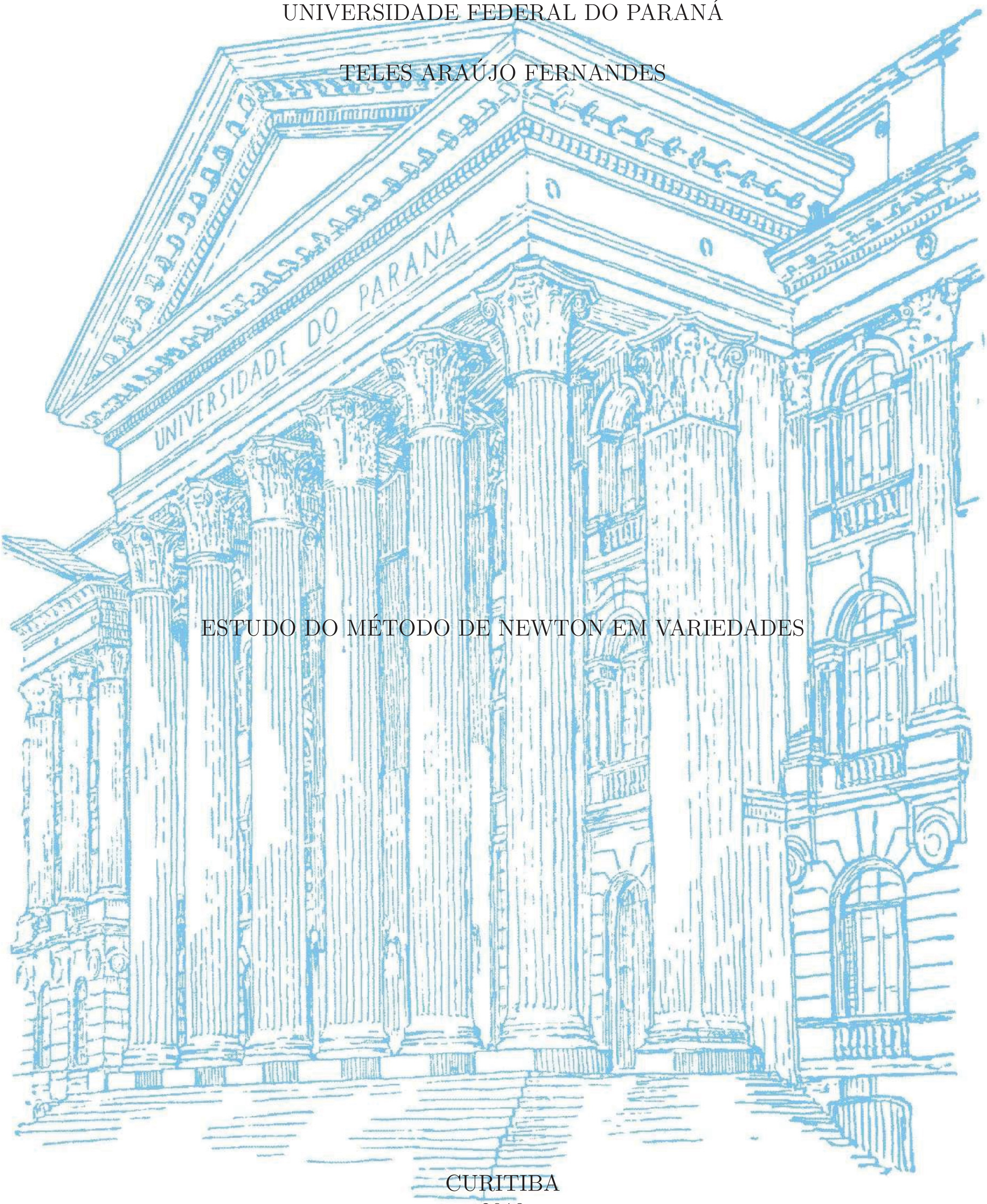
UNIVERSIDADE FEDERAL DO PARANÁ

TELES ARAÚJO FERNANDES



ESTUDO DO MÉTODO DE NEWTON EM VARIEDADES

CURITIBA
2018

TELES ARAÚJO FERNANDES

ESTUDO DO MÉTODO DE NEWTON EM VARIEDADES

Tese apresentada ao curso de Pós-Graduação em Matemática Aplicada, Setor de Ciências Exatas, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Doutor em Matemática Aplicada.

Orientador: Prof. Dr. Jinyun Yuan
Coorientador: Prof. Dr. Orizon P. Ferreira

CURITIBA
2018

FICHA CATALOGRÁFICA ELABORADA PELO SISTEMA DE BIBLIOTECAS/UFPR
BIBLIOTECA DE CIÊNCIA E TECNOLOGIA

F363d Fernandes, Teles Araújo
Damped Newton's Method on Riemannian Manifolds / Teles Araújo Fernandes. – Curitiba, 2018.

Tese (Doutorado) - Universidade Federal do Paraná, Setor de Ciências Exatas, Programa de Pós-Graduação em Matemática Aplicada, 2018.

Orientador: Prof. Dr. Jinyun Yuan.
Coorientador: Prof. Dr. Orizon P. Ferreira.

1. Variedade Riemanniana. 2. Método de Newton. 3. Convergência local. 4. Convergência global. I. Universidade Federal do Paraná. II. Yuan, Jinyun. III. Ferreira, Orizon P. IV. Título.

CDD: 530.092

Bibliotecária: Romilda Santos - CRB-9/1214

MINISTÉRIO DA EDUCAÇÃO
SETOR CIÊNCIAS EXATAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO MATEMÁTICA

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em MATEMÁTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **TELES ARAUJO FERNANDES** intitulada: **Estudo do método de Newton em Variedades**, após terem inquirido o aluno e realizado a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

Curitiba, 16 de Março de 2018.

ALFREDO N. IUSEM
Presidente da Banca Examinadora

YUAN JINYUN
Avaliador Interno

JOSÉ MARIO MARTÍNEZ
Avaliador Externo

GEOVANI NUNES GRAPIGLIA
Avaliador Interno

ORIZON PEREIRA FERREIRA
Avaliador Externo

JOÃO XAVIER DA CRUZ NETO
Avaliador Externo

AGRADECIMENTOS

Primeiramente, a Deus. A minha esposa Mirella e meu filho Davi. Meus orientadores, Yuan Jin Yun e Orizon Pereira Ferreira aos quais tenho profunda admiração por serem pessoas humildes ao mesmo tempo que são altamente competentes como matemáticos. Estes, ensinaram-me matemática e com eles também me tornei um ser humano melhor. A UESB pelo recurso financeiro. Ao professor Geovani Grapilhia, pelos valiosos conselhos acadêmicos e por ter sido um grande amigo nessa etapa. Ao meu compadre Márcio e sua esposa Roberta, sem estes não chegaria até aqui. A minha cunhada e comadre Manuella Cajaiba sempre presente como um anjo nos momentos mais difíceis dessa batalha. Aos amigos (A galerinha) que fizeram meus dias em Curitiba muito mais felizes: Fernando e Pâmela, Ana, Monique, Lay e Diego, Camila, Geovani e Stela. A Odisnei e Leila, Eduardo e Janaine por terem nos acolhidos na cidade de Curitiba. Finalmente, aqueles que emitiram energia positiva para mim na realização desta conquista.

Quando você admite a ignorância, você está abrindo a porta da sabedoria.
Sócrates

RESUMO

Nesta tese, nós estudamos método de Newton para encontrar singularidade de campo de vetor definido em variedade Riemanniana. Obtemos uma importante propriedade do transporte paralelo de vetor para estabelecer (sob hipótese mínima) convergência super-linear da sequência gerada pelo clássico método de Newton para encontrar zero de campo de vetor. Além disso, nós propomos um método de Newton amortecido e apresentamos sua análise global de convergência usando busca linear e uma função mérito. Asseguramos que, após um número finito de iteradas, uma sequência gerada pelo proposto método de Newton amortecido reduz-se a uma sequência gerada pelo método de Newton. Portanto, a taxa de convergência do método proposto é super-linear/quadrática. Nós implementamos ambos os métodos para encontrar minimizadores globais de uma família de funções definidas no cone de matrizes simétricas definidas positivas. Nossos experimentos mostram que a performance do método de Newton amortecido é superior a performance do clássico método de Newton, indicando que o comportamento dos métodos no espaço Euclidiano permanecem neste novo cenário. Para proceder com os experimentos, nós primeiro equipamos o cone de matrizes simétricas definidas positivas com uma estrutura de variedade Riemanniana. Então, definimos as iteradas das sequências geradas pelo método de Newton e pelo método de Newton amortecido usando a curva geodésica nesta variedade. Contudo, computar geodésica involve significante custo numérico. Devido a isso, nós propomos dois novos algoritmos, a saber, *método de Newton com retração* e *método de Newton amortecido com retração*, para encontrar singularidade de campo de vetor definido em variedade Riemanniana. Nós apresentamos análises de convergências desses novos métodos por extender os resultados obtidos para estabelecer o método de Newton e o método de Newton amortecido. Finalmente, nós implementamos o método de Newton amortecido com retração para encontrar minimizadores da família de funções acima mencionada. Neste caso, nossos experimentos mostram que o método de Newton amortecido possui performance similar do método de Newton amortecido com retração. As principais contribuições desta tese são como seguem. 1) Sob hipótese mínima, isto é, invertibilidade da derivada covariante do campo de vetor em sua singularidade, nós mostramos que o método de Newton está bem definido em uma vizinhança aceitável de sua singularidade e que a sequência gerada por este método converge com taxa super-linear, (veja (27)). 2) Nós propomos um *método de Newton amortecido* no contexto de variedade Riemanniana e estabelecemos sua convergência global para uma singularidade do campo de vetor preservando as taxas de convergências super-linear e quadrática do método de Newton (veja (16)). 3) Propomos o *método de Newton amortecido com retração* e estudamos suas propriedades de convergência, obtendo os mesmos resultados do método de Newton e do método de Newton amortecido.

Keywords: *variedade Riemanniana · método de Newton · método de Newton amortecido · convergência local · convergência global · taxa super-linear · taxa quadratica · busca linear · exponential mapping · retração.*

ABSTRACT

In this thesis, we study Newton's method for finding a singularity of a differentiable vector field defined on a Riemannian manifold. We obtain an important property of the parallel transport of a vector which allows to establish (under a mild assumption) super-linear convergence of the sequence generated by the classical Newton's method for finding a zero of a vector field. Moreover, we propose a *damped Newton's method* and we present its global analysis of convergence using a linear search together with a merit function. We ensure that the sequence generated by the proposed damped Newton's method reduces to a sequence generated by the classical iteration of Newton's method after a finite number of iterations. Thus, the convergence rate of the proposed method is super-linear/quadratic. We implement both methods for finding global minimizers of a family of functions defined on the cone of symmetric positive definite matrices. Our experiments show that the performance of the damped Newtons method is superior to that of the classical Newton's method, indicating that the behavior of the methods in Euclidean space persists in this new setting. To proceed with the experiments, we first endow the cone of symmetric positive definite matrices with a Riemannian manifold structure. Then, we define the iterations of the sequences generated by Newton's method and damped Newtons method using the geodesic curve of this manifold. However, in general, performing this task in a computationally efficient manner involves significant numerical challenges. Therefore, we propose two new algorithms, namely *Newton's method with retraction* and *damped Newton's method with retraction*, to find a singularity of a vector field defined on a Riemannian manifold. We present the convergence analysis of these new methods by extending the results of Newton's method and the damped Newton's method. Finally, we implement the damped Newton's method with retraction for finding global minimizers of the aforementioned family of functions. For this case, our experiments show that damped Newton method with retraction does not present a better performance than the damped Newton method. The main contributions of this thesis are the following. 1) Under a mild assumption, i.e., invertibility of the covariant derivative of the vector field at its singularity, we show that Newton's method is well defined in a suitable neighborhood of this singularity and that the sequence generated by this method converges to the solution at a super-linear rate (see (27)). 2) We propose the *damped Newton's method* in the Riemannian manifold context and establish its global convergence to a singularity of a vector field preserving the super-linear convergence rates of Newton's method (see (16)). 3) We propose the *damped Newton's method with retraction* and study its convergence properties, obtaining the same results as those of Newton's method and the damped Newton's method.

Keywords: *Riemannian manifold · Newton's method · damped Newton's method · local convergence · global convergence · super-linear rate · quadratic rate · linesearch · exponential mapping · retraction.*

LIST OF FIGURE

LIST OF TABLES

NOTATIONS

$\mathbb{M}$ - finite-dimensional Riemannian manifold

$T_p\mathbb{M}$ - tangent space of $\mathbb{M}$ at p

$\bigcup_{p\in\mathbb{M}} T_p\mathbb{M}$ - tangent bundle of $\mathbb{M}$

X - vector field defined on $\mathbb{M}$

∇X - covariant derivative of X

$P_{\gamma,a,b}$ - parallel transport of v along γ from a to b

grad$f$ - gradient of f

Hess$f$ - Hessian of f

$\exp_p v$ - exponential mapping at p in the direction v

i_p - injectivity radius of $\mathbb{M}$ at p

$N_X(p)$ - Newton's iterate mapping for exponential at p

$N_{R,X}(p)$ - Newton's iterate mapping for retraction at p

CONTENTS

1. Introduction

Iterative methods on manifolds arise in the context of optimizing a real-valued function, dating back to the work of Luenberger (51) in the early 1970s, if not earlier. Luenberger proposed the idea of performing a line search along geodesics that are computationally feasible. Around 1990, the main research issue was to exploit differential-geometric objects in order to formulate optimization strategies on abstract nonlinear manifolds. Gabay was the first to focus on optimization on manifolds by minimizing a differentiable function defined on a Riemannian manifold (30). In the 1990s, the field of optimization on manifolds gained considerable popularity, especially with the work of Edelman et al. (25). Recent years have witnessed a growing interest in the development of numerical algorithms for nonlinear manifolds, as there are many numerical problems posed in manifolds arising in various natural contexts. For example, finding the largest eigenvalue of a symmetric matrix may be posed as maximizing Rayleigh's quotient defined on a sphere, invariant subspace computations, etc. (5; 25; 30; 67; 70). For such problems, the solutions of a system of equations often have to be computed or the zeros of a vector field have to be found. Because these problems are naturally posed on Riemannian manifolds, we can use the specific underlying geometric and algebraic structures in order to try significantly reduce the computational cost of finding the zeros of a vector field. Moreover, it is preferable to treat certain constrained optimization problems as problems of finding singularities of gradient vector fields on Riemannian manifolds rather than using Lagrange multipliers or projection methods by exploiting the Riemannian geometric structure of the constrained set (3; 68). Thus, algorithms that exploit the differential structure of nonlinear manifolds play an important role in optimization (1; 7; 13; 14; 15; 18; 31; 32; 33; 34; 36; 38; 43; 47; 49; 53; 61; 64; 74). In this thesis, instead of focusing on finding singularities of gradient vector fields on Riemannian manifolds, which includes finding local minimizers, we consider the more general problem of finding singularities of vector fields.

Newton's method is known to be a powerful tool for finding the zeros of nonlinear functions in Banach spaces. It also serves as a powerful theoretical tool with a wide range of applications in pure and applied mathematics (5; 56; 57). These factors have motivated several studies to investigate the issue of generalizing Newton's method from a linear setting to the Riemannian setting (3; 6; 28; 29; 48; 50; 65; 72). In all these previous studies, the analysis of the Riemannian version of Newton's method involved Lipschitz or Lipschitz-like conditions for the covariant derivative of the vector field. In fact, all these studies were concerned with establishing a quadratic rate of convergence for the method. It seems that some type of control on the covariant derivative of the vector field in a suitable neighborhood of its singularity is required for obtaining the quadratic convergence rate of the sequence generated by Newton's method. Thus, the convergence analysis of Newton's method for finding a singularity of a vector field in a Riemannian manifold under Lipschitz or Lipschitz-like conditions is well known (1; 28; 29; 48; 68). However, in a linear context, whenever the derivative of the function that defines the equation is nonsingular at the solution, Newton's method shows local convergence at a super-linear rate (58, chapter 8, Theorem 8.1.10, p. 148). Previously, we performed a local super-linear convergence analysis of Newton's method under a mild assumption (i.e., invertibility of the covariant derivative of the vector field at its singularity) (27). Chapter 3 presents this result.

Although Newton's method shows fast local convergence, it is highly sensitive to the initial iterate and may diverge if the initial iterate is not sufficiently close to the solution. Thus, Newton's method does not converge in general. To overcome this drawback, some strategies have been introduced for using Newton's method in optimization problems, such as the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm which has global convergence by using linesearch, Levenberg–Marquardt algorithm, and the trust region algorithm(10; 21). When the objective function is twice continuously differentiable and strongly convex, the Newton direction is a descent direction of the objective function. Hence, by adjusting the step size in the Newton direction using, e.g., the Armijo rule, we can ensure convergence of Newton's method. This strategy of dumping the Newton step size to globalize Newton's method is known as the damped Newton's method. For a comprehensive study of this method, see (10; 17; 21; 42). For the problem of finding a zero of a nonlinear equation in a Euclidean setting, this strategy of dumping the Newton step size can also be adopted by using a merit function for which the Newton direction is a descent direction. The merit function measures the quality of the approximation to a zero of the nonlinear function being considered; see (21; 22; 59). In Chapter 4, we present a globalization strategy for Newton's method to solve the problem of finding singularities of vector fields defined on Riemannian manifolds. Thus far, studies on globalization strategies in a Riemannian setting have been restricted to optimization problems, such as Newton's method with the Hessian of the objective function updated by the BFGS family, trust region methods (2; 39), and Levenberg–Marquardt methods (3, Chapter 8, Section 8.4.2). To the best of our knowledge, the analysis presented herein is a novel global analysis of a damped Newton's method for finding singularities of vector fields defined on Riemannian manifolds using a linear search together with a merit function. A particular case is finding critical points of real-valued functions defined on Riemannian manifolds. On the basis of the idea presented in (26, Section 4) for the nonlinear complementarity problem, we propose a damped Newton's method in a Riemannian setting. Moreover, we show its global convergence to a singularity of the vector field while preserving the same convergence rate as that of the classical Newton's method. For instance, we can use the proposed method to find a minimizer of a strongly convex function with a super-linear/quadratic rate in Riemannian settings. Our numerical experiments show the properties of global convergence of the proposed damped Newtons method and that this method is superior to the classical Newton's method for the problem of finding global minimizers of some functions defined on the cone of symmetric positive definite matrices.

To establish Newton's method for finding a singularity of a vector field defined on a

Riemannian manifold, essentially, we use the notion of continuously moving in the Newton direction while staying on a geodesic curve in the manifold until we reach a point where the vector field vanishes. By using the geodesic curve, we can define the exponential mapping that can be used to give a short notation for a geodesic with a given starting point and initial velocity. However, the geodesic, and consequently the exponential mapping, is defined as the solution of a nonlinear ordinary differential equation, whose efficient computation generally involves significant numerical challenges. Nevertheless, an approximation of the geodesic is sufficient to guarantee the desired convergence properties. Actually, to obtain the next iterate of an iterative method on a manifold, it is sufficient to use the notion of moving in the direction of a tangent vector while staying on the manifold. It is generalized by the notion of a retraction mapping that may generate a curve on the manifold with greater computational efficiency compared to the exponential mapping. The idea of using computationally efficient alternatives to the exponential mapping was introduced in (52). The strategy of using approximations of classical geometric concepts to obtain efficient iterative algorithms has attracted considerable attention lately in the context of Riemannian optimization; see, e.g., (2; 3; 18; 44; 61; 66; 71). Obtaining an iterative method using retraction is becoming increasingly common, as such algorithms are faster and possibly more robust than existing algorithms. Recent studies on the development of geometric optimization algorithms that exploit the mapping retraction on nonlinear manifolds include (35; 37; 40; 41; 75). A toolbox for building retractions on manifolds can be found in (4). On the basis of these considerations, we have sound arguments for generalization of the results obtained in Chapters 3 and 4 in the sense of retraction. We present these generalizations in Chapter 5.

The remainder of this thesis is organized as follows. Chapter 2 presents the notation and basic results. Chapter 3 describes the local super-linear convergence analysis of Newton's method. Chapter 4 presents a damped Newton's method and its global convergence at a super-linear/quadratic rate. Chapter 5 discusses generalization of the results obtained in Chapters 3 and 4 in the sense of retraction. Chapter 6 describes the numerical experiments. Finally, Chapter 7 states our conclusions and briefly explores directions for future work.

2. Basic Definition and Auxiliary Results

In this chapter, we recall some notations, definitions, and auxiliary results of Riemannian geometry used throughout this work. We begin with some basics concepts that can be found in various introductory books on Riemannian geometry, e.g., (24) and (63).

Let $\mathbb{M}$ be a finite-dimensional Riemannian manifold. Denote the *tangent space* of $\mathbb{M}$ at p by $T_p\mathbb{M}$ and the *tangent bundle* of $\mathbb{M}$ by $\bigcup_{p\in\mathbb{M}} T_p\mathbb{M}$. A *vector field* X on $\mathbb{M}$ is a correspondence that associates, to each point $p \in \mathbb{M}$, a vector $X(p) \in T_p\mathbb{M}$. The vector field is differentiable if the mapping $X : \mathbb{M} \to T\mathbb{M}$ is differentiable. Considering a parameterization $x : U \subset \mathbb{R} \to \mathbb{M}$, we can write

$$X(p) = \sum_{i=1}^{n} a_i(p) \frac{\partial}{\partial x_i},$$

where each $a_i : U \to \mathbb{R}$ is a function on U and $\dfrac{\partial}{\partial x_i}$ is the basis associated with x, $i = 1, \ldots, n$. Then, X is differentiable if and only if the functions a_i are differentiable for some, and therefore for any, parameterization. Let $\mathcal{D}$ denote the set of all real-valued differentiable functions on $\mathbb{M}$ and let $\mathcal{F}$ denote the set of all real-valued functions on $\mathbb{M}$. Then, a vector field X is a mapping $X : \mathcal{D} \to \mathcal{F}$ defined as follows:

$$(Xf)(p) = \sum_{i=1}^{n} a_i(p) \frac{\partial f}{\partial x_i}(p),$$

where f denotes, by abuse of notation, the expression of f in the parameterization x. This idea of a vector as a directional derivative is precisely what is used to define the notion of a tangent vector. The function Xf does not depend on the choice of parameterization x.

A Riemannian metric on $\mathbb{M}$ is given by $p \to \langle \cdot\ ,\ \cdot \rangle_p$. We can delete the index p in the function $\langle \cdot\ ,\ \cdot \rangle_p$ whenever there is no possibility of confusion. The metric can be used to define the length l of a piecewise differentiable continuous curve $\gamma : [a, b] \to \mathbb{M}$ by

$$l[\gamma, a, b] := \int_a^b \|\gamma'(t)\| dt.$$

A distance function on a Riemannian manifold $\mathbb{M}$ is as follows. Given two points $p, q \in \mathbb{M}$, consider all the piecewise differentiable curves joining p to q. Since M is connected, such curves exist. The distance $d(p,q)$ is defined as the infimum of the lengths of all curves $f_{p,q}$, where $f_{p,q}$ is a piecewise differentiable curve joining p to q. The distance induces the original topology on $\mathbb{M}$, i.e., $(\mathbb{M}, d)$ is a complete metric space. The corresponding norm associated with the Riemannian metric $\langle \cdot \, , \, \cdot \rangle$ is denoted by $\| \cdot \|$. An open ball of radius $r > 0$ centered at p is defined by

$$B_r(p) := \{q \in \mathbb{M} : d(p,q) < r\}.$$

Let $\Omega \subseteq \mathbb{M}$ be an open set, and let $\mathcal{D}^r(\Omega)$ denote the ring of real-valued functions of class C^r defined on Ω. Further, let $\mathcal{X}^r(\Omega)$ denote the *space of differentiable vector fields* of class C^r on Ω.

Definition 2.0.1. *An affine connection ∇ on a differentiable manifold $\mathbb{M}$ is a mapping*

$$\nabla : \mathcal{X}^1(\Omega) \times \mathcal{X}^0(\Omega) \to \mathcal{X}^0(\Omega),$$

which is denoted by $(X,Y) \overset{\nabla}{\to} \nabla_Y X$ and satisfies the following properties:

(i) $\nabla_{fY+gZ}X = f\nabla_Y X + g\nabla_Z X$, *where* $X \in \mathcal{X}^1(\Omega)$, $Y, Z \in \mathcal{X}^0(\Omega)$, *and* $f, g \in \mathcal{D}^0(\mathbb{M})$;

(ii) $\nabla_Z (X+Y) = \nabla_Z X + \nabla_Z Y$, *where* $X, Y \in \mathcal{X}^1(\Omega)$ *and* $Y, Z \in \mathcal{X}^0(\Omega)$;

(iii) $\nabla_Y fX = f\nabla_Y X + Y(f)X$, *where* $X \in \mathcal{X}^1(\Omega)$, $Y, \in \mathcal{X}^0(\Omega)$, *and* $f \in \mathcal{D}^0(\mathbb{M})$.

This connection leads to a notion of a derivative that is presented next.

Proposition 2.0.1. *Let $\mathbb{M}$ be a differentiable manifold with an affine connection ∇. There exists a unique correspondence that associates to a vector field V along the differentiable curve $\gamma : I \to \mathbb{M}$, another vector field DV/dt along γ, called the covariant derivative of V along γ, such that*

(i) $D/dt\,(V+W) = DV/dt + DW/dt$;

(ii) $D/dt\,(fV) = (Df/dt)\,V + fDV/dt$, *where V is also a vector field along γ and f is a differentiable function on I;*

(iii) if V is induced by a vector field $X \in \mathcal{X}^1(\Omega)$, i.e., $V(t) = X(\gamma(t))$, then $DV/dt(t) = \nabla_{\gamma'(t)}X$.

Proof. See (24, Prop. 2.2, p. 55) □

Definition 2.0.2. *Let $\mathbb{M}$ be a differentiable manifold with an affine connection ∇. A vector field V along a curve $\gamma : I \to \mathbb{M}$ is called parallel when $DV/dt = 0$ for all $t \in I$.*

Proposition 2.0.2. *(Parallel Transport) Let $\mathbb{M}$ be a differentiable manifold with an affine connection ∇. Let $\gamma : I \to \mathbb{M}$ be a differentiable curve in $\mathbb{M}$ and let v be a vector tangent to $\mathbb{M}$ at $\gamma(a)$, $a \in I$ (i.e., $v \in T_{\gamma(a)}\mathbb{M}$). Then, there exists a unique parallel vector field V along γ such that $V(a) = v$. $V(b)$ is called the parallel transport of v along γ from a to b. In fact, we define an isometry relative to $\langle \cdot \, , \, \cdot \rangle$:*

$$\begin{aligned} P_{\gamma,a,b} : T_{\gamma(a)}\mathbb{M} &\to T_{\gamma(b)}\mathbb{M} \\ v &\mapsto P_{\gamma,a,b}v = V(b). \end{aligned}$$

Note that

$$P_{\gamma, b_1, b_2} \circ P_{\gamma, a, b_1} = P_{\gamma, a, b_2} \quad \text{and} \quad P_{\gamma, b, a} = P^{-1}_{\gamma, a, b}.$$

A connection ∇ is said to be compatible with the metric $\langle \cdot \, , \, \cdot \rangle$ when, for any smooth curve γ and any pair of parallel vector fields P_1 and P_2 along γ, we have $\langle P_1 \, , \, P_2 \rangle = c$, where c is a constant. The Lie bracket of two smooth vector fields X and Y is defined as the smooth vector field $[X, Y]$ such that $[X, Y]f := (X(Y(f) - Y(X(f))$ for all $f \in \mathcal{D}^1(\mathbb{M})$. An affine connection ∇ is said to be symmetric when $\nabla_X Y - \nabla_Y X = [X, Y]$. We are now able to state the following theorem.

Theorem 2.0.1. *(Levi-Civita) Given a Riemannian manifold $\mathbb{M}$, there exists a unique affine connection ∇ on $\mathbb{M}$ satisfying the following conditions:*

(i) ∇ is symmetric;

(ii) ∇ is compatible with the Riemannian metric.

Proof. See (63, Th. 1.2, p. 28). □

The connection given by the above-mentioned theorem is referred to as the Levi-Civita connection associated with $(\mathbb{M}, \langle \cdot \, , \, \cdot \rangle)$. Hereafter, this connection will be referred to as the Levi-Civita (or Riemannian) connection on $\mathbb{M}$. The covariant derivative of $X \in \mathcal{X}^0(\Omega)$ determined by ∇ defines, at each $p \in \Omega$, a linear map

$$\nabla X(p) : T_p\mathbb{M} \rightarrow T_p\mathbb{M} \tag{2.1}$$

$$v \mapsto \nabla X(p)v := \nabla_Y X(p), \tag{2.2}$$

where Y is a vector field such that $Y(p) = v$. The *norm of a linear map* $A : T_p\mathbb{M} \rightarrow T_p\mathbb{M}$ is defined by

$$\|A\| := \sup \{\|Av\| \ : \ v \in T_p\mathbb{M}, \ \|v\| = 1\}.$$

For $f : \mathbb{M} \rightarrow \mathbb{R}$, a twice-differentiable function, the Riemannian metric induces the mappings $f \mapsto \text{grad} f$ and $f \mapsto \text{Hess} f$, which associate its *gradient* and *Hessian* via the rules

$$\langle \text{grad} f, X \rangle := df(X), \qquad \langle \text{Hess} f X, X \rangle := d^2 f(X, X), \qquad \forall \ X \in \mathcal{X}(\Omega),$$

respectively. Therefore, the last equalities imply that

$$\text{Hess} f X = \nabla_X \text{grad} f, \qquad \forall \ X \in \mathcal{X}(\Omega).$$

Definition 2.0.3. *Let $f : \mathbb{M} \rightarrow \mathbb{R}$ be a mapping and $p \in \mathbb{M}$. We say that f is coercive at p iff $\lim_{d(p,q) \rightarrow +\infty} f(q) = +\infty$.*

Let $\mathbb{M}$ be a Riemannian manifold. A parameterized curve $\gamma : I \rightarrow \mathbb{M}$ is a geodesic at $t_0 \in I$ iff

$$\gamma''(t) := \frac{D}{dt}\left(\frac{d\gamma}{dt}\right) = \nabla_{\gamma'(t)}\gamma'(t) = 0$$

at t_0. When γ is a geodesic at t, for all $t \in I$, we say that γ is a *geodesic*. If $[a, b] \subset I$ and $\gamma : I \to M$ is a geodesic, the restriction of γ to $[a, b]$ is called a *geodesic segment joining* $\gamma(a)$ *to* $\gamma(b)$. Note that, if γ is a geodesic, then

$$\frac{d}{dt}\langle\frac{d\gamma}{dt} \ , \ \frac{d\gamma}{dt}\rangle = 2\langle\frac{D}{dt}\frac{d\gamma}{dt} \ , \ \frac{d\gamma}{dt}\rangle = 0.$$

The latter equality implies that the length of the tangent vector $d\gamma/dt$ is constant. Therefore, the parameter of a geodesic is proportional to the arc length, i.e., the arc length l of the curve γ, starting from a fixed origin, say $t = t_0$, is given by $l[\gamma, t_0, t] = \|\gamma'(t_0)\|(t - t_0)$. When $\|\gamma'(t)\| = 1$, we say that γ is arc-length parameterized or normalized.

Proposition 2.0.3. *Given* $p \in \mathbb{M}$*, there exist a neighborhood* V *of* p *in* $\mathbb{M}$*, a number* $\epsilon > 0$*, and a* C^∞ *mapping* $\gamma : (-2, 2) \times U \to \mathbb{M}$*,* $U = \{(q, w) \in T\mathbb{M} \ : \ q \in V, \ w \in T_q\mathbb{M}, \ \|w\| < \epsilon\}$ *such that* $t \to \gamma(t, q, w)$*,* $t \in (-2, 2)$*, is the unique geodesic of* $\mathbb{M}$ *that, at the instant* $t = 0$*, passes through* q *with velocity* w *for every* $q \in V$ *and for every* $w \in T_q\mathbb{M}$*, with* $\|w\| < \epsilon$*.*

Proof. See (24, Prop. 2.7, p. 72) □

The above-mentioned proposition permits us to introduce the concept of the exponential map.

Definition 2.0.4. *Let* $\mathbb{M}$ *be a complete Riemannian manifold,* $p \in \mathbb{M}$*, and* $U \subset T\mathbb{M}$ *be an open set given by Proposition 2.0.3. Then, the exponential map on* U*,* $\exp : U \to \mathbb{M}$*, is given by*

$$\exp(q, v) = \gamma(1, q, v) = \gamma\left(\|v\|, q, v/\|v\|\right), \qquad (q, v) \in U.$$

We apply the restriction of exp to an open subset of the tangent space $T_p\mathbb{M}$, i.e.,

$$\exp_p : B_\epsilon(0_p) \subset T_p\mathbb{M} \to \mathbb{M},$$

by $\exp_p(v) := \exp(p, v)$. Next, we define a complete Riemannian manifold.

Definition 2.0.5. *A Riemannian manifold* $\mathbb{M}$ *is (geodesically) complete if for all* $p \in \mathbb{M}$*, the exponential map* $\exp_p$ *is defined for all* $v \in T_p\mathbb{M}$*, i.e., if any geodesic* $\gamma(t)$ *starting from* p *is defined for all values of parameter* $t \in \mathbb{R}$*.*

Theorem 2.0.2. *(Hopf-Rinow) Let* $\mathbb{M}$ *be a Riemannian manifold and let* $p \in \mathbb{M}$*. The following assertions are equivalent:*

(i) $\exp_p$ *is defined on all* $T_p\mathbb{M}$*.*

(ii) The closed and bounded sets of $\mathbb{M}$ *are compact.*

(iii) $\mathbb{M}$ *is geodesically complete.*

(iv) There exists a sequence of compact subsets $K_n \subset \mathbb{M}$*,* $K_n \subset K_{n+1}$*, and* $\bigcup_n K_n = \mathbb{M}$ *such that if* $q_n \neq K_n$ *then* $d(p, q_n) \to \infty$*. In addition, any of the statements above implies the following.*

(v) For any $q \in \mathbb{M}$, there exists a geodesic segment γ joining p to q, whose length is $d(p,q)$.

Proof. See (63, Th. 1.1, p. 84). □

When there exists a unique geodesic segment joining p to q, then we denote it by γ_{pq}, and if there is no confusion, we will consider the notation P_{pq} instead of $P_{\gamma,a,b}$ when γ is the unique geodesic segment joining p and q. A geodesic γ_{pq} is a *minimal geodesic* if $d(p,q) = l(\gamma)$. The Hopf–Rinow theorem asserts that any pair of points in a complete Riemannian manifold $\mathbb{M}$ can be joined by a (not necessarily unique) minimal geodesic segment.

Let $p \in \mathbb{M}$. The *injectivity radius* of $\mathbb{M}$ at p is defined by

$$i_p := \sup\left\{ r > 0 \ : \ \exp_{p|_{B_r(o_p)}} \text{ is a diffeomorphism} \right\},$$

where o_p denotes the origin of the tangent plane $T_p\mathbb{M}$ and $B_r(0_p) := \{v \in T_p\mathbb{M} : \| v - 0_p \| < r\}$.

Remark 2.0.1. *Let $\bar{p} \in \mathbb{M}$. The above definition implies that if $0 < \delta < i_{\bar{p}}$, then $\exp_{\bar{p}} B_\delta(0_{\bar{p}}) = B_\delta(\bar{p})$. Moreover, for all $p \in B_\delta(\bar{p})$, there exists a unique geodesic segment γ joining p to $\bar{p}$, which is given by $\gamma_{p\bar{p}}(t) = \exp_p(t \exp_p^{-1} \bar{p})$, for all $t \in [0,1]$.*

Let $p \in \mathbb{M}$ and $\delta_p := \min\{1, i_p\}$. Consider the quantity K_p, introduced in (20), which measures the rate at which the geodesics spread apart in $\mathbb{M}$:

$$K_p := \sup\left\{ \frac{d(\exp_q u, \exp_q v)}{\| u - v \|} \ : \ q \in B_{\delta_p}(p),\ u,\ v \in T_q\mathbb{M},\ u \neq v,\ \| v \| \leq \delta_p,\ \| u - v \| \leq \delta_p \right\}. \quad (2.3)$$

Remark 2.0.2. *In particular, when $u = 0$, or more generally, when u and v are on the same line through 0, $d(\exp_q u, \exp_q v) = \| u - v \|$. Hence, $K_p \geq 1$ for all $p \in \mathbb{M}$. Moreover, when $\mathbb{M}$ has non-negative sectional curvature, the geodesics spread apart to a smaller extent than the rays (24, chapter 5), i.e., $d(\exp_p u, \exp_p v) \leq \|u - v\|$; in this case, $K_p = 1$ for all $p \in \mathbb{M}$.*

Let $X \in \mathcal{X}(\Omega)$ and $\bar{p} \in \Omega$. Assume that $0 < \delta < \delta_{\bar{p}}$. Since $\exp_{\bar{p}} B_\delta(0_{\bar{p}}) = B_\delta(\bar{p})$, there exists a unique geodesic joining each $p \in B_\delta(\bar{p})$ to $\bar{p}$. Moreover, using (29, equality 2.3), we obtain

$$X(p) = P_{\bar{p}p} X(\bar{p}) + P_{\bar{p}p} \nabla X(\bar{p}) \exp_{\bar{p}}^{-1} p + d(p,\bar{p}) r(p), \qquad \lim_{p \to \bar{p}} r(p) = 0. \quad (2.4)$$

Now, we will define Newton's method in the Riemannian context. For any $X \in \mathcal{X}(\Omega)$ and $p \in \mathbb{M}$ such that $\nabla X(p)$ is nonsingular, *Newton's iterate mapping for exponential N_X* at p is given by

$$N_X(p) := \exp_p(-\nabla X(p)^{-1} X(p)), \quad (2.5)$$

and Newton's method is defined by

$$p_{k+1} = \exp_{p_k}\left(-\nabla X(p_k)^{-1} X(p_k)\right), \qquad k = 0, 1, \ldots$$

The next result establishes the quadratic convergence of Newton's method. It is a particular case of (28, Theorem 7.1).

Theorem 2.0.3. *Let $X : \Omega \to T\mathbb{M}$ be a continuously differentiable vector field and let $\bar{p} \in \mathbb{M}$ be a singularity of X. Suppose that ∇X is locally Lipschitz continuous at $\bar{p}$ with constant $L > 0$, and $\nabla X(\bar{p})$ is nonsingular. Then, there exists $r > 0$ such that the sequence*

$$p_{k+1} = \exp_{p_k}\left(-\nabla X(p_k)^{-1} X(p_k)\right), \qquad k = 0, 1, \ldots,$$

starting in $p_0 \in B_r(\bar{p}) \setminus \{\bar{p}\}$ is well defined, contained in $B_r(\bar{p})$, and converges to $\bar{p}$, and it holds that

$$d(\bar{p}, p_{k+1}) \leq \frac{L\, K_{\bar{p}}^2 \parallel \nabla X(\bar{p})^{-1} \parallel}{2\left[K_{\bar{p}} - d(p_0, \bar{p}) L \|\nabla X(\bar{p})^{-1}\|\right]}\, d(\bar{p}, p_k)^2, \qquad k = 0,\, 1, \ldots$$

Corollary 2.0.1. *Let $f \in \mathcal{C}^1(\Omega)$, and let $\bar{p} \in \mathbb{M}$ be a critical point of f. Suppose that* Hess f *is locally Lipschitz continuous at $\bar{p}$ with constant $L > 0$ and* Hess $f(\bar{p})$ *is nonsingular. Then, there exists $r > 0$ such that the sequence*

$$p_{k+1} = \exp_{p_k}\left(-\operatorname{Hess} f(p_k)^{-1} \operatorname{grad}(p_k)\right), \qquad k = 0, 1, \ldots,$$

starting in $p_0 \in B_r(\bar{p}) \setminus \{\bar{p}\}$ is well defined, contained in $B_r(\bar{p})$, and converges to $\bar{p}$, and it holds that

$$d(\bar{p}, p_{k+1}) \leq \frac{L\, K_{\bar{p}}^2 \parallel \operatorname{Hess} f(\bar{p})^{-1} \parallel}{2\left[K_{\bar{p}} - d(p_0, \bar{p}) L \|\operatorname{Hess} f(\bar{p})^{-1}\|\right]}\, d(\bar{p}, p_k)^2, \qquad k = 0,\, 1, \ldots$$

We end this section with the well-known Banach's lemma.

Lemma 2.0.1. *Let B be a linear operator and let I_p be the identity operator in $T_p\mathbb{M}$. Assume that $\parallel B - I_p \parallel < 1$; then, B is invertible and $\parallel B^{-1} \parallel \leq \dfrac{1}{1 - \parallel B - I_p \parallel}$.*

Proof. See (59, Lemma 2.3.2, p. 45). □

3. Super-linear Convergence of Newton's Method on Riemannian Manifold

In this chapter, we study Newton's method for finding a singularity of a differentiable vector field defined on a Riemannian manifold. Our main goal is to show that, under the only hypotheses of nonsingularity of the covariant derivative of the vector field, Newton's method is well defined and converges to a zero of this field at a super-linear rate. Formally, the problem researched is as follows. To find a point $p \in \Omega$ satisfying the equation

$$X(p) = 0, \tag{3.1}$$

where $X : \Omega \to T\mathbb{M}$ is a differentiable vector field and $\Omega \subset \mathbb{M}$ is an open set. To solve (4.1), Newton's method formally generates a sequence with an initial point $p_0 \in \Omega$ as described in the following algorithm.

Algorithm 3.0.1.

Newton with Exponential

Step 0. *Take an initial point $p_0 \in \mathbb{M}$, and set $k = 0$. If $\|X(p_0)\| = 0$, stop.*

Step 1. *Compute search direction $v_k \in T_{p_k}\mathbb{M}$ by*

$$v_k = -\nabla X(p_k)^{-1} X(p_k). \tag{3.2}$$

Step 2. *Compute the next iterate by*

$$p_{k+1} := \exp_{p_k} v_k. \tag{3.3}$$

Step 3. *Set $k \leftarrow k + 1$ and go to* **Step 1**.

We remark that, if the covariant derivative is singular in iterate k, then v_k given by (3.2) does not exist. Hence, Newton's Algorithm 3.0.1 stops and does not converge to a solution of (4.1). Our aim is to prove that, under the assumption of nonsingularity of the covariant derivative in a solution p_*, if p_0 belongs to a suitable neighborhood of p_*, then v_k given by (3.2) exists for all k. Moreover, the sequence given by (3.3) is well defined and converges super-linearly to p_*.

3.1 Preliminaries

To achieve our goal, we begin by stating an important property of parallel transport in our context. It is worth noting that, to ensure this property, we use the same ideas as those presented in the proof of (47, Lemma 2.4, item (iv)) for Hadamard manifolds, with some minor necessary technical adjustments to fit any Riemannian manifold.

Lemma 3.1.1. *Let* $\bar{p} \in \mathbb{M}$, $0 < \delta < \delta_{\bar{p}}$, *and* $u \in T_{\bar{p}}\mathbb{M}$. *Then, the vector field* $F : B_\delta(\bar{p}) \to T\mathbb{M}$ *defined by* $F(p) := P_{\bar{p}p}u$ *is continuous.*

Proof. Assume that $\mathbb{M}$ is n-dimensional. Let $p \in B_\delta(\bar{p})$ and γ_p be the unique geodesic segment joining $\bar{p}$ to p (according to Remark 2.0.1). Let $u \in T_{\bar{p}}\mathbb{M}$. From the definition of parallel transport, there is a unique continuously differentiable vector field Y_p along γ_p such that $Y_p(\gamma_p(0)) = u$, $Y_p(\gamma_p(1)) = P_{\bar{p}p}u$, and

$$\nabla_{\gamma_p'(t)} Y_p(\gamma_p(t)) = 0, \qquad \forall\, t \in [0,1]; \tag{3.4}$$

see (63, p. 29). The definition of $\delta_{\bar{p}}$ implies that $\varphi := \exp_{\bar{p}}^{-1} : B_\delta(\bar{p}) \to B_{r_{\bar{p}}}(0_{\bar{p}})$ is a diffeomorphism; hence, $(B_\delta(\bar{p}), \varphi)$ is a local chart at $\bar{p}$. For each $j = 1, 2, ..., n$, define $y^j : B_\delta(\bar{p}) \to \mathbb{R}$ by $y^j = \pi^j \circ \varphi$, where $\pi^j : T_{\bar{p}}\mathbb{M} \to \mathbb{R}$ is the projection defined by $\pi^j(a_1, ..., a_j, ..., a_n) = a_j$ for all $(a_1, ..., a_j, ..., a_n) \in T_{\bar{p}}\mathbb{M}$. Then, $(B_\delta(\bar{p}), \varphi, y^j)$ is a local coordinate system at $\bar{p}$. Let $\{\partial/\partial y^j\}$ be the associated correspondent natural basis to $(B_\delta(\bar{p}), \varphi, y^j)$. Since $\gamma_p(t) \in B_\delta(\bar{p})$ for all $t \in [0,1]$ and $Y_p(\gamma_p(t)) \in T_{\gamma_p(t)}\mathbb{M}$, we can write

$$Y_p(\gamma_p(t)) = \sum_j Y_p^j(t) \frac{\partial}{\partial y^j}|_{\gamma_p(t)} \quad \forall\, t \in [0,1],$$

where each coordinate function $Y_p^j : [0,1] \to \mathbb{R}$ is continuously differentiable for all $j = 1, 2, ..., n$. For simplicity, we set $y_p^j := y^j \circ \gamma_p(\cdot)$ for each $j = 1, 2, ..., n$. Thus, (3.4) is equivalent to the ordinary differential equation

$$\frac{dY_p^k}{dt} + \sum_{i,j} \Gamma_{i,j}^k(\gamma_p) \frac{dy_p^i}{dt} Y_p^j = 0, \quad k = 1, 2, ..., n,$$

where $\Gamma_{i,j}^k$ are the Christoffel symbols of the connection ∇; see (63, p. 29). Hence, the last equality implies that $\{Y_p^k : k = 1, 2, ..., n\}$ is the unique solution of the following system of $p-$parameter linear differential equations:

$$\begin{cases} \dfrac{dY_p^k}{dt} = -\displaystyle\sum_j^n a_{k,j} Y_p^j, \quad k = 1, 2, ..., n, \\ \displaystyle\sum_j Y_p^j(0) \frac{\partial}{\partial y^j}|_{\gamma_p(0)} = u, \end{cases}$$

where, for (k,j), $k, j = 1, ..., n$, the continuous function $a_{k,j} : [0,1] \times B_\delta(\bar{p}) \to \mathbb{R}$ is given by

$$a_{k,j}(t,p) = \sum_{i=1}^n \Gamma_{i,j}^k(\gamma_p(t)) \frac{dy_p^i(t)}{dt}.$$

Thus, from the continuity on parameters for differential equations (see, e.g., (23, Theorem 10.7.1, p. 353)), the solution $\{Y^k_{\cdot}(\cdot)\}$ is continuous on $[0,1] \times B_\delta(\bar{p})$, and, equivalently, $Y_{\cdot}(\gamma_{\cdot}(\cdot))$ is continuous on $[0,1] \times B_\delta(\bar{p})$. Furthermore, we have $F(p) = P_{\bar{p}p}u = Y_p(\gamma_p(1))$ for any $p \in B_\delta(\bar{p})$. Therefore, F is continuous on $B_\delta(\bar{p})$ and the proof is complete. □

Next, we present an immediate consequence of Lemma 3.1.1.

Corollary 3.1.1. *Let $\bar{p} \in \mathbb{M}$, $0 < \delta < \delta_{\bar{p}}$, and $u \in T_{\bar{p}}\mathbb{M}$. If the vector field $Z : B_\delta(\bar{p}) \to T\mathbb{M}$ is continuous at $\bar{p}$, then the mapping $G : B_\delta(\bar{p}) \to T_{\bar{p}}\mathbb{M}$ defined by $G(p) := P_{p\bar{p}}Z(p)$ is also continuous at $\bar{p}$.*

Proof. Since parallel transport is an isometry, it follows from the definition of the vector field G that

$$\|G(p) - G(\bar{p})\| = \|Z(p) - P_{\bar{p}p}Z(\bar{p})\|.$$

Considering that Z is continuous at $\bar{p}$ and $P_{\bar{p}\bar{p}} = I_{\bar{p}} = P_{p\bar{p}}P_{\bar{p}p}$, we conclude from Lemma 3.1.1 that

$$\lim_{p\to\bar{p}} \|Z(p) - P_{\bar{p}p}Z(\bar{p})\| = 0.$$

Therefore, the desired result follows by a simple combination of the last two equalities. □

The next result ensures that, if $\nabla X(\bar{p})$ is nonsingular, then there exists a neighborhood of $\bar{p}$ where ∇X is also nonsingular. Moreover, in this neighborhood, ∇X^{-1} is bounded.

Lemma 3.1.2. *Assume that ∇X is continuous at $\bar{p}$. Then,*

$$\lim_{p\to\bar{p}} \|P_{p\bar{p}}\nabla X(p)P_{\bar{p}p} - \nabla X(\bar{p})\| = 0. \tag{3.5}$$

Moreover, if $\nabla X(\bar{p})$ is nonsingular, then there exists $0 < \bar{\delta} < \delta_{\bar{p}}$ such that $B_{\bar{\delta}}(\bar{p}) \subset \Omega$, and for each $p \in B_{\bar{\delta}}(\bar{p})$, the following statements hold:

(i) $\nabla X(p)$ *is nonsingular;*

(ii) $\|\nabla X(p)^{-1}\| \leq 2\,\|\nabla X(\bar{p})^{-1}\|$.

Proof. Let $0 < \delta < \delta_{\bar{p}}$ such that $B_\delta(\bar{p}) \subset \Omega$. For each $u \in T_{\bar{p}}\mathbb{M}$, define $Z : B_\delta(\bar{p}) \to T\mathbb{M}$ by

$$Z(p) = \nabla X(p)P_{\bar{p}p}u.$$

By applying Lemma 3.1.1, we conclude that $P_{\bar{p}p}u$ is continuous on $B_\delta(\bar{p})$. Thus, because ∇X is continuous, Z is also continuous on $B_\delta(\bar{p})$. Hence, using Corollary 3.1.1, we conclude that the mapping $F : B_\delta(\bar{p}) \to T_{\bar{p}}\mathbb{M}$ defined by

$$F(p) = P_{p\bar{p}}Z(p)$$

is also continuous at $\bar{p}$. Considering that $P_{\bar{p}\bar{p}} = I_{\bar{p}}$ as well as the definitions of the mappings F and Z, we conclude that $\lim_{p\to\bar{p}} F(p) = \nabla X(\bar{p})u$. Now, define the mapping

$$B_\delta(\bar{p}) \ni p \mapsto [P_{p\bar{p}}\nabla X(p)P_{\bar{p}p} - \nabla X(\bar{p})] \in \mathcal{L}(T_{\bar{p}}\mathbb{M}, T_{\bar{p}}\mathbb{M}),$$

where $\mathcal{L}(T_{\bar{p}}\mathbb{M}, T_{\bar{p}}\mathbb{M})$ denotes the space consisting of all linear operators from $T_{\bar{p}}\mathbb{M}$ to $T_{\bar{p}}\mathbb{M}$. Since $\lim_{p\to\bar{p}} F(p) = \nabla X(\bar{p})u$, for each $u \in T_{\bar{p}}\mathbb{M}$, the definition of F implies that

$$\lim_{p\to\bar{p}} \left[P_{p\bar{p}}\nabla X(p)P_{\bar{p}p} - \nabla X(\bar{p})\right] u = 0, \quad u \in T_{\bar{p}}\mathbb{M}.$$

Since $T_{\bar{p}}\mathbb{M}$ is finite-dimensional and $[P_{p\bar{p}}\nabla X(p)P_{\bar{p}p} - \nabla X(\bar{p})] \in \mathcal{L}(T_{\bar{p}}\mathbb{M}, T_{\bar{p}}\mathbb{M})$, for each $p \in B_\delta(\bar{p})$, the above equality implies that the equality (3.5) holds. Now, we proceed with the proof of item (i) . The equality (3.5) implies that there exists $0 < \bar{\delta} < \delta$ such that

$$\|P_{p\bar{p}}\nabla X(p)P_{\bar{p}p} - \nabla X(\bar{p})\| \leq \frac{1}{2\,\|\nabla X(\bar{p})^{-1}\|}, \qquad \forall\, p \in B_{\bar{\delta}}(\bar{p}).$$

Thus, from the last inequality and the property of the operator norm defined in $\mathcal{L}(T_{\bar{p}}\mathbb{M}, T_{\bar{p}}\mathbb{M})$, for all $p \in B_{\bar{\delta}}(\bar{p})$, we obtain

$$\left\|\nabla X(\bar{p})^{-1}P_{p\bar{p}}\nabla X(p)P_{\bar{p}p} - I_{\bar{p}}\right\| \leq \|\nabla X(\bar{p})^{-1}\|\,\|P_{p\bar{p}}\nabla X(p)P_{\bar{p}p} - \nabla X(\bar{p})\| \leq \frac{1}{2}. \tag{3.6}$$

Hence, from Lemma 2.0.1, we conclude that $\nabla X(\bar{p})^{-1}P_{p\bar{p}}\nabla X(p)P_{\bar{p}p}$ is a nonsingular operator for each $p \in B_{\bar{\delta}}(\bar{p})$. Since $\nabla X(\bar{p})$ and the parallel transport are nonsingular, $\nabla X(p)$ is also nonsingular for each $p \in B_{\bar{\delta}}(\bar{p})$, and the proof of the first item is complete. To prove item (ii), we first note that, from (3.6) and Lemma 2.0.1, it follows that for all $p \in B_{\bar{\delta}}(\bar{p})$,

$$\left\|\left[\nabla X(\bar{p})^{-1}P_{p\bar{p}}\nabla X(p)P_{\bar{p}p}\right]^{-1}\right\| \leq \frac{1}{1 - \|\nabla X(\bar{p})^{-1}P_{p\bar{p}}\nabla X(p)P_{\bar{p}p} - I_{\bar{p}}\|}.$$

Since parallel transport is an isometry, by combining (3.6) with the above inequality, we obtain

$$\left\|\nabla X(p)^{-1}P_{\bar{p}p}\nabla X(\bar{p})\right\| \leq 2.$$

Thus, using the properties of the norm and the fact that the parallel transport is an isometry, the last inequality implies that, for all $p \in B_{\bar{\delta}}(\bar{p})$, we have

$$\left\|\nabla X(p)^{-1}\right\| \leq \left\|\nabla X(p)^{-1}P_{\bar{p}p}\nabla X(\bar{p})\right\|\left\|\nabla X(\bar{p})^{-1}\right\| \leq 2\left\|\nabla X(\bar{p})^{-1}\right\|,$$

which is the desired inequality in the second item. Thus, the proof of the lemma is complete. □

Lemma 3.1.2 establishes the non-singularity of ∇X in a neighborhood of $\bar{p}$. It ensures that there exists a neighborhood of $\bar{p}$ where the Newton's iterate (3.3) is well defined, but it does not guarantee that it belongs to this neighborhood. In the next lemma, we will establish this fact. For stating the next result, since $\bar{\delta}$ is given by Lemma 3.1.2, we remark that *Newton's iterate mapping* $N_X : B_{\bar{\delta}}(\bar{p}) \to \mathbb{M}$ given by (2.5) is well defined.

Lemma 3.1.3. *Let $p_* \in \Omega$ such that p_* is a singularity of X. Assume that ∇X is continuous at p_* and $\nabla X(p_*)$ is nonsingular. Then,*

$$\lim_{p\to p_*} \frac{d(N_X(p), p_*)}{d(p, p_*)} = 0.$$

Proof. Let $\bar\delta$ be given by Lemma 3.1.2 and $p \in B_{\bar\delta}(p_*)$. Some algebraic manipulations show that

$$\nabla X(p)^{-1}X(p) + \exp_p^{-1} p_* = \nabla X(p)^{-1}\big[X(p) - P_{p_*p}X(p_*) - P_{p_*p}\nabla X(p_*)\exp_{p_*}^{-1} p + [P_{p_*p}\nabla X(p_*) - \nabla X(p)P_{p_*p}]\exp_{p_*}^{-1} p\big].$$

Define $r(p) := [X(p) - P_{\bar p p}X(p_*) - P_{p_*p}\nabla X(p_*)\exp_{p_*}^{-1} p]/d(p,p_*)$ for $p \in B_{\bar\delta}(p_*)$. From (5.8), we have $\lim_{p\to p_*} r(p) = 0$. Thus, using the above equality, the definition of r, $d(p,p_*) = \|\exp_{p_*}^{-1} p\|$, and some properties of the norm, we conclude that

$$\left\|\nabla X(p)^{-1}X(p) + \exp_p^{-1} p_*\right\| \le \left\|\nabla X(p)^{-1}\right\| [\|r(p)\| + \|P_{p_*p}\nabla X(p_*) - \nabla X(p)P_{p_*p}\|]\, d(p,p_*).$$

Since $p \in B_{\bar\delta}(p_*)$ and the parallel transport is an isometry, item (ii) of Lemma 3.1.2 implies that

$$\left\|\nabla X(p)^{-1}X(p) + \exp_p^{-1} p_*\right\| \le 2\left\|\nabla X(p_*)^{-1}\right\| \big[\,\|r(p)\| + \|P_{pp_*}\nabla X(p)P_{p_*p} - \nabla X(p_*)\|\,\big]\, d(p,p_*). \quad (3.7)$$

From (3.5) and since $\lim\limits_{p\to p_*} r(p) = 0$, the right-hand side of the last inequality tends to zero as p tends to p_*. Recalling that $\delta_{p_*} = \min\{1, i_{p_*}\}$, we can shrink $\bar\delta$, if necessary, to obtain

$$\left\|\nabla X(p)^{-1}X(p) + \exp_p^{-1} p_*\right\| \le \delta_{p_*}, \qquad \forall\, p \in B_{\bar\delta}(p_*).$$

Hence, from the definition of Newton's iterate mapping for exponential N_X in (2.5) and the definition of K_{p_*} in (5.7), we have

$$d(N_X(p), p_*) \le K_{p_*}\left\|-\nabla X(p)^{-1}X(p) - \exp_p^{-1} p_*\right\|, \qquad \forall\, p \in B_{\bar\delta}(p_*).$$

Therefore, by combining (5.12) with the last inequality, we conclude that for all $p \in B_{\bar\delta}(p_*)$,

$$\frac{d(N_X(p), p_*)}{d(p,p_*)} \le 2K_{p_*}\left\|\nabla X(p_*)^{-1}\right\| [\|\, r(p)\,\| + \|\, P_{pp_*}\nabla X(p)P_{p_*p} - \nabla X(p_*)\,\|].$$

By letting p tend to $p*$ in the last inequality, and by considering (3.5) and the fact that $\lim_{p\to p_*} r(p) = 0$, we get the desired result. □

3.2 Convergence Analysis

Now, we are ready to establish the main result in this chapter, and its proof is a combination of the two previous lemmas.

Theorem 3.2.1. *Let $\Omega \subset \mathbb{M}$ be an open set, $X : \Omega \to T\mathbb{M}$ be a differentiable vector field, and $\bar p \in \Omega$. Suppose that p_* is a singularity of X, ∇X is continuous at p_*, and $\nabla X(p_*)$ is nonsingular. Then, there exists $\bar\delta > 0$ such that, for all $p_0 \in B_{\bar\delta}(p_*)$, the Newton sequence*

$$p_{k+1} = \exp_{p_k}(-\nabla X(p_k)^{-1}X(p_k)), \qquad k = 0, 1, \ldots, \quad (3.8)$$

is well defined, contained in $B_{\bar\delta}(p_*)$, and converges super-linearly to p_*.

Proof. Let $\bar{\delta}$ be given by Lemma 3.1.2. From Lemma 3.1.3, we can shrink $\bar{\delta}$, if necessary, to conclude that

$$d(N_X(p), p_*) < \frac{1}{2} d(p, p_*), \qquad \forall\, p \in B_{\bar{\delta}}(p_*). \tag{3.9}$$

Thus, $N_X(p) \in B_{\bar{\delta}}(p_*)$, for all $p \in B_{\bar{\delta}}(p_*)$. Note that (2.5) and (5.16) imply that $\{p_k\}$ satisfies

$$p_{k+1} = N_X(p_k), \qquad k = 0, 1, \dots \tag{3.10}$$

which is indeed an equivalent definition of this sequence. Since $N_X(p) \in B_{\bar{\delta}}(p_*)$ for all $p \in B_{\bar{\delta}}(p_*)$, it follows from (5.17) and Lemma 3.1.2 item (i) that, for all $p_0 \in B_{\bar{\delta}}(p_*)$, the Newton sequence $\{p_k\}$ is well defined and contained in $B_{\bar{\delta}}(p_*)$. Moreover, using (5.18) and (5.17), we obtain

$$d(p_{k+1}, p_*) < \frac{1}{2} d(p_k, p_*), \qquad k = 0,\, 1, \dots$$

The above inequality implies that $\{p_k\}$ converges to p_*. Thus, by combining (5.17) with Lemma 3.1.3, we conclude that

$$\lim_{k \to +\infty} \frac{d(p_{k+1}, p_*)}{d(p_k, p_*)} = 0.$$

Therefore, $\{p_k\}$ converges super-linearly to p_* and the proof is complete. □

Next, we present an application of Theorem 3.2.1 for finding the critical points of a twice-differentiable function defined on a Riemannian manifold.

Corollary 3.2.1. *Let $\Omega \subset \mathbb{M}$ be an open set, $f : \Omega \to \mathbb{R}$ be a twice-differentiable function, and $p_* \in \Omega$. Suppose that p_* is a critical point of f,* Hess f *is continuous at p_*, and* Hess $f(p_*)$ *is nonsingular. Then, there exists $\bar{\delta} > 0$ such that, for all $p_0 \in B_{\bar{\delta}}(p_*)$, the Newton sequence*

$$p_{k+1} = \exp_{p_k}(-\text{Hess}\, f(p_k)^{-1} \text{grad}\, f(p_k)), \qquad k = 0, 1, \dots,$$

is well defined, contained in $B_{\bar{\delta}}(p_*)$, and converges super-linearly to p_*.

Proof. Letting $X = \text{grad}\, f$, the result follows by applying Theorem 3.2.1. □

4. Damped Newton Method on Riemannian Manifold

In this chapter, we consider the problem of finding a singularity of a differentiable vector field. The formal statement of the problem is as follows: Find a point $p \in \mathbb{M}$ such that

$$X(p) = 0, \tag{4.1}$$

where $X : \Omega \subseteq \mathbb{M} \to T\mathbb{M}$ is a differentiable vector field. The Newton's method applied to this problem has *local* super-linear convergence rate, whenever the covariant derivative in the singularity of X is non-singular, see (27), and, additionally, if the covariant derivative is Lipschitz around the singularity then the method has Q-quadratic convergence rate, see (28). In order to globalize Newton's method keeping the same properties aforementioned, we will use the same idea of the Euclidean setting by introducing a linear search strategy. This modification of Newton's method called *damped Newton's method* perform a linear search decreasing the merit function $\varphi : \mathbb{M} \to \mathbb{R}$,

$$\varphi(p) = \frac{1}{2} \parallel X(p) \parallel^2, \tag{4.2}$$

in order reaches the super-linear convergence region of the Newton's method. In the following we present the formal description of the damped Newton's algoritm.

Algorithm 4.0.1.

Damped Newton with Exponential

Step 0. *Choose a scalar $\sigma \in (0, 1/2)$, let an initial point $p_0 \in \mathbb{M}$, and set $k = 0$.*

Step 1. *Compute search direction $v_k \in T_{p_k}\mathbb{M}$ as a solution of the linear equation*

$$X(p_k) + \nabla X(p_k)v = 0. \tag{4.3}$$

If v_k exists, go to **Step** 2. *Otherwise, set the search direction as $v_k = -\operatorname{grad} \varphi(p_k)$, where φ is defined by* (4.2), *i.e.,*

$$v_k = -\nabla X(p_k)^T X(p_k). \tag{4.4}$$

If $v_k = 0$, stop.

Step 2. *Compute the step size by the rule*

$$\alpha_k := \max\left\{2^{-j} \ : \ j \in \mathbb{N},\ \varphi\left(\exp_{p_k}(2^{-j}v_k)\right) \le \varphi(p_k) + \sigma 2^{-j}\left\langle \text{grad}\ \varphi(p_k),\ v_k\right\rangle\right\}, \quad (4.5)$$

and set the next iterate as

$$p_{k+1} := \exp_{p_k}(\alpha_k v_k). \quad (4.6)$$

Step 3. *Set $k \leftarrow k+1$ and go to step 1.*

4.1 Preliminaries

In this section, we present some preliminary results to ensure that the sequence generated by the damped Newton's method (Algorithm 4.0.1) is well defined and converges. We begin with a result that will be useful to establish that the sequence is well defined.

Lemma 4.1.1. *Let $p \in \Omega$ such that $X(p) \neq 0$. Assume that $v = -\nabla X(p)^T X(p)$ or that v is a solution of the equation*

$$X(p) + \nabla X(p)v = 0. \quad (4.7)$$

If $v \neq 0$, then v is the descent direction for φ from p, i.e., $\langle \text{grad}\ \varphi(p), v\rangle < 0$.

Proof. First we assume that v satisfies (4.7). Since grad $\varphi(p) = \nabla X(p)^T X(p)$, we obtain that $\langle \text{grad}\ \varphi(p), v\rangle = \langle X(p), \nabla X(p)v\rangle$. Thus, considering that $X(p) \neq 0$, $v \neq 0$, and v satisfies (4.7), we conclude that

$$\langle \text{grad}\ \varphi(p), v\rangle = -\parallel X(p) \parallel^2 < 0,$$

which implies that v is a descent direction for φ from p. Now, we assume that $v = -\nabla X(p)^T X(p)$. Hence, $\langle \text{grad}\ \varphi(p), v\rangle = -\|\nabla X(p)^T X(p)\|^2 < 0$, and we also have that v is the descent direction for φ from p. □

Under suitable assumptions, the following result guarantees that the damped Newton's method, after a finite number of iterates, reduces to the classical iteration of Newton's method.

Lemma 4.1.2. *Let $p_* \in \mathbb{M}$ be a solution of (4.1). If ∇X is continuous at p_* and $\nabla X(p_*)$ is nonsingular, then there exists $0 < \hat{\delta} < \delta_{p_*}$ such that $B_{\hat{\delta}}(p_*) \subset \Omega$, $\nabla X(p)$ is nonsingular for each $p \in B_{\hat{\delta}}(p_*)$, and*

$$\lim_{p \to p_*} \frac{\varphi\left(N_X(p)\right)}{\parallel X(p) \parallel^2} = 0. \quad (4.8)$$

As a consequence, there exists $\delta > 0$ such that, for all $\sigma \in (0, 1/2)$ and $\delta < \hat{\delta}$, it holds that

$$\varphi\left(N_X(p)\right) \le \varphi\left(p\right) + \sigma\left\langle \text{grad}\ \varphi(p), -\nabla X(p)^{-1}X(p)\right\rangle, \qquad \forall\ p \in B_\delta p_*). \quad (4.9)$$

Proof. Using item (i) of Lemma 3.1.2, we obtain that there exists $0 < \hat{\delta} < \delta_{p_*}$ such that $B_{\hat{\delta}}(p_*) \subset \Omega$ and $\nabla X(p)$ is nonsingular for each $p \in B_{\hat{\delta}}(p_*)$. We proceed to prove (5.23). To simplify the notation, we define

$$v_p = -\nabla X(p)^{-1} X(p), \qquad p \in B_{\hat{\delta}}(p_*).$$

Since $X(p_*) = 0$ and ∇X is continuous at p_*, we have $\lim_{p \to p_*} v_p = 0$. Moreover, using (5.8) and the isometry of the parallel transport, and considering that $\| \exp_{p_*}^{-1} \exp_p(v_p) \| = d(\exp_p(v_p), p_*)$, we have

$$\varphi\left(N_X(p)\right) = \frac{1}{2} \parallel X\left(\exp_p(v_p)\right) - P_{p_* \exp_p(v_p)} X(p_*) \parallel^2 \leq \left[\|\nabla X(p_*)\| + \|r(\exp_p(v_p))\|\right]^2 d^2(\exp_p(v_p), p_*).$$

Hence, after some simples algebraic manipulations, we can conclude from the last inequality that

$$\frac{\varphi\left(\exp_p(v_p)\right)}{\parallel X(p) \parallel^2} \leq \left[\|\nabla X(p_*)\| + \|r(\exp_p(v_p))\|\right]^2 \frac{d^2(\exp_p(v_p), p_*)}{d^2(p, p_*)} \frac{d^2(p, p_*)}{\parallel X(p) \parallel^2}, \quad \forall\, p \in B_{\hat{\delta}}(p_*) \backslash \{p_*\}. \tag{4.10}$$

On the other hand, since $X(p_*) = 0$ and $\nabla X(p_*)$ is nonsingular, it is easy to see that

$$\exp_{p_*}^{-1} p = -\nabla X(p_*)^{-1}\left[P_{pp_*} X(p) - X(p_*) - \nabla X(p_*) \exp_{p_*}^{-1} p\right] + \nabla X(p_*)^{-1} P_{pp_*} X(p). \tag{4.11}$$

From (5.8) and since $\nabla X(p_*)$ is nonsingular, we conclude that there exists $0 < \bar{\delta} < \delta_{p_*}$ such that

$$\left\|\nabla X(p_*)^{-1}\left[P_{pp_*} X(p) - X(p_*) - \nabla X(p_*) \exp_{p_*}^{-1} p\right]\right\| \leq \frac{1}{2} d(p, p_*), \qquad \forall\, p \in B_{\bar{\delta}}(p_*).$$

Combining (5.26) with the last inequality and considering that $d(p_*, p) = \| \exp_{p_*}^{-1} p \|$, we conclude that $d(p_*, p) \leq d(p, p_*)/2 + \|\nabla X(p_*)^{-1} P_{pp_*} X(p)\|$ for all $p \in B_{\bar{\delta}}(p_*)$, which implies that

$$d^2(p_*, p) \leq \left[2\|\nabla X(p_*)^{-1}\|\right]^2 \|X(p)\|^2, \qquad \forall\, p \in B_{\bar{\delta}}(p_*).$$

Letting $\tilde{\delta} = \min\{\hat{\delta}, \bar{\delta}\}$, we conclude from (5.25) and the last inequality that $\forall\, p \in B_{\tilde{\delta}}(p_*) \backslash \{p_*\}$,

$$\frac{\varphi\left(\exp_p(v_p)\right)}{\parallel X(p) \parallel^2} \leq \left[2\|\nabla X(p_*)^{-1}\| \left(\|\nabla X(p_*)\| + \|r(\exp_p(v_p))\|\right)\right]^2 \frac{d^2(\exp_p(v_p), p_*)}{d^2(p, p_*)}.$$

Therefore, from Lemma 3.1.3 and considering that $\lim_{p \to p_*} r(\exp_p(v_p)) = 0$, the equality (5.23) follows by taking the limit as p tends to p_* in the above inequality. For proving (5.24), we first use (5.23) for concluding that there exists $\delta > 0$ such that $\delta < \hat{\delta}$ and for $\sigma \in (0, 1/2)$, we have

$$\varphi\left(N_X(p)\right) \leq \frac{1 - 2\sigma}{2} \parallel X(p) \parallel^2, \quad \forall\, p \in B_\delta(p_*).$$

Since $\text{grad}\ \varphi(p) = \nabla X(p)^T X(p)$, we obtain $\langle \text{grad}\ \varphi(p), -\nabla X(p)^{-1} X(p) \rangle = -\|X(p)\|^2$. Then, the last inequality is equivalent to (5.24) and the proof is complete. □

In the next result, we show that whenever the vector field is continuous and has a nonsingular covariant derivative at a singularity, there exist a neighborhood around it that is invariant by Newton's iterate mapping for the associated exponential.

Lemma 4.1.3. *Let $p_* \in \mathbb{M}$ be a solution of (4.1). If ∇X is continuous at p_* and $\nabla X(p_*)$ is nonsingular, then there exists $0 < \hat{\delta} < \delta_{p_*}$ such that $B_{\hat{\delta}}(p_*) \subset \Omega$ and $\nabla X(p)$ is nonsingular for each $p \in B_{\hat{\delta}}(p_*)$. Moreover, $N_X(p) \in B_{\hat{\delta}}(p_*)$ for all $p \in B_{\hat{\delta}}(p_*)$.*

Proof. It follows from Lemma 3.1.2 that there exists $0 < \hat{\delta} < \delta_{p_*}$ such that $B_{\hat{\delta}}(p_*) \subset \Omega$ and $\nabla X(p)$ is nonsingular for each $p \in B_{\hat{\delta}}(p_*)$. Thus, shrinking $\hat{\delta}$ if necessary, we can use Lemma 3.1.3 to conclude that $d(N_X(p), p_*) < d(p, p_*)/2$ for all $p \in B_{\hat{\delta}}(p_*)$, which implies the last statement of the lemma. □

4.2 Convergence Analysis

Finally, we are ready to establish the main result, i.e., the damped Newton's method with exponential (Algorithm 4.0.1) converges globally to a solution of (4.1), preserving the fast convergence rates of Newton's method (3.3).

Lemma 4.2.1. *The sequence $\{p_k\}$ generated by Algorithm 4.0.1 is well defined.*

Proof. If $v_k \neq 0$ and $X(p_k) \neq 0$, then by using Lemma 4.1.1, we conclude that v_k satisfying (4.3) or (4.4) is the descent direction for φ from p_k. Hence, by using a standard argument, we conclude that α_k defined in (4.5) is well defined, and hence the sequence generated by the damped Newton's method (Algorithm 4.0.1) is well defined. □

It is worth noting that, if the sequence generated by Algorithm 4.0.1 is finite, then the last point generated is a solution of (4.1) or it is a critical point of φ defined in (4.2). Thus, we can assume that $\{p_k\}$ is infinite, $v_k \neq 0$, and $X(p_k) \neq 0$ for all k.

Theorem 4.2.1. *Let $X : \Omega \subseteq \mathbb{M} \to T\mathbb{M}$ be a differentiable vector field. Assume that $\{p_k\}$ generated by the damped Newton's method (Algorithm 4.0.1) has an accumulation point $\bar{p} \in \Omega$ such that ∇X is continuous at $\bar{p}$ and $\nabla X(\bar{p})$ is nonsingular. Then, $\bar{p}$ is a singularity of X and $\{p_k\}$ converges super-linearly to $\bar{p}$. Moreover, the convergence rate is quadratic provided that ∇X is locally Lipschitz continuous at $\bar{p}$.*

Proof. We begin by showing that $\bar{p}$ is a singularity of X. Since $\{p_k\}$ is infinite, without loss of generality, we can assume that grad $\varphi(p_k) \neq 0$ and $X(p_k) \neq 0$ for all $k = 0, 1, \ldots$. Hence, from (4.5) and (4.6), it follows that

$$\varphi(p_k) - \varphi(p_{k+1}) \geq -\sigma\alpha_k \langle \text{grad } \varphi(p_k), v_k \rangle = \begin{cases} \sigma\alpha_k \|X(p_k)\|^2 > 0, & \text{if } v_k \text{ satisfies (4.3);} \\ \sigma\alpha_k \|\text{ grad } \varphi(p_k)\|^2 > 0, & \text{else.} \end{cases}$$

Then, the sequence $\{\varphi(p_k)\}$ is a strictly decreasing sequence, and because it is bounded below by zero, it converges. Hence, taking the limit as k tends to infinity in the above inequality, we conclude that $\lim_{k\to\infty}[\alpha_k \langle \text{grad } \varphi(p_k), v_k\rangle] = 0$. Let $\{p_{k_j}\}$ be a subsequence of sequence $\{p_k\}$ such that $\lim_{k_j\to+\infty} p_{k_j} = \bar{p}$. Thus, we have

$$\lim_{k_j\to\infty} \left[\alpha_{k_j} \left\langle \text{grad } \varphi(p_{k_j}), v_{k_j}\right\rangle\right] = 0. \tag{4.12}$$

Since ∇X is continuous at $\bar{p}$ and $\nabla X(\bar{p})$ is nonsingular, using Lemma 3.1.2 we can also assume that $\nabla X(p_{k_j})$ is nonsingular for all $j = 0, 1, \ldots$. Hence, (4.3) has a solution, and then,

$$v_{k_j} = -\nabla X(p_{k_j})^{-1} X(p_{k_j}), \qquad \left\langle \text{grad } \varphi(p_{k_j}), v_{k_j}\right\rangle = -\|X(p_{k_j})\|^2 \qquad j = 0, 1, \ldots. \tag{4.13}$$

For analyzing the consequences of (4.12), we consider the following two possible cases:

a) $\liminf_{j\to\infty}\alpha_{k_j} > 0$;

b) $\liminf_{j\to\infty}\alpha_{k_j} = 0$.

First, we assume that item a) holds. From (4.12), passing onto a further subsequence if necessary, we can assume that

$$\lim_{j\to\infty} \left\langle \text{grad } \varphi(p_{k_j}), v_{k_j}\right\rangle = 0. \tag{4.14}$$

Taking the limit in the above equality as j tends to infinity, and considering that $\lim_{j\to+\infty} p_{k_j} = \bar{p}$ and X is continuous at $\bar{p}$, we conclude from (4.14) and (4.13) that $X(\bar{p}) = 0$; in this case, $\bar{p}$ is a singularity of X. Now, we assume that item b) holds. Hence, given $s \in \mathbb{N}$, we can take j to be sufficiently large such that $\alpha_{k_j} < 2^{-s}$. Thus, 2^{-s} does not satisfy Armijo's condition (4.5), i.e.,

$$\varphi\left(\exp_{p_{k_j}}(2^{-s}v_k)\right) > \varphi(p_{k_j}) + \sigma 2^{-s} \left\langle \text{grad } \varphi(p_{k_j}),\ v_{k_j}\right\rangle.$$

Letting j tend to infinity, considering that the exponential mapping is continuous, $\lim_{j\to+\infty} p_{k_j} = \bar{p}$, and since ∇X and X are continuous at $\bar{p}$, it follows from (4.13) and the last inequality that

$$\varphi\left(\exp_{\bar{p}}(2^{-s}\bar{v})\right) > \varphi(\bar{p}) + \sigma 2^{-s} \langle \text{grad } \varphi(\bar{p}),\ \bar{v}\rangle,$$

where $\bar{v} = -\nabla X(\bar{p})^{-1}X(\bar{p})$, which implies that $[\varphi\left(\exp_{\bar{p}}(2^{-s}\bar{v})\right) - \varphi(\bar{p})]/2^{-s} \geq \sigma \langle \text{grad } \varphi(\bar{p}),\ \bar{v}\rangle$. Then, letting s tend to infinity, we conclude that $\langle \text{grad } \varphi(\bar{p}),\ \bar{v}\rangle \geq \sigma \langle \text{grad } \varphi(\bar{p}),\ \bar{v}\rangle$, or equivalently, $\|X(\bar{p})\|^2 \leq \sigma\|X(\bar{p}\|^2$. Thus, since $\sigma \in (0, 1/2)$, we have $\| X(\bar{p}) \|= 0$; in this case, we also obtain that $\bar{p}$ is a singularity of X.

We proceed to prove that there exists k_0 such that $\alpha_k = 1$ for all $k \geq k_0$. Since $\nabla X(\bar{p})$ is nonsingular and $X(\bar{p}) = 0$, Lemma 3.1.2 and Lemma 4.1.3 imply that there exists $\hat{\delta} > 0$ such that $\nabla X(p)$ is nonsingular and $N_X(p) \in B_\delta(\bar{p})$ for all $p \in B_\delta(\bar{p})$ and all $\delta \leq \hat{\delta}$. Thus, considering that $\bar{p}$ is an accumulation point of $\{p_k\}$, there exists k_0 such that $p_{k_0} \in B_{\hat{\delta}}(\bar{p})$, and shrinking $\hat{\delta}$ if necessary, we can also conclude from Lemma 4.1.2 that

$$\varphi\left(N_X(p_{k_0})\right) \leq \varphi\left(p_{k_0})\right) + \sigma \langle \text{grad } \varphi(p_{k_0}), v_{k_0}\rangle,$$

with $v_{k_0} = -\nabla X(p_{k_0})^{-1}X(p_{k_0})$. Hence, the last inequality, (2.5), and (4.5) imply that $\alpha_{k_0} = 1$, and using (4.6), we conclude that $p_{k_0+1} = N_X(p_{k_0})$. Since $N_X(p_{k_0}) \in B_{\hat{\delta}}(\bar{p})$, we also have $p_{k_0+1} \in B_{\hat{\delta}}(\bar{p})$. Then, an induction step is completely analogous, yielding

$$\alpha_k = 1, \qquad p_{k+1} = N_X(p_k) \in B_{\hat{\delta}}(\bar{p}), \qquad \forall\, k \geq k_0. \tag{4.15}$$

To obtain super-linear convergence of the entire sequence $\{p_k\}$, let $\bar{\delta} > 0$ be given by Theorem 3.2.1. Thus, shrinking $\hat{\delta}$ if necessary so that $\hat{\delta} < \bar{\delta}$, we can apply Theorem 3.2.1 to conclude from (5.27) that $\{p_k\}$ converges super-linearly to $\bar{p}$.

To prove that $\{p_k\}$ converges quadratically to $\bar{p}$, we first make $\hat{\delta} < r$, where r is given by Theorem 2.0.3. Then, considering that ∇X is locally Lipschitz continuous at $\bar{p}$, the result follows from the combination of (5.27) and Theorem 2.0.3. □

An important issue is the existence of accumulation points of iterative sequences generated by Algorithm 4.0.1. This is guaranteed when φ given by (4.2) is coercive. Indeed, if φ is coercive, it is easy see that the set $L_0 := \{p \in \Omega \ : \ \varphi(p) \leq \varphi(p_0)\}$ is compact. Moreover, if $\{p_k\}$ is a sequence generated by damped Newton's method 1 (Algorithm 4.0.1), then $\{p_k\} \subset L_0$.

5. On Newton's Method with Retraction

In this chapter, we generalize Theorem 3.2.1 and Theorem 4.2.1 for a general retraction. As mentioned in the introduction, there are sound arguments to establish these generalizations because efficient computation of the exponential, in general, involves significant numerical challenges. As we shall see, the statement and proof of the results of this section follow arguments similar to those in the previous sections.

5.1 Basic Definition and Auxiliary Results

In this section, we introduce some preliminary results that will be useful for achieving the aim of this chapter. We begin by defining the concept of retraction, which was introduced by (52); see also (5, Section 3).

Definition 5.1.1. *A retraction on a manifold $\mathbb{M}$ is a smooth mapping R from the tangent bundle $T\mathbb{M}$ onto $\mathbb{M}$ with the following properties. If R_p denotes the restriction of R to $T_p\mathbb{M}$, then*

(i) $R_p 0_p = p$, where 0_p denotes the origin of $T_p\mathbb{M}$;

(ii) with the canonical identification $T_{0_p}T_p\mathbb{M} \simeq T_p\mathbb{M}$, $R'_p 0_p = I_p$, where I_p is the identity mapping on $T_p\mathbb{M}$ and R'_p denotes the differential of R_p.

We remark that for every tangent vector v in $T_p\mathbb{M}$, the curve $\gamma : t \to R_p tv$ satisfies $\gamma'(0) = v$. Thus, moving along this curve γ is thought of as moving in the direction v while constrained to the manifold $\mathbb{M}$. We can see that Definition 5.1.1 implies that the *exponential map* is a retraction. In the following, we present some examples of retractions.

Example 5.1.1. *Let $\mathbb{P}^n_{++}$ be the cone of real symmetric positive definite matrices of order $n \times n$, and let $T_P\mathbb{P}^n_{++}$ be the space of symmetric matrices of order $n \times n$. A retraction on $\mathbb{P}^n_{++}$ is given by*

$$R_P V := P + V + \frac{1}{2} V P^{-1} V, \qquad P \in \mathbb{P}^n_{++} \quad V \in T_P\mathbb{P}^n_{++}; \tag{5.1}$$

see (73).

Example 5.1.2. *Let* $\mathbb{S}^{n-1} := \{p = (p_1, ..., p_{n+1}) \in \mathbb{R}^{n+1} : \| p \| = 1\}$ *be a sphere. A retraction on* $\mathbb{S}^{n-1}$ *is given by*

$$R_p v = \frac{p+v}{\|p+v\|}, \qquad p \in \mathbb{S}^{n-1}, \quad v \in T_p\mathbb{S}^n := \left\{v \in \mathbb{R}^{n+1} : \langle p, v\rangle = 0\right\};$$

see (3, Ch. 4).

Example 5.1.3. *Let* $\mathbb{S}(p,n) := \left\{P \in \mathbb{R}^{n\times p} : \ P^T P = I_p\right\}$ *be the Stiefel manifold, where* I_p *denotes the identity matrix of dimension* p *and* $T_P\mathbb{S}(p,n) := \left\{P \in \mathbb{R}^{n\times p} : \ P^T V + V^T P = 0\right\}$ *denotes its tangent space. A retraction on* $\mathbb{S}(p,n)$ *is given by*

$$R_P V = (P+V)\left(I_p + V^T V\right)^{-1/2}, \qquad P \in \mathbb{S}(p,n), \quad V \in T_P\mathbb{S}(p,n),$$

where V^T *denotes the transpose of matrix* V*; see (3, Ch. 4).*

As mentioned in Chapter 2, if $\gamma : [a,b] \to \mathbb{M}$ is a differentiable curve, then for each $t \in [a,b]$, the derivative covariant ∇ induces an isometry relative to the inner product of $\mathbb{M}$, $P_{\gamma,a,t} \colon T_{\gamma(a)}\mathbb{M} \to T_{\gamma(t)}\mathbb{M}$, i.e., *parallel transport* along a segment of curve γ joining the points $\gamma(a)$ and $\gamma(t)$. Thus, we have the following important remark.

Remark 5.1.1. *Let* $\gamma(t) = \exp_p(tv)$ *be a geodesic. Then, in a suitable neighborhood of* p*, we obtain the equality*

$$\exp_p^{-1}\gamma(t) = -P_{\gamma,t,0}\exp_{\gamma(t)}^{-1} p. \tag{5.2}$$

This property of parallel transport along a geodesic curve allows us establish Lemma 3.1.3. However, this property of parallel transport is not ensured for a curve given by any retraction. This hampers the generalization of Theorem 3.2.1 and Theorem 4.2.1.

Since $R'_p(0_p) = I_p$, by the inverse function theorem, R_p is a local diffeomorphism. Hence, as defined for the exponential mapping in Chapter 2, we also can define, for $p \in \mathbb{M}$, the *injectivity radius* of $\mathbb{M}$ at p with respect to R_p as follows:

$$i_{R_p} := \sup\left\{r > 0 \ : \ R_{p|_{B_r(0_p)}} \text{ is a diffeomorphism}\right\}, \tag{5.3}$$

where $B_r(0_p) := \{v \in T_p\mathbb{M} : \| v - 0_p \| < r\}$ is called the neighborhood of injectivity of p with respect to R.

Remark 5.1.2. *Let* $\bar{p} \in \mathbb{M}$*. The definition* (5.3) *implies that if* $0 < \delta < i_{R_{\bar{p}}}$*, then* $R_{\bar{p}}B_\delta(0_{\bar{p}}) = B_\delta(\bar{p})$*. Let* $v \in B_\delta(0_{\bar{p}})$*, and consider a curve* $[0,1] \ni t \to \gamma(t) = R_{\bar{p}}tv$ *such that* $\gamma(t) \subset B_\delta(\bar{p})$ *for all* t*. Hence, for all* $p = \gamma(t)$*,* $R_{\bar{p}}^{-1}p$ *is well defined. For all* $t \in [0,1]$ *satisfying* $p = \gamma(t)$ *such that for some* $0 < \hat{\delta} < i_{R_p}$*,* $\bar{p} \in B_{\hat{\delta}}(p)$ *holds, we can conclude that* $R_p^{-1}\bar{p}$ *is well defined.*

Consider γ as given in Remark 5.1.2. *Hereafter, if* $0 < \delta < i_{R_{\bar{p}}}$ *and* $p \in B_\delta(\bar{p})$*, we denote by* $P_{\gamma,p,\bar{p}}$ *parallel transport along this segment of curve* γ *joining* p *to* $\bar{p}$*; its inverse is denoted by* $P_{\gamma,\bar{p},p}$. Let $\bar{p} \in \mathbb{M}$ and $0 < \delta < i_{R_{\bar{p}}}$. Consider γ as the curve given by Remark 5.1.2. We will assume that the following equality holds.

$$\lim_{p\to\bar{p}} \frac{P_{\gamma,p,\bar{p}}R_p^{-1}\bar{p} - R_{\bar{p}}^{-1}p}{d(p,\bar{p})} = 0, \qquad \forall p \in B_\delta(\bar{p}). \tag{5.4}$$

Note that, (5.2) implies that the exponential map exp satisfies the equality (5.4). The following result establishes an important relation between the retraction and the Riemannian distance; see (35, Lemma 3).

Corollary 5.1.1. *Let $\mathbb{M}$ be a Riemannian manifold endowed with retraction R and $\bar{p} \in \mathbb{M}$. Then, there exist $a_0 > 0$, $a_1 > 0$, and δ_{a_0,a_1} such that for all p in a sufficiently small neighborhood of $\bar{p}$ and all $v \in T_p\mathbb{M}$ with $\|v\| \leq \delta_{a_0,a_1}$, the following inequality holds.*

$$a_0\|v\| \leq d\left(p,\ R_p(v)\right) \leq a_1\|v\|. \tag{5.5}$$

Consider $\bar{p} \in \mathbb{M}$ and $i_{R_{\bar{p}}}$ as given by (5.3). We define the following quantity.

$$\delta_{R_{\bar{p}}} := \min\{1, i_{R_{\bar{p}}}\}$$

Let $a_0 > 0$, $a_1 > 0$, and δ_{a_0,a_1} be given as in Corollary 5.1.1. Shrink δ_{a_0,a_1}, if necessary, such that $\delta_{a_0,a_1} < \delta < \delta_{R_{\bar{p}}}$. Let $v \in B_{\delta_{a_0,a_1}}(0_{\bar{p}})$ and $p = R_{\bar{p}}(v)$. Then, by the inequality (5.5), we can conclude that

$$a_0 \left\|R_{\bar{p}}^{-1}(p)\right\| \leq d\left(p,\ \bar{p}\right) \leq a_1 \left\|R_{\bar{p}}^{-1}(p)\right\|, \qquad \forall\ p \in B_\delta(\bar{p}). \tag{5.6}$$

Using the last inequality, we can ensure that the number $K_{R,p}$ as follows is well defined.

$$K_{R,p} = \sup\left\{\frac{d(R_q u, R_q v)}{\|u-v\|} : q \in B_{R_{\delta_p}}(p), u, v \in T_q\mathbb{M},\ u \neq v, \|v\| \leq \delta_{R_p}, \|u-v\| \leq \delta_{R_p}\right\}. \tag{5.7}$$

Let $X \in \mathcal{X}(\Omega)$ and $\bar{p} \in \Omega$, and assume that $0 < \delta < \delta_{R_{\bar{p}}}$. In this condition, using (29, Eq. (2.3)) and (5.5), we can conclude that

$$X(p) = P_{\gamma,\bar{p},p}X(\bar{p}) + P_{\gamma,\bar{p},p}\nabla X(\bar{p})R_{\bar{p}}^{-1}p + \|R_{\bar{p}}^{-1}p\|r(p), \qquad \lim_{p\to\bar{p}} r(p) = 0, \tag{5.8}$$

where $p \in B_\delta(\bar{p})$ and $[0,1] \ni t \to \gamma(t) = R_{\bar{p}}tR_{\bar{p}}^{-1}p$.

5.2 Newton's Method with Retraction

In this section, we propose a generalization of Newton's algorithm 3.0.1 for retraction for solving (4.1) with a differentiable vector field X. The formal description of the algorithm is as follow.

Algorithm 5.2.1.

Newton with Retraction

Step 0. *Take an initial point* $p_0 \in \mathbb{M}$*, and set* $k = 0$*. If* $\|X(p_0)\| = 0$*, stop.*

Step 1. *Compute the search direction* $v_k \in T_{p_k}\mathbb{M}$ *as a solution of the linear equation*

$$X(p_k) + \nabla X(p_k)v = 0. \tag{5.9}$$

If v_k*, exists go to* **Step** 2*. Otherwise, stop.*

Step 2. *Compute*

$$p_{k+1} := R_{p_k} v_k. \tag{5.10}$$

Step 3. *Set* $k \leftarrow k + 1$ *and go to* **Step 1***.*

We remark that, when the covariant derivative is nonsingular at the singularity of the vector field, and if the retraction R in Step 2 is the exponential mapping *exp*, Lemma 3.1.2 ensures that the solution of (5.9) is given by $v_k = -\nabla X(p_k)^{-1}X(p_k)$. Hence, the above algorithm becomes Newton's method given by (3.3). In this case, Theorem 3.2.1 ensures that the sequence generated by Algorithm 5.2.1 starting at a suitable point converges to a singularity of the vector field at a super-linear rate.

5.2.1 Convergence Analysis

In this section, we present the local convergence properties of the sequence generated by Algorithm 5.2.1. Toward this end, we introduce *Newton's iterate mapping for retraction.* Consider $\bar{p} \in \Omega$, where Ω is an open set of $\mathbb{M}$. Assume that $\nabla X(\bar{p})$ is nonsingular and $\bar{\delta}$ is given as in Lemma 3.1.2. Then, *Newton's iterate mapping for retraction* $N_{R,X} : B_{\bar{\delta}}(\bar{p}) \to \mathbb{M}$ is given by

$$N_{R,X}(p) := R_p\left(-\nabla X(p)^{-1}X(p)\right). \tag{5.11}$$

Note that $N_{R,X}$ is well defined since Lemma 3.1.2 ensures that $\nabla X(p)$ is nonsingular for all $p \in B_{\bar{\delta}}(\bar{p})$. The proof of the next lemma is similar to the proof of Lemma 3.1.3. However, we will present it here to highlight the technical details.

Lemma 5.2.1. *Let* $p_* \in \Omega$ *such that* $X(p_*) = 0$*. If* ∇X *is continuous at* p_* *and* $\nabla X(p_*)$ *is nonsingular, then*

$$\lim_{p \to p_*} \frac{d(N_{R,X}(p), p_*)}{d(p, p_*)} = 0.$$

Proof. Consider the following equality, obtained using some algebraic manipulations.

$$\nabla X(p)^{-1}X(p) + P_{\gamma,p_*,p}R_{p_*}^{-1}p = \nabla X(p)^{-1}\big[X(p) - P_{\gamma,p_*,p}X(p_*) - P_{\gamma,p_*,p}\nabla X(p_*)R_{p_*}^{-1}p + \left[P_{\gamma,p_*,p}\nabla X(p_*) - \nabla X(p)P_{\gamma,p_*,p}\right]R_{p_*}^{-1}p\big].$$

From the above equality and some properties of the norm, we have for each $p \in B_{\bar{\delta}}(p_*)$, where $\bar{\delta}$ is given by Lemma 3.1.2, that

$$\frac{\left\|\nabla X(p)^{-1}X(p) + P_{\gamma,p_*,p}R_{p_*}^{-1}p\right\|}{d(p,\ p_*)} \leq \left\|\nabla X(p)^{-1}\right\|\left[\,\|r(p)\| + \|P_{\gamma,p_*,p}\nabla X(p_*) - \nabla X(p)P_{\gamma,p_*,p}\|\,\right]\frac{\left\|R_{p_*}^{-1}p\right\|}{d(p,\ p_*)},$$

where r is given by (5.8). From (5.6), it follows that $\left\|R_{p_*}^{-1}p\right\|/d(p,\ p_*) \leq 1/a_0$. Combining the last two inequalities, we have

$$\frac{\left\|\nabla X(p)^{-1}X(p) + P_{\gamma,p_*,p}R_{p_*}^{-1}p\right\|}{d(p,\ p_*)} \leq \frac{\|\nabla X(p)^{-1}\|}{a_0}\left[\,\|r(p)\| + \|P_{\gamma,p_*,p}\nabla X(p_*) - \nabla X(p)P_{\gamma,p_*,p}\|\,\right].$$

Combining the last inequality with item (ii) of Lemma 3.1.2, we can obtain the following:

$$\left\|\nabla X(p)^{-1}X(p) + P_{\gamma,p_*,p}R_{p_*}^{-1}p\right\| \leq \frac{2}{a_0}\left\|\nabla X(p_*)^{-1}\right\|\left[\,\|r(p)\| + \|P_{\gamma,p,p_*}\nabla X(p)P_{\gamma,p_*,p} - \nabla X(p_*)\|\,\right]d(p,\ p_*). \tag{5.12}$$

Taking the limit as p tends to p_* on the right-hand side of the last inequality and using (3.5) and $\lim_{p \to p_*} r(p) = 0$, we have

$$\lim_{p \to p_*}\left\|\nabla X(p)^{-1}X(p) + P_{\gamma,p_*,p}R_{p_*}^{-1}p\right\| = 0. \tag{5.13}$$

Since $R_p^{-1}p_*$ is continuous, by using Corollary 3.1.1, we can conclude that $P_{\gamma,p,p_*}R_p^{-1}p_*$ is continuous. Thus,

$$\lim_{p \to p_*}\left\|P_{\gamma,p,p_*}R_p^{-1}p_* - R_{p_*}^{-1}p\right\| = 0. \tag{5.14}$$

Using the triangular inequality and since parallel transport is an isometry, we conclude that

$$\left\|\nabla X(p)^{-1}X(p) + R_p^{-1}p_*\right\| \leq \left\|\nabla X(p)^{-1}X(p) + P_{\gamma,p_*,p}R_{p_*}^{-1}p\right\| + \left\|P_{pp_*}R_p^{-1}p_* - R_{p_*}^{-1}p\right\|. \tag{5.15}$$

Thus, from (5.13) and (5.14), the left-hand side of the last inequality tends to zero as p tends to p_*. Hence, $\left\|\nabla X(p)^{-1}X(p) + R_p^{-1}p_*\right\| \leq \delta_{R_{p_*}}$ for all $p \in B_{\bar{\delta}}(p_*)$. From the definition of $N_{R,X}$ in (5.11) and the definition of K_{R,p_*} in (5.7), it follows that

$$d(N_{R,X}(p), p_*) \leq K_{R,p_*}\left\|-\nabla X(p)^{-1}X(p) - R_p^{-1}p_*\right\|, \qquad \forall\, p \in B_{\bar{\delta}}(p_*).$$

From (5.12) and (5.15), combined with the last inequality, we can conclude the following:

$$\frac{d(N_{R,X}(p), p_*)}{d(p, p_*)} \leq \frac{2}{a_0} K_{p_*} \left\| \nabla X(p_*)^{-1} \right\| \left[\|r(p)\| + \|P_{\gamma,p,p_*} \nabla X(p) P_{\gamma,p_*,p} - \nabla X(p_*)\| \right] + \frac{\left\| P_{\gamma,p,p_*} R_p^{-1} p_* - R_{p_*}^{-1} p \right\|}{d(p, p_*)}.$$

The desired result follows by letting p tend to p_* in the last inequality and by considering (3.5), (5.4), and that $r(p)$ tends to zero as p tends to p_*. □

In the following, we present the main result of this section, i.e., a generalization of Theorem 3.2.1 for retraction. Its proof is similar to the proof of Theorem 3.2.1, and for the sake of completeness, we will present it here.

Theorem 5.2.1. *Let $\mathbb{M}$ be a Riemannian manifold with a retraction R and let $\Omega \subset \mathbb{M}$ be an open set. Let $X : \Omega \to T\mathbb{M}$ be a differentiable vector field and $p_* \in \Omega$. Consider the Newton sequence given by*

$$p_{k+1} = R_{p_k}(-\nabla X(p_k)^{-1} X(p_k)), \qquad k = 0, 1, \ldots \tag{5.16}$$

Suppose that p_* is a singularity of X, ∇X is continuous at p_*, and $\nabla X(p_*)$ is nonsingular. Then, there exists $\bar{\delta} > 0$ such that, for all $p_0 \in B_{\bar{\delta}}(p_*)$, the sequence (5.16) is well defined, contained in $B_{\bar{\delta}}(p_*)$, and converges super-linearly to p_*.

Proof. First, note that from (5.11) and (5.16), we can obtain the following definition of Newton's sequence:

$$p_{k+1} = N_{R,X}(p_k), \qquad k = 0, 1, \ldots \tag{5.17}$$

From Lemma 3.1.3 and by considering $\bar{\delta}$ as given by Lemma 3.1.2, we can conclude that

$$d(N_{R,X}(p), p_*) < \frac{1}{2} d(p, p_*), \qquad \forall\, p \in B_{\bar{\delta}}(p_*). \tag{5.18}$$

Thus, from (5.17), (5.18), and Lemma 3.1.2 item (i), it follows that for all $p_0 \in B_{\bar{\delta}}(p_*)$, the Newton sequence $\{p_k\}$ defined by (5.16) is well defined and contained in $B_{\bar{\delta}}(p_*)$. Moreover, $\{p_k\}$ converges to p_* since (5.18) and (5.17) imply that $d(p_{k+1}, p_*) < 1/2 d(p_k, p_*)$, $k = 0, 1, \ldots$. The super-linear convergence of Newton's sequence follows by combining (5.17) with Lemma 3.1.3. □

The convergence of the sequence generated by Algorithm 5.2.1 depends on the initial point p_0. Hence, this method is a local method. In general, we cannot guarantee that equation (5.9) has a solution. If not, the method stops and it does not converge to a singularity of the vector field in consideration. In the next section, we will consider a modification of Algorithm 5.2.1 to globalize this algorithm in addition to improving its efficiency.

5.3 Damped Newton Method with Retraction

In this section, we present and study the convergence properties of a new algorithm for solving (4.1) with a differentiable vector field X, which is a generalization of Algorithm 4.0.1 with retraction. We point out that the results of the next two sections are similar to those in Sections 4.1 and 4.2, respectively. However, we have decided to present them for completeness. Let φ be the function defined in (4.2) associated with X. The formal description of the method is as follows.

Algorithm 5.3.1.

Damped Newton with Retraction

Step 0. *Choose a scalar* $\sigma \in (0, 1/2)$*, take an initial point* $p_0 \in \mathbb{M}$*, and set* $k = 0$.

Step 1. *Compute search direction* $v_k \in T_{p_k}\mathbb{M}$ *as a solution of the linear equation*

$$X(p_k) + \nabla X(p_k)v = 0. \tag{5.19}$$

If v_k *exists, go to* **Step** 2*. Otherwise, set the search direction as* $v_k = -\operatorname{grad} \varphi(p_k)$*, where* φ *is defined by* (4.2)*, i.e.,*

$$v_k = -\nabla X(p_k)^* X(p_k). \tag{5.20}$$

If $v_k = 0$*, stop.*

Step 2. *Compute the step size by the rule*

$$\alpha_k := \max\left\{2^{-j} \ : \ \varphi\left(R_{p_k}(2^{-j}v_k)\right) \leq \varphi(p_k) + \sigma 2^{-j} \left\langle \operatorname{grad} \varphi(p_k),\ v_k\right\rangle,\ j \in \mathbb{N}\right\}, \tag{5.21}$$

and set the next iterate as

$$p_{k+1} := R_{p_k}(\alpha_k v_k). \tag{5.22}$$

Step 3. *Set* $k \leftarrow k+1$ *and go to* **Step 1**.

We remark that Lemma 4.1.1 ensures that the above algorithm is well defined, since v_k defined by (5.19) or (5.20) is a descent direction for φ. In the next section, we study the convergence properties of this algorithm.

5.3.1 Preliminaries

In this section, we present some preliminary results to ensure that the sequence generated by Algorithm 5.3.1 converges to a solution of the problem (4.1) at a super-linear rate. We begin by showing that the damped Newton's algorithm with retraction generates a sequence that, after a finite number of iterates, reduces to the Newton's sequence given by (5.10). This result is a version of Lemma 5.23 for retraction. Its formal statement is as follows.

Lemma 5.3.1. *Let $p_* \in \mathbb{M}$ such that $X(p_*) = 0$. If ∇X is continuous at p_* and $\nabla X(p_*)$ is nonsingular, then there exists $0 < \hat{\delta} < \delta_{R_{p_*}}$ such that $B_{\hat{\delta}}(p_*) \subset \Omega$, $\nabla X(p)$ is nonsingular for each $p \in B_{\hat{\delta}}(p_*)$, and*

$$\lim_{p \to p_*} \frac{\varphi\left(N_{R,X}(p)\right)}{\| X(p) \|^2} = 0. \tag{5.23}$$

As a consequence, there exists $\delta > 0$ such that, for all $\sigma \in (0, 1/2)$ and $\delta < \hat{\delta}$, it holds that

$$\varphi\left(N_{R,X}(p)\right) \leq \varphi\left(p\right)) + \sigma \left\langle \text{grad } \varphi(p), -\nabla X(p)^{-1} X(p) \right\rangle, \qquad \forall\, p \in B_{\delta}(p_*). \tag{5.24}$$

Proof. Lemma 3.1.2 ensures that there exists $0 < \hat{\delta} < \delta_{R_{p_*}}$ such that $B_{\hat{\delta}}(p_*) \subset \Omega$ and $\nabla X(p)$ is nonsingular for each $p \in B_{\hat{\delta}}(p_*)$. To prove (5.23), we denote $v_p = -\nabla X(p)^{-1} X(p)$ for all $p \in B_{\hat{\delta}}(p_*)$. Since $X(p_*) = 0$ and ∇X is continuous at p_* and nonsingular, we have $\lim_{p \to p_*} v_p = 0$. Moreover, from (5.6) and (5.8), it follows that

$$\varphi\left(N_{R,X}(p)\right) = \frac{1}{2} \left\| X\left(R_p v_p\right) - P_{\gamma, p_*, R_p v_p} X(p_*) \right\|^2 \leq \left[\|\nabla X(p_*)\| + \|r(R_{p_*} v_p)\|\right]^2 d^2(R_p v_p, p_*).$$

Hence, after some simple algebraic manipulations, the last inequality implies the following:

$$\frac{\varphi\left(N_{R,X}(p)\right)}{\| X(p) \|^2} \leq \left[\|\nabla X(p_*)\| + \|r(R_{p_*} v)\|\right]^2 \frac{d^2(R_{p_*} v, p_*)}{d^2(p, p_*)} \frac{d^2(p, p_*)}{\| X(p) \|^2}, \quad \forall\, p \in B_{\hat{\delta}}(p_*) \backslash \{p_*\}. \tag{5.25}$$

Note that, by considering that p_* is a singularity of X and $\nabla X(p_*)$ is nonsingular,

$$R_{p_*}^{-1} p = -\nabla X(p_*)^{-1} \left[P_{\gamma, p, p_*} X(p) - X(p_*) - \nabla X(p_*) R_{p_*}^{-1} p\right] + \nabla X(p_*)^{-1} P_{\gamma, p, p_*} X(p). \tag{5.26}$$

By using (5.6) and (5.8), we can take $\bar{\delta}$ with $0 < \bar{\delta} < \delta_{R_{p_*}}$ to conclude that

$$\begin{aligned} \left\|\nabla X(p_*)^{-1} \left[P_{\gamma, p, p_*} X\left(p\right) - X(p_*) - \nabla X(p_*) R_{p_*}^{-1} p\right]\right\| &\leq \frac{a_0}{2a_1} \left\|R_{p_*}^{-1} p\right\| \\ &\leq \frac{1}{2a_1} d\left(p, p_*\right), \qquad \forall\, p \in B_{\bar{\delta}}(p_*). \end{aligned}$$

From the last inequality, (5.6), and (5.26), we have $d(p, p_*) \leq d(p, p_*)/2 + a_1 \|\nabla X(p_*)^{-1} P_{\gamma, p, p_*} X\left(p\right)\|$ for all $p \in B_{\bar{\delta}}(p_*)$. Hence, we have

$$d^2\left(p_*, p\right) \leq \left[2a_1 \|\nabla X(p_*)^{-1}\|\right]^2 \|X\left(p\right)\|^2, \qquad \forall\, p \in B_{\bar{\delta}}(p_*).$$

From (5.25) and the last inequality, for all $p \in B_{\tilde{\delta}}(p_*) \backslash \{p_*\}$, where $\tilde{\delta} = \min\{\hat{\delta}, \bar{\delta}\}$, we conclude that

$$\frac{\varphi\left(N_{R,X}(p)\right)}{\| X(p) \|^2} \leq \left[2a_1 a_0^{-1} \|\nabla X(p_*)^{-1}\|\right]^2 \left[\ \|\nabla X(p_*)\| + \|r(R_{p_*} v)\|\right]^2 \frac{d^2(R_{p_*} v, p_*)}{d^2(p, p_*)}.$$

The equality (5.23) follows by taking the limit as p tends to p_* in the above inequality and considering Lemma 5.2.1. Moreover, from (5.23), there exists $\delta > 0$ such that $\delta < \hat{\delta}$ with

$$\varphi\left(N_{R,X}(p)\right) \leq \frac{1 - 2\sigma}{2} \| X(p) \|^2, \quad \forall\, p \in B_{\delta}(p_*)$$

when $\sigma \in (0, 1/2)$. Since $\langle \text{grad } \varphi(p), -\nabla X(p)^{-1} X(p) \rangle = -\|X(p)\|^2$, the last inequality is equivalent to (5.24). □

In the next result, we show that whenever the vector field is continuous and has a non-singular covariant derivative at a singularity, there exist a neighborhood around it that is invariant by Newton's iterate mapping for the associated retraction.

Lemma 5.3.2. *Let $p_* \in \mathbb{M}$ such that $X(p_*) = 0$. If ∇X is continuous at p_* and $\nabla X(p_*)$ is nonsingular, then there exists $0 < \hat{\delta} < \delta_{R_{p_*}}$ such that $B_{\hat{\delta}}(p_*) \subset \Omega$ and $\nabla X(p)$ is nonsingular for each $p \in B_{\hat{\delta}}(p_*)$. Moreover, $N_{R,X}(p) \in B_{\hat{\delta}}(p_*)$ for all $p \in B_{\hat{\delta}}(p_*)$.*

Proof. It similar to the proof of Lemma 4.1.3. □

5.3.2 Convergence Analysis

In this section, we establish the main result of this chapter. We will present a result on the global convergence of the damped Newtons method with retraction, preserving the property of super-linear convergence of Algorithm 4.0.1. We begin by proving that the sequence generated by the damped Newton's algorithm with retraction is well defined.

Lemma 5.3.3. *The sequence $\{p_k\}$ generated by the damped Newton's algorithm is well defined.*

Proof. Lemma 4.1.1 ensures that v_0 satisfying (5.19) or (5.20) is the descent direction for φ from p_0. Hence, α_0 in (5.21) is well defined. By induction, we can prove that the sequence $\{p_k\}$ given by (5.22) is well defined. □

The next result is the main result of this section. The proof of this result is similar to the proof of Theorem 4.2.1. Nevertheless, we will show the main idea of this proof for the sake of completeness.

Theorem 5.3.1. *Let $X : \Omega \subseteq \mathbb{M} \to T\mathbb{M}$ be a differentiable vector field. Assume that $\{p_k\}$ generated by the damped Newton's method has an accumulation point $\bar{p} \in \Omega$ such that ∇X is continuous at $\bar{p}$ and $\nabla X(\bar{p})$ is nonsingular. Then, $\bar{p}$ is a singularity of X and $\{p_k\}$ converges super-linearly to $\bar{p}$.*

Proof. Let $\{p_k\}$ be a sequence generated by Algorithm 5.3.1 and $\bar{p}$ be an accumulation point of $\{p_k\}$ such that ∇X is continuous at $\bar{p}$ and $\nabla X(\bar{p})$ is nonsingular. Since the retraction is a continuous mapping, using Lemma 3.1.2, Lemma 5.3.2, Lemma 5.3.1, and the same arguments as those used in the proof of Theorem 4.2.1, we can conclude that $\bar{p}$ is a singularity of X and that there exists $\hat{\delta} > 0$ and k_0 such that

$$\alpha_k = 1, \qquad p_{k+1} = N_{R,X}(p_k) \in B_{\hat{\delta}}(\bar{p}), \qquad \forall\, k \geq k_0.$$

The super-linear convergence of the entire sequence $\{p_k\}$ holds by shrinking $\hat{\delta}$ if necessary so that $\hat{\delta} < \bar{\delta}$, where $\bar{\delta}$ is given by Theorem 5.2.1, and by applying this theorem together with the last equality. □

6. Numerical Experiments

In this chapter, we shall give some numerical experiments to illustrate the performance of the damped Newton method and damped Newton method with retraction for minimizing families of functions defined on the cone of symmetric positive definite matrices. Our goal here is to compare the efficiency of Damped Newton method with Newton method where this last has convergence global. Moreover, we compare the efficiency of the Damped Newton method with damped Newton method with retraction in iterates and time. Before present our numerical experiment, we need to introduce some concepts. Let $\mathbb{P}^n$ be the set of symmetric matrices of order $n \times n$ and $\mathbb{P}^n_{++}$ be the cone of symmetric positive definite matrices. Following Rothaus (62), let $\mathbb{M} := (\mathbb{P}^n_{++}, \langle \, , \, \rangle)$ be the Riemannian manifold endowed with the Riemannian metric defined by

$$\langle U, V \rangle = \mathrm{tr}(V\psi''(P)U) = \mathrm{tr}(VP^{-1}UP^{-1}), \qquad P \in \mathbb{M}, \qquad U, V \in T_P\mathbb{M} \approx \ \mathbb{P}^n, \qquad (6.1)$$

where tr$P$ denotes the trace of $P \in \mathbb{P}^n$. In this case exponential mapping is given by

$$\exp_P V = P^{1/2} e^{\left(P^{-1/2}VP^{-1/2}\right)} P^{1/2}, \qquad P \in \mathbb{M}, \qquad V \in T_P\mathbb{M} \approx \ \mathbb{P}^n. \qquad (6.2)$$

Let X and Y be vector fields in $\mathbb{M}$. Then, by using (63, Theorem 1.2, page 28), we can prove that the Levi-Civita connection of $\mathbb{M}$ is given by

$$\nabla_Y X(P) = X'Y - \frac{1}{2}\left[YP^{-1}X + XP^{-1}Y\right], \qquad (6.3)$$

where $P \in \mathbb{M}$, and X' denotes the Euclidean derivative of X. Therefore, it follows from (6.1) and (6.3) that the Riemannian gradient and hessian of a twice-differentiable function $f : \mathbb{M} \to \mathbb{R}$ are respectively given by:

$$\mathrm{grad} f(P) = Pf'(P)P, \qquad \mathrm{Hess}\, f(P)V = Pf''(P)VP + \frac{1}{2}\left[Vf'(P)P + PF'(P)V\right], \quad (6.4)$$

for all $V \in T_P\mathbb{M}$, where $f'(P)$ and $f''(P)$ are the Euclidean gradient and hessian of f at P, respectively. In this case, by using (6.2), the Newton's iteration for finding $P \in \mathbb{M}$ such that grad$f(P) = 0$ is given by

$$P_{k+1} = P_k^{1/2} e^{P_k^{-1/2} V_k P_k^{-1/2}} P_k^{1/2}, \qquad k = 0, 1, \ldots, \qquad (6.5)$$

where, by using again the equalities in (6.4), V_k is the unique solution of the Newton's linear equation

$$P_k f''(P_k)V_k P_k + \frac{1}{2}\left[V_k f'(P_k)P_k + P_k f'(P_k)V_k\right] = -P_k f'(P_k)P_k, \qquad (6.6)$$

which corresponds to (5.19) for $X = \mathrm{grad} f$. Now, we are going to present concrete examples for (6.5) and (6.6). Let $i = 1, 2$ and $f_i : \mathbb{P}^n_{++} \to \mathbb{R}$ be defined, respectively, by

$$f_1(P) = \mathrm{a}_1 \ln \det P + \mathrm{b}_1 \operatorname{tr} P^{-1}, \qquad f_2(P) = \mathrm{a}_2 \ln \det P - \mathrm{b}_2 \operatorname{tr} P, \tag{6.7}$$

where $\mathrm{b}_1/\mathrm{a}_1 > 0$, $\mathrm{a}_2/\mathrm{b}_2 > 0$, and $\det P$ denotes the determinant of P, respectively. Then, for each $i = 1, 2$, the Euclidean gradient and hessian of f_i are given, respectively, by

$$f_1'(P) = \mathrm{a}_1 P^{-1} - \mathrm{b}_1 P^{-2}, \qquad f_1''(P)V = \mathrm{b}_1 \left(P^{-1}VP^{-2} + P^{-2}VP^{-1}\right) - \mathrm{a}_1 P^{-1}VP^{-1}, \tag{6.8}$$

$$f_2'(P) = \mathrm{a}_2 P^{-1} - \mathrm{b}_2 I, \qquad f_2''(P)V_k = -\mathrm{a}_2 P^{-1}VP^{-1}, \tag{6.9}$$

where I denotes the $n \times n$ identity matrix. It follows from (6.4), (6.7), (6.8) and (6.9) that

$$\operatorname{grad} f_1(P) = \mathrm{a}_1 P - \mathrm{b}_1 I, \qquad \operatorname{grad} f_2(P) = \mathrm{a}_2 P - \mathrm{b}_2 P^2. \tag{6.10}$$

From the last two equalities, we can conclude that the global minimizer of f_i, for each $i = 1, 2$ are $P_1^* = \mathrm{b}_1/\mathrm{a}_1\, I$ and $P_2^* = \mathrm{a}_2/\mathrm{b}_2\, I$, respectively. Our task is to execute explicitly the Damped Newton method to find the global minimizer of f_i, for each $i = 1, 2$. For this purpose, consider $\varphi_i : \mathbb{P}^n_{++} \to \mathbb{R}$ given by $\varphi_i(P) = 1/2 \left\|\operatorname{grad} f_i(P)\right\|^2$, $i = 1, 2$. Thus, by using (6.10) we conclude that

$$\operatorname{grad} \varphi_1(P) = \mathrm{ab} I - \mathrm{b}_1^2 P^{-1}, \qquad \operatorname{grad} \varphi_2(P) = d^2 P^3 - \mathrm{a}_2\mathrm{b}_2 P^2. \tag{6.11}$$

By combining (6.1), (6.2) and (6.10), after some calculations we obtain the following equalities

$$\varphi_1\left(\exp_P V\right) = \frac{1}{2}\operatorname{tr}\left(\mathrm{a}_1 I - \mathrm{b}_1\left(P^{1/2} e^{P^{-1/2}VP^{-1/2}} P^{1/2}\right)^{-1}\right)^2, \tag{6.12}$$

$$\varphi_2\left(\exp_P V\right) = \frac{1}{2}\operatorname{tr}\left(\mathrm{a}_2 I - \mathrm{b}_2 P^{1/2} e^{P^{-1/2}VP^{-1/2}} P^{1/2}\right)^2. \tag{6.13}$$

Finally, by using (6.5), (6.6), definition of f_1 in (6.7), (6.8), left hand sides of (6.10) and (6.11) and (6.12) we give the damped Newton method, for finding the global minimizer of f_1. The formal algorithm to find the global minimizer of function f_2 will not be presented here because it can be obtained similarly by using (6.5), (6.6), the definition of f_2 in (6.7), (6.9), right hand sides of (6.10) and (6.11) and (6.13). The Damped Newton algorithim for the matrix cone is given as follows.

Damped Newton Method in $\mathbb{P}^n_{++}$ (DNM-$\mathbb{P}^n_{++}$)

Step 0. Choose a scalar $\sigma \in (0, 1/2)$, take an initial point $P_0 \in \mathbb{M}$, and set $k = 0$;

Step 1. Compute *search direction* V_k, as a solution of the linear equation

$$P_k V_k + V_k P_k = 2(P_k^2 - \mathrm{a}_1/\mathrm{b}_1 P_k^3). \tag{6.14}$$

If V_k exists go to **Step** 2. Otherwise, set the search direction $V_k = -\operatorname{grad} \varphi_1(p_k)$, i.e,

$$V_k = \mathrm{b}_1^2 P_k^{-1} - \mathrm{a}_1 \mathrm{b}_1 I. \tag{6.15}$$

If $V_k = 0$, stop;

Step 2. Compute the *stepsize* by the rule

$$\alpha_k := \max\left\{2^{-j} : \frac{1}{2} tr\left(\mathrm{a}_1 I - \mathrm{b}_1 \left(P_k^{1/2} e^{2^{-j} P_k^{-1/2} V P_k^{-1/2}} P_k^{1/2}\right)^{-1}\right)^2 \leq \left(\frac{1}{2} - \sigma 2^{-j}\right) tr\left(\mathrm{a}_1 I - \mathrm{b}_1 P_k^{-1}\right)^2\right\} \tag{6.16}$$

and set the *next iterated* as

$$P_{k+1} := P_k^{1/2} e^{\alpha_k P_k^{-1/2} V_k P_k^{-1/2}} P_k^{1/2}; \tag{6.17}$$

Step 3. Set $k \leftarrow k+1$ and go to **Step 1**.

Although the domain of f_1 is a subset of the symmetric matrix set, namely, $\mathbb{P}^n_{++}$, equality (6.17) shows that DNM-$\mathbb{P}^n_{++}$ generates only feasible points without using projections or any other procedure to remain the feasibility. Hence, problem of minimizing f_1 in $\mathbb{P}^n_{++}$ can be seen as unconstrained Riemannian optimization problem. Note that, in this case, the equation (6.14) always has unique solution $V_k \in T_{P_k}\mathbb{P}^n_{++}$ see (19, Th. 8.2.1), and consequently the direction V_k in (6.15) does not play any role here. Thus, in this case, we can compare the Newton and damped Newton methods in number of iterations and CPU time to reach the solution. The Newton method is formally described as follow.

Newton Method in $\mathbb{P}^n_{++}$ (NM-$\mathbb{P}^n_{++}$)

Step 0. Take an initial point $P_0 \in \mathbb{M}$, and set $k = 0$;

Step 1. Compute *search direction* V_k as a solution of the linear equation

$$P_k V_k + V_k P_k = 2(P_k^2 - \mathrm{a}_1/\mathrm{b}_1 P_k^3).$$

Step 2. Compute the *next iterated* by

$$P_{k+1} := P_k^{1/2} e^{P_k^{-1/2} V_k P_k^{-1/2}} P_k^{1/2};$$

Step 3. Set $k \leftarrow k+1$ and go to **Step 1**.

In order to compare the efficiency of the damped Newton method with the damped Newton method with retraction, we present damped Newton method with retraction in $\mathbb{P}^n_{++}$ which follows by combining (5.1), (6.1), (6.10), (6.14) and (6.15).

Damped Newton Method with Retraction $\mathbb{P}^n_{++}$ (DNMR-$\mathbb{P}^n_{++}$)

Step 0. Choose a scalar $\sigma \in (0, 1/2)$, take an initial point $P_0 \in \mathbb{M}$, and set $k = 0$;

Step 1. Compute *search direction* V_k, as a solution of the linear equation

$$P_k V_k + V_k P_k = 2(P_k^2 - a/bP_k^3). \tag{6.18}$$

If V_k exists go to **Step** 2. Otherwise, set the search direction $V_k = -\operatorname{grad} \varphi_1(p_k)$, i.e,

$$V_k = \mathrm{b}_1^2 P_k^{-1} - \mathrm{ab}I. \tag{6.19}$$

If $V_k = 0$, stop;

Step 2. Compute the *stepsize* by the rule

$$\alpha_k := \max \left\{ 2^{-j} : \frac{1}{2} tr \left(\mathrm{a}_1 I - \mathrm{b}_1 \left(P_k + V_k + \frac{1}{2} V_k P_k^{-1} V_k \right)^{-1} \right)^2 \leq \left(\frac{1}{2} - \sigma 2^{-j} \right) tr \left(\mathrm{a}_1 I - \mathrm{b}_1 P_k^{-1} \right)^2 \right\} \tag{6.20}$$

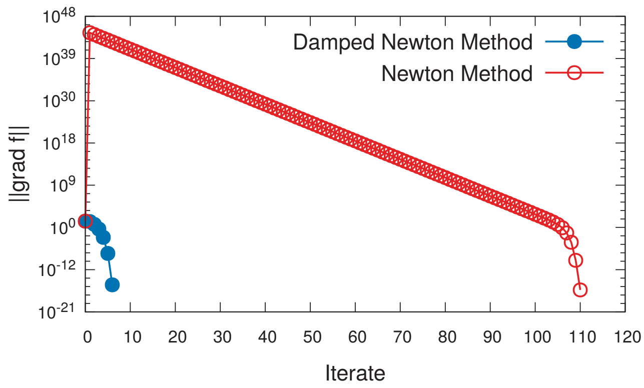
and set the *next iterated* as

$$P_{k+1} := P_k + V_k + \frac{1}{2} V_k P_k^{-1} V_k; \tag{6.21}$$

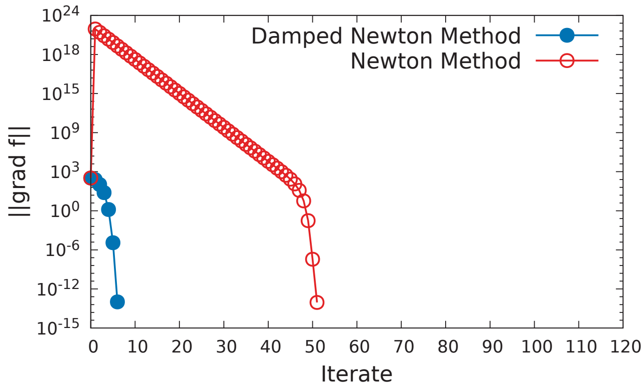
Step 3. Set $k \leftarrow k + 1$ and go to **Step 1**.

Although the domain of f_1 is a subset of the symmetric matrix set, namely, $\mathbb{P}^n_{++}$, equality (6.21) ensures that DNM-$\mathbb{P}^n_{++}$ generates only feasible points, (see (73, Prop. 3.1)). Hence, without using projections or any other procedure to remain the feasibility the problem of minimizing f_1 onto $\mathbb{P}^n_{++}$ can be seen as unconstrained Riemannian optimization problem.

We have implemented the above three algorithms by using MATLAB R2015b. The experiments are performed on an Intel Core 2 Duo Processor 2.26 GHz , 4 GB of RAM, and OSX operating system. We compare the iterative behavior of NM-$\mathbb{P}^n_{++}$ and DNM-$\mathbb{P}^n_{++}$ applied to f_1 and f_2 with randomly chosen initial guesses. In all experiments the stop criterio at the iterate $P_k \in \mathbb{P}^n_{++}$ is $\|\operatorname{grad} f(P_k)\| \leq 10^{-8}$ where $\|\cdot\|$ is norm associated to the metric given by (6.1). All codes are freely available at `http://www2.uesb.br/professor/telesfernandes/wp-content/uploads/2018/04/codigo_tese.zip`. We run the above two



(a) f_1 with $\mathrm{b}_1/\mathrm{a}_1 = 0.1$ and $n = 1000$.



(b) f_2 with $\mathrm{b}_2/\mathrm{a}_2 = 0.002$ and $n = 1000$.

Figure 6.1: Performance of DNM-$\mathbb{P}^n_{++}$ and NM-$\mathbb{P}^n_{++}$ to f_1 and f_2

algorithms in a huge number of problems by varying the dimensions and parameters $\mathrm{b}_1/\mathrm{a}_1$ and $\mathrm{b}_2/\mathrm{a}_2$ in functions f_1 and f_2, respectively. In general, DNM-$\mathbb{P}^n_{++}$ is superior to NM-$\mathbb{P}^n_{++}$ in number of iterations and CPU time. As we can see in figures (a) and (b) NM-$\mathbb{P}^n_{++}$ performs huge number of iterations before reaching the superlinear convergence region while the line-search of DNM-$\mathbb{P}^n_{++}$ decreasing gradient norm ensuring a superior performance in number of iterations. Note that in these figures DNM-$\mathbb{P}^n_{++}$ and NM-$\mathbb{P}^n_{++}$ have the same behavior when the iterations close to the solution, which is the consequence of Lemma 5.3.1. In part the efficiency of the DNM-$\mathbb{P}^n_{++}$ can be explained due to the linearsearch decreasing the merit function.

Table 6.1 shows that DNM-$\mathbb{P}^n_{++}$ is superior to NM-$\mathbb{P}^n_{++}$ in number of iterations and CPU time, besides showing the number of evaluation of the function in the linear-search to compute the stepsize. In the columns of this table we read: n is the dimension of $\mathbb{P}^n$, $\mathrm{b}_1/\mathrm{a}_1$ and $\mathrm{b}_2/\mathrm{a}_2$ are the parameters defining the functions f_1 and f_2, respectively, NIT is the number of iterates to reach the minimizer of the function, FE is the number of function evaluations in the Armijo's rule and Time is CPU time in seconds. We knows that, in general, the linear-search to compute the stepsize is expensive, but the computational effort did not grow significantly with the dimension of of $\mathbb{P}^n$ in the DNM-$\mathbb{P}^n_{++}$.

	b_i/a_i	n	**NM-$\mathbb{P}^n_{++}$** NIT	Time (sec.)	**DNM-$\mathbb{P}^n_{++}$** NIT	FE	Time (sec.)
f_1	0.1	1	41	0.1879	4	17	0.1012
		100	109	1.9406	6	24	0.5875
		1000	110	1239.7	6	24	107.19
	1.0	1	7	0.0964	4	13	0.1059
		100	13	0.4369	4	14	0.4341
		1000	13	147.57	4	14	64.929
	1.5	1	6	0.1038	4	13	0.1222
		100	10	0.2161	5	17	0.2472
		1000	10	112.69	6	21	97.08
f_2	0.01	1	26	0.3075	5	18	0.1659
		100	12	0.9245	5	17	0.2086
		1000	12	137.70	5	17	85.003
	0.002	1	125	0.2898	6	24	0.1022
		100	51	0.9040	6	23	0.2740
		1000	51	626.86	6	23	108.72
	0.001	1	249	0.3749	6	25	0.1739
		100	100	1.6574	7	27	0.4238
		1000	100	1131.3	7	27	130.53

Table 6.1: Performance of DNM-$\mathbb{P}^n_{++}$ and NM-$\mathbb{P}^n_{++}$ to f_1 and f_2 and the effort to compute the step-size.

	b_i/a_i	n	**DNMR-$\mathbb{P}^n_{++}$** NIT	FE	Time (sec.)	**DNM-$\mathbb{P}^n_{++}$** NIT	FE	Time (sec.)
f_1	0.1	1	41	123	0.0251	4	17	0.0174
		100	109	327	0.7863	6	24	0.1432
		1000	110	330	421.47	6	24	28.688
	1.0	1	7	21	0.0182	4	13	0.0173
		100	13	39	0.1682	4	14	0.0535
		1000	13	39	48.450	4	14	17.133
	1.5	1	6	19	0.0167	4	13	0.0178
		100	10	30	0.0836	5	17	0.0582
		1000	10	30	36.793	6	21	24.403
f_2	0.01	1	1	5	0.0142	5	18	0.0202
		100	6	19	0.0915	5	17	0.0550
		1000	6	19	19.998	5	17	17.941
	0.002	1	5	18	0.0178	6	24	0.0208
		100	6	20	0.0536	6	23	0.0628
		1000	6	20	19.309	6	23	23.767
	0.001	1	6	21	0.0169	6	25	0.0182
		100	5	18	0.0500	7	27	0.0716
		1000	5	18	16.976	7	27	27.539

Table 6.2: Performance of DNMR-$\mathbb{P}^n_{++}$ and DNM-$\mathbb{P}^n_{++}$ to f_1 and f_2 and the effort to compute the step-size.

7. Conclusions and Further Research

In this doctoral thesis, under non-singularity of the covariant derivative of the vector field at its zero and without any additional conditions on this derivative, we established the super-linear local convergence of Newtons method with exponential mapping, i.e., Algorithm 3.0.1, as well as the super-linear global convergence of the damped Newton's method with exponential mapping, i.e., Algorithm 4.0.1. Moreover, we presented generalizations of these algorithms with retraction. However, we assumed the equality (5.4), which is true for exponential mapping, but we did not prove it in the sense of any retraction. It is worth noting that we assumed that the Riemannian manifold is finite-dimensional in order to establish Lemma 3.1.1 and Lemma 3.1.2. On the other hand, we know that Newton's method converges super-linearly on infinite-dimensional Banach spaces under non-singularity of the derivative of the operator at its zero. Since many important problems arise in infinite-dimensional Riemannian manifolds as problems of finding singularities of vector fields, e.g., see (45; 54), it would be interesting to extend our results to such manifolds. We also remark that a global convergence analysis of a damped Newton's method without assuming non-singularity of the Hessian of the objective function at its critical points was established in the Euclidean context (10, Chapter 1, Section 1.3.3). Since many important problems in the context of a Riemannian manifold become the problem of finding a singularity of a vector field (5; 25), it would be interesting to obtain a similar global analysis for a damped Newton's method to find a singularity of a vector field defined on a Riemannian manifold.

We implemented Newton's method with exponential, i.e., Algorithm 3.0.1, and the damped Newtons method with exponential, i.e., Algorithm 4.0.1, for the problem of finding the global minimizers of an academic family of functions defined on the cone of symmetric positive definite matrices. Our numerical results showed that the damped Newtons method with exponential is superior to the classical Newton's method given by Algorithm 3.0.1. Moreover, we also implemented damped Newtons method with retraction for aforementioned problem. This algorithm present the same iterates numbers and functions evaluate to damped Newtons method with exponential. Thus, in this sense the damped Newton's method with retraction does not present computational advantages over the damped Newtons method with exponential mapping. Theses numerical experiments motivated us to apply Algorithm 5.3.1 (since it is a generalization of Algorithm 4.0.1 for retraction) to the Karcher problem (presented below), which is defined on the cone of symmetric positive definite matrices. Furthermore, we intend to compare the performance of Algorithm 5.3.1 with that of efficient methods for solving this problem in the Euclidean space. In the next section, we present the damped Newton's method with retraction to minimize the Rayleigh quotient.

7.1 Rayleigh Quotient on the Sphere

Computing the eigenvalues of a symmetric $n\times n$ matrix A has several important applications, see for instance, (9; 60). This problem is equivalent to finding a minimizer of the Rayleigh quotient of A on the sphere. We intend to evaluate the performance of Algorithm5.3.1 to solve this problem with efficient methods in the Euclidean space. Toward this end, we begin by establishing Algorithm 4.0.1. Before we present this algorithm, we make some considerations.

Let $\langle\ ,\ \rangle$ be the Euclidean inner product, with the corresponding norm denoted by $\|\cdot\|$. The $n-$dimensional Euclidean sphere and its tangent hyperplane at a point p are respectively denoted by

$$\mathbb{S}^n := \left\{p=(p_1,...,p_{n+1})\in\mathbb{R}^{n+1} :\| p \|= 1\right\}, \qquad T_p\mathbb{S}^n := \left\{v\in\mathbb{R}^{n+1} : \langle p, v\rangle = 0\right\}.$$

Let $\mathbb{M} := (\mathbb{S}^n, \langle\ ,\ \rangle)$. Let X and Y be vector fields in $\mathbb{M}$. Then, using (46, Theorem 1.2, p. 325), we can prove that the Levi-Civita connection of $\mathbb{M}$ is given by

$$\nabla_Y X(p) = \left[I - pp^T\right] X'(p)Y, \tag{7.1}$$

where $X'(p)$ is the usual derivative of X at p and I denotes the $(n+1)\times(n+1)$ identity matrix. Hence, for $f:\mathbb{M}\to\mathbb{R}$, a twice-differentiable function, using the Euclidean inner product and (7.1), the *gradient* and *Hessian* of f at p are, respectively, given by

$$\operatorname{grad} f(p) = \left[I - pp^T\right] f'(p), \qquad \operatorname{Hess} f(p) = [I - pp^T]\,[f''(p) - \langle f'(p), p\rangle I], \tag{7.2}$$

where $f'(p)$ and $f''(p)$ are the Euclidean gradient and Hessian of f at p, respectively. In particular, let A be an $n\times n$ symmetric matrix and $f:\mathbb{S}^n\to\mathbb{R}$ be the Rayleigh quotient function, i.e.,

$$f(p) = p^T Ap.$$

Then, from (2.2), (7.1), and (7.2), Newtons equation for unknown $v\in T_p\mathbb{M}$ is given by

$$\nabla_Y X(p) = \nabla \operatorname{grad} f(p)v = \left[I - pp^T\right][Av - \langle p, Ap\rangle v] = -\left[I - pp^T\right] Ap.$$

Since $v\in T_p\mathbb{M}$, $\left[I - pp^T\right] v = v$. Hence, the Newton equation can be written as follows:

$$\left[I - pp^T\right] A \left[I - pp^T\right] v - vp^T Ap = -\left[I - pp^T\right] Ap.$$

Finally, by using the last equality and the retraction defined in (5.1.2), we can present the proposed algorithm.

Algorithm 7.1.1.

Damped Newton with retraction for Rayleigh Quotient in $\mathbb{M}=\mathbb{S}^n$

Step 0. *Choose a scalar $\sigma\in(0,1/2)$, take an initial point $p_0\in\mathbb{M}$, and set $k=0$.*

Step 1. *Compute search direction* v_k *as a solution of the equation linear system*

$$\begin{cases} \left[I - p_k p_k^T\right] A \left[I - p_k p_k^T\right] v_k - v_k p_k^T A p_k = -\left[I - p_k p_k^T\right] A p_k, \\ p_k^T v_k = 0. \end{cases}$$

If v_k *exists, go to* **Step** 2*. Otherwise, set the search direction as* $v_k = -\operatorname{grad}\varphi(p_k)$*, where* φ *is defined by* (4.2)*, i.e.,*

$$v_k = -4\left([I - p_k p_k^T][A - \langle Ap_k, p_k\rangle I]\right)^T \left[I - p_k p_k^T\right] A p_k.$$

If $v_k = 0$*, stop.*

Step 2. *Compute the step size by the rule*

$$\alpha_k := \max\left\{2^{-j} \ : \left\|\left(I - q_k q_k^T\right) A q_k\right\|^2 \le \left\|\left(I - p_k p_k^T\right) A p_k\right\|^2 + \sigma 2^{2-j}\left\langle\left([I - p_k p_k^T][A - \langle Ap_k, p_k\rangle]\right)^T \left[I - p_k p_k^T\right] A p_k,\ v_k\right\rangle,\ j \in \mathbb{N}\right\},$$

where $q_k = \dfrac{p_k + 2^{-j} v_k}{\|p_k + 2^{-j} v_k\|}$ *and set the next iterate as*

$$p_{k+1} := \frac{p_k + \alpha_k v_k}{\|p_k + \alpha_k v_k\|}.$$

Step 3. *Set* $k \leftarrow k + 1$ *and go to* **Step 1**.

7.2 Karcher Problem

The Karcher problem is an important problem arising when one has to represent, through a single matrix H, the results of several experiments involving a set of many $n \times n$ symmetric positive matrices $A_1, \ldots, A_m$. This problem has several applications; see (12; 55; 60; 69) and the references therein. The geometric structure on $\mathbb{P}^n_{++}$, the cone of symmetric positive definite matrices of order $n \times n$, with the metric given by (6.1), makes $\mathbb{P}^n_{++}$ a complete Riemannian manifold with negative curvature; see (11). Every compact set in $\mathbb{P}^n_{++}$ has a unique center of mass; see (8, Sec. 6.1.5). Representing through a single matrix H the results of several experiments involving a set of many $n \times n$ symmetric positive matrices $A_1, \ldots, A_m$ can be defined as finding their center of mass. This problem is formally defined by

$$\min_{P \in \mathbb{P}^n_{++}} \sum_{i=1}^{m} d^2\left(A_i, P\right), \tag{7.-2}$$

where, for P and Q points in $\mathbb{P}^n_{++}$, $d(P,Q) := \left\|log\left(P^{-1/2} Q P^{-1/2}\right)\right\|_F$ with $\|\cdot\|_F$ denoting the Frobenius norm. The unique solution of (7.2) is referred to as the "least squares geometric mean," or more frequently as the "Karcher mean." Numerical methods have been introduced to solve the problem (7.2), for instance, fixed point iteration; see (55). To the best of our knowledge, thus far, a Riemannian Newton's method has not been adopted to solve (7.2).

8. Reference

[1] P.-A. Absil, L. Amodei, and G. Meyer. Two Newton methods on the manifold of fixed-rank matrices endowed with Riemannian quotient geometries. *Comput. Statist.*, 29(3-4):569–590, 2014.

[2] P.-A. Absil, C. G. Baker, and K. A. Gallivan. Trust-region methods on Riemannian manifolds. *Found. Comput. Math.*, 7(3):303–330, 2007.

[3] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, Princeton, NJ, 2008.

[4] P.-A. Absil and J. Malick. Projection-like retractions on matrix manifolds. *SIAM J. Optim.*, 22(1):135–158, 2012.

[5] R. L. Adler, J.-P. Dedieu, J. Y. Margulies, M. Martens, and M. Shub. Newton's method on Riemannian manifolds and a geometric model for the human spine. *IMA J. Numer. Anal.*, 22(3):359–390, 2002.

[6] I. K. Argyros and S. Hilout. Newton's method for approximating zeros of vector fields on Riemannian manifolds. *J. Appl. Math. Comput.*, 29(1-2):417–427, 2009.

[7] G. C. Bento, O. P. Ferreira, and J. G. Melo. Iteration-complexity of gradient, subgradient and proximal point methods on Riemannian manifolds. *J. Optim. Theory Appl.*, 173(2):548–562, 2017.

[8] M. Berger. *A panoramic view of Riemannian geometry*. Springer-Verlag, Berlin, 2003.

[9] M. W. Berry, Z. Drmaˇ c, and E. R. Jessup. Matrices, vector spaces, and information retrieval. *SIAM Rev.*, 41(2):335–362, 1999.

[10] D. P. Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Computer Science and Applied Mathematics. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 2014.

[11] R. Bhatia. *Positive definite matrices*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2007.

[12] D. A. Bini and B. Iannazzo. Computing the Karcher mean of symmetric positive definite matrices. *Linear Algebra Appl.*, 438(4):1700–1710, 2013.

[13] T. Bittencourt and O. Ferreira. Kantorovich?s theorem on newton?s method under majorant condition in riemannian manifolds. *Journal of Global Optimization*, pages 1–25, 2016.

[14] T. Bittencourt and O. P. Ferreira. Kantorovich's theorem on Newton's method under majorant condition in Riemannian manifolds. *J. Global Optim.*, 68(2):387–411, 2017.

[15] P. B. Borckmans, S. Easter Selvan, N. Boumal, and P.-A. Absil. A Riemannian subgradient algorithm for economic dispatch with valve-point effect. *J. Comput. Appl. Math.*, 255:848–866, 2014.

[16] M. Bortoloti, T. Fernandes, O. Ferreira, and J. Yuan. Damped newton's method on riemannian manifolds. *arXiv preprint arXiv:1803.05126*, 2018.

[17] O. Burdakov. Some globally convergent modifications of newton's method for solving systems of nonlinear equations. In *Soviet mathematics-Doklady*, volume 22, pages 376–378, 1980.

[18] L. Cambier and P.-A. Absil. Robust low-rank matrix completion by Riemannian optimization. *SIAM J. Sci. Comput.*, 38(5):S440–S460, 2016.

[19] B. N. Datta. *Numerical methods for linear control systems: design and analysis*, volume 1. Academic Press, 2004.

[20] J.-P. Dedieu, P. Priouret, and G. Malajovich. Newton's method on Riemannian manifolds: convariant alpha theory. *IMA J. Numer. Anal.*, 23(3):395–419, 2003.

[21] J. E. Dennis, Jr. and R. B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*, volume 16 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996. Corrected reprint of the 1983 original.

[22] P. Deuflhard. *Newton methods for nonlinear problems*, volume 35 of *Springer Series in Computational Mathematics*. Springer, Heidelberg, 2011. Affine invariance and adaptive algorithms, First softcover printing of the 2006 corrected printing.

[23] J. Dieudonné. *Foundations of modern analysis*. Academic Press, New York-London, 1969. Enlarged and corrected printing, Pure and Applied Mathematics, Vol. 10-I.

[24] M. P. do Carmo. *Riemannian geometry*. Mathematics: Theory & Applications. Birkhäuser Boston, Inc., Boston, MA, 1992. Translated from the second Portuguese edition by Francis Flaherty.

[25] A. Edelman, T. A. Arias, and S. T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.*, 20(2):303–353, 1999.

[26] F. Facchinei and C. Kanzow. A nonsmooth inexact Newton method for the solution of large-scale nonlinear complementarity problems. *Math. Programming*, 76(3, Ser. B):493–512, 1997.

[27] T. A. Fernandes, O. P. Ferreira, and J. Yuan. On the Superlinear Convergence of Newton's Method on Riemannian Manifolds. *J. Optim. Theory Appl.*, 173(3):828–843, 2017.

[28] O. P. Ferreira and R. C. M. Silva. Local convergence of Newton's method under a majorant condition in Riemannian manifolds. *IMA J. Numer. Anal.*, 32(4):1696–1713, 2012.

[29] O. P. Ferreira and B. F. Svaiter. Kantorovich's theorem on Newton's method in Riemannian manifolds. *J. Complexity*, 18(1):304–329, 2002.

[30] D. Gabay. Minimizing a differentiable function over a differential manifold. *J. Optim. Theory Appl.*, 37(2):177–219, 1982.

[31] P. Grohs and S. Hosseini. Nonsmooth trust region algorithms for locally Lipschitz functions on Riemannian manifolds. *IMA J. Numer. Anal.*, 36(3):1167–1192, 2016.

[32] P. Grohs and S. Hosseini. ε-subgradient algorithms for locally lipschitz functions on riemannian manifolds. *Adv. Comput. Math.*, 42(2):333–360, 2016.

[33] S. Hosseini and A. Uschmajew. A Riemannian gradient sampling algorithm for nonsmooth optimization on manifolds. *SIAM J. Optim.*, 27(1):173–189, 2017.

[34] Y. Hu, C. Li, and X. Yang. On convergence rates of linearized proximal algorithms for convex composite optimization with applications. *SIAM J. Optim.*, 26(2):1207–1235, 2016.

[35] W. Huang, P.-A. Absil, and K. A. Gallivan. A Riemannian symmetric rank-one trust-region method. *Math. Program.*, 150(2, Ser. A):179–216, 2015.

[36] W. Huang, P.-A. Absil, and K. A. Gallivan. A riemannian bfgs method for nonconvex optimization problems. In *Numerical Mathematics and Advanced Applications ENUMATH 2015*, pages 627–634. Springer, 2016.

[37] W. Huang, P.-A. Absil, and K. A. Gallivan. Intrinsic representation of tangent vectors and vector transports on matrix manifolds. *Numer. Math.*, 136(2):523–543, 2017.

[38] W. Huang, K. A. Gallivan, and P.-A. Absil. A Broyden class of quasi-Newton methods for Riemannian optimization. *SIAM J. Optim.*, 25(3):1660–1685, 2015.

[39] W. Huang, K. A. Gallivan, and P.-A. Absil. A Broyden class of quasi-Newton methods for Riemannian optimization. *SIAM J. Optim.*, 25(3):1660–1685, 2015.

[40] W. Huang, K. A. Gallivan, A. Srivastava, and P.-A. Absil. Riemannian optimization for registration of curves in elastic shape analysis. *J. Math. Imaging Vision*, 54(3):320–343, 2016.

[41] W. Huang, K. A. Gallivan, and X. Zhang. Solving PhaseLift by low-rank Riemannian optimization methods for complex semidefinite constraints. *SIAM J. Sci. Comput.*, 39(5):B840–B859, 2017.

[42] A. F. Izmailov and M. V. Solodov. *Newton-type methods for optimization and variational problems*. Springer Series in Operations Research and Financial Engineering. Springer, Cham, 2014.

[43] A. Kovnatsky, K. Glashoff, and M. M. Bronstein. Madmm: A generic algorithm for non-smooth optimization on manifolds. *arXiv preprint arXiv:1505.07676*, 2015.

[44] D. Kressner, M. Steinlechner, and B. Vandereycken. Low-rank tensor completion by riemannian optimization. *BIT Numerical Mathematics*, 54(2):447–468, Jun 2014.

[45] A. Kristály, V. t. D. R˘ adulescu, and C. G. Varga. *Variational principles in mathematical physics, geometry, and economics*, volume 136 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2010. Qualitative analysis of nonlinear equations and unilateral problems, With a foreword by Jean Mawhin.

[46] S. Lang. *Fundamentals of differential geometry*, volume 191 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1999.

[47] C. Li, G. López, and V. Martín-Márquez. Monotone vector fields and the proximal point algorithm on Hadamard manifolds. *J. Lond. Math. Soc. (2)*, 79(3):663–683, 2009.

[48] C. Li and J. Wang. Newton's method on Riemannian manifolds: Smale's point estimate theory under the γ-condition. *IMA J. Numer. Anal.*, 26(2):228–251, 2006.

[49] C. Li and J. Wang. Newton's method for sections on Riemannian manifolds: generalized covariant α-theory. *J. Complexity*, 24(3):423–451, 2008.

[50] C. Li, J.-H. Wang, and J.-P. Dedieu. Smale's point estimate theory for Newton's method on Lie groups. *J. Complexity*, 25(2):128–151, 2009.

[51] D. G. Luenberger. The gradient projection method along geodesics. *Management Sci.*, 18:620–631, 1972.

[52] J. H. Manton. Optimization algorithms exploiting unitary constraints. *IEEE Transactions on Signal Processing*, 50(3):635–650, 2002.

[53] J. H. Manton. A framework for generalising the Newton method and other iterative methods from Euclidean space to manifolds. *Numer. Math.*, 129(1):91–125, 2015.

[54] J. Milnor. *Morse theory*. Based on lecture notes by M. Spivak and R. Wells. Annals of Mathematics Studies, No. 51. Princeton University Press, Princeton, N.J., 1963.

[55] M. Moakher. On the averaging of symmetric positive-definite tensors. *Journal of Elasticity*, 82(3):273–296, 2006. Cited By :83.

[56] J. Moser. A new technique for the construction of solutions of nonlinear differential equations. *Proc. Nat. Acad. Sci. U.S.A.*, 47:1824–1831, 1961.

[57] J. Nash. The imbedding problem for Riemannian manifolds. *Ann. of Math. (2)*, 63:20–63, 1956.

[58] J. M. Ortega. *Numerical analysis. A second course.* Academic Press, New York-London, 1972. Computer Science and Applied Mathematics.

[59] J. M. Ortega and W. C. Rheinboldt. *Iterative solution of nonlinear equations in several variables*, volume 30 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000. Reprint of the 1970 original.

[60] Y. Rathi, A. Tannenbaum, and O. Michailovich. Segmenting images on the tensor manifold. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.

[61] W. Ring and B. Wirth. Optimization methods on Riemannian manifolds and their application to shape space. *SIAM J. Optim.*, 22(2):596–627, 2012.

[62] O. S. Rothaus. Domains of positivity. *Abh. Math. Sem. Univ. Hamburg*, 24:189–235, 1960.

[63] T. Sakai. *Riemannian geometry*, volume 149 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1996. Translated from the 1992 Japanese original by the author.

[64] H. Sato and T. Iwai. A Riemannian optimization approach to the matrix singular value decomposition. *SIAM J. Optim.*, 23(1):188–212, 2013.

[65] V. H. Schulz. A Riemannian view on shape optimization. *Found. Comput. Math.*, 14(3):483–501, 2014.

[66] U. Shalit, D. Weinshall, and G. Chechik. Online learning in the embedded manifold of low-rank matrices. *Journal of Machine Learning Research*, 13:429–458, 2012.

[67] S. T. Smith. *Geometric optimization methods for adaptive filtering.* ProQuest LLC, Ann Arbor, MI, 1993. Thesis (Ph.D.)–Harvard University.

[68] S. T. Smith. Optimization techniques on Riemannian manifolds. In *Hamiltonian and gradient flows, algorithms and control*, volume 3 of *Fields Inst. Commun.*, pages 113–136. Amer. Math. Soc., Providence, RI, 1994.

[69] S. Sra and R. Hosseini. Conic geometric optimization on the manifold of positive definite matrices. *SIAM J. Optim.*, 25(1):713–739, 2015.

[70] C. Udriste. *Convex functions and optimization methods on Riemannian manifolds*, volume 297 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1994.

[71] B. Vandereycken. Low-rank matrix completion by riemannian optimization. *SIAM Journal on Optimization*, 23(2):1214–1236, 2013.

[72] J.-H. Wang and C. Li. Kantorovich's theorems for Newton's method for mappings and optimization problems on Lie groups. *IMA J. Numer. Anal.*, 31(1):322–347, 2011.

[73] X. Yuan, W. Huang, P.-A. Absil, and K. A. Gallivan. A riemannian quasi-newton method for computing the karcher mean of symmetric positive definite matrices. Technical Report FSU17-02, Florida State University, 2017.

[74] Z. Zhao, Z.-J. Bai, and X.-Q. Jin. A Riemannian Newton algorithm for nonlinear eigenvalue problems. *SIAM J. Matrix Anal. Appl.*, 36(2):752–774, 2015.

[75] X. Zhu. A Riemannian conjugate gradient method for optimization on the Stiefel manifold. *Comput. Optim. Appl.*, 67(1):73–110, 2017.