Universidade Federal de Goiás
IME - Instituto de Matemática e Estatística
Programa de Pós Graduação em Matemática

Maurício Silva Louzeiro

# Optimization methods on Riemannian manifolds with lower bound curvature: gradient for scalar and multi-objective functions and subgradient for scalar functions

Doctoral Thesis
Funded by CAPES

Goiânia
Fevereiro de 2019

**TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG**

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou *download*, a título de divulgação da produção científica brasileira, a partir desta data.

**1. Identificação do material bibliográfico:** **[ ] Dissertação** **[ X ] Tese**

**2. Identificação da Tese ou Dissertação:**

Nome completo do autor: Maurício Silva Louzeiro

Título do trabalho: Optimization methods on Riemannian manifolds with lower bound curvature: gradient for scalar and multi-objective functions and subgradient for scalar functions

**3. Informações de acesso ao documento:**

Concorda com a liberação total do documento [ X ] SIM [ ] NÃO[1]

Havendo concordância com a disponibilização eletrônica, torna-se imprescindível o envio do(s) arquivo(s) em formato digital PDF da tese ou dissertação.

Maurício Silva Louzeiro

Assinatura do(a) autor(a)[2]

Ciente e de acordo:

Assinatura do(a) orientador(a)[2]

Data: 07 / 03 / 2019

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. A extensão deste prazo suscita justificativa junto à coordenação do curso. Os dados do documento não serão disponibilizados durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

[2] A assinatura deve ser escaneada.

Maurício Silva Louzeiro

# Optimization methods on Riemannian manifolds with lower bound curvature: gradient for scalar and multi-objective functions and subgradient for scalar functions

Tese apresentada ao Programa de Pós-Graduação do Instituto de Matemática e Estatística da Universidade Federal de Goiás, como requisito parcial para obtenção do título de Doutor em Matemática.

**Área de concentração**: Otimização
**Orientador**: Prof. Dr. Orizon Pereira Ferreira
**Coorientador**: Prof. Dr. Leandro da Fonseca Prudente

Goiânia
Fevereiro de 2019

Ficha de identificação da obra elaborada pelo autor, através do
Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Louzeiro, Maurício
Optimization methods on Riemannian manifolds with lower bound curvature: gradient for scalar and multi-objective functions and subgradient for scalar functions [manuscrito] / Maurício Louzeiro. - 2019.
83 f.: il.

Orientador: Prof. Dr. Orizon Ferreira; co-orientador Dr. Leandro Prudente.
Tese (Doutorado) - Universidade Federal de Goiás, Instituto de Matemática e Estatística (IME), Programa de Pós-Graduação em Matemática, Goiânia, 2019.
Bibliografia.

1. Métodos de otimização. 2. programação convexa. 3. variedade Riemanniana. 4. curvatura limitada inferiormente. 5. complexidade. I. Ferreira, Orizon, orient. II. Título.

CDU 51

UFG
UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA
Campus Samambaia – Caixa Postal 131 – CEP: 74.001-970 – Goiânia-GO.
Fones: (62) 3521-1208 e 3521-1137 www.ime.ufg.br

**ATA DA REUNIÃO DA BANCA EXAMINADORA DA DEFESA DE TESE DE MAURÍCIO SILVA LOUZEIRO** – Ao vigésimo sexto dia do mês de fevereiro do ano de dois mil e dezenove (26/02/2019), às 11:00 horas, reuniram-se os componentes da Banca Examinadora: Prof. Orizon Pereira Ferreira - Orientador, Prof. Glaydston de Carvalho Bento, Prof. João Xavier da Cruz Neto, Prof. Paulo Sérgio Marques dos Santos e Luis Roman Lucambio Perez, sob a presidência do primeiro, e em sessão pública realizada no auditório do Instituto de Matemática e Estatística, procederem a avaliação da defesa de tese intitulada: **"Optimization methods on Riemannian manifolds with lower bound curvature: Gradient for scalar and multi-objective functions and Subgradient for scalar functions"**, em nível de Doutorado, área de concentração em Otimização, de autoria de Maurício Silva Louzeiro, discente do Programa de Pós-Graduação em Matemática da Universidade Federal de Goiás. A sessão foi aberta pelo Presidente da Banca, Prof. Orizon Pereira Ferreira que fez a apresentação formal dos membros da Banca. A seguir, a palavra foi concedida ao autor da tese que, em 45 minutos procedeu a apresentação de seu trabalho. Terminada a apresentação, cada membro da Banca arguiu o examinando, tendo-se adotado o sistema de diálogo sequencial. Terminada a fase de arguição, procedeu-se a avaliação da defesa. Tendo-se em vista o que consta na Resolução nº. 1513 do Conselho de Ensino, Pesquisa, Extensão e Cultura (CEPEC), que regulamenta o Programa de Pós-Graduação em Matemática e procedidas às correções recomendadas, a tese foi **APROVADA** por unanimidade, considerando-se integralmente cumprido este requisito para fins de obtenção do título de **DOUTOR EM MATEMÁTICA**, na área de concentração em Otimização pela Universidade Federal de Goiás. A conclusão do curso dar-se-á quando da entrega na secretaria do PPGM da versão definitiva da tese, com as devidas correções supervisionadas e aprovadas pelo orientador. Cumpridas as formalidades de pauta, às 12:00 horas a presidência da mesa encerrou esta sessão de defesa de tese e para constar eu, Ana Maria Pereira Pinto, secretária do PPGM, lavrei a presente Ata que, depois de lida e aprovada, será assinada pelos membros da Banca Examinadora em quatro vias de igual teor.

Prof. Dr. Orizon Pereira Ferreira
Presidente - IME/UFG

Prof. Dr. Glaydston de Carvalho Bento
Membro – IME/UFG

Prof. Dr. João Xavier da Cruz Neto
Membro – DM/UFPI

Prof. Dr. Paulo Sérgio Marques dos Santos
Membro – DM/UFPI

Prof. Dr. Luis Roman Lucambio Perez
Membro – IME/UFG

MAURÍCIO SILVA LOUZEIRO

# Optimization methods on Riemannian manifolds with lower bound curvature: Gradient for scalar and multi-objective functions and Subgradient for scalar functions

Tese defendida no Programa de Pós–Graduação do Instituto de Matemática e Estatística da Universidade Federal de Goiás como requisito parcial para obtenção do título de Doutor em Matemática, aprovada em 26 de fevereiro de 2019, pela Banca Examinadora constituída pelos professores:

**Prof. Dr. Orizon Pereira Ferreira**
Instituto de Matemática e Estatística - UFG
Presidente da Banca

**Prof. Dr. Glaydston de Carvalho Bento**
Instituto de Matemática e Estatística - UFG

**Prof. Dr. Luis Roman Lucambio Perez**
Instituto de Matemática e Estatística - UFG

**Prof. Dr. João Xavier da Cruz Neto**
Departamento de Matemática – UFPI

**Prof. Dr. Paulo Sérgio Marques dos Santos**
Departamento de Matemática - UFPI

**Dedicado a:**

minha família.

# Agradecimentos

A Deus pelos momentos que me fizeram acreditar que eu estava no caminho certo, e pela força que obtive ao superar alguns momentos difíceis durante o meu doutorado.

Aos meus pais Ivonete e Nelson pelos conselhos, pela atenção, pelo apoio incondicional, pelo carinho e por estarem sempre dispostos a oferecerem o melhor para eu e meus irmãos, colocando nossos sonhos à frente dos seus.

Ao meu orientador Prof. Dr. Orizon Pereira Ferreira pela paciência, pelo interesse em entender as minhas ideias, por estar sempre disposto a discutir problemas matemáticos e a sanar dúvidas, pelos conselhos, pelo esforço feito para que eu participasse de congressos importantes da nossa área (por exemplo, o ELAVIO2018 no Chile e o BRAZOP2018 em Foz do Iguaçu) e pela imensa contribuição dada a esta tese.

Ao meu coorientador Prof. Dr. Leandro da Fonseca Prudente por ter aceitado me coorientar e pela grande contribuição dada nos artigos que compõe esta tese, principalmente nos experimentos numéricos.

Aos professores: Dr. Glaydston de Carvalho Bento, Dr. João Xavier da Cruz Neto, Dr. Luis Román Lucambio Pérez e Dr. Paulo Sérgio Marques dos Santos, por aceitarem o convite para participar da banca examinadora, e pelas sugestões dadas.

Aos amigos e colegas pelas descontrações e pelas boas discussões sobre matemática.

A CAPES pelo auxílio financeiro.

# Abstract

Let $\mathcal{M}$ a Riemannian manifolds with lower bounded curvature. In this thesis, we consider first-order iterative methods to solve optimization problems on $\mathcal{M}$. The gradient method to solve the problem $\min\{f(p) : p \in \mathcal{M}\}$, where $f : \mathcal{M} \to \mathbb{R}$ is a continuously differentiable convex function is presented with Lipschitz step-size, adaptive step-size and Armijo's step-size. The first procedure requires that the objective function has Lipschitz continuous gradient, which is not necessary for the other approaches. Convergence of the whole sequence to a minimizer, without any level set boundedness assumption, is proved. Iteration-complexity bound for functions with Lipschitz continuous gradient is also presented. In addition, all these approaches are considered in the multiobjective setting. Here we also consider the subgradient method to solve the problem $\min\{f(p) : p \in \mathcal{M}\}$, where $f : \mathcal{M} \to \overline{\mathbb{R}}$ is a convex function. Iteration-complexity bounds of the subgradient method with exogenous step-size and Polyak's step size are stablished, completing and improving recent results on the subject. Finally, some examples and numerical experiments are presented.

***Keywords:*** Optimization methods, convex programming, Riemannian manifold, lower bounded curvature, complexity.

# Resumo

Seja $\mathcal{M}$ uma variedade Riemanniana com curvatura limitada inferiormente. Nesta tese, consideramos métodos iterativos de primeira ordem para resolver problemas de otimização sobre variedades Riemannianas com curvatura limitada inferiormente. O método do gradiente para resolver o problema $\min\{f(p) : p \in \mathcal{M}\}$, onde $f : \mathcal{M} \to \mathbb{R}$ é uma função convexa continuamente diferenciável, é apresentado com tamanho de passo Lipshitz, tamanho de passo adaptativo e tamanho de passo de Armijo. O primeiro tipo de passo requer que a função objetivo tenha gradiente continuamente Lipshitz, o que não é necessário para os outros. A convergência total da sequência para um minimizador, sem qualquer hipótese de limitação do conjunto de nível, é provada. Limitantes para a complexidade na iteração para funções com gradiente continuamente Lipschitz também são apresentados. Além disso, todas essas abordagens são consideradas no contexto de otimização multiobjetivo. Aqui também consideramos o método do subgradiente para resolver o problema $\min\{f(p) : p \in \mathcal{M}\}$, onde $f : \mathcal{M} \to \overline{\mathbb{R}}$ é uma função convexa. Limitantes para a complexidade na iteração do método do subgradiente com tamanho de passo exógeno e tamanho de passo de Polyak são estabelecidos, completando e melhorando os resultados recentes sobre o assunto. Finalmente, alguns exemplos e experimentos numéricos são apresentados.

***Palavras-chave :*** Métodos de otimização, programação convexa, variedade Riemmaniana, curvatura limitada inferiormente, complexidade.

# Contents

# Chapter 1

# Introduction

Endowing a manifold $\mathcal{M}$ with a suitable Riemannian metric, an Euclidean non-convex constrained problem can be seen as a Riemannian convex unconstrained problem. This property is already well known and, can be exploited in order to find global minimizers [14, 15, 25, 61] and to reduce the iteration-complexity of finding such solutions [12, 75]. In this work we will present some examples showing that endowing the set of constraints with a suitable Riemannian metric the objective function can be also *Riemannian Lipschitz gradient*. Consequently, the geometric and algebraic structures that come from the Riemannian metric make possible to greatly reduce the computational cost for solving such problems. Indeed, it is also widely known that, in several contexts, the iteration complexity of the gradient method for convex optimization problems with Lipschitz gradient is much lower than for general nonconvex problems; see for example [12, 44, 61, 66, 75] and references therein. Furthermore, many Euclidean optimization problems are naturally posed on the Riemannian context; see [29, 44, 65, 66]. Then, to take advantage of the Riemannian geometric structure, it is preferable to treat these problems as the ones of finding singularities of gradient vector fields on Riemannian manifolds rather than using Lagrange multipliers or projection methods; see [51, 65, 67]. The Riemannian structures can also opens up new research directions that aid in developing competitive algorithms; see [1,29,44,58,65,66]. For this purpose, extensions of concepts and techniques of optimization from Euclidean space to Riemannian context have been quite frequently in recent years. Papers dealing with this subject include, but are not limited to [31, 47, 48, 52, 67–69, 71, 75, 76].

In [15] the subgradient method was introduced to solve convex feasibility problems on complete Riemannian manifolds with non-negative sectional curvatures. In this work, the authors propose the question: how to solve the feasibility problem on a Riemannian manifold with negative sectional curvature? Recently, this question was answered in [69,71], where the convex feasibility problem is analyzed in manifolds with lower bounded sectional curvatures. The extension contained in these works is based on auxiliary results obtained from the Toponogov theorem and algebraic manipulations. Some of these auxiliary results contributed greatly to this thesis.

In the chapter 3, we consider the gradient method to minimize a *continuously differentiable convex function* $f : \mathcal{M} \to \mathbb{R}$, whereas $\mathcal{M}$ is endowed with a structure of *complete Riemannian manifold with lower bounded curvature*. The gradient method is one of the oldest methods for the minimization of a differentiable function in Euclidean space. Despite having slow convergence rate, the simplicity of implementation, the low memory requirements and cost per iteration, make the gradient method quite attractive to solve large-scale optimization problems. Indeed, the computational cost per iteration is mildly dependent on the dimension of the problem, yielding computational efficiency for this method; see [44, 57, 62]. In addition, the gradient method is the starting point for designing many more sophisticated and efficient algorithms, including fast gradient method, accelerated gradient method and Barzilai-Borwein method; see [56, 74] for a comprehensive study on this subject. To the best of our knowledge the gradient method was the first optimization method to be considered in a Riemannian setting. In order to deal with constrained optimization problems in the Euclidean space, Luenberger [51] proposed and established important convergence properties of gradient method by using the Riemannian structure of the constraint set induced by the Euclidean structure. Since then, the gradient method has been studied in general Riemannian manifold. Some early works dealing with this method include [40, 61, 65, 67]. However, the obtained convergence results in these previous works demand that the initial points of the sequence belong to a bounded level set of the objective function establishing only that all its cluster points are stationary. By assuming convexity of the objective function and that the manifold has *non-negative curvature*, it has been proven in [23] that, for a suitable choice of the step-size and without any level set boundedness assumption, the whole sequence converges to a solution. Recently new important properties of the gradient method in Riemannian settings have been obtained. For instance, in [76] the authors provided iteration-complexity bounds for convex optimization problems on Hadamard manifolds. In [18], the authors established iteration-complexity bounds without any assumption on the convexity of the problem and curvature of the manifold. In [17] the gradient method is considered to compute the Karcher mean, which is a strong convex function in the cone of symmetric positive definite matrices endowed with a suitable Riemannian metric. In [2] is studied properties of the gradient method for the problem of finding the global Riemannian center of mass of a set of data points on a Riemannian manifold. In [8] is extended the convergence analysis of the gradient method to the Hadamard setting for continuously differentiable functions which satisfy the Kurdyka-Lojasiewicz inequality.

By the aforementioned we see that the gradient method remains a subject of considerable interest. The full convergence of the sequence generated by the gradient method under convexity of the objective function and *lower boundedness of the curvature* of the Riemannian manifold is a new contribution of this work, which adds important results in the available convergence theory of this method. The analysis of the method is presented with three different finite procedures for determining the step-size, namely, Lipschitz step-size, adaptive step-size and Armijo's step-size. It should be noted that we use a recent inequality established in [69, 71]. Numerical experiments are provided to illustrate the effectiveness of the method in this new setting and certify the obtained theoretical results. In particular, we consider the problem of finding the Riemannian mass center and the so-called Karcher mean. Our experiments indicate that adaptive size is a promising scheme that is worth considering.

In chapter 4 of this thesis, we study the steepest descent method for multiobjective optimization on Riemannian manifolds. A constrained multiobjective optimization problem with constraint set $\mathcal{M}$, consists of $m$ objective functions $f_1, \ldots, f_m$, that have to be optimized at the same time on $\mathcal{M}$. In recent years, there has been a significant increase in the number of papers addressing this class of problems; for example, see [10, 20, 36, 50, 53, 54]. This method, was proposed in [35] and since of then several variants have been considered, including but not limited to [6, 7, 28, 38, 39, 42]. Recently some iteration-complexity results to gradient method for unconstrained multi-objective optimization problem were presented in [37]. These results have been shown to be the same global rates as for steepest descent method in scalar objective optimization.

The aim of the chapter 4 is twofold. First, asymptotic analysis will be done for quasi-convex and convex vectorial functions. In fact, in [13] asymptotic analysis of this method has already been done in Riemannian context; see also [11]. However, the analysis asymptotic presented in these previous works is just to step-size given by Armijo rule and it demand that the Riemannian manifolds have nonnegative sectional curvature. The asymptotic analysis presented in the present chapter increase the previous ones in two different aspects. *It is provided an analysis with three different finite procedures for determining the step-size, namely, Lipschitz step-size, adaptive step-size and Armijo-type step-size* and only *lower boundedness of the curvature* of the Riemannian manifold is assumed. The second aim is to present *iteration-complexity bounds for steepest descent method for multiobjective optimization on Riemannian manifolds*. It is worth noting that, our results generalize to the Riemannian context the results obtained in [37]. Besides, we present one iteration-complexity bound that is new even in Euclidean setting. In addition, some examples are presented to emphasize the importance of working in this new context.

In chapter 5 of this thesis, we consider the subgradient method to solve the optimization problem $\min\{f(p) \ : \ p \in \mathcal{M}\}$, where the constraint set $\mathcal{M}$ is endowed with a structure of a *complete Riemannian manifold with lower bounded curvature* and $f : \mathcal{M} \to \overline{\mathbb{R}}$ is a *convex function*, where $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ denotes the extended real set numbers, see [32]. In this chapter we establish an iteration-complexity bound of the subgradient method with exogenous step-size and Polyak's step-size, for convex optimization problems on complete Riemannian manifolds with lower bounded sectional curvatures. Our results increase the range of applicability of the method compared to the respective results obtained in [12, 75, 76]. Moreover, in the asymptotic analysis with exogenous step-size, we do not assume that the solution set is nonempty, completing the result of [70, Theorem 3.1]. It should be noted that our analysis use a recently inequality obtained in [69, 71].

The subgradient method is a very simple algorithm for solving convex optimization problems and, besides it is the departure point for many other more sophisticated and efficient algorithms, including $\epsilon$-subgradient methods, bundle methods and cutting-plane algorithm; see [22] for a comprehensive study on this subject. The subgradient method was originally developed by Shor and others in the 1960s and 1970s and since of then, it and its variants have been applied to a far wider variety of problems in optimization theory; see [41,60]. In order to deal with non-smooth convex optimization problems on complete Riemanian manifolds with non-negative sectional curvature, [33] extended and analyzed the subgradient method which, as in the Euclidean context, is quite simple and possess nice convergence properties. After this pioneering work, the subgradient method in the Riemannian setting has been studied in different contexts; see, for instance, [9, 15, 43, 69, 71]. Recently, an asymptotic analysis of the subgradient method with exogenous step-size and dynamic step-size for convex optimization was considered in the context of manifolds with lower bounded sectional curvatures, see [70].

# Chapter 2

# Preliminaries

In this chapter, we recall some concepts, notations, and basics results about Riemannian manifolds. For more details we refer the reader to [26, 64, 67].

We denote by $T_p\mathcal{M}$ the *tangent space* of a finite dimensional Riemannian manifold $\mathcal{M}$ at $p$. The corresponding norm associated to the Riemannian metric $\langle \cdot, \cdot \rangle$ is denoted by $\| \cdot \|$. We use $\ell(\alpha)$ to denote the length of a piecewise smooth curve $\alpha : [a, b] \to \mathcal{M}$. The Riemannian distance between $p$ and $q$ in $\mathcal{M}$ is denoted by $d(p, q)$, which induces the original topology on $\mathcal{M}$. Denote by $\mathcal{X}(\mathcal{M})$, the space of smooth vector fields on $\mathcal{M}$. Let $\nabla$ be the Levi-Civita connection associated to $(\mathcal{M}, \langle \cdot, \cdot \rangle)$. For each $t \in [a, b]$ and a piecewise smooth curve $\alpha : [a, b] \to \mathcal{M}$, $\nabla$ induces an isometry relative to $\langle \cdot, \cdot \rangle$, $P_{\alpha,a,t} \colon T_{\alpha(a)}\mathcal{M} \to T_{\alpha(t)}\mathcal{M}$ defined by $P_{\alpha,a,t}\, v = V(t)$, where $V$ is the unique vector field on the curve $\alpha$ such that $\nabla_{\alpha'(t)}V(t) = 0$ and $V(a) = v$. The isometry $P_{\alpha,a,t}$ is called *parallel transport* along of $\alpha$ joining $\alpha(a)$ to $\alpha(t)$ and, when there is no confusion, it will be denoted by $P_{\alpha,p,q}$. A vector field $V$ along a smooth curve $\gamma$ is said to be *parallel* iff $\nabla_{\gamma'}V = 0$. If $\gamma'$ itself is parallel, we say that $\gamma$ is a *geodesic*. The restriction of a geodesic to a closed bounded interval is called a *geodesic segment*.

**Definition 2.0.1** *A geodesic segment joining $p$ to $q$ in $\mathcal{M}$ is said to be minimal if its length is equal to $d(p, q)$.*

A Riemannian manifold is *complete* if the geodesics are defined for any values of $t \in \mathbb{R}$. Hopf-Rinow's theorem asserts that any pair of points in a complete Riemannian manifold $\mathcal{M}$ can be joined by a (not necessarily unique) minimal geodesic segment. Owing to the completeness of the Riemannian manifold $\mathcal{M}$, for each $p \in \mathcal{M}$, the *exponential map* $\exp_p : T_p\mathcal{M} \to \mathcal{M}$ is given by $\exp_p v = \gamma(1)$, where $\gamma(0) = p$ and $\gamma'(0) = v$. *In this thesis, all manifolds are assumed to be Riemannian connected, finite dimensional, and complete.* For $f : \mathcal{M} \to \mathbb{R}$ a differentiable function on the open set $\mathcal{D} \subset \mathcal{M}$, the Riemannian metric induces the mapping $f \mapsto \operatorname{grad} f$ which associates its *gradient* via the following rule

$$\langle \operatorname{grad} f(p), V(p) \rangle := df(p)V(p), \qquad \forall\; p \in \mathcal{D},\; V \in \mathcal{X}(\mathcal{D}).$$

For a twice-differentiable function, the mapping $f \mapsto \text{hess} f$ associates its *hessian* via the rule $\langle \text{hess} f\, V, V\rangle := d^2 f(V, V)$, for all $V \in \mathcal{X}(\mathcal{D})$, where the last equalities imply that

$$\text{hess} f\, V = \nabla_V \operatorname{grad} f, \qquad \forall\, V \in \mathcal{X}(\mathcal{D}).$$

Consider two geodesic segments $\gamma_1, \gamma_2 : [0, +\infty) \to \mathcal{M}$ with $\gamma_1(0) = \gamma_2(0) = p$. Denote by $\angle_p(\gamma_1, \gamma_2)$ the angle between $\gamma_1$ and $\gamma_2$ at $p$, which is defined to be the angle between the tangent vectors $\gamma_1'(0)$ and $\gamma_2'(0)$. The geodesic triangle $\Delta(p_1p_2p_3)$ in $\mathcal{M}$ is a figure consisting of three points $p_1, p_2, p_3$ (the vertices of $\Delta(p_1p_2p_3)$) and three geodesic segments $\gamma_i$ (the edges of $\Delta(p_1p_2p_3)$) that join $p_{i-1}$ to $p_{i+1}$ with $i = 1, 2, 3(\mod 3)$. For each $i = 1, 2, 3(\mod 3)$, the inner angle of $\Delta(p_{i-1}p_ip_{i+1})$ at $p_i$ is denoted by $\angle(p_{i-1}p_ip_{i+1})$, which equals $\angle_{p_i}(-\gamma_{i-1}, \gamma_{i+1})$. Let $\mathcal{M}^m_\kappa$ be an m-dimensional complete simply connected Riemannian manifolds of constant curvature $\kappa$. The following proposition is known in [64, p.138]. *From now on we will assume*

$$\kappa < 0, \qquad \hat{\kappa} := \sqrt{|\kappa|}. \tag{2.1}$$

**Proposition 2.0.2** *Let $\Delta(p_1p_2p_3)$ be a geodesic triangle in $\mathcal{M}^2_\kappa$. Then, the following "law of cosines" holds*

$$\cosh(\hat{\kappa}\ell_2) = \cosh(\hat{\kappa}\ell_1)\cosh(\hat{\kappa}\ell_3) - \sinh(\hat{\kappa}\ell_1)\sinh(\hat{\kappa}\ell_3)\cos\angle(p_1p_2p_3),$$

*where $\ell_i = d(p_{i-1}, p_{i+1})$ for each $i = 1, 2, 3(\mod 3)$.*

Following [64, p.161], a generalized geodesic hinge $\Lambda(p; \gamma, \beta)$ in $\mathcal{M}$ is a figure consisting of a point $p \in \mathcal{M}$ (the vertex of the hinge) and two geodesic segments $\gamma$, $\beta$ (the edges of the hinge) emanating from $p$ with one being minimal. Moreover, a hinge $\Lambda(\overline{p}; \overline{\gamma}, \overline{\beta})$ in $\mathcal{M}^2_\kappa$ is called a comparison hinge of $\Lambda(p; \gamma, \beta)$ if it satisfies

$$\ell(\overline{\gamma}) = \ell(\gamma), \quad \ell(\overline{\beta}) = \ell(\beta) \quad \text{and} \quad \angle_{\overline{p}}(\overline{\gamma}, \overline{\beta}) = \angle_p(\gamma, \beta).$$

The following proposition follows from the *Toponogov comparison theorem*, see [64, p.161].

**Proposition 2.0.3** *Let $\Lambda(p; \gamma, \beta)$ be a generalized geodesic hinge in $\mathcal{M}$. Then, there exists a comparison hinge $\Lambda(\overline{p}; \overline{\gamma}, \overline{\beta})$ in $\mathcal{M}^2_\kappa$ such that $d(q_\gamma, q_\beta) \leq d(\overline{q}_{\overline{\gamma}}, \overline{q}_{\overline{\beta}})$, where $q_\gamma$, $q_\beta$, $\overline{q}_{\overline{\gamma}}$ and $\overline{q}_{\overline{\beta}}$ denote the end points of $\gamma$, $\beta$, $\overline{\gamma}$ and $\overline{\beta}$, respectively.*

In the proof of the next result, we use the same ideas as those presented in the proof of [69, Lemma 3.2], with some minor technical adjustments needed to settle it to our goals.

**Lemma 2.0.4** *Let $\mathcal{M}$ be a complete Riemannian manifolds with sectional curvature $K \geq \kappa$. Let $p, q \in \mathcal{M}$, $p \neq q$, $v \in T_p\mathcal{M}$, $\gamma : [0, +\infty) \longrightarrow \mathcal{M}$ be defined by $\gamma(t) = exp_p(tv)$ and a minimizing geodesic $\beta : [0, 1] \to \mathcal{M}$ with $\beta(0) = p$ and $\beta(1) = q$. Then, for any $t \in [0, +\infty)$ there holds*

$$\cosh(\hat{\kappa}d(\gamma(t), q)) \leq \cosh(\hat{\kappa}d(p, q)) + \hat{\kappa}\cosh(\hat{\kappa}d(p, q))\sinh(t\hat{\kappa}\|v\|)\left(\frac{t\|v\|}{2} - \frac{\tanh(\hat{\kappa}d(p, q))}{\hat{\kappa}d(p, q)}\frac{\langle v, \beta'(0)\rangle}{\|v\|}\right), \tag{2.2}$$

*and, consequently, the following inequality holds*

$$d^2(\gamma(t), q) \leq d^2(p, q) + \frac{\sinh(\hat{\kappa} t \|v\|)}{\hat{\kappa}} \left( t\|v\| \frac{\hat{\kappa} d(p, q)}{\tanh(\hat{\kappa} d(p, q))} - \frac{2\langle v, \beta'(0)\rangle}{\|v\|} \right). \tag{2.3}$$

*Proof.* Consider the generalized geodesic hinge $\Lambda(p; \gamma, \beta)$ in $\mathcal{M}$, and let $\theta := \angle_p(\gamma, \beta)$. Since $\ell(\beta) = d(p, q) = \|\beta'(0)\|$, we have

$$\langle v, \beta'(0)\rangle = \|v\|\, d(p, q) \cos\theta. \tag{2.4}$$

Let $t \in [0, +\infty)$ and set $y := \gamma(t)$. Then, by Proposition 2.0.4, exist a comparison geodesic hinge $\Lambda(\overline{p}; \overline{\gamma}, \overline{\beta})$ of the hinge $\Lambda(p; \gamma, \beta)$ in $M^2_\kappa$ such that

$$d(y, q) \leq d(\overline{y}, \overline{q}), \tag{2.5}$$

where $\overline{y}$ and $\overline{q}$ are the end points of $\overline{\gamma}$ and $\overline{\beta}$, respectively. Thus, considering the triangle geodesic $\Delta(\overline{y}\,\overline{p}\,\overline{q})$ in $M^2_\kappa$, we have

$$d(\overline{p}, \overline{y}) = t\,\|v\|\,, \qquad d(\overline{p}, \overline{q}) = d(p, q), \qquad \angle_{\overline{p}}(\overline{\gamma}, \overline{\beta}) = \theta. \tag{2.6}$$

It follows from (2.5) that $\cosh(\hat{\kappa} d(y, q)) \leq \cosh(\hat{\kappa} d(\overline{y}, \overline{q}))$. Thus, using Proposition 2.0.2 to the geodesic triangle $\Delta(\overline{y}\,\overline{p}\,\overline{q})$ in $M^2_\kappa$ together with (2.6), we obtain that

$$\cosh(\hat{\kappa} d(y, q)) \leq \cosh(\hat{\kappa} d(p, q)) \cosh(\hat{\kappa} t\,\|v\|) - \sinh(\hat{\kappa} d(p, q)) \sinh(\hat{\kappa} t\,\|v\|) \cos\theta.$$

By using (2.4) and taking into account that $\cosh(\hat{\kappa} t\,\|v\|) \leq 1 + (\hat{\kappa} t\,\|v\|\,/2) \sinh(\hat{\kappa} t\,\|v\|)$, for all $t \geq 0$, after some algebraic manipulation the last inequality becomes (2.2). To prove (2.3), first note that (2.2) is equivalent to

$$\frac{2\hat{\kappa} d(p, q)\,[\cosh(\hat{\kappa} d(y, q)) - \cosh(\hat{\kappa} d(p, q))]}{\sinh(\hat{\kappa} d(p, q))} \leq \hat{\kappa} \sinh(\hat{\kappa} t\,\|v\|) \left( t\,\|v\| \frac{\hat{\kappa} d(p, q)}{\tanh(\hat{\kappa} d(p, q))} - \frac{2\langle v, \beta'(0)\rangle}{\|v\|} \right).$$

On the other hand, from [69, Lemma 3.1] we have $s^2 - t^2 \leq 2t[\cosh(s) - \cosh(t)]/\sinh(t)$, for all $s\,, t \geq 0$. Thus,

$$\hat{\kappa}^2 d^2(y, q) - \hat{\kappa}^2 d^2(p, q) \leq \frac{2\hat{\kappa} d(p, q)\,[\cosh(\hat{\kappa} d(y, q)) - \cosh(\hat{\kappa} d(p, q))]}{\sinh(\hat{\kappa} d(p, q))}.$$

Therefore, (2.3) follows by combining the two last inequalities, which completes the proof.

■

We proceeded to recall some concepts and basic properties about convexity in the Riemannin context. For more details see, for example, [61, 64, 67, 69]. For any two points $p, q \in \mathcal{M}$, $\Gamma_{pq}$ denotes the set of all geodesic segments $\gamma : [0, 1] \to \mathcal{M}$ with $\gamma(0) = p$ and $\gamma(1) = q$. Let the nonempty subset $\Omega \subset \mathcal{M}$. We use $\Gamma^\Omega_{pq}$ to denote the set of all $\gamma \in \Gamma_{pq}$ such that $\gamma(t) \in \Omega$, for all $t \in [0, 1]$.

**Definition 2.0.5** *A nonempty subset $\Omega \subset \mathcal{M}$ is said to be weakly convex if, for any $p, q \in \Omega$, there is a minimal geodesic segment joining $p$ to $q$ belonging to $\Omega$.*

**Definition 2.0.6** *A function $f : \mathcal{M} \to \mathbb{R}$ is said to be convex on the set $\Omega \subset \mathcal{M}$ if $\Omega$ is weakly convex and, for any $p, q \in \Omega$ and $\gamma \in \Gamma_{pq}^{\Omega}$, the composition $f \circ \gamma : [0,1] \to \mathbb{R}$ is a convex function on $[0,1]$, i.e.,*

$$(f \circ \gamma)(t) \leq (1-t)f(p) + tf(q), \qquad \forall\, t \in [0,1].$$

For $f$ a differentiable function on $\mathcal{D} \subset \mathcal{M}$ and a weakly convex set $\Omega \subset \mathcal{D}$, we have the following characterization: $f$ is convex on $\Omega$ iff there holds

$$f(\gamma(t)) \geq f(p) + \langle \operatorname{grad} f(p), \gamma'(0) \rangle, \qquad \forall\, \gamma \in \Gamma_{pq}^{\Omega}, \;\; p, q \in \Omega, \tag{2.7}$$

see [67, Theorem 5.1].

The following concept will be useful in the analysis of the sequence generated by the iterative methods discussed in this thesis.

**Definition 2.0.7** *A sequence $\{y_k\}$ in the complete metric space $(\mathcal{M}, d)$ is quasi-Fejér convergent to a set $W \subset \mathcal{M}$ if, for every $w \in W$, there exists a sequence $\{\epsilon_k\} \subset \mathbb{R}$ such that $\epsilon_k \geq 0$, $\sum_{k=1}^{\infty} \epsilon_k < +\infty$, and $d^2(y_{k+1}, w) \leq d^2(y_k, w) + \epsilon_k$, for all $k = 0, 1, \ldots$.*

The main property of a quasi-Fejér sequence is stated in the next result, and its proof is similar to the one proved in [19], by replacing the Euclidean distance by the Riemannian.

**Theorem 2.0.8** *Let $\{y_k\}$ be a sequence in the complete metric space $(\mathcal{M}, d)$. If $\{y_k\}$ is quasi-Fejér convergent to a nonempty set $W \subset \mathcal{M}$, then $\{y_k\}$ is bounded. Furthermore, if a cluster point $\bar{y}$ of $\{y_k\}$ belongs to $W$, then $\lim_{k\to\infty} y_k = \bar{y}$.*

# Chapter 3

# Gradient method for optimization on Riemannian manifolds with lower bounded curvature

In this chapter, we consider the gradient method to solve the following optimization problem:

$$\min\{f(p)\colon p \in \mathcal{M}\}, \tag{3.1}$$

where the constraint set $\mathcal{M}$ is endowed with a structure of a *complete Riemannian manifold with lower bounded curvature* and $f : \mathcal{M} \to \mathbb{R}$ is a *continuously differentiable convex function*. The analysis of the gradient method is presented with three different finite procedures for determining the step-size, namely, Lipschitz step-size, adaptive step-size and Armijo's step-size. The first procedure requires that the objective function has Lipschitz continuous gradient, which is not necessary for the other approaches. Convergence of the whole sequence to a minimizer, without any level set boundedness assumption, is proved. Iteration-complexity bound for functions with Lipschitz continuous gradient is also presented. Numerical experiments are provided to illustrate the effectiveness of the method in this new setting and certify the obtained theoretical results. In particular, we consider the problem of finding the Riemannian center of mass and the so-called Karcher's mean.

## 3.1 Notations and auxiliary results

The following lemma plays an important role in next sections and its proof can be obtained, with some minor technical adjustments, following the ideas of Lemma 2.0.4 together with the characterization given in (2.7).

**Lemma 3.1.1** *Let $\mathcal{M}$ be a Riemannian manifolds with sectional curvature $K \geq \kappa$, and $\hat{\kappa}$ be defined in* (2.1)*. Assume that $f$ is differentiable and convex on the set $\Omega \subset \mathcal{M}$, $p \in \Omega$*

*and* $\gamma : [0, +\infty) \to \mathcal{M}$ *is defined by* $\gamma(t) = \exp_p(-t \operatorname{grad} f(p))$. *Then, for any* $t \in [0, +\infty)$ *and* $q \in \Omega$ *there holds*

$$\cosh(\hat{\kappa} d(\gamma(t), q)) \leq \cosh(\hat{\kappa} d(p, q)) + \\ \hat{\kappa} \cosh(\hat{\kappa} d(p, q)) \sinh(\hat{\kappa} t \|\operatorname{grad} f(p)\|) \left[ \frac{t \|\operatorname{grad} f(p)\|}{2} - \frac{\tanh(\hat{\kappa} d(p, q))}{\hat{\kappa} d(p, q)} \frac{f(p) - f(q)}{\|\operatorname{grad} f(p)\|} \right]$$

*and, consequently, the following inequality holds*

$$d^2(\gamma(t), q) \leq d^2(p, q) + \\ \frac{\sinh(\hat{\kappa} t \| \operatorname{grad} f(p)\|)}{\hat{\kappa}} \left[ t \| \operatorname{grad} f(p)\| \frac{\hat{\kappa} d(p, q)}{\tanh(\hat{\kappa} d(p, q))} - \frac{2}{\|\operatorname{grad} f(p)\|} (f(p) - f(q)) \right].$$

Next we present the definition of Lipschitz continuous gradient vector field; see [24].

**Definition 3.1.2** *Let* $f : \mathcal{M} \to \mathbb{R}$ *be a differentiable function on the set* $\mathcal{D} \subset \mathcal{M}$. *The gradient vector field of* $f$ *is said to be Lipschitz continuous on* $\mathcal{D}$ *with constant* $L \geq 0$ *if, for any* $p, q \in \mathcal{D}$ *and* $\gamma \in \Gamma_{pq}^{\mathcal{D}}$, *it holds that* $\|P_{\gamma,p,q} \operatorname{grad} f(p) - \operatorname{grad} f(q)\| \leq L\ell(\gamma)$.

It is well known that the covariant derivative can be expressed in terms of parallel transport. Namely, for $p \in \mathcal{M}$, $v \in T_p\mathcal{M}$ and $\alpha$ a differentiable curve such that $\alpha(0) = p$ and $\alpha'(0) = v$ we have

$$\nabla_v V(p) = \lim_{t \to 0} \frac{1}{t} \left[ P_{\alpha,\alpha(t),p} V(\alpha(t)) - V(p) \right], \qquad \forall\, V \in \mathcal{X}(\mathcal{M}). \tag{3.2}$$

Thus, from (3.2) the "fundamental theorem of calculus" can be stated by following

$$P_{\alpha,q,p} V(q) - V(p) = \int_0^1 P_{\alpha,\alpha(t),p} \nabla_{\alpha'(t)} V(\alpha(t))\, dt, \tag{3.3}$$

where $\alpha(1) = q$, see [34, eq.(2.4)]. The *norm of the hessian* $\operatorname{hess} f$ at $p \in \mathcal{M}$ is given by

$$\|\operatorname{hess} f(p)\| := \sup \{ \|\operatorname{hess} f(p) v\| \; : \; v \in T_p\mathcal{M}, \; \|v\| = 1 \}. \tag{3.4}$$

In the following result we present a characterization for twice continuously differentiable functions with Lipschitz continuous gradient vector field.

**Lemma 3.1.3** *Let* $f : \mathcal{M} \to \mathbb{R}$ *be a twice continuously differentiable function on the set* $\mathcal{D} \subset \mathcal{M}$. *The gradient vector field of* $f$ *is Lipschitz continuous on* $\mathcal{D}$ *with constant* $L \geq 0$ *if, and only if, there exists* $L \geq 0$ *such that* $\|\operatorname{hess} f(p)\| \leq L$, *for all* $p \in \mathcal{D}$.

*Proof.* First we assume that $\operatorname{grad} f$ is Lipschitz continuous with constant $L \geq 0$. Let $p \in \mathcal{D}$ and $v \in T_p\mathcal{M}$ with $\|v\| = 1$. Consider $\gamma$ the geodesic given by $\gamma(t) = \exp_p(tv)$. Thus, considering that $\|v\| = 1$, it follows from Definition 3.1.2 that

$$\left\| \frac{1}{t} \left[ P_{\gamma,\gamma(t),p} \operatorname{grad} f(\gamma(t)) - \operatorname{grad} f(p) \right] \right\| \leq L,$$

for all $t$ small enough. Thus, letting $t$ goes to 0 in the last inequality and taking into account (3.2) and that $\gamma'(0) = v$ we obtain that $\|\mathrm{hess} f(p)v\| \leq L$. Therefore, from (3.4) we conclude that the norm of the hessian of $f$ is bounded by $L$. Reciprocally, assume that the norm of the hessian of $f$ is bounded by $L$. Let $p$ and $q \in \mathcal{D}$ and consider $\gamma : [0,1] \to \mathcal{D}$ a geodesic segment joining $p$ to $q$ with $\gamma(0) = p$. By using (3.3) we have

$$P_{\gamma,q,p}\, \mathrm{grad}\, f(q) - \mathrm{grad}\, f(p) = \int_0^1 P_{\gamma,\gamma(t),p} \mathrm{hess} f(\gamma(t))\, \gamma'(t)\, dt.$$

Since the parallel transport is an isometry and considering that $\|\mathrm{hess} f(p)\| \leq L$, for all $p \in \mathcal{M}$, we conclude that

$$\|P_{\gamma,q,p}\, \mathrm{grad}\, f(q) - \mathrm{grad}\, f(p)\| \leq L \int_0^1 \|\gamma'(t)\|\, dt = L\ell(\gamma)).$$

Therefore, from the Definition 3.1.2 follows that $\mathrm{grad}\, f$ is Lipschitz continuous on $\mathcal{D}$ with constant $L$, which conclude the proof. ■

The next lemma can be found in [12, Corollary 2.1] with minor adjustments. Its proof follows from the fundamental theorem of calculus.

**Lemma 3.1.4** *Let $f : \mathcal{M} \to \mathbb{R}$ be a differentiable function on the set $\mathcal{D} \subset \mathcal{M}$ and $a > 0$. Assume that $\mathrm{grad}\, f$ is Lipschitz continuous on $\mathcal{D}$ with constant $L \geq 0$ and $p \in \mathcal{D}$. If $\exp_p(-t\, \mathrm{grad}\, f(p)) \in \mathcal{D}$, for all $t \in [0,a]$, then there holds*

$$f(\exp_p(-t\, \mathrm{grad}\, f(p))) \leq f(p) - \left(1 - \frac{L}{2}t\right) t\, \|\mathrm{grad}\, f(p)\|^2, \qquad \forall\, t \in [0,a].$$

Note that if $\mathcal{D} = \mathcal{M}$, then the condition $\exp_p(-t\, \mathrm{grad}\, f(p)) \in \mathcal{D}$ for all $t \in [0,a]$, in Lemma 3.1.4, plays no role. In the following example we present a function satisfying all the assumptions of Lemma 3.1.4 for the case $\mathcal{D} \neq \mathcal{M}$.

**Example 3.1.5** Let $\mathcal{M} = \{p \in \mathbb{R}^n : \|p\| = 1\}$ the Euclidean sphere and $q \in \mathcal{M}$. Define $\varphi_q(p) := d^2(p,q)/2$, for all $p \in \mathcal{M}$. The function $\varphi_q$ is differentiable in $\mathcal{D} := \{p \in \mathcal{M} : d(p,q) < 5\pi/6\}$ and convex in $\Omega := \{p \in \mathcal{M} : d(p,q) \leq \pi/2\}$. Furthermore, $\mathrm{grad}\, \varphi_q$ is Lipschitz continuous on $\mathcal{D}$, because hess $\varphi_q$ is continuous in $\mathcal{M}\backslash\{-q\} \supset \mathcal{D}$. Indeed, combining [30, Lemma 3] with Lemma 3.1.3 we conclude that

$$L = \sup_{p \in \mathcal{D}} \frac{|\langle p,q\rangle \arccos \langle p,q\rangle|}{\sqrt{1 - \langle p,q\rangle^2}} = \frac{5\pi}{6}\sqrt{3}.$$

Since $\mathrm{grad}\, \varphi_q(p) = -\exp_p^{-1} q$ for all $p \in \mathcal{M}\backslash\{-q\}$, after some calculations, we conclude that $d(\exp_p(-t\, \mathrm{grad}\, \varphi_q(p)), p) \leq td(p,q)$, for all $p \in \mathcal{D}$. Hence, letting $p \in \Omega$ we have

$$d((\exp_p(-t\, \mathrm{grad}\, \varphi_q(p)), q) \leq d(\exp_p(-t\, \mathrm{grad}\, \varphi_q(p)), p) + d(p,q) \leq (t+1)\frac{\pi}{2},$$

and then $\exp_p(-t\, \mathrm{grad}\, \varphi_q(p)) \in \mathcal{D}$, for all $t \in [0, 1/L]$. For more details about the function $\varphi_q$; see [30].

The study of the gradient method for convex functions is well understood for Riemannian manifold with nonnegative sectional curvature and Hadamard manifolds; see [24, 75, 76]. In order to increase the domain of applications of the method, *hereafter, we assume that $\mathcal{M}$ is a complete Riemannian manifolds with sectional curvature $K \geq \kappa$, where $\kappa < 0$*, unless the contrary is explicitly stated.

## 3.2 The Riemannian gradient method

In this section we state the Riemannian gradient method to solve (3.1) and the strategies for choosing the step-size that will be used in our analysis.

**Assumption 3.2.1** Let $\mathcal{D} \subset \mathcal{M}$ be an open set, $f : \mathcal{M} \to \mathbb{R}$ be continuously differentiable in $\mathcal{D}$, $\Omega^*$ be the *solution set* of the problem (3.1), $f^* := \inf_{p\in\mathcal{D}} f(p)$ be the *optimum value* of $f$, and $c \in \mathbb{R}$. *From now on, we assume that $\Omega^*$ is non-empty and $f$ is convex on the sub-level set $\mathcal{L}_c f$*, where

$$\mathcal{L}_c f := \{p \in \mathcal{M} : \; f(p) \leq c\} \subset \mathcal{D}.$$

The statement of *Riemannian gradient algorithm* to solve (3.1) is as follows.

---

**Algorithm 1:** Gradient algorithm in a Riemanian manifold $\mathcal{M}$

---

**Step 0.** Let $p_0 \in \mathcal{L}_c f$. Set $k = 0$.

**Step 1.** If $\operatorname{grad} f(p_k) = 0$, then **stop**; otherwise, choose a step-size $t_k > 0$ and compute

$$p_{k+1} := \exp_{p_k}\left(-t_k \operatorname{grad} f(p_k)\right). \tag{3.5}$$

**Step 2.** Set $k \leftarrow k + 1$ and proceed to **Step 1**.

---

In the following we present three different strategies for choosing the step-size $t_k > 0$ in Algorithm 1. In the first strategy we assume that $\operatorname{grad} f$ is Lipschitz continuous.

**Strategy 1 (Lipschitz step-size)** *Assume that* $\operatorname{grad} f$ *is Lipschitz continuous on $\mathcal{D}$ with constant $L \geq 0$ and that* $\exp_p(-t \operatorname{grad} f(p)) \in \mathcal{D}$, *for all $p \in \mathcal{L}_c f$ and $t \in [0, 1/L]$. Let $\varepsilon > 0$ and take*

$$\varepsilon < t_k \leq \frac{1}{L}. \tag{3.6}$$

Despite knowing that $\operatorname{grad} f$ is Lipschitz continuous, in general, the Lipschitz constant is not computable. Next strategy can be used to compute the step-size without any Lipschitz condition. However, as we shall show, if $\operatorname{grad} f$ is Lipschitz with constant $L > 0$, then the step-size computed is an approximation to the step-size $1/L$; see [5].

**Strategy 2 (adaptive step-size)** *Take $\zeta \in (0, 1/2]$, $L_0 > 0$, and $0 < \eta < 1$. Set $t_k := L_k^{-1}$, where $L_k := \eta^{-i_k} L_{k-1}$ and*

$$i_k := \min\left\{i\colon\ f(\gamma_k(\tau_i)) \leq f(p_k) - \zeta\tau_i \left\|\operatorname{grad} f(p_k)\right\|^2,\ i = 0, 1, \ldots\right\}, \tag{3.7}$$

*where $\tau_i := \eta^i L_{k-1}^{-1}$ and $\gamma_k(\tau_i) := \exp_{p_k}(-\tau_i\ \operatorname{grad} f(p_k))$.*

**Strategy 3 (Armijo's step-size)** *Choose $\delta \in (0, 1)$ and take*

$$t_k := \max\left\{2^{-i}\colon\ f\left(\gamma_k(2^{-i})\right) \leq f(p_k) - \delta 2^{-i} \left\|\operatorname{grad} f(p_k)\right\|^2,\ i = 0, 1, \ldots\right\}, \tag{3.8}$$

*where $\gamma_k(2^{-i}) := \exp_{p_k}(-2^{-i}\ \operatorname{grad} f(p_k))$.*

**Remark 3.2.2** If $\mathcal{D} = \mathcal{M}$, then the condition $\exp_p(-t \operatorname{grad} f(p)) \in \mathcal{D}$ for all $t \in [0, a]$, in Strategy 1, plays no role. Recall that the function in Example 3.1.5 satisfies this condition for $\mathcal{D} \neq \mathcal{M}$.

**Remark 3.2.3** Strategy *2* can be seen as an Armijo-type line search where the first trial step-size at iteration $k$ is set to be equal to $t_{k-1}$. Indeed, taking $L_0 = 1$, and $\eta = 1/2$ the inequality in (3.7) can be equivalently rewritten as

$$f(\gamma_k(2^{-i} t_{k-1})) \leq f(p_k) - \delta 2^{-i} t_{k-1} \left\|\operatorname{grad} f(p_k)\right\|^2.$$

The proof of the well-definedness of Strategies 2 and 3 follows the usual arguments and will be omitted. Hence, the sequence $\{p_k\}$ generated by Algorithm 1 with Strategies 1, 2 or 3 is well-defined. Finally we remark that, due to $f$ satisfies the Assumption 3.2.1, $\operatorname{grad} f(p) = 0$ if only if $p \in \Omega^*$. Therefore, *from now on we assume that* $\operatorname{grad} f(p_k) \neq 0$*, or equivalently, $p_k \notin \Omega^*$, for all $k = 0, 1, \ldots$.*

### 3.2.1 Asymptotic convergence analysis

In this section our goal is to prove that the sequence $\{p_k\}$, generated by the gradient method with Strategies 1, 2 or 3, converges to a solution of problem (3.1).

**Lemma 3.2.4** *Let $\{p_k\}$ be generated by Algorithm 1 with Strategies 1, 2 or 3. Then,*

$$f(p_{k+1}) \leq f(p_k) - \nu t_k \left\|\operatorname{grad} f(p_k)\right\|^2, \qquad k = 0, 1, \ldots, \tag{3.9}$$

*where $\nu = 1/2$ for Strategy 1, $\nu = \zeta$ for Strategy 2 and $\nu = \delta$ for Strategy 3. Consequently, $\{f(p_k)\}$ is non-increasing sequence and $\lim_{k\to+\infty} t_k \|\operatorname{grad} f(p_k)\|^2 = 0$.*

*Proof.* For Strategies 2 and 3, inequality (3.9) follows directly from (3.7) and (3.8), respectively. Now, we assume that $\{p_k\}$ is generated by using Strategy 1. In this case, Lemma 3.1.4 implies that

$$f(p_{k+1}) = f(\exp_{p_k}(-t_k\ \operatorname{grad} f(p_k))) \leq f(p_k) - \left(1 - \frac{L}{2} t_k\right) t_k \left\|\operatorname{grad} f(p_k)\right\|^2,$$

for all $k = 0, 1, \ldots$. Hence, taking into account (3.6) we have $1/2 \leq (1 - Lt_k/2)$ and then (3.9) follows. Therefore, (3.9) holds for $\{p_k\}$ generated by using one of the three strategies. It is immediate from (3.9) that $\{f(p_k)\}$ is non-increasing. Moreover, (3.9) implies that

$$\sum_{k=0}^{\ell} t_k \left\|\operatorname{grad} f(p_k)\right\|^2 \leq \frac{1}{\nu} \sum_{k=0}^{\ell} f(p_k) - f(p_{k+1}) \leq \frac{1}{\nu} \left(f(p_0) - f^*\right),$$

for each nonnegative integer $\ell$, which implies that $t_k \left\|\operatorname{grad} f(p_k)\right\|^2$ goes to zero, as $k$ goes to infinity, completing the proof. ■

**Remark 3.2.5** Whenever $\operatorname{grad} f$ is Lipschitz continuous on $\mathcal{D}$ with constant $L \geq 0$, the step-size in Strategy 2 can be seen as an approximation for $1/L$. Indeed, since $L_0 > 0$ and $0 < \eta < 1$ in Strategy 2, we conclude that $t_k := L_k^{-1} \leq L_{k-1}^{-1} = t_{k-1}$, for all $k = 0, 1, \ldots$. Thus $t_k \leq 1/L_0$, for all $k = 0, 1, \ldots$. If $L_0 \geq L$, then it follows from (3.9) that $t_k = 1/L_0$, for all $k = 0, 1, \ldots$. Therefore, for Strategy 2 we assume $L_0 \leq L$. In this case, (3.9) holds for $t_k = 1/L$ and then (3.7) implies that $\eta/L \leq t_k$. Hence,

$$\frac{\eta}{L} \leq t_k \leq \frac{1}{L_0}, \qquad k = 0, 1, \ldots. \tag{3.10}$$

Let $p_0 \in \mathcal{L}_c f$. By Lemma 3.2.4, we define the constant $\rho > 0$ as follows

$$\sum_{k=0}^{\infty} t_k^2 \left\|\operatorname{grad} f(p_k)\right\|^2 \leq \rho := \begin{cases} 2[f(p_0) - f^*]/L, & \text{for Strategy 1;} \\ [f(p_0) - f^*]/(\zeta L_0), & \text{for Strategy 2;} \\ [f(p_0) - f^*]/\delta, & \text{for Strategy 3.} \end{cases} \tag{3.11}$$

In the following result, in particular, we bound the sequence $\{p_k\}$ generated by Algorithm 1 with Strategies 1, 2 or 3.

**Lemma 3.2.6** *Let $q \in \Omega^*$ and $\{p_k\}$ the sequence generated by Algorithm 1 with Strategies 1, 2 or 3. Then, there holds*

$$d(p_{k+1}, q) \leq \frac{1}{\hat{\kappa}} \cosh^{-1}\left(\cosh(\hat{\kappa} d(p_0, q)) e^{\frac{1}{2}\left(\hat{\kappa}\sqrt{\rho}\right) \sinh\left(\hat{\kappa}\sqrt{\rho}\right)}\right), \qquad k = 0, 1, \ldots. \tag{3.12}$$

*Proof.* Applying the first inequality of Lemma 3.1.1, with $t = t_k$ and $p = p_k$ , we have $p_{k+1} = \gamma(t_k)$, and taking into account that $q \in \Omega^*$, we conclude that

$$\cosh(\hat{\kappa} d(p_{k+1}, q)) \leq \cosh(\hat{\kappa} d(p_k, q)) \left[1 + (\hat{\kappa} t_k \left\|\operatorname{grad} f(p_k)\right\|)^2 \frac{\sinh(\hat{\kappa} t_k \left\|\operatorname{grad} f(p_k)\right\|)}{2\hat{\kappa} t_k \left\|\operatorname{grad} f(p_k)\right\|}\right],$$

for all $k = 0, 1, \ldots$, where $\hat{\kappa}$ is defined in (2.1). Since (3.11) implies $t_k \left\|\operatorname{grad} f(p_k)\right\| \leq \sqrt{\rho}$, for all $k = 0, 1, \ldots$, and the map $(0, +\infty) \ni t \mapsto \sinh(t)/t$ is increasing, we conclude that

$$\cosh(\hat{\kappa} d(p_{k+1}, q)) \leq \cosh(\hat{\kappa} d(p_k, q)) \left[1 + a \left(t_k \left\|\operatorname{grad} f(p_k)\right\|\right)^2\right], \qquad k = 0, 1, \ldots,$$

where $a := \hat{\kappa}(\sinh(\hat{\kappa}\sqrt{\rho}))/(2\sqrt{\rho})$. Now note that the last inequality implies that

$$\cosh(\hat{\kappa} d(p_{k+1}, q)) \leq \cosh(\hat{\kappa} d(p_k, q)) e^{a(t_k \|\operatorname{grad} f(p_k)\|)^2}, \qquad k = 0, 1, \ldots,$$

Therefore, by using (3.11), it follows that $\cosh(\hat{\kappa} d(p_{k+1}, q)) \leq \cosh(\hat{\kappa} d(p_0, q)) e^{a\rho}$, which is equivalent to (4.15) by considering the definition of $\hat{\kappa}$ in (2.1). ■

Let us define the following auxiliary constant

$$\mathcal{C}^q_{\rho,\kappa} := \frac{\sinh\left(\hat{\kappa}\sqrt{\rho}\right)}{\hat{\kappa}\sqrt{\rho}} \left[1 + \cosh^{-1}\left(\cosh(\hat{\kappa} d(p_0, q)) e^{\frac{1}{2}\left(\hat{\kappa}\sqrt{\rho}\right) \sinh\left(\hat{\kappa}\sqrt{\rho}\right)}\right)\right], \tag{3.13}$$

where $\rho$ in defined in (3.11).

**Lemma 3.2.7** *Let $\{p_k\}$ be generated by Algorithm 1 with Strategies 1, 2 or 3. Then, for each $q \in \Omega^*$, there holds*

$$d^2(p_{k+1}, q) \leq d^2(p_k, q) + \frac{t_k}{\nu} \mathcal{C}^q_{\rho,\kappa} \left[f(p_k) - f(p_{k+1})\right] + 2t_k \left[f^* - f(p_k)\right], \tag{3.14}$$

*for all $k = 0, 1, \ldots$, where $\nu = 1/2$ for Strategy 1, $\nu = \zeta$ for Strategy 2 and $\nu = \delta$ for Strategy 3.*

*Proof.* Define $\gamma_k(t) = \exp_{p_k}(-t \operatorname{grad} f(p_k))$, for all $t \in [0, +\infty)$. Then, $\gamma_k(0) = p_k$ and, from (3.5), we obtain $\gamma_k(t_k) = p_{k+1}$. Applying second inequality of Lemma 3.1.1 with $\gamma = \gamma_k$, after some manipulations, we conclude that

$$d^2(p_{k+1}, q) \leq d^2(p_k, q) + \\ \frac{\sinh\left(\hat{\kappa} t_k \|\operatorname{grad} f(p_k)\|\right)}{\hat{\kappa} t_k \|\operatorname{grad} f(p_k)\|} \left[t_k^2 \|\operatorname{grad} f(p_k)\|^2 \frac{\hat{\kappa} d(p_k, q)}{\tanh\left(\hat{\kappa} d(p_k, q)\right)} + 2t_k \left[f^* - f(p_k)\right]\right], \tag{3.15}$$

for all $k = 0, 1, \ldots$. On the other hand, $t/\tanh(t) \leq 1 + t$, for all $t \geq 0$, and the map $(0, +\infty) \ni t \mapsto \sinh(t)/t$ is increasing and bounded below by 1. Thus, taking into account that (3.11) implies $t_k \|\operatorname{grad} f(p_k)\| \leq \sqrt{\rho}$ for all $k = 0, 1, \ldots$, and considering $f^* - f(p_k) \leq 0$ for all $k = 0, 1, \ldots$, we conclude from (3.15) that

$$d^2(p_{k+1}, q) \leq d^2(p_k, q) + \frac{\sinh\left(\hat{\kappa}\sqrt{\rho}\right)}{\hat{\kappa}\sqrt{\rho}} t_k^2 \|\operatorname{grad} f(p_k)\|^2 \left[1 + \hat{\kappa} d(p_k, q)\right] + 2t_k \left[f^* - f(p_k)\right],$$

for all $k = 0, 1, \ldots$, where $\rho$ is defined in (3.11). Thus, by Lemma 3.2.4, we obtain

$$d^2(p_{k+1}, q) \leq d^2(p_k, q) + \frac{t_k}{\nu} \frac{\sinh\left(\hat{\kappa}\sqrt{\rho}\right)}{\hat{\kappa}\sqrt{\rho}} \left[1 + \hat{\kappa} d(p_k, q)\right] \left[f(p_k) - f(p_{k+1})\right] + 2t_k \left[f^* - f(p_k)\right],$$

for all $k = 0, 1, \ldots$. Therefore, by Lemma 3.2.6 and (3.13), we have (3.14), which concludes the proof. ■

Finally, we are ready to prove the full convergence of $\{p_k\}$ to a minimizer of $f$.

**Theorem 3.2.8** *Let $\{p_k\}$ be generated by Algorithm 1 with Strategies 1, 2 or 3. Then $\{p_k\}$ converges to a solution of the problem in* (3.1).

*Proof.* First note that (3.6), (3.8) and (3.10) imply $0 < t_k \leq 1/L$ or $0 < t_k \leq 1$ or $0 < t_k \leq 1/L_0$, for all $k = 0, 1, \ldots$, for Strategies 1, 3 or 2, respectively. Let $\Gamma \coloneqq \max\{1/L, 1, 1/L_0\}$. Thus, for Strategies 1, 2 or 3 we conclude from Lemma 3.2.7 that

$$d^2(p_{k+1}, q) \leq d^2(p_k, q) + \frac{1}{\nu}\Gamma \mathcal{C}^q_{\rho,\kappa}\left[f(p_k) - f(p_{k+1})\right], \qquad k = 0, 1, \ldots,$$

for all $q \in \Omega^*$. Considering that $\sum_{i=0}^{\infty}[f(p_k) - f(p_{k+1})] \leq [f(p_0) - f^*]$, we conclude that $\{p_k\}$ is quasi-Fejér convergent to $\Omega^*$. Therefore, since $\Omega^*$ is non-empty, the sequence $\{p_k\}$ is bounded. Let $\bar{p}$ be an cluster point of $\{p_k\}$ and $\left\{p_{k_j}\right\}$ be a subsequence of $\{p_k\}$ such that $\lim_{j\to\infty} p_{k_j} = \bar{p}$. It follows from Lemma 3.2.4 that $\lim_{k\to\infty} t_k \left\|\operatorname{grad} f(p_k)\right\|^2 = 0$, and due to $\{t_k\}$ has a cluster point $\bar{t} \in [0, \Gamma]$, we analyze the following two possibilities

$$(\mathbf{a})\ \bar{t} > 0, \qquad\qquad (\mathbf{b})\ \bar{t} = 0.$$

Assume that (**a**) holds. In this case, considering that $\lim_{k\to\infty} t_k \left\|\operatorname{grad} f(p_k)\right\|^2 = 0$ and grad $f$ is continuous, we conclude that

$$0 = \lim_{j\to\infty} t_{k_j} \left\|\operatorname{grad} f(p_{k_j})\right\| = \bar{t}\left\|\operatorname{grad} f(\bar{p})\right\|.$$

Hence, $\operatorname{grad} f(\bar{p}) = 0$ and then $\bar{p} \in \Omega^*$. Note that if Strategy 1 is used, then $\bar{t}$ satisfies only **(a)**. Now, we assume that (**b**) holds. In this case Strategies 2 or 3 is used. First assume Algorithm 1 with Strategy 2. Since $\{t_{k_j}\}$ converges to $\bar{t} = 0$ and $\{t_k\}$ is non-increasing, it follows that $\{t_k\}$ converges to $\bar{t} = 0$. Hence, taking $r \in \mathbb{N}$, we can conclude that $t_k < \eta^r L_0^{-1}$ for $k$ sufficiently large. Thus, (3.7) implies

$$f\left(\exp_{p_k}\left(\eta^r L_0^{-1}[-\operatorname{grad} f(p_{k_j})]\right)\right) > f(p_k) - \eta^r L_0^{-1}\zeta\left\|\operatorname{grad} f(p_k)\right\|^2.$$

Letting $k$ goes to $+\infty$ in the above inequality and taking into account that grad $f$ and the exponential mapping are continuous, we obtain

$$f\left(\exp_{\bar{p}}\left(\eta^r L_0^{-1}[-\operatorname{grad} f(\bar{p})]\right)\right) \geq f(\bar{p}) - \eta^r L_0^{-1}\zeta\left\|\operatorname{grad} f(\bar{p})\right\|^2.$$

The last inequality is equivalent to

$$-\frac{f\left(\exp_{\bar{p}}\left(\eta^r L_0^{-1}[-\operatorname{grad} f(\bar{p})]\right)\right) - f(\bar{p})}{\eta^r L_0^{-1}} \leq \zeta\left\|\operatorname{grad} f(\bar{p})\right\|^2.$$

Thus, letting $r$ goes to $+\infty$ we obtain $\left\|\operatorname{grad} f(\bar{p})\right\|^2 \leq \zeta\left\|\operatorname{grad} f(\bar{p})\right\|^2$ which implies $\operatorname{grad} f(\bar{p}) = 0$, i.e., $\bar{p} \in \Omega^*$. Therefore, since $\{p_k\}$ is quasi-Fejér convergent to $\Omega^*$, we conclude from Theorem 2.0.8 that $\{p_k\}$ converges to $\bar{p}$. Finally, assume that Strategy 3 is used. Since $\{t_{k_j}\}$ converges to $\bar{t} = 0$, taking $r \in \mathbb{N}$, we conclude that $t_{k_j} < 2^{-r}$ for $j$ sufficiently large. Thus Armijo's condition (3.8) is not satisfied for $2^{-r}$, i.e.,

$$f(\exp_{p_{k_j}}(2^{-r}[-\operatorname{grad} f(p_{k_j})])) > f(p_{k_j}) - 2^{-r}\delta \left\|\operatorname{grad} f(p_{k_j})\right\|^2.$$

Letting $j$ goes to $+\infty$ in the above inequality and taking into account that $\operatorname{grad} f$ and the exponential mapping are continuous, we obtain

$$f(\exp_{\bar{p}}(2^{-r}[-\operatorname{grad} f(\bar{p})])) \geq f(\bar{p}) - 2^{-r}\delta \left\|\operatorname{grad} f(\bar{p})\right\|^2.$$

The last inequality is equivalent to

$$-\frac{f(\exp_{\bar{p}}(2^{-r}[-\operatorname{grad} f(\bar{p})])) - f(\bar{p})}{2^{-r}} \leq \delta \left\|\operatorname{grad} f(\bar{p})\right\|^2.$$

Thus, letting $r$ goes to $+\infty$ we obtain $\|\operatorname{grad} f(\bar{p})\|^2 \leq \delta \|\operatorname{grad} f(\bar{p})\|^2$, which implies $\operatorname{grad} f(\bar{p}) = 0$, i.e., $\bar{p} \in \Omega^*$. Therefore, since $\{p_k\}$ is quasi-Fejér convergent to $\Omega^*$, we conclude from Theorem 2.0.8 that $\{p_k\}$ converges to $\bar{p}$ and the proof is completed. ■

### 3.2.2 Iteration-complexity analysis

In this section, we present an iteration-complexity bound related to the gradient method for minimizing a convex function with Lipschitz continuous gradient with constant $L > 0$. In the following, as an application of Lemma 3.2.7, we obtain the iteration-complexity bound for the gradient method with Strategy *2*.

**Theorem 3.2.9** *Let $\{p_k\}$ be generated by Algorithm 1 with Strategy 2 for $\zeta = 1/2$. Then, for every $N \in \mathbb{N}$, there holds*

$$f(p_N) - f^* \leq L\frac{L_0\ d^2(p_0, q) + 2\left(\mathcal{C}^q_{\rho,\kappa} - 1\right)[f(p_0) - f^*]}{2NL_0\eta}, \tag{3.16}$$

*for each $q \in \Omega^*$. As a consequence, given a tolerance $\epsilon > 0$, the number of iterations required to obtain $p_N \in \mathcal{M}$ such that $f(p_N) - f^* < \epsilon$, is bounded by*

$$L\left[L_0\ d^2(p_0, q) + 2\left(\mathcal{C}^q_{\rho,\kappa} - 1\right)[f(p_0) - f^*]\right]/(2L_0\epsilon\eta) = \mathcal{O}\left(1/\epsilon\right).$$

*Proof.* Take $q \in \Omega^*$. After some algebraic manipulations, Lemma 3.2.7 implies

$$2t_k\left(f(p_{k+1}) - f^*\right) \leq \left[d^2(p_k, q) - d^2(p_{k+1}, q)\right] + 2t_k\left[\mathcal{C}^q_{\rho,\kappa} - 1\right]\left[f(p_k) - f(p_{k+1})\right],$$

for all $k = 0, 1, \ldots$. Using (3.10) and taking into account that $\mathcal{C}^q_{\rho,\kappa} \geq 1$, $f(p_{k+1}) - f^* \geq 0$ and $f(p_k) - f(p_{k+1}) \geq 0$, for all $k = 0, 1, \ldots$, it follows that

$$\frac{2\eta}{L}\left[f(p_{k+1}) - f^*\right] \leq \left[d^2(p_k, q) - d^2(p_{k+1}, q)\right] + \frac{2}{L_0}\left[\mathcal{C}^q_{\rho,\kappa} - 1\right]\left[f(p_k) - f(p_{k+1})\right],$$

Summing both sides of the above inequality for $k = 0, 1, \ldots, N-1$, we obtain

$$\frac{2\eta}{L}\sum_{i=0}^{N-1}[f(p_{i+1}) - f^*] \leq \left[d^2(p_0, q) - d^2(p_N, q)\right] + \frac{2}{L_0}\left[\mathcal{C}^q_{\rho,\kappa} - 1\right][f(p_0) - f(p_N)].$$

Since $\{f(x_k)\}$ is a decreasing sequence, we conclude that

$$\frac{2\eta}{L}N\left(f(p_N) - f^*\right) \leq \left[d^2(p_0, q) - d^2(p_N, q)\right] + \frac{2}{L_0}\left[\mathcal{C}^q_{\rho,\kappa} - 1\right][f(p_0) - f(p_N)],$$

which is equivalent to (3.16). The second statement of the theorem follows as an immediate consequence of the first part. ■

Whenever the Lipschitz constant $L > 0$ is computable, we can take a constant step-size and Theorem 3.2.9 trivially implies the following result.

**Theorem 3.2.10** *Let $\{p_k\}$ be generated by Algorithm 1 with Strategy* 1 *for $t_k = 1/L$, for all $k = 0, 1, \ldots$ . Then, for every $N \in \mathbb{N}$, there holds*

$$f(p_N) - f^* \leq \frac{L\ d^2(p_0, q) + 2\left(\mathcal{C}^q_{\rho,\kappa} - 1\right)[f(p_0) - f^*]}{2N}, \tag{3.17}$$

*for each $q \in \Omega^*$. As a consequence, given a tolerance $\epsilon > 0$, the number of iterations required by the gradient method to obtain $p_N \in \mathcal{M}$ such that $f(p_N) - f^* < \epsilon$, is bounded by*

$$\left[L\ d^2(p_0, q) + 2\left(\mathcal{C}^q_{\rho,\kappa} - 1\right)[f(p_0) - f^*]\right]/(2\epsilon) = \mathcal{O}\left(1/\epsilon\right).$$

We remark that, if $\kappa = 0$ then $\mathcal{C}^q_{\rho,\kappa} = 1$. As a consequence, Theorem 3.2.10 merges into [12, Theorem 3.2].

**Corollary 3.2.11** *Let $\{p_k\}$ be generated by Algorithm 1 with Strategy* 1 *for $t_k = 1/L$, for all $k = 0, 1, \ldots$ . Then, for every $N \in \mathbb{N}$, there holds*

$$\min\left\{\|\operatorname{grad} f(p_k)\|: k = 0, 1, \ldots, N\right\} \leq \frac{2\sqrt{L\left[L\ d^2(p_0, q) + 2\left(\mathcal{C}^q_{\rho,\kappa} - 1\right)[f(p_0) - f^*]\right]}}{N}, \tag{3.18}$$

*for each $q \in \Omega^*$. As a consequence, given a tolerance $\epsilon > 0$, the number of iterations required by the gradient method to obtain $p_N \in \mathcal{M}$ such that $\|\operatorname{grad} f(p_N)\| < \epsilon$, is bounded by $\sqrt{L\left[L\ d^2(p_0, q) + 2\left(\mathcal{C}^q_{\rho,\kappa} - 1\right)[f(p_0) - f^*]\right]}/\epsilon = \mathcal{O}\left(1/\epsilon\right)$.*

*Proof.* Let $N \in \mathbb{N}$. Using the notation $\lceil N/2 \rceil$ for the least integer that is greater than or equal to $N/2$, we have

$$f(p_{N+1}) - f^* + \sum_{j=\lceil N/2\rceil}^{N}[f(p_j) - f(p_{j+1})] = f(p_{\lceil N/2\rceil}) - f^*. \tag{3.19}$$

Thus, combining the last inequality with Theorem 3.2.10, we conclude that

$$f(p_{N+1}) - f^* + \sum_{j=\lceil N/2\rceil}^{N} [f(p_j) - f(p_{j+1})] \leq \frac{L\ d^2(p_0, q) + 2\left(\mathcal{C}^q_{\rho,\kappa} - 1\right)[f(p_0) - f^*]}{2\lceil N/2\rceil}. \quad (3.20)$$

On the other hand, using Lemma 3.2.4 and considering that $t_k = 1/L$, we obtain

$$\frac{1}{2L}\sum_{j=\lceil N/2\rceil}^{N} \|\mathrm{grad}\, f(p_j)\|^2 \leq \sum_{j=\lceil N/2\rceil}^{N} [f(p_j) - f(p_{j+1})] \leq f(p_{\lceil N/2\rceil}) - f^*.$$

In view of $N/2 \leq \lceil N/2\rceil$, the above inequality together with (3.19) and (3.20) yield

$$\frac{1}{2L}\sum_{j=\lceil N/2\rceil}^{N} \|\mathrm{grad}\, f(p_j)\|^2 \leq \frac{L\ d^2(p_0, q) + 2\left(\mathcal{C}^q_{\rho,\kappa} - 1\right)[f(p_0) - f^*]}{N}.$$

Therefore,

$$\min\{\|\,\mathrm{grad}\, f(p_k)\|^2 :\ k = \lceil N/2\rceil, \ldots, N\} \leq \frac{4L\left[L\ d^2(p_0, q) + 2\left(\mathcal{C}^q_{\rho,\kappa} - 1\right)[f(p_0) - f^*]\right]}{N^2},$$

which implies the desired inequality. The second statement of the corollary follows as an immediate consequence of the first one. ■

We end this section by recalling an iteration-complexity bound for non-convex functions defined in a general Riemannian manifolds, which appeared in [18].

**Theorem 3.2.12** *Let $\{p_k\}$ be generated by Algorithm 1 with Strategy* 1. *Then, for every $N \in \mathbb{N}$, there holds*

$$\min\{\|\,\mathrm{grad}\, f(p_k)\| : k = 0, 1, \ldots, N\} \leq \frac{\sqrt{2L(f(p_0) - f^*)}}{\sqrt{N+1}}.$$

*As a consequence, given a tolerance $\epsilon > 0$, the number of iterations required to obtain $p_N \in \mathcal{M}$ such that $\|\,\mathrm{grad}\, f(p_N)\| < \epsilon$ is bounded by $\mathcal{O}(L(f(p_0) - f^*)/\epsilon^2)$.*

It is worth to point out that results on iteration-complexity bound to the gradient method on Riemannian manifold with non-negative curvature and in Hadamard manifolds with lower bounded curvature has already appeared [12, 75, 76]. The result of this section present a contribution to the systematic study of the iteration-complexity of the gradient methods in the Riemannian setting.

## 3.3 Examples

In the following sections, we present some examples of functions satisfying the assumptions of our results in the previous sections. In particular, we show that, by endowing the constrained set with a suitable Riemannian metric, a constrained Euclidean optimization problem with non-convex objective function having non-Lipschitz gradient can be seen as unconstrained Riemanian optimization problem with convex objective function having Lipschitz gradient. *Throughout the next sections we denote*

$$\mathbb{R}^n_{++} := \left\{ x := (x_1, \ldots, x_n)^T \in \mathbb{R}^{n\times 1} \quad : \quad x_i > 0, \quad i = 1, \ldots, n \right\},$$

*the positive orthant,* $\mathbb{P}^n$ *the set of symmetric matrices of order* $n \times n$ *and* $\mathbb{P}^n_{++}$ *the cone of symmetric positive definite matrices.*

### 3.3.1 Examples in the positive orthant

In this section, we present examples in the positive orthant endowed with a new Riemannian metric. To present this examples we need some definitions and results of Riemannian geometry. *From now on, for each* $x \in \mathbb{R}^n$ *we will denote by diag*$(x)$ *the diagonal matrix in* $\mathbb{R}^{n\times n}$ *that satisfies* $(diag(x))_{ii} = x_i$ *for all* $i = 1, \ldots, n$.

Endowing $\mathbb{R}^n_{++}$ with the Riemannian metric $\langle \cdot, \cdot \rangle$ defined by $\langle u, v \rangle := u^T G(x) v$, for all $x \in \mathbb{R}^n_{++}$ and $u, v \in T_x\mathbb{R}^n_{++} \equiv \mathbb{R}^n$, where $G : \mathbb{R}^n_{++} \to \mathbb{P}^n_{++}$ is given by

$$G(x) := \operatorname{diag}(x_1^{-2}, \ldots, x_n^{-2}) \in \mathbb{R}^{n\times n}, \tag{3.21}$$

we obtain a complete Riemannian manifold with zero curvature, which will be denoted by $\mathcal{M} := (\mathbb{R}^n_{++}, G)$. The Christoffel symbols of $\mathcal{M}$ are given by

$$\Gamma^k_{ij} := \begin{cases} -x_k^{-1}, & \text{if} \quad i = j = k, \\ \ \ 0\ , & \text{if} \quad \textit{otherwise}. \end{cases} \tag{3.22}$$

Let $f : \mathcal{M} \to \mathbb{R}$ be a twice differentiable function. We denote by $f'(x)$ and $f''(x)$ the Euclidean gradient and hessian of $f$ at $x$, respectively. Thus, (3.21) implies that the Riemannian *gradient* of $f$ is given by

$$\operatorname{grad} f(x) = \operatorname{diag}(x)^2 f'(x), \qquad x \in \mathcal{M}. \tag{3.23}$$

Moreover, (3.21) and (3.22) imply that the Riemannian *hessian* of $f$ is given by

$$\operatorname{hess} f(x)v = \left[\operatorname{diag}(x)^2 f''(x) + \operatorname{diag}(x)\operatorname{diag}\left(f'(x)\right)\right] v, \qquad x \in \mathcal{M}. \tag{3.24}$$

Next we present two examples of convex functions with Lipschitz gradient in $\mathcal{M} := (\mathbb{R}^n_{++}, G)$.

**Example 3.3.1** Consider the function $f : \mathbb{R}^n_{++} \to \mathbb{R}$ defined by

$$f(x) := \sum_{i=1}^{n} f_i(x_i), \qquad f_i(x_i) := -a_i \mathrm{e}^{-b_i x_i} + c_i \ln(x_i)^2 + d_i \ln(x_i), \quad i = 1, \ldots n, \tag{3.25}$$

where $a_i, b_i, d_i \in \mathbb{R}_+$ and $c_i \in \mathbb{R}_{++}$ satisfy $c_i > a_i$. Since $f$ is coercive, it has a minimum. By using (3.25) the first and second derivative of $f$ at $x \in \mathbb{R}^n_{++}$ are given by $f'(x) := (f'_1(x_1), \ldots, f'_n(x_n))$ and $f''(x) := \operatorname{diag}(f''_1(x_1), \ldots, f''_n(x_n))$, where

$$f'_i(x_i) = a_i b_i \mathrm{e}^{-b_i x_i} + 2c_i \frac{\ln(x_i)}{x_i} + \frac{d_i}{x_i}, \quad f''_i(x_i) = -a_i b_i^2 \mathrm{e}^{-b_i x_i} + 2c_i \left[\frac{1 - \ln(x_i)}{x_i^2}\right] - \frac{d_i}{x_i^2}, \tag{3.26}$$

for all $i = 1, \ldots, n$. Note that $f''_i(1) < 0$, for all $i = 1, \ldots, n$, and then $f$ is not Euclidean convex. Using (3.24) and (3.26) the hessian of $f$ in $\mathcal{M} := (\mathbb{R}^n_{++}, G)$ is given by

$$\operatorname{hess} f(x)v := \left(a_1 b_1 \mathrm{e}^{-b_1 x_1} \left(x_1 - b_1 x_1^2\right) + 2c_1, \ldots, a_n b_n \mathrm{e}^{-b_n x_n} \left(x_n - b_n x_n^2\right) + 2c_n\right) v.$$

Since $c_i > a_i$, we have $a_i b_i \mathrm{e}^{-b_i x_i} (x_i - b_i x_i^2) + 2c_i \geq 0$, for all $i = 1, \ldots, n$. Hence, by using the definition of the metric, for $v = (v_1, \ldots, v_n)^T \in \mathbb{R}^n$ and $x \in \mathbb{R}^n_{++}$, we have

$$\langle \operatorname{hess} f(x)v, v\rangle = \sum_{i=1}^n \left[a_i b_i \mathrm{e}^{-b_i x_i} \left(x_i - b_i x_i^2\right) + 2c_i\right] \frac{v_i^2}{x_i^2} \geq 0,$$

concluding that $f$ is convex in $\mathcal{M}$. Since $\|v\| = v^T G(x) v = 1$, we have $v_i^2 \leq x_i^2$ and owing that $a_i b_i \mathrm{e}^{-b_i x_i} (x_i - b_i x_i^2) + 2c_i < a_i + 2c_i$, for all $i = 1, \ldots, n$ , we obtain

$$\|\operatorname{hess} f(x)v\|^2 = \sum_{i=1}^n \left[a_i b_i \mathrm{e}^{-b_i x_i} \left(x_i - b_i x_i^2\right) + 2c_i\right]^2 \frac{v_i^2}{x_i^2} < \sum_{i=1}^n (a_i + 2c_i)^2, \qquad x \in \mathbb{R}_{++}.$$

Therefore, (3.4) and Lemma 3.1.3 imply that $\operatorname{grad} f$ is Lipschitz continuous with constant $L < \sum_{i=1}^n (a_i + 2c_i)^2$.

**Example 3.3.2** Consider the function $f : \mathbb{R}^n_{++} \to \mathbb{R}$ defined by

$$f(x) := \sum_{i=1}^n h_i(x_i), \qquad f_i(x_i) := a_i \ln\left(x_i^{d_i} + b_i\right) - c_i \ln(x_i), \qquad i = 1, \ldots n, \tag{3.27}$$

where $a_i, b_i, c_i, d_i \in \mathbb{R}_{++}$ satisfy $c_i < a_i d_i$ and $d_i \geq 2$, for all $i = 1, \ldots n$. The minimizer of $f$ is $x^* = (x_1^*, \ldots, x_n^*)$, where $x_i^* = \sqrt[d_i]{b_i c_i / (a_i d_i - c_i)}$, for $i = 1, \ldots, n$. By using (3.25) the first and second derivative of $f$ at $x \in \mathbb{R}^n_{++}$ are given by $f'(x) := (f'_1(x_1), \ldots, f'_n(x_n))$ and $f''(x) := \operatorname{diag}(f''_1(x_1), \ldots, f''_n(x_n))$, respectively, where

$$f'_i(x_i) = a_i d_i \frac{x_i^{d_i - 1}}{x_i^{d_i} + b_i} - \frac{c_i}{x_i}, \qquad f''_i(x_i) = a_i d_i \frac{(d_i - 1) b_i x_i^{d_i - 2} - x_i^{2d_i - 2}}{(x_i^{d_i} + b_i)^2} + \frac{c_i}{x_i^2} \qquad i = 1, \ldots, n.$$

Since $c_i < a_i d_i$, for $x_i$ sufficiently big we have $f''_i(x_i) < 0$, for all $i = 1, \ldots n$. Thus, $f$ is not Euclidean convex. Hence, using (3.24) we obtain that the hessian of $f$ is given by

$$\operatorname{hess} f(x)v := \operatorname{diag}\left(a_1 b_1 d_1^2 \frac{x_1^{d_1}}{\left(x_1^{d_1} + b_1\right)^2}, \ldots, a_n b_n d_n^2 \frac{x_n^{d_n}}{(x_n^{d_n} + b_n)^2}\right) v, \tag{3.28}$$

Thus, considering that $\langle \operatorname{hess} f(x)v, v\rangle = v^T G(x) \operatorname{hess} f(x)v$ we conclude that

$$\langle \operatorname{hess} f(x)v, v\rangle = \sum_{i=1}^{n} a_i b_i d_i^2 \frac{x_i^{d_i-2}}{\left(x_i^{d_i}+b_i\right)^2} v_i^2 \geq 0,$$

which implies that $f$ is convex in $\mathcal{M}$. Since $\|v\| = v^T G(x) v = 1$ we have $v_i^2 \leq x_i^2$, for all $i = 1, \ldots, n$. Hence, (3.28) and the definition of the metric yield

$$\|\operatorname{hess} f(x)v\|^2 = \sum_{i=1}^{n} a_i^2 b_i^2 d_i^4 \frac{x_i^{2d_i-2}}{\left(x_i^{d_i}+b_i\right)^4} v_i^2 \leq \sum_{i=1}^{n} a_i^2 b_i^2 d_i^4 \frac{x_i^{2d_i}}{\left(x_i^{d_i}+b_i\right)^4} < \sum_{i=1}^{n} a_i^2 d_i^4.$$

Therefore, (3.4) and Lemma 3.1.3 imply that $\operatorname{grad} f$ is Lipschitz continuous with constant $L < \sum_{i=1}^{n} a_i^2 d_i^4$.

We end this section by presenting two more examples of convex functions with Lipschitz gradients in $\mathcal{M} := (\mathbb{R}^n_{++}, G)$.

**Example 3.3.3** Let the function $f : \mathbb{R}^n_{++} \longrightarrow \mathbb{R}$ defined by

$$f(x) := a \ln\left(x^T x + b\right) - c \sum_{i=1}^{n} \ln(x_i),$$

where $a, b, c \in \mathbb{R}_{++}$ and $nc < 2a$. Note that, the first and the second euclidian derivatives of $f$ are given, respectively, by

$$f'(x) = \frac{2a}{x^T x + b} x - c\left(x_1^{-1}, \ldots, x_n^{-1}\right), \tag{3.29}$$

$$f''(x)v = \frac{2a}{x^T x + b} v - \frac{4a x^T v}{(x^T x + b)^2} x + c \operatorname{diag}\left(x_1^{-2}, \ldots, x_n^{-2}\right) v, \tag{3.30}$$

for all $v \in \mathbb{R}^n$ and $x \in \mathbb{R}^n_{++}$. A minimizer of $f$ is $x_* := \sqrt{bc/(2a - nc)}(1, \ldots, 1) \in \mathbb{R}^n_{++}$. Hence, using (3.24), (3.29) and (3.30), we obtain

$$\operatorname{hess} f(x)v = \frac{4a}{x^T x + b} \operatorname{diag}(x)^2 \left[v - \frac{x^T v}{x^T x + b} x\right].$$

Thus, $f$ is convex in $\mathcal{M}$. Indeed, because for all $x \in \mathcal{M}$ and $v \in T_x\mathcal{M}$ holds

$$\langle \operatorname{hess} f(x)v, v\rangle = \frac{4a}{x^T x + b}\left[v^T v - \frac{(x^T v)^2}{x^T x + b}\right] \geq 0.$$

Now, we will prove that $f$ has gradient lipschitz in $\mathcal{M}$. First, note that

$$\|\operatorname{hess} f(x)v\|^2 = \left(\frac{4a}{x^T x + b}\right)^2 \left[v - \frac{x^T v}{x^T x + b} x\right]^T \operatorname{diag}(x)^2 \left[v - \frac{x^T v}{x^T x + b} x\right].$$

By doing simple algebraic manipulations it follows that $\|\text{hess}\, f(x)v\|^2$ is equal to

$$\left(\frac{4a}{x^Tx+b}\right)^2\left[v^T\text{diag}(x)^2v-2\frac{x^Tv}{x^Tx+b}x^T\text{diag}(x)^2v+\left(\frac{x^Tv}{x^Tx+b}\right)^2x^T\text{diag}(x)^2x\right]. \quad (3.31)$$

Since $\|v\|=v^TG(x)v=1$ we have $|v_i|\leq x_i$ for all $i=1,\ldots,n$. Consequently,

$$v^T\text{diag}(x)^2v\leq\left(x^Tx\right)^2, \quad \left|v^T\text{diag}(x)^2x\right|\leq\left(x^Tx\right)^2, \quad x^T\text{diag}(x)^2x\leq\left(x^Tx\right)^2.$$

Hence, using that $|x^Tv|\leq x^Tx$ and (3.31) , we obtain

$$\|\text{hess}\, f(x)v\|^2\leq\left(\frac{4ax^Tx}{x^Tx+b}\right)^2\left[1+\frac{2x^Tx}{x^Tx+b}+\frac{\left(x^Tx\right)^2}{(x^Tx+b)^2}\right]<64a^2.$$

Therefore, (3.4) and Lemma 3.1.3 imply that $\text{grad}\, f$ is Lipschitz continuous with constant $L<8a$.

**Example 3.3.4** Consider the function $f:\mathbb{R}^n_{++}\longrightarrow\mathbb{R}$ defined by

$$f(x):=a\ln\left(\Pi_x^2+b\right)-c\sum_{i=1}^n\ln\left(x_i\right).$$

where $\Pi_x:=x_1\cdot\ldots\cdot x_n$ and $a,b,c\in\mathbb{R}_{++}$. Note that, the first euclidian derivative of $f$ is given by

$$f'(x)=2a\left(\frac{\Pi_x^2x_1^{-1}}{\Pi_x^2+b},\ldots,\frac{\Pi_x^2x_n^{-1}}{\Pi_x^2+b}\right)-c\left(x_1^{-1},\ldots,x_n^{-1}\right), \qquad \forall x\in\mathbb{R}^n_{++}.$$

Hence, we can say that the second euclidean derivative of $f$ for all $x\in\mathbb{R}^n_{++}$, denoted by $f''(x)$, is given by the following expression

$$2a\begin{bmatrix}\frac{\Pi_x^2x_1^{-2}(\Pi_x^2+b)-2\Pi_x^4x_1^{-2}}{(\Pi_x^2+b)^2} & \cdots & \frac{2\Pi_x^2x_1^{-1}x_n^{-1}(\Pi_x^2+b)-2\Pi_x^4x_n^{-1}x_1^{-1}}{(\Pi_x^2+b)^2}\\ \vdots & \ddots & \vdots\\ \frac{2\Pi_x^2x_n^{-1}x_1^{-1}(\Pi_x^2+b)-2\Pi_x^4x_n^{-1}x_1^{-1}}{(\Pi_x^2+b)^2} & \cdots & \frac{\Pi_x^2x_n^{-2}(\Pi_x^2+b)-2\Pi_x^4x_n^{-2}}{(\Pi_x^2+b)^2}\end{bmatrix}+c\,\text{diag}\left(x_1^{-2},\ldots,x_n^{-2}\right).$$

Given this, using (3.24) we obtain

$$\text{hess}\, f(x)v=\frac{4ab\Pi_x^2}{(\Pi_x^2+b)^2}\text{diag}(x)^2\begin{bmatrix}x_1^{-2} & \cdots & x_1^{-1}x_n^{-1}\\ \vdots & \ddots & \vdots\\ x_n^{-1}x_1^{-1} & \cdots & x_n^{-2}\end{bmatrix}v=\frac{4ab\Pi_x^2}{(\Pi_x^2+b)^2}\left[\left(x^{-1}\right)^Tv\right]x,$$

where $x^{-1}:=(x_1^{-1},...,x_n^{-1})$. Thus, $f$ is convex in $\mathcal{M}$. Indeed, because for all $x\in\mathcal{M}$ and $v\in T_x\mathcal{M}$ holds

$$\langle\text{hess}\, f(x)v,v\rangle=\frac{4ab\Pi_x^2}{(\Pi_x^2+b)^2}\left[\left(x^{-1}\right)^Tv\right]^2>0.$$

Now, we will prove that $f$ has gradient lipschitz in $\mathcal{M}$. For this purpose, calculate

$$\|\text{hess}\, f(x)v\|^2 = \frac{16a^2b^2\Pi_x^4}{(\Pi_x^2+b)^4}\left[\left(x^{-1}\right)^T v\right]^2 n.$$

Since $v^TG(x)v = 1$ we have $|v_i| \leq x_i$, for all $i = 1, \ldots, n$. Hence, the last equality implies

$$\|\text{hess}\, f(x)v\|^2 < \frac{16a^2b^2\Pi_x^4}{(\Pi_x^2+b)^4}n^3 < 16a^2n^3.$$

Therefore, (3.4) and Lemma 3.1.3 imply that grad $f(x)$ is Lipschitz continuous with constant $L < 4an\sqrt{n}$.

### 3.3.2 Examples in the SPD matrices cone

In this section, we present examples in the cone of symmetric positive definite matrices with new Riemannian metric. Following Rothaus [63], let $\mathcal{M} := (\mathbb{P}^n_{++}, \langle\cdot,\cdot\rangle)$ be the Riemannian manifold endowed with the Riemannian metric given by

$$\langle U, V\rangle := \text{tr}(VX^{-1}UX^{-1}), \qquad X \in \mathcal{M}, \qquad U, V \in T_X\mathcal{M}, \tag{3.32}$$

where $\text{tr}(X)$ denotes the trace of $X \in \mathbb{P}^n$ and $T_X\mathcal{M} \approx \mathbb{P}^n$. In fact, $\mathcal{M}$ is a Hadamard manifold, see for example [45, Theorem 1.2. p. 325] and its curvature is bound below; see [46]. The *gradient* and *hessian* of $f : \mathcal{M} \longrightarrow \mathbb{R}$ are given by

$$\text{grad}\, f(X) = Xf'(X)X, \tag{3.33}$$

$$\text{hess}\, f(X)V = Xf''(X)VX + \frac{1}{2}\left[Vf'(X)X + Xf'(X)V\right], \tag{3.34}$$

where $V \in T_X\mathcal{M}$, $f'(X)$ and $f''(X)$ are the Euclidean gradient and hessian of $f$ at $X$, respectively. In the following, we present two examples of convex functions with Lipschitz gradient in $\mathcal{M} := (\mathbb{P}^n_{++}, \langle\cdot,\cdot\rangle)$.

**Example 3.3.5** Consider the function $f : \mathbb{P}^n_{++} \longrightarrow \mathbb{R}$ defined by

$$f(X) = a\ln(\det(X))^2 - b\ln\left(\det(X)\right), \tag{3.35}$$

where $a, b \in \mathbb{R}_{++}$. The Euclidean gradient and hessian of $f$ are given, respectively, by

$$f'(X) = \left[2a\ln(\det(X)) - b\right]X^{-1}, \tag{3.36}$$

$$f''(X)V = 2a\,\text{tr}(X^{-1}V)X^{-1} - \left[2a\ln(\det(X)) - b\right]X^{-1}VX^{-1}, \tag{3.37}$$

for all $X \in \mathbb{P}^n_{++}$ and $V \in \mathbb{P}^n$. It follows from (3.36) that each $X \in \mathcal{M}$ satisfying $\det X = e^{b/(2a)}$ is a critical point of $f$. Thus, letting $V = I_n$ and $X = tI_n$ with $t \in \mathbb{R}_{++}$ in (3.37) we obtain $f''(tI_n)I_n = [2ant^{-2} - 2an\ln t + b]I_n$. Thus, $f''(tI_n)$ is not positive definite for

$t$ sufficiently large. Hence, $f$ is not Euclidean convex. Moreover, $f''$ is not bounded and, consequently, $f'$ is not Lipschitz. On the other hand, combining (3.34), (3.36) and (3.37), after some calculations, we obtain

$$\operatorname{hess} f(X)V = 2a\,\mathrm{tr}(X^{-1}V)X, \qquad \langle \operatorname{hess} f(X)V, V\rangle = 2a\mathrm{tr}(X^{-1}V)^2 \geq 0, \tag{3.38}$$

for all $X \in \mathcal{M}$ and $V \in T_X\mathcal{M}$. Thus, $f$ is convex in $\mathcal{M}$. Moreover, (3.32) with (3.38) yield $\|\operatorname{hess} f(X)V\| = 2a\mathrm{tr}(X^{-1}V)$, for all $X \in \mathcal{M}$ and $V \in T_X\mathcal{M}$. If we assume that $\|V\|^2 = \mathrm{tr}(VX^{-1}VX^{-1}) = 1$, then $\mathrm{tr}(X^{-1}V) \leq \sqrt{n}$. Hence,

$$\|\operatorname{hess} f(X)V\| \leq 2a\sqrt{n}, \qquad X \in \mathcal{M}, \qquad V \in T_X\mathcal{M}, \quad \|V\| = 1.$$

Therefore, (3.4) and Lemma 3.1.3 imply that $\operatorname{grad} f$ is Lipschitz with constant $L \leq 2a\sqrt{n}$.

**Example 3.3.6** Consider the function $f : \mathbb{P}^n_{++} \longrightarrow \mathbb{R}$ by

$$f(X) = a\ln\left(\det(X)^{b_1} + b_2\right) - c\ln\left(\det X\right). \tag{3.39}$$

where $a, b_1, b_2, c \in \mathbb{R}_{++}$ with $c < ab_1$ The Euclidean gradient and hessian of $f$ are given, respectively, by

$$f'(X) = \frac{ab_1\left(\det X\right)^{b_1}}{\left(\det X\right)^{b_1} + b_2}X^{-1} - cX^{-1}, \tag{3.40}$$

$$f''(X)V = \frac{ab_2b_1^2\left(\det X\right)^{b_1}}{[\left(\det X\right)^{b_1} + b_2]^2}\mathrm{tr}(X^{-1}V)X^{-1} + \left[c - \frac{ab_1\left(\det X\right)^{b_1}}{\left(\det X\right)^{b_1} + b_2}\right]X^{-1}VX^{-1}. \tag{3.41}$$

The equality (3.40) implies that all $X \in \mathcal{M}$ such that $\det X = \sqrt[b_1]{b_2c/(ab_1 - c)}$ are critical points for $f$. Letting $V = I_n$ and $X = tI_n$ with $t \in \mathbb{R}_{++}$ in (3.41) we obtain that

$$f''(tI_n)I_n = \left[\frac{(c - ab_1)\,t^{2nd} + (ab_2b_1^2n + 2b_2c - ab_2)\,t^{nd} + b_2^2c}{\left(t^{nd} + b_2\right)^2 t^2}\right]I_n. \tag{3.42}$$

Thus, $f''(tI_n)$ is not positive definite for $t$ conveniently chosen and $f$ is not Euclidean convex. Moreover, from (3.42) we obtain also that $f''$ is not bounded and consequently $f'$ is not Lispchtiz. The combination of (3.34), (3.40), (3.41) and (3.32) yield

$$\operatorname{hess} f(X)V = \frac{ab_2b_1^2\left(\det X\right)^{b_1}}{\left[\left(\det X\right)^{b_1} + b_2\right]^2}\mathrm{tr}(X^{-1}V)X, \tag{3.43}$$

for all $X \in \mathcal{M}$ and $V \in T_X\mathcal{M}$. Hence, by using (3.32) and (3.43) we conclude that

$$\langle \operatorname{hess} f(X)V, V\rangle = \frac{ab_2b_1^2\left(\det X\right)^{b_1}}{\left[\left(\det X\right)^{b_1} + b_2\right]^2}\mathrm{tr}(X^{-1}V)^2 \geq 0, \qquad X \in \mathcal{M}, \quad V \in T_X\mathcal{M},$$

which implies that $f$ is convex in $\mathcal{M}$. Moreover, it follows from (3.32) and (3.43) that

$$\|\text{hess}\, f(X)V\| = \sqrt{n}\frac{ab_2b_1^2\,(\det X)^{b_1}}{\left[(\det X)^{b_1} + b_2\right]^2}\text{tr}(X^{-1}V), \qquad X \in \mathcal{M}, \quad V \in T_X\mathcal{M}. \tag{3.44}$$

Therefore, by assuming that $\|V\|^2 := \text{tr}(VX^{-1}VX^{-1}) = 1$ we obtain from (3.44) that

$$\|\text{hess}\, f(X)V\| \leq \frac{ab_2b_1^2\,(\det X)^{b_1}}{\left[(\det X)^{b_1} + b_2\right]^2} n < ab_1^2 n, \qquad X \in \mathcal{M}, \quad V \in T_X\mathcal{M}.$$

Therefore, (3.4) and Lemma 3.1.3 imply that $f$ has Lipschitz gradient with constant $L < ab_1^2 n$.

## 3.4 Numerical experiments

In this section, we present some numerical experiments to illustrate the behavior of the Riemannian gradient method for minimizing convex functions onto the positive orthant and the cone of symmetric positive definite matrices. We implemented Algorithm 1 with Strategies 1, 2 and 3, and tested it on the examples of Section 3.3. Additionally, we consider the application of the method to compute the Riemannian center of mass, which is a specific instance of a geometric mean for points in a Riemannian manifold. In due course, we will describe this problem in more detail.

Considering the positive orthant, the *exponential mapping* $\exp_x : T_x\mathcal{M} \to \mathcal{M}$ in the Riemannian manifold $\mathcal{M} := (\mathbb{R}^n_{++}, G)$ is assigned by

$$\exp_x(v) = \left(x_1 e^{\frac{v_1}{x_1}}, \ldots, x_n e^{\frac{v_n}{x_n}}\right), \tag{3.45}$$

for each $v := (v_1, \ldots, v_n)^T \in \mathbb{R}^{n\times 1}$ and $x := (x_1, \ldots, x_n)^T \in \mathbb{R}^n_{++}$, see [58]. By using the gradient in (3.23) and the definition of metric, we obtain

$$\|\text{grad}\, f(x)\|^2 = \text{grad}\, f(x)^T G(x)\, \text{grad}\, f(x) = \sum_{i=1}^{n} \left[x_i \frac{\partial f}{\partial x_i}(x)\right]^2.$$

For the cone of symmetric positive definite matrices, the *exponential mapping* $\exp_X : T_X\mathcal{M} \to \mathcal{M}$ in the Riemannian manifold $\mathcal{M} := (\mathbb{P}^n_{++}, \langle\cdot,\cdot\rangle)$ is given by

$$\exp_X(V) = X^{1/2} e^{\left(X^{-1/2} V X^{-1/2}\right)} X^{1/2}, \tag{3.46}$$

for each $V \in \mathbb{P}^n$ and $X \in \mathbb{P}^n_{++}$. By using (3.33), for each $X \in \mathcal{M}$, we have

$$\|\text{grad}\, f(X)\|^2 = \text{tr}\left(\left[X f'(X)\right]^2\right).$$

In both cases, although (3.1) is a constrained optimization problem, by (3.45) and (3.46), Algorithm 1 generates only feasible points without using projections or any other procedure to remain the feasibility. Hence, problem (3.1) can be seen as unconstrained Riemannian optimization problem.

Our numerical experience indicates that it is advantageous to perform a reasonably stringent line search. Therefore, we used $\zeta = 1/2$ for Strategies 2 and $\delta = 1/2$ for Strategies 3. Additionally, we set $L_0 = 1$ and $\eta = 1/2$ for Strategy 2. We stopped the execution of Algorithm 1 at $p_k$ declaring convergence if

$$\|f'(p_k)\|_\infty \leq 10^{-5}.$$

Since, by (3.23) and (3.33), grad $f(p_k) = 0$ if only if $f'(p_k) = 0$, this is a reasonable stopping criterion. The maximum number of allowed iterations was set to 1000. Codes are written in Matlab and are freely available at `https://orizon.mat.ufg.br/`.

### 3.4.1 Academic problems

We begin the numerical experiments by testing the Riemannian gradient method on the problems of minimizing the functions of the examples 3.3.1, 3.3.2, 3.3.5 and 3.3.6. We call these problems by Problem 1, 2, 3 and 4, respectively.

#### Academic problems in the positive orthant

In this section, we compare the performance of the Riemannian with the Euclidian gradient methods on Problems 1 and 2. We considered Algorithm 1 with Strategy 3 and implemented the Euclidian gradient method also using the Armijo rule with the same algorithmic parameters. It is worth mentioning that, in principle, the Euclidian method can generate iterates out of the positive orthant. Thus, in order to keep the feasibility, in each iteration, we simply determine the maximum step-size to remain within the feasible set and perform a convenient line search by shrinking the step-size until the Armijo condition is satisfied.

We generated several instances of Problems 1 and 2 by considering functions (3.25) and (3.27), respectively, with $n = 100$ and different parameters. In all cases, for each $i = 1, \ldots, n$, we set parameters $a_i$ with the same value. Equivalently for parameters $b_i$, $c_i$, and $d_i$.

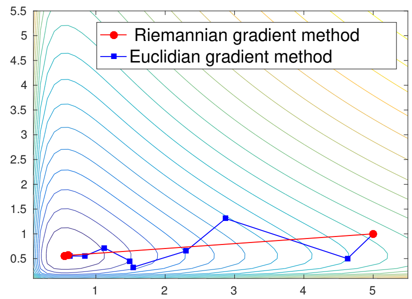
**Problem 1.** First, parameters $a_i$, $b_i$, and $d_i$ were randomly generated between 0 and 10. Then, in order to guarantee that $c_i > a_i$, we randomly generated parameters $c_i$ between $1.1a_i$ and $5.0a_i$. All problems were solved 100 times using starting points from a uniform random distribution inside the box $[0,\ 20]^n$. For each method, Table 3.1 informs the percentage of runs that has reached a critical point (%), the average numbers of iterations (it) and functions evaluations (nfev) of the successful runs.

| | | | | | Riemannian Gradient method | | | Euclidian Gradient method | | |
|---|---|---|---|---|---|---|---|---|---|---|
| # | $a_i$ | $b_i$ | $c_i$ | $d_i$ | % | it | nfev | % | it | nfev |
| 1 | 3.77 | 8.17 | 11.10 | 5.92 | 100.0 | 14.1 | 85.5 | 100.0 | 72.3 | 255.4 |
| 2 | 7.88 | 5.49 | 17.95 | 3.01 | 100.0 | 21.1 | 148.5 | 100.0 | 56.9 | 208.2 |
| 3 | 8.96 | 1.72 | 42.11 | 7.18 | 100.0 | 17.0 | 137.1 | 100.0 | 56.0 | 203.3 |
| 4 | 3.14 | 1.30 | 13.77 | 9.32 | 100.0 | 9.0 | 55.0 | 100.0 | 76.3 | 232.6 |
| 5 | 5.49 | 1.72 | 6.82 | 0.83 | 100.0 | 10.0 | 51.0 | 100.0 | 65.1 | 227.3 |
| 6 | 4.59 | 4.25 | 13.31 | 8.11 | 100.0 | 11.0 | 67.0 | 100.0 | 71.2 | 228.5 |
| 7 | 2.10 | 3.80 | 4.31 | 0.10 | 100.0 | 21.2 | 107.0 | 100.0 | 54.1 | 184.7 |
| 8 | 8.69 | 7.47 | 28.54 | 4.77 | 100.0 | 8.0 | 57.1 | 100.0 | 61.1 | 255.3 |
| 9 | 9.85 | 2.24 | 44.60 | 0.57 | 100.0 | 16.0 | 129.0 | 100.0 | 52.0 | 201.9 |
| 10 | 2.60 | 1.71 | 9.65 | 2.07 | 100.0 | 18.0 | 109.2 | 100.0 | 57.1 | 185.6 |
| 11 | 6.03 | 1.40 | 13.57 | 8.94 | 100.0 | 9.0 | 55.0 | 100.0 | 79.2 | 238.1 |
| 12 | 5.71 | 4.99 | 9.37 | 3.22 | 100.0 | 20.1 | 121.7 | 100.0 | 59.2 | 191.2 |
| 13 | 1.38 | 6.07 | 6.78 | 4.86 | 100.0 | 9.0 | 46.1 | 100.0 | 73.5 | 219.6 |
| 14 | 2.22 | 0.24 | 5.58 | 9.04 | 100.0 | 14.0 | 71.0 | 100.0 | 141.8 | 408.6 |
| 15 | 4.19 | 6.24 | 7.73 | 9.48 | 100.0 | 7.0 | 36.0 | 100.0 | 105.8 | 315.4 |
| 16 | 8.27 | 2.42 | 10.96 | 3.02 | 100.0 | 17.0 | 103.0 | 100.0 | 66.3 | 237.4 |
| 17 | 4.72 | 0.64 | 19.35 | 0.62 | 100.0 | 18.0 | 127.0 | 100.0 | 55.6 | 204.1 |
| 18 | 2.99 | 1.63 | 11.15 | 6.44 | 100.0 | 14.0 | 85.1 | 100.0 | 75.8 | 250.8 |

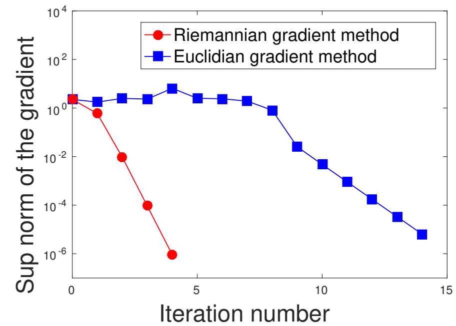
Table 3.1: Parameters of function (3.25) as well as the performance of the Riemannian and Euclidian gradient methods.

As can be seen, the Riemannian gradient method is clearly more efficient than the Euclidian gradient method in this set of problems. In *all* 18 problem instances considered, the Riemannian version required fewer iterations and function evaluations. Overall, on average, the Riemannian gradient method performed 19.8% of iterations and 37.5% of function evaluations required by the Euclidian method.

Figure 3.1 (a) shows a typical behavior of the methods on Problem 1. This case corresponds to $n = 2$, $a_i = 1$, $b_i = c_i = d_i = 2$ for $i = 1, 2$, and the initial point $p_0 = [5,\ 1]^T$. The stopping criterion was satisfied with 4 and 14 iterations for the Riemannian and Euclidian gradient methods, respectively. The *zig-zag* path of the Euclidian gradient method can be seen clearly. In contrast, the Riemannian method rapidly approaches the minimizer. In Figure 3.1 (b), the sup-norm of the euclidean gradient is displayed as a function of the iteration number, which clearly shows the distinction between the methods. While the Euclidian method required 10 iterations for $\|f'(p_k)\|_\infty$ to reach order of $10^{-2}$, the Riemannian algorihtm required only 2 iterations.




(a)




(b)

Figure 3.1: (a) A typical behavior of the Riemannian and the Euclidian gradient methods for which the *zigzag* pattern appears for the Euclidian algorihtm. (b) Sup-norm of the euclidean gradients per iteration.

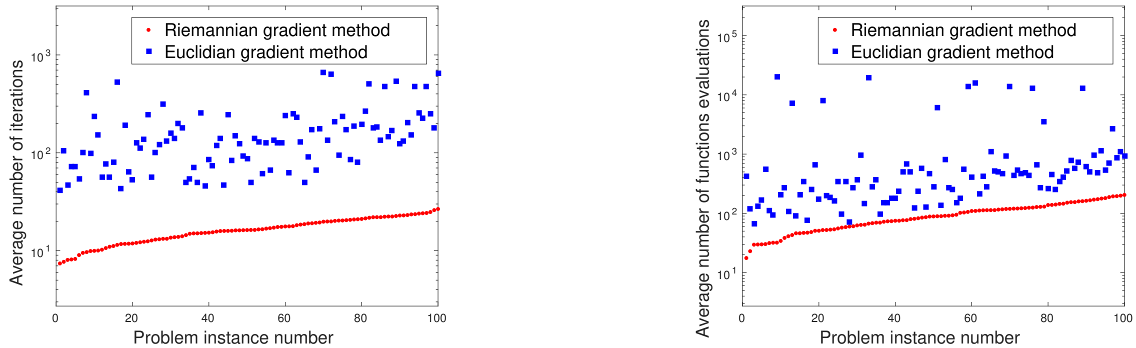
**Problem 2.** We tested the algorithms on a set of 100 instances of Problem 2. We randomly generated parameters $a_i$ and $b_i$ between 0 and 10, parameters $d_i$ between 2 and 10, and a constant $\mu_i$ belonging to the interval $(0,\ 1)$. Then, we set $c_i = \mu_i a_i d_i$, fulfilling the conditions $c_i < a_i d_i$ and $d_i \geq 2$, for all $i = 1, \ldots, n$. As for Problem 1, each instance was solved 100 times using starting points from a uniform random distribution inside the box $[0,\ 20]^n$. The results are given in the following form: for each problem instance, Figure 3.2 (a) informs the average number of iterations, and Figure 3.2 (b) informs the average number of functions evaluations. As a matter of aesthetics, the graphs are independent and were organized in an increasing way.

Figure 3.2 shows that the Riemannian gradient method required fewer iterations and function evaluations than the Euclidian gradient method in *all* problem instances. In terms of percentages, on average, the Riemannian algorithm performed 9.7% and 5.6% of the number of iterations and functions evaluations required by the Euclidian algorithm, respectively.

The results of this section allow us to conclude that there are problems for which the introduction of a suitable metric makes it possible to explore their geometric and algebraic structures, resulting in a large reduction in the computational cost of obtaining their solution. In fact, by introducing a suitable Riemannian metric, a constrained optimization problem with non-convex objective function and non-Lipschitz gradient can be transformed into an optimization problem with convex objective function and Lipschitz gradient.

### Academic problems in the SPD matrices cone

In this section we illustrate the practical applicability of the Riemannian gradient method for minimizing convex functions onto the cone of symmetric positive definite matrices. We



(a) (b)

Figure 3.2: (a) Average number of iterations and (b) average number of functions evaluations required for each of 100 instances of Problem 2 for the Riemannian and the Euclidian gradient methods.

used Problem 3 to test the Riemannian gradient method varying the dimension and the domain of the starting points, while Problem 4 was used to compare the different line search strategies. For Problem 3, we adopted Strategy 3.

**Problem 3.** We set $a = b = 1$ in function (3.35). In the first set of tests, we assigned the following values to the dimension: $n = 10$, 20, 50, 100, and 150. For each specific value of $n$, we run the Riemannian gradient method 100 times using random starting points with eigenvalues belonging to the interval (0, 20). In the second set of tests, we set $n = 50$ and varied the interval that contains the eigenvalues of the starting points. Again, for each combination, the method was run 100 times using random starting points. The results for the first and second set of tests are in Table 3.2 (a) and (b), respectively. First column of Table 3.2 (a) informs the considered dimension, while the first column of Table 3.2 (b) contains the interval for the eigenvalues of the starting points. Columns "%", "it", and "nfev" are as in Table 3.1.

| $n$ | % | it | nfev |
|---|---|---|---|
| 10 | 100.0 | 18.2 | 110.2 |
| 20 | 100.0 | 19.9 | 140.4 |
| 50 | 100.0 | 14.2 | 114.9 |
| 100 | 100.0 | 15.2 | 138.2 |
| 150 | 100.0 | 27.1 | 271.5 |

(a)

| $\lambda_i(X_0)$ | % | it | nfev |
|---|---|---|---|
| (0 10) | 98.0 | 14.2 | 114.6 |
| (0 100) | 99.0 | 14.6 | 117.6 |
| (0 500) | 99.0 | 15.0 | 121.0 |
| (0 1000) | 100.0 | 15.1 | 121.6 |
| (0 2000) | 100.0 | 15.2 | 122.4 |

(b)

Table 3.2: Performance of the Riemannian gradient method in Problem 3 varying: (a) the dimension; (b) the domain of the starting points.

The highlight of Table 3.2 is that the Riemannian gradient method was robust with respect to the dimension and to the choice of the starting points. Furthermore, except for the case where $n = 150$, it was not very sensitive to the variation of the dimension or to the domain of the starting points.

For comparative purposes, we implemented and tested the Euclidean method in Problem 3. For $n = 5$ (respectively, $n = 10$), 15 (respectively, 96) out of the 100 considered starting points resulted in an iteration history that reached the maximum number of iterations allowed. Finally, we observe that, by using (3.46) and the function (3.35), the Riemannian and the Euclidian gradient iteration becomes, respectively,

$$X_{k+1} = \left[\det(X_k)^{2a} e^b\right]^{-t_k} X_k \qquad k = 0, 1, \ldots,$$

and

$$X_{k+1} = X_k - t_k \left[2a \ln(\det(X_k)) - b\right] X_k^{-1}, \qquad k = 0, 1, \ldots,$$

where the steep-size $t_k > 0$ is computed according to the adopted line search strategy. Thus, we can see that the Riemannian gradient iterations are simpler and have a lower computational cost to be performed.

**Problem 4.** We set $n = 100$, $a = b_1 = b_2 = 1$ and $c = 0.5$ in function (3.39), fulfilling $c < ab_1$. We tested the Riemannian gradient method with each of the three strategies by running each combination 100 times using random starting points with eigenvalues belonging to the interval $(0,\ 20)$. The results in Table 3.3 are given as in the previous tables.

| Strategy 1 | | | Strategy 2 | | | Strategy 3 | | |
|---|---|---|---|---|---|---|---|---|
| % | it | nfev | % | it | nfev | % | it | nfev |
| 100.0 | 452.5 | 453.5 | 99.0 | 15.3 | 21.3 | 100.0 | 15.3 | 70.9 |

Table 3.3: Performance of the Riemannian gradient method with the different line search strategies.

For Strategy 1, since the Lipschitz gradient constant satisfies $L < ab_1^2 n$, we used the Lipschitz step-size $t_k = 1/(ab_1^2 n) < 1/L$, for all $k = 1, 2, \ldots$. Overall, as can be seen in Table 3.3, the Riemannian method with Lipschitz step-sizes is clearly the least efficient, requiring an exceedingly large number of iterations. In this case the method performs one function evaluation per iteration. The poor performance is due to the short step-sizes in all iterations. On the other hand, we point out that the efficiency of the Riemannian gradient method with Lipschitz step-size is closely related to an accurate estimate of the Lipschitz gradient constant.

Remark 3.2.3 helps to explain the results of Table 3.3 for Strategies 2 and 3. Regardless of the starting point, Algorithm 1 with both strategies performed exactly the same number of iterations. Additionally, in a typical run, the step-sizes were non-increasing. Therefore, overall, by Remark 3.2.3, the adaptive scheme in Strategy 2 required fewer function evaluations per iteration then the Armijo line search of Strategy 3.

Despite the simple line search mechanisms employed here, the numerical results indicate that, as it has to be expected, the efficient implementation of line search algorithms can significantly improve the Riemannian gradient method.

### 3.4.2 The Riemannian center of mass

The Riemannian center of mass and so called Karcher mean is a specific instance of a geometric mean for points in Riemannian manifolds. It has several practical applications and has appeared in many papers, we refer the reader to [17,44,66] and the references therein.

#### The center of mass on the SPD matrices cone

Denotes by $\|\cdot\|_F$ the Frobenius norm associated to the inner product $\langle U, V\rangle_F := \mathrm{tr}(VU)$, for all $U, V \in \mathbb{P}^n_{++}$. Let $d$ be the Riemannian distance defined in $\mathcal{M} := (\mathbb{P}^n_{++}, \langle\cdot,\cdot\rangle)$, i.e.,

$$d(A, X) = \left\|\ln\left(X^{-1/2}AX^{-1/2}\right)\right\|_F, \qquad A, X \in \mathbb{P}^n_{++}, \tag{3.47}$$

see [58]. The Karcher mean of $m$ positive definite matrices $A_1, \ldots, A_m \in \mathbb{P}^n_{++}$ is the unique solution of the optimization problem

$$\min\left\{f(X) := \frac{1}{2}\sum_{j=1}^{m}\left\|\ln\left(X^{-1/2}A_iX^{-1/2}\right)\right\|_F^2 \ : \ X \in \mathbb{P}^n_{++}\right\}. \tag{3.48}$$

Indeed, $f$ is a strong convex function in $\mathcal{M}$ due to the square of the distance (3.47) be strongly convex in $\mathcal{M}$, see for example [24]. Since $f$ is a strong convex function, all sub-level sets of $f$ are bounded. As a consequence, $f$ has Riemannian Lipschitz gradient on each sublevel set of $f$. Finally, we remark that (3.47) is not an Euclidean convex function. By [44] and using (3.33), we conclude that

$$\operatorname{grad} f(X) = \sum_{i=1}^{m} X^{1/2}\ln\left(X^{1/2}A_i^{-1}X^{1/2}\right)X^{1/2}. \tag{3.49}$$

Thus, by using (3.46) and (3.49), the Riemannian gradient iteration for solving (3.48) is

$$X_{k+1} = X_k^{1/2}e^{-t_k\sum_{i=1}^{n}\ln\left(X_k^{1/2}A_i^{-1}X_k^{1/2}\right)}X_k^{1/2}, \qquad k = 0, 1, \ldots.$$

see, for example, [76].

In [2], Afsari *et al.* studied the convergence of the Riemannian gradient method with a Lipschitz step-size for the center of mass problem in a manifold with curvature bounded from above and below. The step-size is defined from a local estimate for the Lipschitz gradient constant. Consider problem (3.48), and let $r > 0$ be such that $A_1, \ldots, A_m \subset B(X_0, r)$, where $B(X_0, r)$ is the open ball with center $X_0$ and radius $r$. They showed that it is possible to achieve convergence with $t_k = t$ for all $k = 0, 1, \ldots$, where $t \in (0, 2\bar{t})$ and

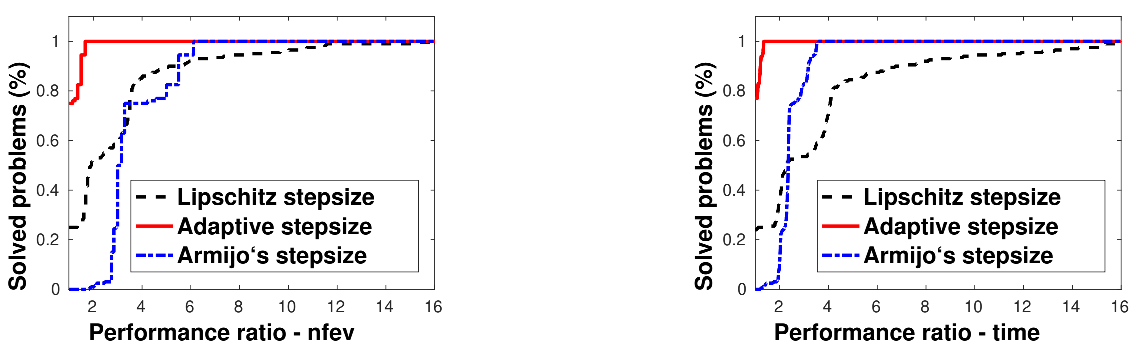
$$\bar{t} = \frac{1}{4r \coth(4r)}. \tag{3.50}$$

Recently, Bento *et al.* [8] extended the convergence of the gradient method to the Hadamard setting for continuously differentiable functions which satisfy the Kurdyka-Lojasiewicz inequality. In particular, they proposed a Riemannian gradient method with Armijo line search for problem (3.48). Basically, their proposal coincides with Algorithm 1 with Strategy 3.

We tested Algorithm 1 with each strategy on a set of 200 randomly generated problems (3.48) with $n = 200$ and $m = 5$, 10, 20 or 50. For each value of $m$ we considered 50 problem instances. Let us clarify how a matrix $A$ was defined. First, we randomly generated an orthonormal matrix $B$ and a diagonal matrix $D$ with elements belonging to $(0, 100)$. Then, we set $A = BDB^T$ ensuring that $A \in \mathbb{P}^n_{++}$. Given a problem instance with data $A_1, \ldots, A_m \in \mathbb{P}^n_{++}$, we defined the starting point $p_0$ as the *explog* geometric mean given by

$$X_0 \coloneqq \exp\left(\frac{1}{m}\sum_{i=1}^{m} \ln(A_i)\right),$$

see, for example, [3]. For Strategy 1 the Lipschitz step-size $t$ was defined according to [2]. We set $t = 1.99\bar{t}$, where $\bar{t}$ is given by (3.50). Radius $r$ can be calculated by computing the maximum distance of $p_0$ to each matrix $A_i$, $i = 1, \ldots, m$. Numerical comparisons are reported in Figure 3.3 using performance profiles [27]. We adopted the number of functions evaluations and CPU time as performance measurements.




(a) (b)

Figure 3.3: Performance profile comparing the Riemannian gradient method with different line search strategies using as performance measurement: (a) number of function evaluations; (b) CPU time.

As can be seen, Algorithm 1 with Strategy 2 is the most efficient on the chosen set of test problems. Efficiencies of the methods are 25.0% (respectively, 24.0%), 75.0% (respectively, 76.0%), and 0.0% (respectively, 0.0%) respectively, considering the number of function evaluations (respectively, CPU time) as performance measurement. Efficiency of Algorithm 1 with Strategy 3 is 0.0% because Strategy 2 outperformed Strategy 3 in all considered instances. Curiously, $m$ is equal to 20 in all problems for which Strategy 1 was the most efficient. Robustness are 99.5%, 100.0%, and 100.0% respectively, see Table 3.4. Only in a problem instance Algorithm 1 with Strategy 1 reached the maximum number of iterations allowed.

| | Efficiency (nfev – CPU time) (%) | Robustness (%) |
|---|---|---|
| Strategy 1 | 25.0 – 24.0 | 99.5 |
| Strategy 2 | 75.0 – 76.0 | 100.0 |
| Strategy 3 | 0.0 – 0.0 | 100.0 |

Table 3.4: Efficiency and Robustness of Algorithm 1 with different line search strategies on a set of 200 randomly generated Riemannian center of mass problems.

The similarity of the Figures 3.3 (a) and (b) suggests that the number of function evaluations is a good indicator of performance. Indeed, evaluating function $f$ is computationally expensive, since it involves inverting $p$ and computing $m$ matrix logarithms. This implies that line search schemes must be carefully formulated for the center of mass problem. Overall, the naive implementation of the Armijo line search in Strategy 3 was overcome by the method with Lipschitz step-size. On the other hand, the results indicate that the adaptive search proposed in Strategy 2 is a promising scheme worth to consider.

**The center of mass on the positive orthant**

Let $\mathcal{M} := (\mathbb{R}^n_{++}, G)$ be the Riemannian manifolds defined in Section 3.3.1 and $d$ the Riemannian distance associated. Hence, we have

$$d^2(y,x) = \sum_{i=1}^{n} \ln^2\left(\frac{y_i}{x_i}\right), \qquad y = (y_1, \ldots, y_n),\ x = (x_1, \ldots, x_n) \in \mathbb{R}^n_{++}. \tag{3.51}$$

The center of mass of $m$ points $w^1, \ldots, w^m \in \mathbb{R}^n_{++}$ is the unique solution of the optimization problem

$$\min \left\{ f(x) := \frac{1}{2} \sum_{j=1}^{m} d^2(w^j, x) \colon x \in \mathbb{R}^n_{++} \right\}. \tag{3.52}$$

Since the square of the distance (3.51) is strongly convex in $\mathcal{M}$, then $f$ is a strong convex function in $\mathcal{M}$, see for example [24]. By using (3.23), we conclude that

$$\operatorname{grad} f(x) = \left( x_1 \sum_{j=1}^{m} \ln\left(\frac{x_1}{w_1^j}\right), \ldots, x_n \sum_{j=1}^{m} \ln\left(\frac{x_n}{w_n^j}\right) \right),$$

where $x = (x_1, \ldots, x_n) \in \mathbb{R}^n_{++}$. Problem (3.52) has closed solution $x^* = (x^*_1, \ldots, x^*_n) \in \mathbb{R}^n_{++}$ given by

$$x^*_i = \left( \prod_{j=1}^{m} w^j_i \right)^{\frac{1}{m}},$$

for all $i = 1, \ldots, m$. Indeed, direct calculations show that $\operatorname{grad} f(x^*) = 0$.

Due to the closed-form solution, we use problem (3.52) to illustrate the results on iteration-complexity bound of Section 3.2.2. We consider the Riemannian gradient algorithm with Lipschitz step-size. Note that the set of positive definite diagonal matrices can be identified with $\mathbb{R}^n_{++}$. Thus, problem (3.52) can be seen as a particular case of problem (3.48) for positive definite diagonal matrices. Given $w^1, \ldots, w^m \in \mathbb{R}^n_{++}$ and defining $A_i = \operatorname{diag}(w^i)$ for all $i = 1, \ldots, m$, we defined the Lipschitz step-size as in Section 3.4.2.

We set $n = 100$, $m = 5$ and randomly generated the elements of $w^1, \ldots, w^m$ and initial point $x_0$ from a uniform distribution on $(0, 100)$. The computed Lipschitz step-size was set to $t \approx 0.06$. The Riemannian gradient algorithm stopped declaring "solution was found" with 30 iterations. Figures 3.4 (a) and (b) report the function values of the left and right hand sides of inequalities (3.17) and (3.18), respectively.

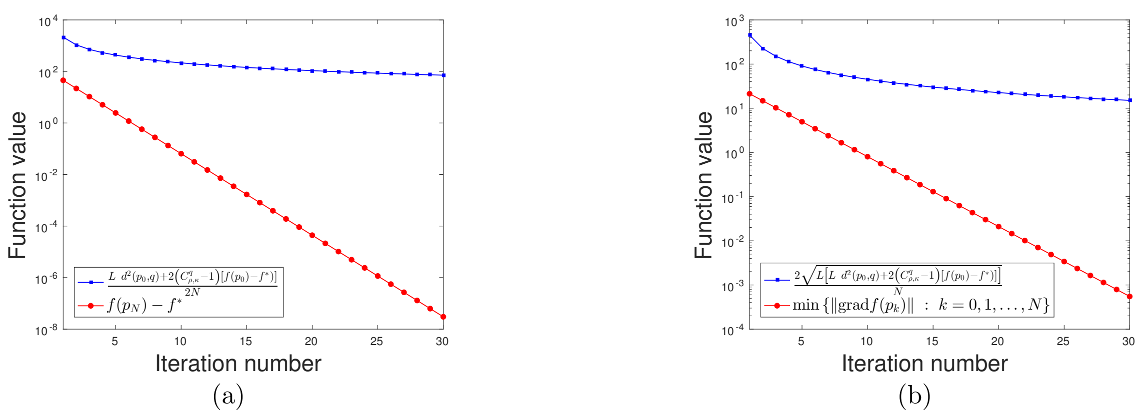



Figure 3.4: Iteration-complexity bound for the Riemannian gradient method with Lipschitz step-size related to: (a) objective function value – Theorem 3.2.10 ; (b) norm of the Riemannian gradient – Corollary 3.2.11.

As can be seen in Figure 3.4, the iteration-complexity bounds related to the objective function value and the norm of the Riemannian gradient are always respected, see Theorem 3.2.10 and Corollary 3.2.11. This illustrate the practical reliability of our iteration-complexity results.

## 3.5 Conclusions

In this chapter, the behavior of the gradient method for convex optimization problems on Riemannian manifolds with lower bounded sectional curvature was analyzed. We considered three different finite procedures for determining the step-size, namely, constant step-size, adaptive procedure and Armijo's procedure. As far as we know, the full convergence of the sequence generated by this method with these three strategies is a new contribution of this chapter, which adds important results in the available convergence theory. Besides, under mild assumptions, we showed that the iteration-complexity bound related to the method is $\mathcal{O}\left(1/\epsilon\right)$ for finding a point $p_N \in \mathcal{M}$ such that $f(p_N) - f^* < \epsilon$. The numerical experiments provided illustrate the effectiveness of the method in this new setting and certify the conclusions suggested by the theoretical results. Despite the simple line search mechanisms employed here, the numerical results indicate that, as it has to be expected, the efficient implementation of line search algorithms can significantly improve the Riemannian gradient method. In particular, the effectiveness of the method to find the Riemannian mass center and the so-called Karcher's mean is presented, indicating that the adaptive procedure is a promising scheme that is worth considering. For this reason, it would be interesting to analyze stochastic versions of the the gradient method by using adaptive procedures.

We remark that the assumption of boundedness from below of the sectional curvature, is just sufficient to obtain the convergence of the gradient method. Indeed, in [72] has been shown that all continuous convex functions on a complete manifold with finite volume are constants. Therefore, letting $\mathcal{M}$ be a Riemannian manifold with sectional curvature unbounded from below and finite volume, we conclude that for any differentiable convex function $f : \mathcal{M} \to \mathbb{R}$, the gradient method applied to minimizer it converges trivially to its minimizer (since all convex function $f$ is constant). Note that the set of Riemannian manifolds with sectional curvature unbounded from below and finite volume is nonempty. In fact, in [1, p. 457] is presented a two dimensional manifold with no lower bounded for Gaussian curvature and finite area. Moreover, in [8] is given a class of non-trivial convex functions on Hadamard manifolds whose sequences generated by the gradient method for these functions converge to for their respective minimizers, see [8, Theorem 5].

Since, the boundedness from below of the Ricci curvature is a more natural geometrical assumption (see [77] and [21]), one interesting question would be: Is it possible to obtain the results of the present chapter by assuming only that Ricci curvature is bounded from bellow? An affirmative answer to this question would increase the applicability domain of the results of this thesis.

Finally, we expect that this chapter will contribute to the development of studies of optimization methods in the Riemannian setting, including to answer the above questions.

# Chapter 4

# Iteration-complexity and asymptotic analysis of steepest descent method for multiobjective optimization on Riemannian manifolds

The steepest descent method for multiobjective optimization on Riemannian manifolds with lower bounded sectional curvature is analyzed in this chapter. The aim of the chapter is twofold. Firstly, an asymptotic analysis of the method is presented with three different finite procedures for determining the step-size, namely, Lipschitz step-size, adaptive step-size and Armijo-type step-size. The second aim is to present, by assuming that the objective function has Jacobian componentwise Lipschitz continuous, iteration-complexity bounds for the method with these three step-sizes. In addition, some examples are presented to emphasize the importance of working in this new context.

## 4.1 Notations and Auxiliary Results

Letting $\mathcal{I} := \{1, \ldots, m\}$ define $\mathbb{R}^m_+ := \{x \in \mathbb{R}^m : \; x_i \geq 0, \quad i \in \mathcal{I}\}$ and $\mathbb{R}^m_{++} := \{x \in \mathbb{R}^m : x_i > 0, \quad i \in \mathcal{I}\}$. For $x, y \in \mathbb{R}^m_+$, $y \succeq x$ (or $x \preceq y$) means that $y - x \in \mathbb{R}^m_+$ and $y \succ x$ (or $x \prec y$) means that $y - x \in \mathbb{R}^m_{++}$. Let $F := (f_1, \ldots, f_m) : \mathcal{M} \to \mathbb{R}^m$ be a differentiable function. We denote the *Riemannian jacobian* of $F$ at a point $p \in \mathcal{M}$ by

$$\nabla F(p)v := \left(\langle \operatorname{grad} f_1(p), v\rangle, \ldots, \langle \operatorname{grad} f_m(p), v\rangle\right), \qquad v \in T_p\mathcal{M},$$

and the image of the Riemannian jacobian of $F$ at $p$ by $\operatorname{Im}(\nabla F(p)) := \{\nabla F(p)v \; : v \in T_pM\}$. A vectorial function $F : \mathcal{M} \to \mathbb{R}^m$ is said to be *convex* on $\mathcal{M}$ if for any $p, q \in \mathcal{M}$ and $\gamma \in \Gamma_{pq}$ (see notation in Chapter 2) the composition $F \circ \gamma : [0, 1] \to \mathbb{R}$ satisfies

$$F \circ \gamma(t) \preceq (1 - t)F(p) + tF(q), \qquad \forall\, t \in [0, 1].$$

By convexity of $F$ follows $\nabla F(p)\gamma'(0) \preceq F(q) - F(p)$. A vectorial function $F$ is called *quasi-convex* on $\mathcal{M}$ if for every $p, q \in \mathcal{M}$ and $\gamma \in \Gamma_{pq}$ holds $F(\gamma(t)) \preceq \max\{F(p), F(q)\}$, for all $t \in [0,1]$, where the maximum is considered coordinate by coordinate. It is immediate of the above definitions that if $F$ is convex then it is quasi-convex. Moreover, if $F$ is a quasi-convex function we have that $F(q) \preceq F(p)$ implies $\nabla F(p)\gamma'(0) \preceq 0$.

In the following we present a concept of Lipschitz continuity for the Riemannian Jacobian of a vectorial function.

**Definition 4.1.1** *Let $F := (f_1, \ldots, f_m) : \mathcal{M} \to \mathbb{R}^m$ be a differentiable function. If for each $f_i : \mathcal{M} \to \mathbb{R}$ there exists a $L_i \geq 0$ such that, for any $p, q \in \mathcal{M}$ and $\gamma \in \Gamma_{pq}$, there holds $\|P_{\gamma,p,q} \operatorname{grad} f_i(p) - \operatorname{grad} f_i(q)\| \leq L_i \ell(\gamma)$, then we say that $\nabla F$ is componentwise Lipschitz continuous on $\mathcal{M}$ with constant $L := \max_{i=1,\ldots,m} L_i$.*

The proof of the next lemma follows, with appropriate adjustments, the same idea of proof of the scalar version presented in [12, Corollary 2.1]. *Throughout of this chapter we will use the following notation*

$$e := (1, \ldots, 1) \in \mathbb{R}^m.$$

**Lemma 4.1.2** *Let $F := (f_1, \ldots, f_m) : \mathcal{M} \to \mathbb{R}^m$ be a differentiable function. Assume that $\nabla F$ is componentwise Lipschitz continuous on $\mathcal{M}$ with constant $L \geq 0$ and $p \in \mathcal{M}$. Then there holds*

$$F(\exp_p(tv)) \preceq F(p) + t\nabla F(p)v + t^2 \frac{L}{2} \|v\|^2 e, \qquad \forall\, t \in [0, +\infty), \quad v \in T_p\mathcal{M}.$$

Whereas the convergence analysis of the steepest descent method for convex and quasi-convex vector functions on Riemannian manifold with nonnegative sectional curvature are well understood; see for example [11, 13]. Without losing the generality, *hereafter, we assume that $\mathcal{M}$ is a complete Riemannian manifolds with sectional curvature $K \geq \kappa$, where $\kappa < 0$.*

## 4.2 Steepest descent for multiobjective optimization

Let $F := (f_1, \ldots, f_m) : \mathcal{M} \to \mathbb{R}^m$ be a continuously differentiable function. The problem of finding a optimum pareto point of $F$, we denote by

$$\min\{F(p) \ : \ p \in \mathcal{M}\}. \tag{4.1}$$

A point $p \in \mathcal{M}$ satisfying $\operatorname{Im}(\nabla F(p)) \cap (-\mathbb{R}^m_{++}) = \emptyset$ is called *critical Pareto*. A *optimum Pareto point* of $F$ is a point $p_* \in \mathcal{M}$ such that there exists no other $p \in \mathcal{M}$ with $F(p) \preceq F(p_*)$ and $F(p) \neq F(p_*)$. Moreover, a point $p_* \in \mathcal{M}$ is a *weak optimal Pareto* of $F$ if there is no $p \in \mathcal{M}$ with $F(p) \prec F(p_*)$. Consider the following problem

$$\min_{v \in T_p\mathcal{M}} \left\{ \max_{i \in \mathcal{I}} \langle \operatorname{grad} f_i(p), v \rangle + \frac{1}{2}\|v\|^2 \right\}, \qquad \mathcal{I} = \{1, \ldots, m\}. \tag{4.2}$$

Whenever $p \in \mathcal{M}$ is not critical Pareto, the optimization problem (4.2) has only one solution, which is called *steepest descent direction* for $F$ in $p$ and denoted by

$$v_p := \arg\min_{v \in T_p\mathcal{M}} \left\{ \max_{i\in\mathcal{I}} \langle \operatorname{grad} f_i(p), v\rangle + \frac{1}{2}\|v\|^2 \right\}. \tag{4.3}$$

In the next lemma we state an important property of the steepest descent direction. Its proof can be found in [13, Lemma 5.1].

**Lemma 4.2.1** *The steepest descent direction mapping $\mathcal{M} \ni p \mapsto v_p \in T_pM$, is a continuous vector field.*

Moreover, the vector $v_p$ is the solution of the problem (4.2) if and only if there exist $\mu_j \geq 0$, for $j \in \mathcal{I}(v_p) := \{j \in \mathcal{I} :\ \langle \operatorname{grad} f_j(p), v_p\rangle = \max_{i\in\mathcal{I}}\langle \operatorname{grad} f_i(p), v_p\rangle\}$, such that

$$v_p = -\sum_{j\in\mathcal{I}(v_p)} \mu_j \operatorname{grad} f_j(p), \qquad\qquad \sum_{j\in\mathcal{I}(v_p)} \mu_j = 1, \tag{4.4}$$

see [13, Lemma 4.1]. In the following lemma we state an important inequality for our convergence analysis and an equivalence for a point $p \in \mathcal{M}$ to be a critical Pareto.

**Lemma 4.2.2** *Let $p \in \mathcal{M}$ and $v_p$ as defined* (4.3)*. Then,*

$$\max_{i\in\mathcal{I}}\langle \operatorname{grad} f_i(p), v_p\rangle = -\|v_p\|^2. \tag{4.5}$$

*Consequently, $\nabla F(p)v_p \preceq -\|v_p\|^2 e$. In addition, $p$ is critical Pareto point of $F$ if, and only if, $\|v_p\| = 0$.*

*Proof.* Let $p \in \mathcal{M}$ and $v_p$ as defined (4.3). Thus, from the first equality in (4.4) we have

$$-\|v_p\|^2 = \langle -v_p, v_p\rangle = \left\langle \sum_{j\in\mathcal{I}(v_p)} \mu_j \operatorname{grad} f_j(p), v_p \right\rangle = \sum_{j\in\mathcal{I}(v_p)} \mu_j \langle \operatorname{grad} f_j(p), v_p\rangle .$$

Hence, using the definition of $\mathcal{I}(v_p)$ and the second equality in (4.4) the last equality becomes to (4.5), which is the first statement of the lemma. The second statement follows by considering $\nabla F(p)v_p = (\langle \operatorname{grad} f_1(p), v_p\rangle, \ldots, \langle \operatorname{grad} f_m(p), v_p\rangle)$ and the definition fo the set $\mathcal{I}(v_p)$. We proceed with the prove of the third statement of the lemma. Assuming that $p$ is a critical Pareto, it follows from the definition that there exist $i \in \mathcal{I}$ such that $\langle \operatorname{grad} f_i(p), v_p\rangle \geq 0$. Then, the by first part of lemma we have $\|v_p\| = 0$. The converse follows from [13, Lemma 4.2] and the proof is concluded. ∎

The proof of the next lemma is a straight combination of Lemma 4.1.2 with first part of Lemma 4.2.2 and will be omited.

**Lemma 4.2.3** *Assume that $\nabla F$ is componentwise Lipschitz continuous on $\mathcal{M}$ with constant $L \geq 0$. Let $p \in \mathcal{M}$ and $v_p$ as defined* (4.3). *Then there holds*

$$F(exp_p(t\,v_p)) \preceq F(p) + \left(\frac{Lt^2}{2} - t\right)\|v_p\|^2 e, \qquad \forall\ t \in [0, +\infty).$$

Next we state the steepest descent algorithm in Riemanian manifold to solve (4.1).

---

**Algorithm 2:** Steepest descent algorithm in a Riemanian manifold $\mathcal{M}$

---

**Step 0.** Let $p_0 \in \mathcal{M}$. Set $k = 0$.

**Step 1.** Compute $v_k := v_{p_k}$, where $v_{p_k}$ is defined in (4.3). If $v_{p_k} = 0$, then **stop**; otherwise, choose a step-size $t_k > 0$ and compute

$$p_{k+1} \coloneqq \exp_{p_k}(t_k\, v_k)\,. \tag{4.6}$$

**Step 2.** Set $k \leftarrow k + 1$ and proceed to **Step 1**.

---

Our goal is to analyze Algorithm 2 with three different strategies for choosing the step-size $t_k > 0$, an analogous analysis done in the scalar case can be found in [31]. In the first strategy we assume that $\nabla F$ is componentwise Lipschitz continuous and in the last two without any Lipschitz condition. The statements of the strategies are as follows:

**Strategy 4 (Lipschitz step-size)** *Assume that $\nabla F$ is componentwise Lipschitz continuous on $\mathcal{M}$ with constant $L \geq 0$. Let $\varepsilon > 0$ and take*

$$\varepsilon < t_k \leq \frac{1}{L}. \tag{4.7}$$

Despite knowing that $\nabla F$ is componentwise Lipschitz continuous, in general the Lipschitz constant is not computable. Then, the next strategy can be used to compute the step-size without any Lipschitz condition. However, as we shall show, if $\nabla F$ is componentwise Lipschitz continuous with constant $L > 0$ the step-size computed is an approximation to $1/L$; see the scalar case in [5, 31].

**Strategy 5 (adaptive step-size)** *Take $\zeta \in (0, 1/2]$, $L_0 > 0$, $t_0 \coloneqq L_0^{-1}$, and $0 < \eta < 1$. Consider $v_k$ is defined as in* (4.3). *Set $t_k \coloneqq \eta^{i_k} t_{k-1}$, where*

$$i_k \coloneqq \min\left\{i:\ F\left(\exp_{p_k}\left(\eta^i t_{k-1} v_k\right)\right) \preceq F(p_k) - \zeta\eta^i t_{k-1}\|v_k\|^2 e,\ i = 0, 1, \ldots\right\}. \tag{4.8}$$

In the next remark we show that if $\nabla F$ is componentwise Lipschitz continuous on $\mathcal{M}$, the adaptive step-size can be seen as an approximation for $1/L$.

**Remark 4.2.4** Suppose that $\nabla F$ is componentwise Lipschitz continuous on $\mathcal{M}$ with constant $L > 0$ and $L_0 > 0$ is an estimate for $L$. Assume $v_k$ is defined as in (4.3). Taking $t = 1/L$, using Lemma 4.2.2 and taking into account that $\zeta \leq 1/2$, we obtain

$$F(\exp_{p_k}(v_k/L)) \preceq F(p_k) - (\zeta\|v_k\|^2/L)e.$$

Hence, it follows that $t_k = 1/L$ it is always accepted for Strategies 5. Therefore, if $L_0 \geq L$ we have $t_k = 1/L_0$, i.e., the step-size is constant. Now, assume $L_0 \leq L$. Since $L_0 > 0$ and $\eta < 1$ we conclude that $t_k$ in Strategies 5 satisfies

$$\frac{\eta}{L} \leq t_k \leq \frac{1}{L_0}. \tag{4.9}$$

**Strategy 6 (Armijo-type step-size)** *Let $t_{\max} > t_{\min} > 0$ and $\delta \in (0,1)$. Let $v_k = v_{p_k}$ be defined as in* (4.3)*. The step-size $t_k$ is chosen according the following algorithm:*

---

STEP 0. $0 < \omega_1 < \omega_2$ *and* $\omega_2 \in (0,1)$*. Set* $\ell = 0$ *and take* $\hat{t}_{k_0} \in (t_{\min}, t_{\max}]$.

STEP 1. *If*

$$F\left(\exp_{p_k}(\hat{t}_{k_\ell} v_k)\right) \preceq F(p_k) - \delta\hat{t}_{k_\ell}\|v_k\|^2 e, \tag{4.10}$$

*then set* $t_k := \hat{t}_{k_\ell}$ *and* **stop***; otherwise proceed to* STEP 2.

STEP 2. *Set* $\ell \leftarrow \ell + 1$*, choose a step-size* $\hat{t}_{k_{\ell+1}} \in [\omega_1\hat{t}_{k_\ell}, \omega_2\hat{t}_{k_\ell}]$ *and proceed to* STEP 1.

---

In the next remark we show that, for $\nabla F$ componentwise Lipschitz continuous on $\mathcal{M}$, the step-sizes in Strategy 6 are bounded below by a positive constant.

**Remark 4.2.5** Asume that $\nabla F$ is componentwise Lipschitz continuous on $\mathcal{M}$ with constant $L \geq 0$, $t_{\max} > 2[1-\delta]/L$ and $t_{\min} < 2\omega_1(1-\delta)/L$. Thus, Lemma 4.2.3 imply

$$F(\exp_{p_k}(t\, v_k)) \preceq F(p_k) + \left(\frac{Lt}{2} - 1\right) t\,\|v_k\|^2\, e, \qquad \forall\, t \in [0, t_{\max}].$$

Hence, by taking any $t \in (0, 2[1-\delta]/L]$ we conclude from last inequality that

$$F(\exp_{p_k}(t\, v_k)) \preceq F(p_k) - \delta t\|v_k\|^2 e.$$

Therefore, $t_k$ in Strategies 6 satisfies the inequality $t_k > t_{\min}$, for all $k = 0, 1, \ldots$.

Since well-definedness of Strategies 5 and 6 follows by using ordinary arguments, we will omitted its proof here. Hence, the sequence $\{p_k\}$ generated by Algorithm 2 with Strategies 4, 5 or 6 is well-defined. Finally we remind that, $p$ is critical pareto if, and only if, $\|v_p\| = 0$. Therefore, *from now on we assume that $\|v_k\| \neq 0$, for all $k$. Moreover, $\{p_k\}$ denotes the sequence generated by Algorithm 2.*

### 4.2.1 Asymptotic convergence analysis

In this section we analyze asymptotic convergence of the sequence $\{p_k\}$ generated by Algorithm 2 with Strategies 4, 5 and 6. For that, we assume in this section that $\{p_k\}$ is a infinite sequence, and consider the following set

$$\mathcal{A} := \{p \in \mathcal{M} : F(p) \preceq F(p_k), \;\; k = 0, 1, \ldots\}.$$

To proceed with our analysis *from now on we will we also assume that the set* $\mathcal{A}$ *is non-empty.* The following remark gives a condition guaranteeing that $\mathcal{A}$ is different from empty.

**Remark 4.2.6** A condition guaranteeing that $\mathcal{A}$ is non-empty is existence of accumulation point for the sequence $\{p_k\}$.

We will begin our analysis with an inequality that will play an important role in the following.

**Lemma 4.2.7** *Let $\{p_k\}$ be generated with any of Strategies 4, 5 or 6. Then,*

$$F(p_{k+1}) \preceq F(p_k) - \nu t_k \|v_k\|^2 e, \qquad k = 0, 1, \ldots, \tag{4.11}$$

*where $\nu = 1/2$ for Strategy 4, $\nu = \zeta$ for Strategy 5 and $\nu = \delta$ for Strategy 6. As a consequence, there holds* $\lim_{k \to +\infty} t_k \|v_k\|^2 = 0$.

*Proof.* The inequality (4.11) for Strategies 5 and 6 follows from (4.6), (4.8) and (4.10), respectively. Now, assume that $\{p_k\}$ is generated by using Strategies 4. In this case, using (4.6), Lemma 4.2.3 and (4.7) we conclude that

$$F(p_{k+1}) = F(\exp_{p_k}(t_k v_k)) \preceq F(p_k) + \left(\frac{Lt_k}{2} - 1\right) t_k \|v_k\|^2 e, \qquad k = 0, 1, \ldots.$$

Hence, taking into account that (4.7) implies $(Lt_k/2 - 1) \leq -1/2$, (4.11) follows with $\nu = 1/2$. To proceed with the proof of the last statement, take $q \in \mathcal{A}$. Thus, (4.11) yields

$$0 \preceq \sum_{k=0}^{\ell} t_k \|v_k\|^2 e \preceq \frac{1}{\nu} \sum_{k=0}^{\ell} \left(F(p_k) - F(p_{k+1})\right) \preceq \frac{1}{\nu} \left(F(p_0) - F(q)\right),$$

with implies the desired result, and the proof of the lemma is conclude. ■

To simplify the statement and proof of the next result we need to define three auxiliary constants. For that, let $p_0 \in \mathcal{M}$ . By using (4.11) together with (4.7), (4.9) and (4.10) define the first constant $\rho > 0$ as follows

$$\sum_{k=0}^{\infty} t_k^2 \|v_k\|^2 \leq \rho := \begin{cases} \min_{i \in \mathcal{I}} \{2[f_i(p_0) - f_i(q)]/L : \; q \in \mathcal{A}\}, & \text{for Strategy 4;} \\ \min_{i \in \mathcal{I}} \{[f_i(p_0) - f_i(q)]/(\zeta L_0) : \; q \in \mathcal{A}\}, & \text{for Strategy 5;} \\ \min_{i \in \mathcal{I}} \{[f_i(p_0) - f_i(q)] t_{\max}/\delta : \; q \in \mathcal{A}\}, & \text{for Strategy 6.} \end{cases} \tag{4.12}$$

The other two auxiliaries constants $\mathcal{C}^q_{\rho,\kappa} > 0$ and $\mathcal{K}^q_{\rho,\kappa} > 0$ are defined as follows

$$\mathcal{C}^q_{\rho,\kappa} := \cosh^{-1}\left(\cosh(\hat{\kappa}d(p_0,q))e^{\frac{1}{2}\left(\hat{\kappa}\sqrt{\rho}\right)\sinh\left(\hat{\kappa}\sqrt{\rho}\right)}\right), \tag{4.13}$$

$$\mathcal{K}^q_{\rho,\kappa} := \frac{\sinh\left(\hat{\kappa}\sqrt{\rho}\right)}{\hat{\kappa}\sqrt{\rho}} \frac{\mathcal{C}^q_{\rho,\kappa}}{\tanh \mathcal{C}^q_{\rho,\kappa}}, \tag{4.14}$$

where the constants $\hat{\kappa}$ and $\rho$, are defined in (2.1) and (4.12), respectively.

**Lemma 4.2.8** *Let $\{p_k\}$ be generated with any of Strategies 4, 5 or 6 and $q \in \mathcal{A}$. Assume that the function $F$ is quasi-convex on $\mathcal{M}$. Then,*

$$d(p_{k+1}, q) \leq \frac{1}{\hat{\kappa}}\mathcal{C}^q_{\rho,\kappa}, \qquad k = 0, 1, \ldots. \tag{4.15}$$

*As a consequence, $\{p_k\}$ is bounded and the following inequality holds*

$$d^2(p_{k+1}, q) \preceq d^2(p_k, q) + \mathcal{K}^q_{\rho,\kappa} t_k^2 \left\|v_k\right\|^2, \qquad k = 0, 1, \ldots. \tag{4.16}$$

*Proof.* For each $k$, let $\gamma_k : [0, \infty) \longrightarrow \mathbb{R}$ be defined by $\gamma_k(t) = \exp_{p_k}(tv_k)$. Let $\beta_k : [0, 1] \to \mathcal{M}$ be a minimizing geodesic with $\beta_k(0) = p_k$ and $\beta_k(1) = q$. By using (4.4), definition of $v_k$, quasi-convexity $F$ and taking into account that $q \in \mathcal{A}$, we have

$$\langle v_k, \beta'(0)\rangle = -\sum_{j\in\mathcal{I}(v_k)} \mu_j \left\langle \operatorname{grad} f_j(p_k), \beta'(0)\right\rangle \geq 0, \qquad \sum_{j\in\mathcal{I}(v_k)} \mu_j = 1. \tag{4.17}$$

Thus, applying the first inequality of Lemma 2.0.4, with $t = t_k$, $\gamma = \gamma_k$ , $\beta = \beta_k$ and $p = p_k$ and using (4.6) and (4.17), we obtain

$$\cosh(\hat{\kappa}d(p_{k+1}, q)) \leq \cosh(\hat{\kappa}d(p_k, q)) \left(1 + \frac{1}{2}\left(\hat{\kappa}t_k \left\|v_k\right\|\right)^2 \frac{\sinh(\hat{\kappa}t_k \left\|v_k\right\|)}{\hat{\kappa}t_k \left\|v_k\right\|}\right).$$

Since (4.12) implies $t_k \left\|v_k\right\| \leq \sqrt{\rho}$, and the map $(0, +\infty) \ni t \mapsto \sinh(t)/t$ is increasing, we conclude that

$$\cosh(\hat{\kappa}d(p_{k+1}, q)) \leq \cosh(\hat{\kappa}d(p_k, q)) \left(1 + \sigma\left(t_k \left\|v_k\right\|\right)^2\right),$$

where $\sigma := \hat{\kappa}(\sinh(\hat{\kappa}\sqrt{\rho}))/(2\sqrt{\rho})$. Now note that the last inequality implies that

$$\cosh(\hat{\kappa}d(p_{k+1}, q)) \leq \cosh(\hat{\kappa}d(p_k, q))e^{\sigma(t_k\|v_k\|)^2}.$$

Therefore, by using (4.12), it follows that $\cosh(\hat{\kappa}d(p_{k+1}, q)) \leq \cosh(\hat{\kappa}d(p_0, q))e^{\sigma\rho}$, which considering the definition of $\sigma$ and (4.13) yield (4.15). The boundedness of $\{p_k\}$ is immediate from (4.15). We proceed with the proof of (4.16). Now, we apply the second inequality of Lemma 2.0.4 and again we take into account (4.6) and (4.17) to conclude that

$$d^2(p_{k+1}, q) \leq d^2(p_k, q) + \frac{\sinh\left(\hat{\kappa}t_k\|v_k\|\right)}{\hat{\kappa}t_k\|v_k\|} \frac{\hat{\kappa}d(p_k, q)}{\tanh\left(\hat{\kappa}d(p_k, q)\right)} t_k^2\|v_k\|^2. \tag{4.18}$$

Since the maps $(0, +\infty) \ni t \mapsto t/\tanh(t)$ and $(0, +\infty) \ni t \mapsto \sinh(t)/t$ are increasing and positive, taking into account (4.15) and that $t_k \|v_k\| \leq \sqrt{\rho}$, the inequality (4.18) becomes

$$d^2(p_{k+1}, q) \leq d^2(p_k, q) + \frac{\sinh\left(\hat{\kappa}\sqrt{\rho}\right)}{\hat{\kappa}\sqrt{\rho}} \frac{\mathcal{C}^q_{\rho,\kappa}}{\tanh \mathcal{C}^q_{\rho,\kappa}} t_k^2 \|v_k\|^2.$$

Therefore, by using (4.14) we have the desired inequality. ∎

In the next result we show that if $F$ is a quasi-convex function on a Riemannian manifolds with lower bounded sectional curvature, then $\{p_k\}$ is in fact convergent to a critical Pareto point of $F$.

**Theorem 4.2.9** *Let $\{p_k\}$ be generated with any of Strategies 1, 2 or 3. If $F$ is quasi-convex, then $\{p_k\}$ converges to a critical Pareto point of $F$.*

*Proof.* Since $\mathcal{A}$ is non-empty, Lemma 4.2.8 and (4.12) imply that $\{p_k\}$ is bounded and quasi-Fejér convergent to set $\mathcal{A}$. Taking into account Lemma 4.2.7 we conclude that $\{f_s(p_k)\}$ is non-increasing, for all $s = 1, \ldots, m$. Thus, we conclude that all cluster points of $\{p_k\}$ belongs to $\mathcal{A}$. Hence, Theorem 2.0.8 implies that $\{p_k\}$ converges to a point $\bar{p} \in \mathcal{A}$. Hence, remais to prove that $\bar{p}$ is a critical Pareto point of $F$. We know that, for any of the three strategies 4, 5 or 6, the sequence $\{t_k\}$ is bounded. Let $\bar{t} \geq 0$ be a cluster point of $\{t_k\}$ and take $\{t_{k_j}\}$ such that $\lim_{j\to\infty} t_{k_j} = \bar{t}$. First we suppose that $\bar{t} > 0$. Since $\lim_{j\to\infty} p_{k_j} = \bar{p}$ and $\lim_{j\to\infty} t_{k_j} = \bar{t}$, (4.12) and Lemma 4.2.1 imply that $0 = \lim_{j\to\infty} t_{k_j} \left\|v_{k_j}\right\| = \bar{t}\left\|v_{\bar{p}}\right\|.$ Thus, considering that we are under the assumption $\bar{t} > 0$, we obtain $v_{\bar{p}} = 0$. Therefore, Lemma 4.2.2 implies that $\bar{p}$ is a critical Pareto point of $F$. Now, we suppose that $\bar{t} = 0$. In this case, we just need to analyze Strategies 5 and 6, due to Strategy 4 we have $\epsilon \leq \bar{t}$. First assume that Strategy 5 is use and take $r \in \mathbb{N}$. Since $\lim_{j\to\infty} t_{k_j} = 0$ we conclude that if $j$ is large enough, $t_{k_j} < \eta^r t_0 =: C_r$. Thus, for each $j$ large enough, from (4.8) we have

$$f_{s_j}\left(\exp_{p_{k_j}}(C_r v_{k_j})\right) > f_{s_j}(p_{k_j}) - \zeta C_r \|v_{k_j}\|^2,$$

for some $s_j \in \{1, \ldots, m\}$. Since the set $\{1, \ldots, m\}$ is finite, without lose of generality, we assume the there exist $\hat{s}$ and a infinite set of index $j$ such that

$$f_{\hat{s}}\left(\exp_{p_{k_j}}(C_r v_{k_j})\right) > f_{\hat{s}}(p_{k_j}) - \zeta C_r \|v_{k_j}\|^2.$$

Since $\lim_{j\to\infty} p_{k_j} = \bar{p}$ and $\lim_{j\to\infty} t_{k_j} = \bar{t}$, letting $j$ goes to $+\infty$ and taking into account that $v_p$ and the exponential map are continuous, we obtain

$$\frac{f_{\hat{s}}\left(\exp_{\bar{p}}(C_r v_{\bar{p}})\right) - f_{\hat{s}}(\bar{p})}{C_r} \geq -\zeta \|v_{\bar{p}}\|^2.$$

Thus, letting $r$ goes to $+\infty$, yields $\langle \operatorname{grad} f_{\hat{s}}(\bar{p}), v_{\bar{p}} \rangle \geq -\zeta \|v_{\bar{p}}\|^2$. Hence, Lemma 4.2.2 implies that $-\|v_{\bar{p}}\|^2 \geq -\zeta \|v_{\bar{p}}\|^2$ and, considering that $\zeta \in (0, 1/2]$, we have $\|v_{\bar{p}}\| = 0$. Consequently, usando again Lemma 4.2.2 we conclude that $\bar{p}$ is a critical pareto of $F$. Finally, assume that Strategy 6 is used. Since $\lim_{j\to\infty} t_{k_j} = 0$ we conclude that if $j$ is large enough we have $t_{k_j} < t_{min}$. Thus, if $j$ is large enough, there exists $0 < \hat{t}_j \leq t_{max}$ such that $0 < \omega_1 \hat{t}_j \leq t_{k_j}$ and

$$f_{s_j}\left(\exp_{p_{k_j}}(\hat{t}_j v_{k_j})\right) > f_{s_j}(p_{k_j}) - \hat{t}_j \delta \|v_{k_j}\|^2,$$

for some $s_j \in \{1,\ldots,m\}$. Since the set $\{1,\ldots,m\}$ is finite, without lose of generality, we assume the there exist $\hat{s}$ and a infinite set of index $j$ such that

$$\frac{f_{\hat{s}}\left(\exp_{p_{k_j}}(\hat{t}_j v_{k_j})\right) - f_{\hat{s}}(p_{k_j})}{\hat{t}_j} > -\delta \|v_{k_j}\|^2.$$

Let $\gamma_j(t) := \exp_{p_{k_j}}(t v_{k_j})$, for $t > 0$, be a geodesic segment. Thus, mean value theorem implies that there exists $\bar{t}_j \in (0, \hat{t}_j)$ such that

$$\left\langle \operatorname{grad} f_{\hat{s}}\left(\gamma_j(\bar{t}_j)\right), P_{\gamma_j,0,\bar{t}_j} v_{k_j} \right\rangle > -\delta \|v_{k_j}\|^2. \tag{4.19}$$

On the other hand, let $B_\epsilon(\overline{p}) \subset \mathcal{M}$ be a totally normal ball. Hence, considering that $\lim_{j\to+\infty} p_{k_j} = \bar{p}$, Lemma 4.2.1 implies that $\lim_{j\to+\infty} v_{k_j} = \bar{v}_{\bar{p}}$. Moreover, $0 < \omega_1 \hat{t}_j \leq t_{k_j}$ implies that $\lim_{j\to+\infty} \hat{t}_j = 0$. Owing to $0 < \bar{t}_j \leq \hat{t}_j$ we obtain that $\lim_{j\to+\infty} \bar{t}_j = 0$. Hence, for all $j$ large enough we have $\{\bar{t}_j\} \subset (0,1)$ and $\gamma_j(\bar{t}_j) \in B_\epsilon(\overline{p})$, which implies

$$P_{\gamma_j,0,\bar{t}_j} v_{k_j} = \frac{1}{1-\bar{t}_j} \exp^{-1}_{\gamma_j(\bar{t}_j)} \exp_{p_{k_j}} v_{k_j}.$$

Thus, taking the limit as $j$ goes to $+\infty$ and using [4, Lemma 1.1], we conclude that $\lim_{j\to+\infty} P_{\gamma_j,0,\bar{t}_j} v_{k_j} = \bar{v}_{\bar{p}}$ (for a more general version of this equality, see [4, Lemma 1.2]). Then, letting $j$ goes to $+\infty$ in (4.19) and taking into account Lemma 4.2.1, that $\operatorname{grad} f_{\hat{s}}$ and the exponential map are continuous, we obtain that $\langle \operatorname{grad} f_{\hat{s}}(\bar{p}), v_{\bar{p}} \rangle \geq -\delta \|v_{\bar{p}}\|^2$. Hence, Lemma 4.2.2 implies that $-\|v_{\bar{p}}\|^2 \geq -\delta \|v_{\bar{p}}\|^2$ and, considering that $\delta \in (0,1)$, we have $\|v_{\bar{p}}\| = 0$. Consequently, usando again Lemma 4.2.2 we conclude that $\bar{p}$ is a critical Pareto of $F$. Therefore, for all Strategies 4, 5 or 6, $\bar{p}$ is a critical Pareto point of $F$, which concludes the proof. ■

**Corollary 4.2.10** *Let $\{p_k\}$ be generated with any of Strategies 4, 5 or 6. If $F$ is convex, then $\{p_k\}$ converges to a weak optimal Pareto of $F$.*

*Proof.* Since $F$ is convex, critical points are weak optimal Pareto of $F$, see [13, Proposition 5.2]. Considering that convex functions are also quasi-convex the result follows from Theorem 4.2.9. ■

### 4.2.2 Iteration-complexity analysis

In this section we present iteration-complexity bounds related to the steepest descent method with Strategies 4, 5 and 6, for $F$ having $\nabla F$ with componentwise Lipschitz continuous constant $L > 0$. For this purpose, by using (4.7), (4.9) and Remark 4.2.5, define

$$\xi := \begin{cases} \epsilon, & \text{for Strategy 4;} \\ \eta/L, & \text{for Strategy 5;} \\ t_{\min}, & \text{for Strategy 6.} \end{cases} \tag{4.20}$$

The following result extends the scalar result [12, Theorem 3.1] to multiobjective settings. Moreover, it also extends to Riemannian context [37, Theorem 3.1].

**Theorem 4.2.11** *Let $\{p_k\}$ be generated with any of Strategies 4, 5 or 6, and set $f_i^* := \inf\{f_i(q) : \ q \in \mathcal{M}\}$, for $i \in \mathcal{I}$. Suppose that $f_i^*$ is bounded from below for some $i \in \mathcal{I}$, and define $i_* \in \mathcal{I}$ such that*

$$f_{i_*}(p_0) - f_{i_*}^* := \min\left\{f_i(p_0) - f_i^* : \ i \in \mathcal{I}\right\}.$$

*Then, for every $N \in \mathbb{N}$, there holds*

$$\min\left\{\|v_k\| : \ k = 0, 1, \ldots, N-1\right\} \leq \left[\frac{f_{i_*}(p_0) - f_{i_*}^*}{\nu\xi}\right]^{\frac{1}{2}} \frac{1}{\sqrt{N}}.$$

*where $\nu = 1/2$ for Strategy 4, $\nu = \zeta$ for Strategy 5 and $\nu = \delta$ for Strategy 6.*

*Proof.* It follows from Lemma 4.2.7 that $\nu t_k \|v_k\|^2 e \preceq F(p_k) - F(p_{k+1})$, for all $k = 0, 1, \ldots$. By summing both sides of this inequality for $k = 0, 1, \ldots, N-1$ and using (4.20), we obtain

$$\nu\xi \sum_{k=0}^{N-1} \|v_k\|^2 e \preceq F(p_0) - F(p_N).$$

Thus, by the definition of $i_*$, we conclude from the last inequality that

$$\nu\xi N \min\left\{\|v_k\|^2 : \ k = 0, 1, \ldots, N-1\right\} \leq f_{i_*}(p_0) - f_{i_*}^*.$$

which implies the statement of the theorem. ■

**Remark 4.2.12** it is worth mentioning that in the above result it was not necessary to use any hypothesis about convexity of $F$ and curvature of $\mathcal{M}$.

Now we are going to prove that under the assumption of convexity Theorem 4.2.11 can be improved. We begin by presenting an auxiliary inequality.

**Lemma 4.2.13** *Let $\{p_k\}$ be generated with any of Strategies 4, 5 or 6. Assume that $F$ is a convex function on $\mathcal{M}$. Then, for $q \in \mathcal{A}$ and each $k$, there exist $\mu_{j's}^k \geq 0$ satisfying $\sum_{j \in \mathcal{I}(v_k)} \mu_j^k = 1$ such that*

$$d^2(p_{k+1}, q) \leq d^2(p_k, q) + \mathcal{K}_{\rho,\kappa}^q t_k^2 \|v_k\|^2 + 2t_k \sum_{j \in \mathcal{I}(v_k)} \mu_j^k [f_j(q) - f_j(p_k)], \tag{4.21}$$

*where $\rho$ is defined in (4.12).*

*Proof.* For each $k$, let $\gamma_k : [0,\infty) \longrightarrow \mathbb{R}$ be defined by $\gamma_k(t) = \exp_{p_k}(tv_k)$ and a minimizing geodesic $\beta_k : [0,1] \to \mathcal{M}$ with $\beta_k(0) = p_k$ and $\beta_k(1) = q$. Using (4.4) and the convexity of $F$ we conclude that exist $\mu^k_{j's} \geq 0$ satisfying $\sum_{j\in\mathcal{I}(v_k)} \mu^k_j = 1$ such that

$$\langle v_k, \beta_k'(0)\rangle = -\sum_{j\in\mathcal{I}(v_k)} \mu^k_j \langle \operatorname{grad} f_j(p_k), \beta_k'(0)\rangle \geq \sum_{j\in\mathcal{I}(v_k)} \mu^k_j (f_j(p_k) - f_j(q)).$$

Applying the second inequality of Lemma 2.0.4 with $\beta = \beta_k$, $\gamma = \gamma_k$ and $t = t_k$ and using the last inequality we obtain

$$d^2(p_{k+1}, q) \leq d^2(p_k, q) + \frac{\sinh(\hat{\kappa} t_k \|v_k\|)}{\hat{\kappa} t_k \|v_k\|}\left(\frac{\hat{\kappa} d(p_k,q)}{\tanh(\hat{\kappa} d(p_k,q))} t_k^2 \|v_k\|^2 + 2t_k \sum_{j\in\mathcal{I}(v_k)} \mu^k_j (f_j(q) - f_j(p_k))\right). \tag{4.22}$$

Considering that $(0,+\infty) \ni t \mapsto t/\tanh(t)$ and $(0,+\infty) \ni t \mapsto \psi(t) := \sinh(t)/t$ are increasing. Furthermore, taking into account that (4.12) implies $t_k \|v_k\| \leq \sqrt{\rho}$, and using (4.15), the inequality (4.22) becomes

$$d^2(p_{k+1}, q) \leq d^2(p_k, q) + \frac{\sinh(\hat{\kappa}\sqrt{\rho})}{\hat{\kappa}\sqrt{\rho}}\left(\frac{\mathcal{C}^q_{\rho,\kappa}}{\tanh \mathcal{C}^q_{\rho,\kappa}} t_k^2 \|v_k\|^2 + 2t_k \sum_{j\in\mathcal{I}(v_k)} \mu^k_j (f_j(q) - f_j(p_k))\right).$$

Therefore, owing $f_j(q) - f_j(p_k) \leq 0$ and $\psi$ bounded from below by 1, using (4.14) the inequality (4.21) follows from the last inequality, which concludes the proof. ■

The next result, with minor adjustments, is a generalization of [37, Theorem 4.1] to Riemannian setting, when the Armijo's type strategy is used.

**Proposition 4.2.14** *Let $\{p_k\}$ be generated with any of Strategies 4, 5 or 6. Assume that $F$ is a convex function on $\mathcal{M}$ and $q \in \mathcal{A}$. Then, for every $N \in \mathbb{N}$, there are non-negative numbers $\lambda_1, \ldots, \lambda_m$ with $\sum_{i=1}^m \lambda_i = 1$, satisfying*

$$\sum_{i=1}^m \lambda_i [f_i(p_N) - f_i(q)] \leq \frac{d^2(p_0, q) + \mathcal{K}^q_{\rho,\kappa}\rho}{2\xi N}, \tag{4.23}$$

*where $\rho$ is defined in* (4.12).

*Proof.* Since $f_i(p_k) - f_i(q) \geq 0$ for all $i$, Lemma 4.2.13 and (4.20) implies there exist $\mu^k_{i's} \geq 0$ such that

$$2\xi \sum_{i=1}^m \mu^k_i (f_i(p_k) - f_i(q)) \leq d^2(p_k, q) - d^2(p_{k+1}, q) + \mathcal{K}^q_{\rho,\kappa} t_k^2 \|v_k\|^2,$$

and $\sum_{i=1}^{m} \mu_i^k = 1$, where for each $k$, define $\mu_i^k := 0$ for all $i \notin \mathcal{I}(v_k)$. By summing both sides of this inequality for $k = 0, 1, \ldots, N-1$, and using (4.12) follows

$$2\xi \sum_{k=0}^{N-1} \sum_{i=1}^{m} \mu_i^k \left( f_i(p_k) - f_i(q) \right) \preceq d^2(p_0, q) + \mathcal{K}_{\rho,\kappa}^{q} \rho.$$

Since $f_i(p_k)$ is a decreasing sequence for each $i \in \{1, \ldots, m\}$, by some algebraic manipulations in the previous inequality we have

$$\sum_{i=1}^{m} \left[ \frac{1}{N} \sum_{k=0}^{N-1} \mu_i^k \right] \left[ f_i(p_N) - f_i(q) \right] \leq \frac{d^2(p_0, q) + \mathcal{K}_{\rho,\kappa}^{q} \rho}{2\xi N}.$$

Defining $\lambda_i := \sum_{k=0}^{N-1} \mu_i^k / N$ we obtain the inequality in (4.23). To complete the proof, we have show that $\sum_{i=1}^{m} \lambda_i = 1$. For that, it is sufficient to note that

$$\sum_{i=1}^{m} \lambda_i = \frac{1}{N} \sum_{i=1}^{m} \sum_{k=0}^{N-1} \mu_i^k = \frac{1}{N} \sum_{k=0}^{N-1} \sum_{i=1}^{m} \mu_i^k,$$

and $\sum_{i=1}^{m} \mu_i^k = 1$ for each $k$. ■

Finally we are ready to present the main result of this section, namely, the improvement of Theorem 4.2.11. We remark that this result is new, even in Euclidean context.

**Theorem 4.2.15** *Let $\{p_k\}$ be generated with any of Strategies 4, 5 or 6. Assume that $F$ is a convex function on $\mathcal{M}$ and $q \in \mathcal{A}$. Then, for every $N \in \mathbb{N}$, there holds*

$$\min \left\{ \|v_k\| : \ k = 0, 1, \ldots, N \right\} \leq \left( \frac{2 \left( d^2(p_0, q) + \mathcal{K}_{\rho,\kappa}^{q} \rho \right)}{\nu \xi^2} \right)^{\frac{1}{2}} \frac{1}{N}.$$

*where $\rho$ is defined in* (4.12) *and $\nu = 1/2$ for Strategy* 4*, $\nu = \zeta$ for Strategy* 5 *and $\nu = \delta$ for Strategy* 6*.*

*Proof.* Let $N \in \mathbb{N}$ and denote by $\lceil N/2 \rceil$ the least integer that is greater than or equal to $N/2$. It follows from Lemma 4.2.7 that $\nu t_k \|v_k\|^2 e \preceq F(p_k) - F(p_{k+1})$, for all $k = 0, 1, \ldots$. Thus, by summing both sides of this inequality for $k = \lceil N/2 \rceil, \ldots, N$ and using (4.20), we obtain

$$\nu \xi \sum_{k=\lceil N/2 \rceil}^{N} \|v_k\|^2 \leq f_i(p_{\lceil N/2 \rceil}) - f_i(p_{N+1}), \qquad \forall \ i \in \mathcal{I}.$$

Hence, taking non-negative numbers $\lambda_1, \ldots, \lambda_m$ as in the Proposition 4.2.14 and considering that $q \in \mathcal{A}$, we conclude from the last inequality that

$$\nu \xi \sum_{k=\lceil N/2 \rceil}^{N} \|v_k\|^2 \leq \sum_{i=1}^{m} \lambda_i \left( f_i(p_{\lceil N/2 \rceil}) - f_i(p_{N+1}) \right) \leq \sum_{i=1}^{m} \lambda_i \left( f_i(p_{\lceil N/2 \rceil}) - f_i(q) \right).$$

Thus, from Proposition 4.2.14 and considering that $N/2 \leq \lceil N/2 \rceil$ it follows that

$$\sum_{k=\lceil N/2 \rceil}^{N} \|v_k\|^2 \leq \frac{d^2(p_0,q)+\mathcal{K}^q_{\rho,\kappa}\rho}{2\nu\xi^2\lceil N/2\rceil} \leq \frac{d^2(p_0,q)+\mathcal{K}^q_{\rho,\kappa}\rho}{\nu\xi^2 N}.$$

Therefore, $\min\{\|v_k\|^2 : \ k=\lceil N/2\rceil,\ldots,N\} \leq 2(d^2(p_0,q)+\mathcal{K}^q_{\rho,\kappa}\rho)/(\nu\xi^2N^2)$, which implies the desired inequality. ■

## 4.3 Examples

In this section we present some examples to illustrate the results obtained in previous sections. In particular, we will present some examples of multiobjective convex functions such that its Riemannian Jacobian is componentwise Lipschitz continuous.

**Example 4.3.1** Let $\mathbb{P}^n_{++}$ be the cone of symmetric positive definite matrices. Define the multiobjective function $F(X) = (f_1(X),\ldots,f_m(X))$, for $f_i : \mathbb{P}^n_{++} \longrightarrow \mathbb{R}$ is defined by

$$f_i(X) = a_i \ln(\det(X))^2 - b_i \ln\left(\det(X)\right),$$

where $a_i, b_i \in \mathbb{R}_{++}$ for all $i=1,\ldots,m$. Consider $\mathcal{M} := (\mathbb{P}^n_{++}, \langle\cdot,\cdot\rangle)$ where

$$\langle U,V\rangle := \mathrm{tr}(VX^{-1}UX^{-1}), \qquad X\in\mathbb{P}^n_{++}, \qquad U,V\in T_X\mathbb{P}^n_{++}.$$

In $\mathcal{M}$, $f_i$ is convex and has Lipschitz gradient with constant $L_i \leq 2a_i\sqrt{n}$, for each $i=1,\ldots,m$, see example 3.3.5. Hence, from Definition 4.1.1 the Jacobian $\nabla F$ is componentwise Lipschitz continuous with constant $L \leq 2\sqrt{n}\max\{a_1,\ldots,a_m\}$. Therefore, from Corollary 4.2.10 we can apply Algorithm 2 with Strategies 4, 5 or 6 to find weak optimal Pareto of $F$.

In the following we present some result to study multiobjective convex functions such that its Riemannian Jacobian is componentwise Lipschitz continuous. We begin with a result that, with some adjustments in the notation, can be found in [49, Lemma 2].

**Lemma 4.3.2** *Let $\bar{\mathcal{M}}$ and $\mathcal{M}$ be Riemannian manifold, $\bar{\nabla}$ be the Levi-Civita connection associated to $\bar{\mathcal{M}}$ and $\varphi : \bar{\mathcal{M}} \to \mathcal{M}$ be an isometry. Then, $\nabla : \mathcal{X}(\mathcal{M})\times\mathcal{X}(\mathcal{M}) \to \mathcal{X}(\mathcal{M})$ defined by*

$$\nabla_V U := d\varphi(\bar{\nabla}_{\bar{V}}\bar{U}), \qquad \forall\ V,U\in\mathcal{X}(\mathcal{M}). \tag{4.24}$$

*is the Levi-Civita connection associated to $\bar{\mathcal{M}}$, where $\bar{V} = d\varphi^{-1}V$ and $\bar{U} = d\varphi^{-1}U$.*

*Proof.* Let $f$ be continuously differentiable, $V$ and $U$ be vector fields in $\mathcal{M}$. Since $\varphi$ is a diffeomorphism, $f\circ\varphi$ is continuously differentiable, $\bar{V} = \mathrm{d}\varphi^{-1}V$ and $\bar{U} = \mathrm{d}\varphi^{-1}U$ are vector fields in $\bar{\mathcal{M}}$. Thus, we can prove that (4.24) satisfies [64, equations (1.9), (1.10), (1.11) and (1.12) on page 27 and 28] and therefore is the Levi-Civita connection associated to $\mathcal{M}$. ■

The next result is the main tool we need to present the following examples of the section.

**Theorem 4.3.3** *Let $\mathcal{M}$ and $\bar{\mathcal{M}}$ be Riemannian manifolds, $f : \mathcal{M} \to \mathbb{R}$ be a twice-differentiable function and $\varphi : \bar{\mathcal{M}} \to \mathcal{M}$ be an isometry. Then, $f$ has gradient vector field Lipschitz continuous with constant $L \geq 0$ if, and only if, $g : \bar{\mathcal{M}} \to \mathbb{R}$ defined by $g := f \circ \varphi$, has gradient vector field Lipschitz continuous with constant $L \geq 0$.*

*Proof.* Let $\bar{V} \in \mathcal{X}(\bar{\mathcal{M}})$ and set $V(\varphi(q)) = \mathrm{d}\varphi(q)\bar{V}(q)$. Thus, by using the definition of the gradient vector field and the chain rule, we have

$$\begin{aligned}
\left\langle \operatorname{grad} g(q), \bar{V}(q) \right\rangle &= \mathrm{d}g(q)\bar{V}(q) \\
&= \mathrm{d}f(\varphi(q))\mathrm{d}\varphi(q)\bar{V}(q) \\
&= \mathrm{d}f(\varphi(q))V(\varphi(q)) \\
&= \left\langle \operatorname{grad} f(\varphi(q)), V(\varphi(q)) \right\rangle.
\end{aligned}$$

Taking into account that $\varphi$ is an isometry and $V(\varphi(q)) = \mathrm{d}\varphi(q)\bar{V}(q)$, we obtain that

$$\begin{aligned}
\left\langle \operatorname{grad} f(\varphi(q)), V(\varphi(q)) \right\rangle &= \left\langle \operatorname{grad} f(\varphi(q)), \mathrm{d}\varphi(q)\bar{V}(q) \right\rangle \\
&= \left\langle \mathrm{d}\varphi(q)\mathrm{d}\varphi(q)^{-1} \operatorname{grad} f(\varphi(q)), \mathrm{d}\varphi(q)\bar{V}(q) \right\rangle \\
&= \left\langle \mathrm{d}\varphi(q)^{-1} \operatorname{grad} f(\varphi(p)), \bar{V}(q) \right\rangle.
\end{aligned}$$

Hence, combining the two above equality we conclude that $\operatorname{grad} f(\varphi(q)) = \mathrm{d}\varphi(q) \operatorname{grad} g(q)$. Moreover, the definition of the hessian of $f$ together with Lemma 4.3.2 yield

$$\begin{aligned}
\operatorname{hess} f(\varphi(q))\mathrm{d}\varphi(q)\bar{V}(q) &= \operatorname{hess} f(\varphi(q))V((\varphi(q)) \\
&= \nabla_{V(\varphi(q))} \operatorname{grad} f(\varphi(q)) \\
&= \mathrm{d}\varphi(p) \left(\bar{\nabla}_{\bar{V}(q)} \operatorname{grad} g(q)\right) \\
&= \mathrm{d}\varphi(q)\operatorname{hess} g(q)\bar{V}(q),
\end{aligned}$$

which implies that $\operatorname{hess} f(\varphi(q))\mathrm{d}\varphi(q) = \mathrm{d}\varphi(q)\operatorname{hess} g(q)$. Then, using again that $\varphi$ is an isometry, we have $\|\operatorname{hess} f(\varphi(q))\| = \|\operatorname{hess} g(q)\|$. Therefore, by using Lemma 3.1.3 the results follows. ■

The next result is an important property of isometries, its prove can be found in [59, Proposition 5.6.1, p. 196].

**Proposition 4.3.4** *Let $\mathcal{M}$ and $\bar{\mathcal{M}}$ be complete Riemannian manifolds. If $\varphi : \bar{\mathcal{M}} \to \mathcal{M}$ is a isometry and $\gamma$ is a geodesic in $\bar{\mathcal{M}}$, then $\varphi \circ \gamma$ is a geodesic in $\mathcal{M}$.*

The following result is a straight consequence of the definition of isometry and Proposition 4.3.4.

**Theorem 4.3.5** *Let $\mathcal{M}$, $\bar{\mathcal{M}}$ be Riemannian manifold and $\varphi : \bar{\mathcal{M}} \to \mathcal{M}$ an isometry. The function $g : \mathcal{M} \to \mathbb{R}$ is convex iff $f : \bar{\mathcal{M}} \to \mathbb{R}$, defined by $f(p) = (g \circ \varphi)(p)$, is convex.*

In the next example we change the metric of the Euclidean space $\mathbb{R}^n$ to prove, in particular, that the *extended Rosenbrock's banana function* is convex and has gradiente Lipschitz in $\mathbb{R}^n$ with this new metric. It is worth to pointed out that the convexity of this function in two dimension has been established in [67, p. 83].

**Example 4.3.6 (Rosenbrock's banana function class)** Let $f_j : \mathbb{R}^{2n} \to \mathbb{R}$ be a variant of the Rosenbrock's banana function, defined by

$$f_j(x_1, \ldots, x_{2n}) := \sum_{i=1}^{n} a_{ij} \left(x_{2i-1}^2 - x_{2i}\right)^2 + \left(x_{2i-1} - b_{ij}\right)^2, \qquad a_{ij} \in \mathbb{R}_{++}, \quad b_{ij} \in \mathbb{R},$$

for $j = 1, \ldots, m$. Denote $\bar{\mathcal{M}}$ as the Euclidean space $\mathbb{R}^{2n}$ with the usual metric. It is well known that $f_j$ is non-convex and its gradient is non-Lipschitz continuous in $\bar{\mathcal{M}}$. Endowing $\mathbb{R}^{2n}$ with the new Riemannian metric $\langle u, v\rangle := u^T G(x) v$, where $u, v \in \mathbb{R}^{2n}$ and $G(x)$ is the $2n \times 2n$ block diagonal matrix $G(x) = \operatorname{diag}(G_1(x), \ldots, G_n(x))$, where the blocks are given by

$$G_i(x) := \begin{pmatrix} 1 + 4x_{2i-1}^2 & -2x_{2i-1} \\ -2x_{2i-1} & 1 \end{pmatrix}, \qquad i = 1, \ldots, n,$$

and $x := (x_1, \ldots, x_{2n})$, we obtain a Riemannian manifold $\mathcal{M} := (\mathbb{R}^{2n}, G)$. Taking into account that the function $\varphi : \bar{\mathcal{M}} \to \mathcal{M}$ defined by

$$\varphi(z_1, \ldots, z_{2n}) = \left(z_1, z_1^2 - z_2, \ldots, z_{2n-1}, z_{2n-1}^2 - z_{2n}\right),$$

is an isometry, the Riemannian manifolds $\mathcal{M}$ is complete and has constant seccional curvature $K = 0$. On the other hand, $g_j : \bar{\mathcal{M}} \to \mathbb{R}$ defined by

$$g_j(z_1, \ldots, z_{2n}) := (f_j \circ \varphi)(z_1, \ldots, z_{2n}) = \sum_{i=1}^{n} a_{ij} z_{2i}^2 + (z_{2i-1} - b_{ij})^2, \qquad j = 1, \ldots, m,$$

is a quadratics function, which is convex with gradient vector field Lipschitz in $\bar{\mathcal{M}}$ with constant $L_j := \max\{2, 2a_{1j}, \ldots, 2a_{nj}\}$. Therefore, Theorem 4.3.5 and Theorem 4.3.3 imply, respectively, that $f_j$ is also convex and has gradient vector field Lipschitz continuous, with constant $L_j$, in $\mathcal{M}$. Let $F(x) = (f_1(x), \ldots, f_m(x))$ be the Rosenbrock's banana multiobjective function. Hence, $F$ is convex and Definition 4.1.1 implies that $\nabla F$ is componentwise Lipschitz continuous with constant $L = \max\{2, 2a_{11}, \ldots 2a_{nm}\}$. To apply Algorithm 2 for solving multiobjective optimization problem $\min\{F(x) \ : \ x \in \mathcal{M}\}$, we will calculate the exponential map and the gradient of $f_j$ in $\mathcal{M}$. The gradient of $f_j$ is given by $\operatorname{grad} f_j(x) = G(x)^{-1} f_j'(x)$, where $f_j'$ is the usual gradient of $f_j$. Given $z \in \bar{\mathcal{M}}$ the exponential map in $\bar{\mathcal{M}}$, $\overline{\exp}_z : T_z\bar{\mathcal{M}} \to \bar{\mathcal{M}}$, is given by $\overline{\exp}_z(\bar{v}) = z + \bar{v}$. Since $\varphi$ is an isometry, Proposition 4.3.4 implies that the exponential map in $\mathcal{M}$, $\exp_x : T_x\mathcal{M} \to \mathcal{M}$, is given by

$$\exp_x(v) = \varphi(\varphi^{-1}(x) + \mathrm{d}\varphi^{-1}(x)v).$$

Therefore, taking into account that $\varphi^{-1}(x) = (x_1, x_1^2 - x_2, \ldots, x_{2n-1}, x_{2n-1}^2 - x_{2n})$ and $\mathrm{d}\varphi^{-1}(x)v = (v_1, 2x_1v_1 - v_2, \ldots, v_{2n-1}, 2x_{2n-1}v_{2n-1} - v_{2n})$, we obtain that

$$\exp_x(v) = \big(x_1 + v_1, v_1^2 + x_2 + v_2, \ldots, p_{2n-1} + v_{2n-1}, v_{2n-1}^2 + x_{2n} + v_{2n}\big)$$

where $v := (v_1, \ldots, v_n) \in T_x\mathcal{M} \equiv \mathbb{R}^{2n}$.

**Example 4.3.7** Let $f_j : \mathbb{R}^n_{++} \to \mathbb{R}$ be defined by

$$f_j(x) := a_j \ln\left(\prod_{i=1}^{n} x_i^{u_{ij}} + b_j\right) - \sum_{i=1}^{n} w_{ij}\ln(x_i) + c_j \sum_{i=1}^{n} \ln^2(x_i),$$

where $x := (x_1, \ldots, x_n) \in \mathbb{R}^n_{++}$, $u_j := (u_{1j}, \ldots, u_{nj})^T \in \mathbb{R}^n_{+}$, $w_j := (w_{1j}, \ldots, w_{nj})^T \in \mathbb{R}^n_{+}$ and $a_j, b_j, c_j \in \mathbb{R}_{++}$, for all $j = 1, \ldots, m$. Denote $\bar{\mathcal{M}}$ as the Euclidean space $\mathbb{R}^n$ with the usual metric. The function $f$ is in general non-convex and its gradient is non-Lipschitz in $\bar{\mathcal{M}}$. Endowing $\mathbb{R}^n_{++}$ with the new Riemannian metric $\langle u, v\rangle := u^T G(x) v$, where $u, v \in T_x\mathcal{M}$ and $G(x)$ is the $n \times n$ diagonal matrix

$$G(x) := \operatorname{diag}\left(x_1^{-2}, x_2^{-2}, \ldots, x_n^{-2}\right),$$

we obtain the Riemannian manifold $\mathcal{M} := (\mathbb{R}^n_{++}, G)$. Since $\varphi : \bar{\mathcal{M}} \to \mathcal{M}$ defined by

$$\varphi(z_1, \ldots, z_n) = (e^{z_1}, \ldots, e^{z_n}), \tag{4.25}$$

is an isometry, then $\mathcal{M}$ is complete and has constant seccional curvature $K = 0$. The function $g_j : \bar{\mathcal{M}} \to \mathbb{R}$ defined by

$$g_j(z) := (f_j \circ \varphi)(z) = a_j \ln\left(e^{u_j^T z} + b_j\right) - w_j^T z + c_j z^T z, \qquad z := (z_1, \ldots, z_n)^T \in \bar{\mathcal{M}},$$

is convex and its gradient is Lipschitz in $\bar{\mathcal{M}}$ with constant $L_j \leq a_j u_j^T u_j / b_j + 2c_j$. Thus, Theorem 4.3.5 and Theorem 4.3.3 imply, respectively, that $f_j$ is also convex and has gradient Lipschitz in $\mathcal{M}$ with constant $L_j$. Therefore, the multiobjective function $F(x) = (f_1(x), \ldots, f_m(x))$ is convex and Definition 4.1.1 implies that $\nabla F$ is componentwise Lipschitz continuous with constant $L = \max\{L_1, \ldots, L_m\}$. The gradient of $f_j$ is given by

$$\operatorname{grad} f_j(x) = \operatorname{diag}(x)^2 f_j'(x), \qquad x \in \mathcal{M}$$

where $\operatorname{diag}(x) := \operatorname{diag}(x_1, \ldots, x_n)$ and $f_j'$ is the usual derivative. Using the isometry (4.25) Proposition 4.3.4 implies that the exponential map in $\mathcal{M}$, $\exp_x : T_x\mathcal{M} \to \mathcal{M}$, is given by $\exp_x(v) = \varphi(\varphi^{-1}(x) + \mathrm{d}\varphi^{-1}(x)v)$. Thus, considering that $\varphi^{-1}(x) = (\ln x_1, \ldots, \ln x_n)$ and $\mathrm{d}\varphi^{-1}(x)v = (x_1^{-1}v_1, \ldots, x_n^{-1}v_n)$, where $v := (v_1, \ldots, v_n)$, we conclude that

$$\exp_x(v) = \left(x_1 e^{\frac{v_1}{x_1}}, \ldots, x_n e^{\frac{v_n}{x_n}}\right), \qquad v := (v_1, \ldots, v_n) \in T_x\mathcal{M} \equiv \mathbb{R}^n.$$

## 4.4 Conclusions

In this chapter, the behavior of the steepest descent method for multiobjective optimization on Riemannian manifolds with lower bounded sectional curvature is analyzed. It would be interesting to study stochastic versions of this method. An interesting question to be also investigated is the extension and analysis of subgradient method in this new setting.

# Chapter 5

# Iteration-complexity of the subgradient method on Riemannian manifolds with lower bounded curvature

In this chapter we consider the subgradient method to solve the optimization problem defined by:

$$\min\{f(p) \ : \ p \in \mathcal{M}\}, \tag{5.1}$$

where the constraint set $\mathcal{M}$ is endowed with a structure of a *complete Riemannian manifold with lower bounded curvature* and $f : \mathcal{M} \to \overline{\mathbb{R}}$ is a *convex function*, where $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ denotes the extended real set numbers. Iteration-complexity bounds of the subgradient method with exogenous step-size and Polyak's step-size are established, completing and improving recent results on the subject.

## 5.1 Notations and auxiliary results

For any two points $p, q \in \mathcal{M}$, $\Gamma_{pq}$ denotes the set of all geodesic segments $\gamma : [0,1] \to \mathcal{M}$ with $\gamma(0) = p$ and $\gamma(1) = q$. The closed metric ball in $\mathcal{M}$ centered at the point $p \in \mathcal{M}$ with radius $r > 0$ is denoted by $B[p, r]$. Let $\Omega$ be a subset of $\mathcal{M}$. We use *int* $\Omega$ to denote the interior of $\Omega$. A function $f : \mathcal{M} \to \overline{\mathbb{R}}$ is said to be proper if its domain $domf = \{p \in \mathcal{M} : f(p) \neq +\infty\}$ is nonempty, where $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ denotes the extended real set numbers. We use $\Gamma^{f}_{pq}$ to denote the set of all $\gamma \in \Gamma_{pq}$ such that $\gamma \subseteq domf$. A proper function $f : \mathcal{M} \to \overline{\mathbb{R}}$ is said to be *convex* on $\mathcal{M}$ if $domf$ is weakly convex and for any $p, q \in domf$ and $\gamma \in \Gamma^{f}_{pq}$ the composition $f \circ \gamma : [0,1] \to \overline{\mathbb{R}}$ is a convex function on $[0,1]$ i.e.,

$$f \circ \gamma(t) \leq (1-t)f(p) + tf(q), \qquad \forall\, t \in [0,1],$$

see [69]. The *subdifferential* of a convex function $f : \mathcal{M} \to \overline{\mathbb{R}}$ at $p \in domf$ is defined by

$$\partial f(p) := \left\{ s \in T_p\mathcal{M} : f(q) \geq f(p) + \langle s, \gamma'(0) \rangle, \quad \forall\, q \in domf,\ \gamma \in \Gamma^f_{pq} \right\}. \tag{5.2}$$

We remark that the subdiffential set $\partial f(p)$ is nonempty in all $p \in int\, domf$; see [69, Proposition 2.5]. *In this chapter all functions $f : \mathcal{M} \to \overline{\mathbb{R}}$ are assumed to be convex and lower semicontinuous on $\mathcal{M}$.* The following result is also proved in [69, Proposition 2.5].

**Proposition 5.1.1** *Let $\{p_k\} \subset \mathcal{M}$ a bounded sequence. If the sequence $\{s_k\}$ is such that $s_k \in \partial f(p_k)$, for each $k \in \mathbb{N}$, then $\{s_k\}$ is also bounded.*

The following lemma plays an important role in the next sections. Its proof will be omitted here, but it can be obtained, with some minor technical adjustments, following the ideas of Lemma 2.0.4 together with the inequality in (5.2). Remember the notation $\hat{\kappa} := \sqrt{|\kappa|}$.

**Lemma 5.1.2** *Let $p \in \operatorname{int} \operatorname{dom} f$, $0 \neq s \in \partial f(p)$, and let $\gamma : [0, +\infty) \to \mathcal{M}$ be the geodesic defined by $\gamma(t) = exp_p\left(-ts/\left\|s\right\|\right)$. Then, for any $t \in [0, +\infty)$ and $q \in \operatorname{dom} f$ there holds*

$$\cosh(\hat{\kappa} d(\gamma(t), q)) \leq \cosh(\hat{\kappa} d(p, q)) + \hat{\kappa} \cosh(\hat{\kappa} d(p, q)) \sinh(t\hat{\kappa}) \left[ \frac{t}{2} - \frac{\tanh(\hat{\kappa} d(p, q))}{\hat{\kappa} d(p, q)} \frac{f(p) - f(q)}{\|s\|} \right]$$

*and, consequently, the following inequality holds*

$$d^2(\gamma(t), q) \leq d^2(p, q) + \frac{\sinh(\hat{\kappa} t)}{\hat{\kappa}} \left[ \frac{\hat{\kappa} d(p, q)}{\tanh(\hat{\kappa} d(p, q))} t - \frac{2}{\|s\|} \left( f(p) - f(q) \right) \right].$$

We end this section by recalling the concept of Lipschitz continuity of a function. A proper function $f : \mathcal{M} \to \overline{\mathbb{R}}$ is said to be *Lipschitz continuous with constant* $\tau \geq 0$ in $\Omega \subset \mathcal{M}$ if $|f(p) - f(q)| \leq \tau\, d(p, q)$, for any $p, q \in \Omega$.

## 5.2 Iteration-complexity of the subgradient method

In this section, we state the Riemannian subgradient method to solve (5.1) and the strategies for choosing the step-size that will be used in our analysis. Let $f : \mathcal{M} \to \overline{\mathbb{R}}$ be a convex function, $\Omega^*$ be the *solution set* of the problem (5.1) and $f^* := \inf_{x \in \mathcal{M}} f(x)$ be the *optimum value* of $f$. *In our analysis we do not assume that $\Omega^*$ is nonempty, except when explicitly stated.* The statement of *Riemannian subgradient algorithm* to solve the problem (5.1) is as follows.

**Algorithm 3:** Subgradient algorithm in a Riemanian manifold $\mathcal{M}$

**Step 0.** Let $p_0 \in int\, dom f$. Set $k = 0$.

**Step 1.** If $s_k = 0$, then **stop**; otherwise, choose a step-size $s_k \in \partial f(p_k)$, $t_k > 0$ and compute

$$p_{k+1} := \exp_{p_k}\left(-t_k \frac{s_k}{\|s_k\|}\right); \tag{5.3}$$

**Step 2.** Set $k \leftarrow k + 1$ and proceed to **Step 1**.

In the following we present two different strategies for choosing the step-size $t_k > 0$ in Algorithm 3.

**Strategy 7 (Exogenous step-size)**

$$t_k > 0, \qquad \sum_{k=0}^{\infty} t_k = +\infty, \qquad \sigma := \sum_{k=0}^{\infty} t_k^2 < +\infty. \tag{5.4}$$

The step-size in Strategy 7 have been used in several chapter for analyzing subgradient method; see, for example, [22, 33, 70].

**Strategy 8 (Polyak's step-size)** *Assume that $p_0 \in \operatorname{int}\operatorname{dom} f$, $\Omega^* \neq \varnothing$ and set*

$$t_k = \alpha \frac{f(p_k) - f^*}{\|s_k\|}, \qquad 0 < \alpha < 2\frac{\tanh(\hat{\kappa} d_0)}{\hat{\kappa} d_0}, \qquad d_0 := d(p_0, \Omega^*), \tag{5.5}$$

*where $d(p_0, \Omega^*) := \inf\{d(p_0, q);\ q \in \Omega^*\} > 0$.*

This step-size in Strategy 8 was introduced in [60] and has been used in [12, 15, 69].

**Remark 5.2.1** Since the function $(0, +\infty) \mapsto \tanh(t)/t$ is decreasing, then given an estimate $\hat{d} > d_0$ we can chose $0 < \alpha < 2\tanh(\hat{\kappa}\hat{d})/(\hat{\kappa}\hat{d}))$ in Strategy 8. For Riemannian manifold with non-negative curvature, the second inequality in (5.5) holds for all $\kappa < 0$. Due to $\lim_{t\to 0} \tanh(t)/t = 1$, letting $\kappa$ goes to 0, we can choose $0 < \alpha < 2$.

*From now on we assume that the sequence $\{p_k\}$ generated by Algorithm 3 with the two above strategies for choosing the step-size is well defined and is infinite.*

**Remark 5.2.2** Note that if $dom f = \mathcal{M}$ then $\partial f(p) \neq \varnothing$, for all $p \in \mathcal{M}$ and, consequently, the sequence $\{p_k\}$ is well defined. In [70, Theorem 3.1] an asymptotic convergence analysis was established, under the assumption that suitable sets are contained in $int\, dom f$ and that the set $\Omega^*$ is nonempty. The author proves that the sequence generated by the Algorithm 3 is well defined and converges to an element of $\Omega^*$. It is worth to point out that our asymptotic convergence analysis of Algorithm 3, with Strategy 7 for choosing the step-size, we do not assume that $\Omega^*$ is nonempty. In this sense, our results improve the ones of [70, Theorem 3.1].

### 5.2.1 Subgradient method with exogenous step-size

In this section we assume that the sequence $\{p_k\}$ is generated by Algorithm 3 with Strategy 7 for choosing the step-size. To proceed with the analysis of Algorithm 3 we need some preliminaries. Firstly we define

$$\Omega := \left\{ q \in \mathcal{M} : \; f(q) \leq \inf_k f(p_k) \right\}.$$

Note that $\Omega \subset dom f$. It is worth mentioning that, in principle, the set $\Omega$ can be empty. Our first task is to prove that the sequence $\{p_k\}$ is bounded.

**Lemma 5.2.3** *If $\Omega \neq \varnothing$ then, for each $q \in \Omega$ there holds*

$$d(p_{k+1}, q) \leq \frac{1}{\hat{\kappa}} \cosh^{-1}\left( \cosh\left(\hat{\kappa} d(p_0, q)\right) e^{\frac{1}{2}\hat{\kappa}\sqrt{\sigma}\sinh\left(\hat{\kappa}\sqrt{\sigma}\right)} \right), \tag{5.6}$$

*for all $k = 0, 1, \ldots$.*

*Proof.* Applying first inequality of Lemma 5.1.2 with $t = t_k$, $p = p_k$ and $p_{k+1} = \gamma(t_k)$ and taking into account that $q \in \Omega$ we conclude that

$$\cosh\left(\hat{\kappa} d(p_{k+1}, q)\right) \leq \cosh\left(\hat{\kappa} d(p_k, q)\right) \left[ 1 + |\kappa| t_k^2 \frac{\sinh\left(\hat{\kappa} t_k\right)}{2\hat{\kappa} t_k} \right], \qquad k = 0, 1, \ldots.$$

Using definition of $\sigma$ in (5.4) we have $t_k \leq \sqrt{\sigma}$, for all $k = 0, 1, \ldots$. Since the map $(0, +\infty) \ni t \mapsto \sinh(t)/t$ is increasing, it follows from the last inequality that

$$\cosh\left(\hat{\kappa} d(p_{k+1}, q)\right) \leq \cosh\left(\hat{\kappa} d(p_k, q)\right) \left[ 1 + a t_k^2 \right], \qquad k = 0, 1, \ldots,$$

where $a := \hat{\kappa}(\sinh(\hat{\kappa}\sqrt{\sigma}))/(2\sqrt{\sigma})$. Note that the last inequality implies that

$$\cosh\left(\hat{\kappa} d(p_{k+1}, q)\right) \leq \cosh\left(\hat{\kappa} d(p_k, q)\right) e^{a t_k^2}, \qquad k = 0, 1, \ldots.$$

Therefore, we have $\cosh(\hat{\kappa} d(p_{k+1}, q)) \leq \cosh(\hat{\kappa} d(p_0, q)) e^{a\sigma}$, which is equivalent to (5.6) and the proof is concluded. ∎

In the next result we apply Lemmas 5.1.2 and 5.2.3 to derive an inequality that plays an important role in our analysis, which is a generalization of the one obtained in [33, Lemma 4.1]. In the linear setting, this inequality is of fundamental importance to analyze the subgradient method; see, for example, [22]. It is worth noting that it was obtained in [73] for a specific function, namely, the mean function. For stating the next result, for each $q \in \Omega$, we define

$$C_{q,\kappa} := \frac{\sinh\left(\hat{\kappa}\sqrt{\sigma}\right)}{\hat{\kappa}\sqrt{\sigma}} \left[ 1 + \cosh^{-1}\left( \cosh(\hat{\kappa} d(p_0, q)) e^{\frac{1}{2}\hat{\kappa}\sqrt{\sigma}\sinh\left(\hat{\kappa}\sqrt{\sigma}\right)} \right) \right]. \tag{5.7}$$

It is important to note that $C_{q,\kappa}$ is well defined only under the assumption $\Omega \neq \varnothing$.

**Lemma 5.2.4** *If $\Omega \neq \varnothing$ then, for each $q \in \Omega$ there holds*

$$d^2(p_{k+1}, q) \leq d^2(p_k, q) + C_{q,\kappa} t_k^2 + 2\,\frac{t_k}{\|s_k\|}[f(q) - f(p_k)], \qquad s_k \in \partial f(p_k), \qquad k = 0, 1, \ldots.$$

*Proof.* Applying first inequality of Lemma 5.1.2 with $t = t_k$, $p = p_k$ and $p_{k+1} = \gamma(t_k)$, and taking into account that $q \in \Omega$, we conclude that

$$d^2(p_{k+1}, q) \leq d^2(p_k, q) + \frac{\sinh(\hat{\kappa} t_k)}{\hat{\kappa} t_k} \left[ \frac{\hat{\kappa} d(p_k, q)}{\tanh(\hat{\kappa} d(p_k, q))} t_k^2 + \frac{2t_k}{\|s_k\|} [f(q) - f(p_k)] \right],$$

for all $k = 0, 1, \ldots$. On the other hand, $t/\tanh(t) \leq 1 + t$, for all $t \geq 0$ and the map $(0, +\infty) \ni t \mapsto \sinh(t)/t$ is increasing and bounded below by 1. Thus, taking into account that $t_k \leq \sqrt{\sigma}$ and $f(q) - f(p_k) \leq 0$, for all $k = 0, 1, \ldots$, we conclude that

$$d^2(p_{k+1}, q) \leq d^2(p_k, q) + \frac{\sinh(\hat{\kappa}\sqrt{\sigma})}{\hat{\kappa}\sqrt{\sigma}} [1 + \hat{\kappa} d(p_k, q)]\, t_k^2 + \frac{2t_k}{\|s_k\|} [f(q) - f(p_k)],$$

for all $k = 0, 1, \ldots$. Therefore, combining Lemma 5.2.3 with (5.7) the desired inequality follows and the proof is concluded. ■

**Remark 5.2.5** For Riemannian manifold with non-negative curvature, the inequality in Lemma 5.2.4 holds for all $\kappa < 0$. Since $\lim_{\kappa \to 0} C_{q,\kappa} = 1$, the inequality of the Lemma 5.2.4 merges into the inequality [33, Lemma 4.1].

Now we are ready to prove the main result of this section.

**Theorem 5.2.6** *Assume that $\Omega^* \neq \varnothing$ and $f : \mathcal{M} \to \overline{\mathbb{R}}$ is Lipschitz continuous with constant $\tau \geq 0$. Then, for all $p_* \in \Omega^*$ and every $N \in \mathbb{N}$, the following inequality holds*

$$\min \{ f(p_k) - f^* \;:\; k = 0, 1, \ldots, N \} \leq \tau \frac{d^2(p_0, p_*) + C_{p_*,\kappa} \sum_{k=0}^{N} t_k^2}{2 \sum_{k=0}^{N} t_k}. \tag{5.8}$$

*Proof.* Let $p_* \in \Omega^*$. Since $\Omega^* \subset \Omega$, applying Lemma 5.2.4 with $q = p_*$, we obtain

$$d^2(p_{k+1}, p_*) \leqslant d^2(p_k, p_*) + C_{p_*,\kappa} t_k^2 + 2\frac{t_k}{\|s_k\|}[f^* - f(p_k)], \qquad s_k \in \partial f(p_k),$$

for all $k = 0, 1, \ldots$. Hence, performing the sum of the above inequality for $k = 0, 1, \ldots, N$, after some algebraic manipulations, we have

$$2 \sum_{k=0}^{N} \frac{t_k}{\|s_k\|} [f(p_k) - f^*] \leq d^2(p_0, p_*) - d^2(p_{N+1}, p_*) + C_{p_*,\kappa} \sum_{k=0}^{N} t_k^2.$$

Since $f$ is Lipschitz continuous with constant $\tau \geq 0$, we have $\|s_k\| \leq \tau$, for all $k = 0, 1, \ldots$. Therefore,

$$\frac{2}{\tau} \min \{ f(p_k) - f^* :\; k = 0, 1, \ldots, N \} \sum_{k=0}^{N} t_k \leq d^2(p_0, p_*) + C_{p_*,\kappa} \sum_{k=0}^{N} t_k^2,$$

which is equivalent to the desired inequality. ■

**Remark 5.2.7** Note that, for Riemannian manifold with non-negative curvature the inequality in (5.8) holds for all $\kappa < 0$. Since $\lim_{\kappa\to 0} C_{q,\kappa} = 1$, Theorem 5.2.6 is reduced to [12, Theorem 3.3].

We remark that in the first part of the next theorem we do not assume that $\Omega^* \neq \varnothing$. Additionally, it is worth to point out that the second part was first obtained in [70]. Since it is an immediate consequence of the first part and Lemma 5.2.4, we decide to include its proof here.

**Theorem 5.2.8** *The following equality holds*

$$\liminf_{k} f(p_k) = f^*. \tag{5.9}$$

*In addition, if $\Omega^* \neq \varnothing$ then the sequence $\{p_k\}$ converges to a point $p_* \in \Omega^*$.*

*Proof.* Assume by contradiction that $\liminf_k f(p_k) > f^*$. In this case, we have $\Omega \neq \varnothing$. Thus, from Lemma 5.2.3, we conclude that $\{p_k\}$ is bounded and, consequently, by using Proposition 5.1.1, the sequence $\{s_k\}$ is also bounded. Let $C_1 > 0$ such that $\|s_k\| < C_1$, for $k = 0, 1, \ldots$. On the other hand, letting $q \in \Omega$, there exist $C_2 > 0$ and $k_0 \in \mathbb{N}$ such that $f(q) < f(p_k) - C_2$, for all $k \geq k_0$. Hence, using Lemma 5.2.4 and considering that $\|s_k\| < C_1$, for $k = 0, 1, \ldots$, we have

$$d^2(p_{k+1}, q) \leq d^2(p_k, q) + C_{q,\kappa} t_k^2 - 2\frac{C_2}{C_1} t_k, \qquad k = k_0, k_0 + 1, \ldots.$$

Consider $\ell \in \mathbb{N}$. Thus, from the last inequality, after some calculations, we conclude that

$$\frac{2\,C_2}{C_1} \sum_{j=k_0}^{\ell+k_0} t_j \leq d^2(p_{k_0}, q) - d^2(p_{k_0+\ell}, q) + C_{q,\kappa} \sum_{j=k_0}^{\ell+k_0} t_j^2 \leq d^2(p_{k_0}, q) + C_{q,\kappa} \sum_{j=k_0}^{\ell+k_0} t_j^2.$$

Since the last inequality holds for all $\ell \in \mathbb{N}$, then using the inequality in (5.4) we have a contraction. Therefore, (5.9) holds.

For proving the last statement, let us assume that $\Omega^* \neq \varnothing$. In this case, we have $\Omega \neq \varnothing$ and, from Lemma 5.2.3, the sequence $\{p_k\}$ is bounded. Moreover, Lemma 5.2.4 implies, in particular, that $\{p_k\}$ is quasi-Féjer convergent to $\Omega$. The equality (5.9) implies that $\{f(p_k)\}$ possesses a decreasing monotonous subsequence $\{f(p_{k_j})\}$ such that $\lim_{j\to\infty} f(p_{k_j}) = f^*$. We can assume that $\{f(p_k)\}$ is decreasing, monotonous and converges to $f^*$. Being bounded, the sequence $\{p_k\}$ possesses a convergent subsequence $\{p_{k_\ell}\}$. Let us say that $\lim_{\ell\to\infty} p_{k_\ell} = p_*$, which by the continuity of $f$ implies $f(p_*) = \lim_{\ell\to\infty} f(p_{k_\ell}) = f^*$, and then $p_* \in \Omega$. Hence, $\{p_k\}$ has an cluster point $p_* \in \Omega$, and due to $\{p_k\}$ be quasi-Féjer convergent to $\Omega$, it follows from Theorem 2.0.8 that the sequence $\{p_k\}$ converges to $p_*$. ■

### 5.2.2 Subgradient method with Polyak step-size

In this section, we assume that $\Omega^* \neq \varnothing$ and $\{p_k\}$ is generated by Algorithm 3 with Strategy 8 for choosing the step-size. Let us define

$$C_{\kappa,d_0} := \frac{2}{\alpha} - \frac{\hat{\kappa} d_0}{\tanh(\hat{\kappa} d_0)} > 0, \tag{5.10}$$

where $\alpha$ and $d_0$ are defined in (5.5).

**Remark 5.2.9** Since $\lim_{t\to 0} \tanh(t)/t = 1$, we conclude that for Riemannian manifolds with nonnegative curvature, namely, for $\kappa = 0$, (5.10) become $C_{\kappa,d_0} \equiv 2/\alpha - 1 > 0$.

In the next result, we apply Lemma 5.1.2 to obtain an inequality that plays an important role in our analysis. Before state this result, we set

$$\bar{q} \in \Omega^* \quad \text{such that} \quad d_0 = d(p_0, \bar{q}). \tag{5.11}$$

**Lemma 5.2.10** *Let $\bar{q} \in \Omega^*$ satisfying* (5.11). *Then the following inequality holds*

$$d^2(p_{k+1}, \bar{q}) \leq d^2(p_k, \bar{q}) - C_{\kappa,d_0} \alpha^2 \frac{[f(p_k) - f^*]^2}{\|s_k\|^2}, \qquad k = 0, 1, \ldots.$$

*Proof.* First we are going to prove that $d(p_k, \bar{q}) \leq d_0$, for all $k = 0, 1, \ldots$. The proof will be made by induction. For $k = 0$ is immediate. Assume that $d(p_k, \bar{q}) \leq d_0$. Using the second inequality of Lemma 5.1.2 with $q = \bar{q}$, $t = t_k$, $p = p_k$, $s = s_k$, $p_{k+1} = \gamma(t_k)$ and considering that $f^* = f(\bar{q})$, we obtain

$$d^2(p_{k+1}, \bar{q}) \leq d^2(p_k, \bar{q}) + \frac{\sinh(\hat{\kappa} t_k)}{\hat{\kappa} t_k} \left[ \frac{\hat{\kappa} d(p_k, \bar{q})}{\tanh(\hat{\kappa} d(p_k, \bar{q}))} t_k^2 + \frac{2 t_k}{\|s_k\|} [f^* - f(p_k)] \right].$$

Since the map $(0, +\infty) \ni t \mapsto t/\tanh(t)$ is increasing, using the assumption $d(p_k, \bar{q}) \leq d_0$ and definition of $t_k$ in (5.5), the last inequality becomes

$$d^2(p_{k+1}, \bar{q}) \leq d^2(p_k, \bar{q}) + \frac{\sinh(\hat{\kappa} t_k)}{\hat{\kappa} t_k} \left[ \frac{\hat{\kappa} d_0}{\tanh(\hat{\kappa} d_0)} - \frac{2}{\alpha} \right] \alpha^2 \frac{[f(p_k) - f^*]^2}{\|s_k\|^2}. \tag{5.12}$$

Thus, the inequalities in (5.5) imply that $d(p_{k+1}, \bar{q}) \leq d(p_k, \bar{q}) \leq d_0$ and the induction is concluded. Hence, $d(p_k, \bar{q}) \leq d_0$, for all $k = 0, 1, \ldots$. Therefore, we can also prove that (5.12) hods, for all $k = 0, 1, \ldots$. Taking into account that $\sinh(\hat{\kappa} t_k)/(\hat{\kappa} t_k) \geq 1$, the combination of second inequality in (5.5), (5.10) and (5.12) yield the desired inequality. ∎

**Remark 5.2.11** Since $\lim_{t\to 0} \tanh(t)/t = 1$ and $\lim_{t\to 0} \sinh(t)/t = 1$, then by using similar idea considered in the proof of Lemma 5.2.10, we can show that, for Riemannian manifolds with nonnegative curvature, holds $d^2(p_{k+1}, q) \leq d^2(p_k, q) - (2/\alpha - 1)t_k^2$, for all $k = 0, 1, \ldots$ and all $q \in \Omega^*$.

The next result presents an iteration-complexity bound for the subgradient method with the Polyak's step-size rule.

**Theorem 5.2.12** *Assume that $f : \mathcal{M} \to \overline{\mathbb{R}}$ is Lipschitz continuous with constant $\tau \geq 0$. Let $\bar{q} \in \Omega^*$ satisfying* (5.11). *Then, for every $N \in \mathbb{N}$, there holds*

$$\sum_{k=0}^{N}[f(p_k) - f^*]^2 \leq \frac{\tau^2 d^2(p_0, \bar{q})}{C_{\kappa,d_0}}. \tag{5.13}$$

*As a consequence,*

$$\min\{f(p_k) - f^* \ : \ k = 0, 1, \ldots, N\} \leq [\tau d(p_0, \bar{q})]/\sqrt{C_{\kappa,d_0}(N+1)}. \tag{5.14}$$

*Proof.* Since $f$ is Lipschitz continuous with constant $\tau \geq 0$, we have $\|s_k\| \leq \tau$, for all $k = 0, 1, \ldots$. Thus, it follows from Lemma 5.2.10 that

$$[f(p_k) - f^*]^2 \leq \frac{\tau^2}{C_{\kappa,d_0}\alpha^2}[d^2(p_k, \bar{q}) - d^2(p_{k+1}, \bar{q})], \qquad k = 0, 1, \ldots.$$

Performing the sum of the above inequality for $k = 0, 1, \ldots, N$, we obtain (5.13). The second statement of the theorem is an immediate consequence of the first one. ∎

**Remark 5.2.13** It is worth noting that if $\kappa = 0$ we have $C_{q,\kappa} = 1$ and then Theorem 5.2.12 merges into the inequality [12, Theorem 3.4].

**Theorem 5.2.14** *The following equality holds* $\lim_{k\to\infty} f(p_k) = f^*$. *Consequently, all cluster point of $\{p_k\}$ is a solution of* (5.1).

*Proof.* Letting $N$ goes to $+\infty$ in (5.13), we conclude that $\lim_{k\to\infty} f(p_k) = f^*$. It follows from Lemma 5.2.10 that $\{p_k\}$ is bounded. For concluding the proof, let $\bar{p}$ be an accumulation point of $\{p_k\}$ and $\{p_{k_i}\}$ a subsequence of $\{p_k\}$ such that $\lim_{k_i\to+\infty} p_{k_i} = \bar{p}$. Therefore, $f(\bar{p}) = \lim_{k_i\to\infty} f(p_{k_i}) = f^*$ and then $\bar{p} \in \Omega^*$. ∎

**Corollary 5.2.15** *For $\kappa = 0$ the sequence $\{p_k\}$ converges to a point $q \in \Omega^*$.*

*Proof.* Lemma 5.2.10 implies that $\{p_k\}$ is bounded. As a consequence, $\{p_k\}$ has at least one cluster point. Thus, by Theorem 5.2.14, it follows that there exists a subsequence $\{p_{k_i}\}$ of $\{p_k\}$ converging to a $q \in \Omega^*$. Hence, $\lim_{k_i\to\infty} d(p_{k_i}, q) = 0$. Since the inequality of Remark 5.2.11 implies that $\{d(p_k, q)\}$ is monotonic decreasing, it holds that $\lim_{k\to\infty} d(p_k, q) = 0$, completing the proof. ∎

## 5.3 Numerical examples

In this section, we numerically illustrate the results on complexity-iteration bounds of Section 5.2. For this aim, we consider the convex feasibility problem in Riemannian setting which consists of finding a point $p \in \mathcal{M}$ such that

$$p \in C := \bigcap_{i=1}^{m} C_i, \qquad C_i := \{p \in \mathcal{M} : \; f_i(p) \leq 0\}, \tag{5.15}$$

where $f_i : \mathcal{M} \to \mathbb{R}$ is convex, for all $i = 1, ..., m$. This problem can be equivalently rewritten as an optimization problem (5.1) where $f : \mathcal{M} \to \mathbb{R}$ is given by

$$f(p) := \max\left\{f_1(p), \ldots, f_m(p), 0\right\}.$$

Note that $f(x) \geq 0$ for all $x \in \mathcal{M}$. If $C \neq \varnothing$, then $C = \{p \in \mathcal{M} : \; f(p) = 0\}$. Thus, $C$ is the solution set of the problem (5.1) and $f^* = 0$. Now, if the interior of $C$ is nonempty, i.e, $\mathrm{int}C \neq \varnothing$, then there exist $\epsilon > 0$ and $\hat{x} \in \mathcal{M}$ such that $f_i(\hat{x}) \leq -\epsilon$, for all $i = 1, \ldots, m$. In this case, defining

$$f(p) := \max\left\{f_1(p), \ldots, f_m(p), -\epsilon\right\}, \tag{5.16}$$

the solution set of the problem (5.1) is contained in $\mathrm{int}C$ and $f^* = -\epsilon$.

Our examples consist of convex feasibility problems (5.15) where $C$ has nonempty interior. Let us explain how the examples were generated. Let $\mathcal{M}$ be a Riemannian Manifold with sectional curvature bounded from above by $K$ and set

$$\rho_K := \frac{1}{2}\min\left\{\mathrm{inj}\,\mathcal{M}, \; \frac{\pi}{2\sqrt{K}}\right\},$$

where $\mathrm{inj}\,\mathcal{M}$ is the injectivity radius of $\mathcal{M}$, with the convention that $1/\sqrt{K} = +\infty$ for $K \leq 0$; see [64, pag. 110]. Let $d$ be the associated Riemannian distance. Set $q \in \mathcal{M}$, and choose $r > 0$ and $v_1, ..., v_m \in T_q\mathcal{M}$ in such a way that

$$a_i := \exp_q\left(r\frac{v_i}{\|v_i\|}\right) \in B(q, \rho), \tag{5.17}$$

for all $i = 1, \ldots, m$, and some $\rho < \rho_K$. Since $d(a_i, q) = r$, we conclude that $a_i \in \partial B[q, r]$, where $\partial B[q, r]$ denotes the boundary of $B[q, r]$, for all $i = 1, \ldots, m$. Let $\epsilon > 0$ and define $f_i : \mathcal{M} \to \mathbb{R}$ by

$$f_i(p) := d(p, a_i) - r - \epsilon,$$

for each $i = 1, \ldots, m$, and consider $f$ given by (5.16). In this case, we have $B[q, \epsilon] \subset C$, $f^* = -\epsilon$, and $d_0 \leq d(p_0, q)$ where $d_0$ is defined in (5.5). Moreover, $f$ is Lipschitz continuous with constant $\tau = 1$. Given $p \in \mathcal{M}$, it follows that

$$-\sum_{j \in I_p} \alpha_j \frac{\exp_p^{-1} a_j}{d(a_j, p)} \in \partial f(p),$$

where $I_p := \{j : p \neq a_j,\ j = 1, \ldots, m\}$ and $\sum_{j \in I_p} \alpha_j = 1$, see [14, 73]. We generated two examples with different types of Riemannian manifolds $\mathcal{M}$ as described below.

**Example 5.3.1 (Positive definite symmetric matrices)** Let $\mathbb{P}^n$ and $\mathbb{P}^n_{++}$ be the set of symmetric matrices and the set of positive definite symmetric matrices, respectively. Let $\mathcal{M} := (\mathbb{P}^n_{++}, \langle \cdot, \cdot \rangle)$ be the Riemannian manifold endowed with the Riemannian metric given by

$$\langle U, V \rangle := \operatorname{tr}(VX^{-1}UX^{-1}), \qquad X \in \mathcal{M}, \qquad U, V \in T_X\mathcal{M} \approx \mathbb{P}^n.$$

Let $d$ be the Riemannian distance defined in $\mathcal{M} := (\mathbb{P}^n_{++}, \langle \cdot, \cdot \rangle)$, see (3.47).

We set $n = 10$, $m = 10$, $r = 1$, and $\epsilon = 0.1$. We random generated matrix $q \in \mathbb{P}^n_{++}$ and the starting point $p_0 \in \mathbb{P}^n_{++}$ with eigenvalues belonging to $(0, 100)$, and matrices $v_1, ..., v_m \in \mathbb{P}^n$ with eigenvalues belonging to $(-100, 100)$. Then, matrices $a_1, ..., a_m \in \mathbb{P}^n_{++}$ were generated according to (5.17).

**Example 5.3.2 (Sphere)** Let $\mathbb{S} := \{x \in \mathbb{R}^n :\ \|x\| = 1\}$ be the $(n-1)$-dimensional unit sphere. Endowing the sphere $\mathbb{S}$ with the Euclidean metric $\langle \cdot, \cdot \rangle$ we obtain a complete Riemannian manifold with curvature equal to 1, which will be also denoted by $\mathbb{S}$. The tangent plane at $x \in \mathbb{S}$ is given by $T_x\mathbb{S} := \{v \in \mathbb{R}^n :\ \langle v, x \rangle = 0\}$ and the exponential mapping $\exp_x : T_x\mathcal{M} \to \mathcal{M}$ in assigned by

$$\exp_x v := \begin{cases} \cos(\|v\|)x + \sin(\|v\|)\frac{v}{\|v\|}, & v \neq 0, \\ x, & v = 0. \end{cases}$$

The inverse of the exponential mapping $\exp_x^{-1} : \mathcal{M} \to T_x\mathcal{M}$ is given by

$$\exp_x^{-1} y := \frac{\arccos \langle x, y \rangle}{\sqrt{1 - \langle x, y \rangle^2}}(I - xx^T)y.$$

The Riemannian distance between $x, y \in \mathbb{S}$ is given by $d(x, y) = \arccos \langle x, y \rangle$, for more details, see, for example, [30].

We set $n = 200$, $m = 50$, $r = \pi/16$, and $\epsilon = 0.001$. We defined $q = (1, \ldots, 1)/\sqrt{n}$ and random generated vectors $v_1, ..., v_m \in T_q\mathbb{S}$. Then, vectors $a_1, ..., a_m \in \mathbb{S}$ were generated according to (5.17). The starting point $p_0 \in \mathbb{S}$ was generated by taking a random vector $v \in T_q\mathbb{S}$ and setting $p_0 = \exp_q\left(\lambda \frac{\pi}{8} \frac{v}{\|v\|}\right)$, where $\lambda \in (0, 1)$.

We coded Algorithm 3 in Matlab and run it on the above examples. For Example 5.3.1 we used the exogenous step-size given by $t_k = 1/(k + 1)$ for all $k = 0, 1, \ldots$, while for Example 5.3.2 we adopted the Polyak's step-size with $\alpha = 1.9999 \times \tanh(d(p_0, q))/d(p_0, q)$. For each example, since $f^* = -\epsilon$, by Theorems 5.2.8 and 5.2.14 respectively, there exists $k_0$ such that $p_k \in C$ for all $k \geq k_0$. Therefore, these convex feasibility problems are solved by Algorithm 3 in a finite number of iterations. Indeed, Algorithm 3 found a feasible point with 55 and 41 iterations for Examples 5.3.1 and 5.3.2, respectively. Figure 5.1 (a) corresponds to Example 5.3.1 and reports the function values of the left and right hand sides of inequality (5.8) for each iteration of Algorithm 3. In its turn, Figure 5.1 (b) is related to Example 5.3.2 and illustrates the iteration-complexity bound given by (5.14).

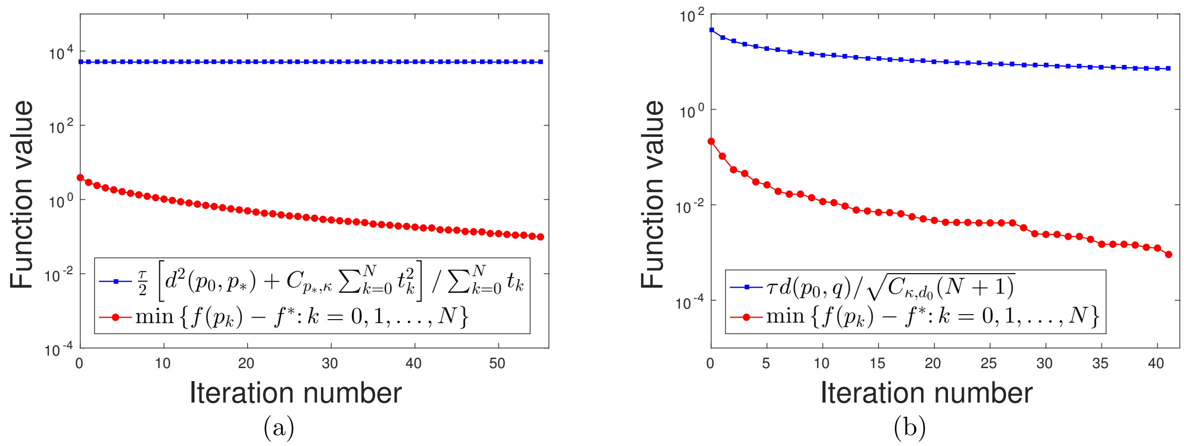



Figure 5.1: Iteration-complexity bound for the Riemannian subgradient method with: (a) exogenous step-size applied to Example 5.3.1 – Theorem 5.2.6; (b) Polyak's step-size applied to Example 5.3.2 – Theorem 5.2.12.

As can be seen in Figure 5.1, inequalities (5.8) and (5.14) in Theorems 5.2.8 and 5.2.14 are met for all iterations of Algorithm 3, illustrating the practical reliability of our iteration-complexity results.

## 5.4 Conclusions

In this chapter, we analyzed the iteration-complexity of subgradient method with exogenous step-size and Polyak's step-size. In general, the Polyak's step-size has a better performance than the exogenous step-size, but the choice of exogenous step-size is also interesting because it does not depend on any data computed during the algorithm, being important in large scale optimization problems. Since the feasibility and optimization problems are close related, this chapter complements the understanding of the subgradient algorithm in this settings. Finally, we remark that for Riemannian manifolds with curvature unbounded below, perhaps another strategy for the step will be need since is not possible to control the distance between the geodesics. Indeed, if the curvature is positive the geodesics emanating from the same point tend to approximate one each other, the contrary occurs if the curvature is negative.

# Chapter 6

# Final remarks

In this thesis, the behavior of the gradient and subgradient methods for convex optimization problems on Riemannian manifolds with lower bounded sectional curvature was analyzed. Most of the theory made here is based on the extension of known results on manifolds with non-negative curvature. This extension is important because there are a huge number of known Riemannian manifolds with lower bounded curvature, for instance: compact manifolds and manifolds with constant curvature. The proofs of the main results in this thesis depend of inequalities obtained through the extension of the "law of cosines" of Euclidean space. From these inequalities and some other auxiliary results obtained here, we believe that it is possible to obtain other convergence results on Riemannian manifolds with lower bounded curvature. Given this, it would be interesting to study in future work: inexact and stochastic versions for gradient and subgradient methods, the subgradient method for multiobjective optimization, as well the incremental gradient and subgradient methods, see [16, 55] and other methods that have similar convergence analysis to the methods considered in this work.

Finally, we remark that for Riemannian manifolds with curvature unbounded below, perhaps another strategy for the step will be need since is not possible to control the distance between the geodesics with the Toponogov comparison theorem. Therefore, a interesting problem for theoretical purposes that can be investigated is find another test strategy that makes it possible to extend the results of this thesis without assuming any restriction on the curvature of the manifold. An advance for this would be interesting to make restriction on the curvature of Ricci instead of the sectional.

# Bibliography


[1] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, Princeton, NJ, 2008. With a foreword by Paul Van Dooren.

[2] B. Afsari, R. Tron, and R. Vidal. On the convergence of gradient descent for finding the Riemannian center of mass. *SIAM J. Control Optim.*, 51(3):2230–2260, 2013.

[3] T. Ando, C.-K. Li, and R. Mathias. Geometric means. *Linear Algebra and its Applications*, 385:305–334, 2004.

[4] E. Batista, G. Bento, and O. Ferreira. An extragradient-type algorithm for variational inequality on hadamard manifolds. *arXiv preprint arXiv:1804.09292*, 2018.

[5] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2(1):183–202, 2009.

[6] J. Y. Bello Cruz. A subgradient method for vector optimization problems. *SIAM J. Optim.*, 23(4):2169–2182, 2013.

[7] J. Y. Bello Cruz and G. Bouza Allende. A steepest descent-like method for variable order vector optimization problems. *J. Optim. Theory Appl.*, 162(2):371–391, 2014.

[8] G. C. Bento, S. D. B. Bitar, J. X. Cruz Neto, P. R. Oliveira, and J. C. Souza. The steepest descent method for computing riemannian center of mass on Hadamard manifolds. *Technical report, submitted*, 2017.

[9] G. C. Bento and J. X. Cruz Neto. A subgradient method for multiobjective optimization on Riemannian manifolds. *J. Optim. Theory Appl.*, 159(1):125–137, 2013.

[10] G. C. Bento, J. X. Cruz Neto, G. López, A. Soubeyran, and J. C. O. Souza. The proximal point method for locally Lipschitz functions in multiobjective optimization with application to the compromise problem. *SIAM J. Optim.*, 28(2):1104–1120, 2018.

[11] G. C. Bento, J. X. da Cruz Neto, and P. S. M. Santos. An inexact steepest descent method for multicriteria optimization on Riemannian manifolds. *J. Optim. Theory Appl.*, 159(1):108–124, 2013.

[12] 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.

[13] G. C. Bento, O. P. Ferreira, and P. R. Oliveira. Unconstrained steepest descent method for multicriteria optimization on Riemannian manifolds. *J. Optim. Theory Appl.*, 154(1):88–107, 2012.

[14] G. C. Bento, O. P. Ferreira, and P. R. Oliveira. Proximal point method for a special class of nonconvex functions on Hadamard manifolds. *Optimization*, 64(2):289–319, 2015.

[15] G. C. Bento and J. G. Melo. Subgradient method for convex feasibility on Riemannian manifolds. *J. Optim. Theory Appl.*, 152(3):773–785, 2012.

[16] D. P. Bertsekas. A new class of incremental gradient methods for least squares problems. *SIAM J. Optim.*, 7(4):913–926, 1997.

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

[18] N. Boumal, P.-A. Absil, and C. Cartis. Global rates of convergence for nonconvex optimization on manifolds. *IMA Journal of Numerical Analysis*, 39(1):1–33, 2018.

[19] R. Burachik, L. M. G. Drummond, A. N. Iusem, and B. F. Svaiter. Full convergence of the steepest descent method with inexact line searches. *Optimization*, 32(2):137–146, 1995.

[20] G. A. Carrizo, P. A. Lotito, and M. C. Maciel. Trust region globalization strategy for the nonconvex unconstrained multiobjective optimization problem. *Math. Program.*, 159(1-2, Ser. A):339–369, 2016.

[21] T. H. Colding. Aspects of Ricci curvature. In *Comparison geometry (Berkeley, CA, 1993–94)*, volume 30 of *Math. Sci. Res. Inst. Publ.*, pages 83–98. Cambridge Univ. Press, Cambridge, 1997.

[22] R. Correa and C. Lemaréchal. Convergence of some algorithms for convex minimization. *Math. Program.*, 62(2, Ser. B):261–275, 1993.

[23] J. X. da Cruz Neto, L. L. de Lima, and P. R. Oliveira. Geodesic algorithms in Riemannian geometry. *Balkan J. Geom. Appl.*, 3(2):89–100, 1998.

[24] J. X. da Cruz Neto, O. P. Ferreira, and L. R. Lucambio Pérez. Contributions to the study of monotone vector fields. *Acta Math. Hungar.*, 94(4):307–320, 2002.

[25] J. X. Da Cruz Neto, O. P. Ferreira, L. R. L. Pérez, and S. Z. Németh. Convex- and monotone-transformable mathematical programming problems and a proximal-like point method. *J. Global Optim.*, 35(1):53–69, 2006.

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

[27] E. D. Dolan and J. J. Moré. Benchmarking optimization software with performance profiles. *Math. program.*, 91(2):201–213, 2002.

[28] L. M. G. n. Drummond and A. N. Iusem. A projected gradient method for vector optimization problems. *Comput. Optim. Appl.*, 28(1):5–29, 2004.

[29] 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.

[30] O. P. Ferreira, A. N. Iusem, and S. Z. Németh. Concepts and techniques of optimization on the sphere. *TOP*, 22(3):1148–1170, 2014.

[31] O. P. Ferreira, M. S. Louzeiro, and L. F. Prudente. Gradient Method for Optimization on Riemannian Manifolds with Lower Bounded Curvature. *ArXiv e-prints*, June 2018.

[32] O. P. Ferreira, M. S. Louzeiro, and L. F. Prudente. Iteration-complexity of the subgradient method on Riemannian manifolds with lower bounded curvature. *Optimization*, 0(0):1–17, 2018.

[33] O. P. Ferreira and P. R. Oliveira. Subgradient algorithm on Riemannian manifolds. *J. Optim. Theory Appl.*, 97(1):93–104, 1998.

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

[35] J. Fliege and B. F. Svaiter. Steepest descent methods for multicriteria optimization. *Math. Methods Oper. Res.*, 51(3):479–494, 2000.

[36] J. Fliege and A. I. F. Vaz. A method for constrained multiobjective optimization based on SQP techniques. *SIAM J. Optim.*, 26(4):2091–2119, 2016.

[37] J. Fliege, A. I. F. Vaz, and L. N. Vicente. Complexity of gradient descent for multiobjective optimization. *Optimization Methods and Software*, 0:1–11, 2018.

[38] E. H. Fukuda and L. M. Graña Drummond. On the convergence of the projected gradient method for vector optimization. *Optimization*, 60(8-9):1009–1021, 2011.

[39] E. H. Fukuda and L. M. Graña Drummond. Inexact projected gradient method for vector optimization. *Comput. Optim. Appl.*, 54(3):473–493, 2013.

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

[41] J.-L. Goffin. Subgradient optimization in nonsmooth optimization (including the Soviet revolution). *Doc. Math.*, (Extra vol.: Optimization stories):277–290, 2012.

[42] L. M. Graña Drummond and B. F. Svaiter. A steepest descent method for vector optimization. *J. Comput. Appl. Math.*, 175(2):395–414, 2005.

[43] P. Grohs and S. Hosseini. $\varepsilon$-subgradient algorithms for locally lipschitz functions on Riemannian manifolds. *Adv. Comput. Math.*, 42(2):333–360, 2016.

[44] B. Jeuris, R. Vandebril, and B. Vandereycken. A survey and comparison of contemporary algorithms for computing the matrix geometric mean. *Electron. Trans. Numer. Anal.*, 39:379–402, 2012.

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

[46] C. Lenglet, M. Rousson, R. Deriche, and O. Faugeras. Statistics on the manifold of multivariate normal distributions: theory and application to diffusion tensor MRI processing. *J. Math. Imaging Vision*, 25(3):423–444, 2006.

[47] C. Li, B. S. Mordukhovich, J. Wang, and J.-C. Yao. Weak sharp minima on Riemannian manifolds. *SIAM J. Optim.*, 21(4):1523–1560, 2011.

[48] C. Li and J.-C. Yao. Variational inequalities for set-valued vector fields on Riemannian manifolds: convexity of the solution set and the proximal point algorithm. *SIAM J. Control Optim.*, 50(4):2486–2514, 2012.

[49] B. Lin, X. He, C. Zhang, and M. Ji. Parallel vector field embedding. *J. Mach. Learn. Res.*, 14:2945–2977, 2013.

[50] L. R. Lucambio Pérez and L. F. Prudente. Nonlinear conjugate gradient methods for vector optimization. *SIAM J. Optim.*, 28(3):2690–2720, 2018.

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

[52] 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.

[53] O. Montonen, N. Karmitsa, and M. M. Mäkelä. Multiple subgradient descent bundle method for convex nonsmooth multiobjective optimization. *Optimization*, 67(1):139–158, 2018.

[54] V. Morovati, L. Pourkarimi, and H. Basirzadeh. Barzilai and Borwein's method for multiobjective optimization problems. *Numer. Algorithms*, 72(3):539–604, 2016.

[55] A. Nedić and D. P. Bertsekas. Incremental subgradient methods for nondifferentiable optimization. *SIAM J. Optim.*, 12(1):109–138, 2001.

[56] Y. Nesterov. *Introductory lectures on convex optimization*, volume 87 of *Applied Optimization*. Kluwer Academic Publishers, Boston, MA, 2004. A basic course.

[57] Y. Nesterov. Gradient methods for minimizing composite functions. *Math. Program.*, 140(1, Ser. B):125–161, 2013.

[58] Y. E. Nesterov and M. J. Todd. On the Riemannian geometry defined by self-concordant barriers and interior-point methods. *Found. Comput. Math.*, 2(4):333–361, 2002.

[59] P. Petersen. *Riemannian geometry*, volume 171 of *Graduate Texts in Mathematics*. Springer, Cham, third edition, 2016.

[60] B. T. Poljak. Subgradient methods: a survey of Soviet research. In *Nonsmooth optimization (Proc. IIASA Workshop, Laxenburg, 1977)*, volume 3 of *IIASA Proc. Ser.*, pages 5–29. Pergamon, Oxford-New York, 1978.

[61] T. Rapcsák. *Smooth nonlinear optimization in* $\mathbf{R}^n$, volume 19 of *Nonconvex Optimization and its Applications*. Kluwer Academic Publishers, Dordrecht, 1997.

[62] M. Raydan. The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem. *SIAM J. Optim.*, 7(1):26–33, 1997.

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

[64] 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.

[65] 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.

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

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

[68] J. Wang, C. Li, G. Lopez, and J.-C. Yao. Proximal point algorithms on Hadamard manifolds: linear convergence and finite termination. *SIAM J. Optim.*, 26(4):2696–2729, 2016.

[69] X. Wang, C. Li, J. Wang, and J.-C. Yao. Linear convergence of subgradient algorithm for convex feasibility on Riemannian manifolds. *SIAM J. Optim.*, 25(4):2334–2358, 2015.

[70] X. M. Wang. Subgradient algorithms on riemannian manifolds of lower bounded curvatures. *Optimization*, 67(1):179–194, 2018.

[71] X. M. Wang, C. Li, and J. C. Yao. Subgradient projection algorithms for convex feasibility on Riemannian manifolds with lower bounded curvatures. *J. Optim. Theory Appl.*, 164(1):202–217, 2015.

[72] T. Yamaguchi. Locally geodesically quasiconvex functions on complete Riemannian manifolds. *Trans. Amer. Math. Soc.*, 298(1):307–330, 1986.

[73] L. Yang. Riemannian median and its estimation. *LMS J. Comput. Math.*, 13:461–479, 2010.

[74] Y.-x. Yuan. Step-sizes for the gradient method. In *Third International Congress of Chinese Mathematicians. Part 1, 2*, volume 2 of *AMS/IP Stud. Adv. Math., 42, pt. 1*, pages 785–796. Amer. Math. Soc., Providence, RI, 2008.

[75] H. Zhang, S. J. Reddi, and S. Sra. Fast stochastic optimization on Riemannian manifolds. *ArXiv e-prints*, pages 1–17, 2016.

[76] H. Zhang and S. Sra. First-order methods for geodesically convex optimization. *JMLR: Workshop and Conference Proceedings*, 49(1):1–21, 2016.

[77] S. Zhu. The comparison geometry of Ricci curvature. In *Comparison geometry (Berkeley, CA, 1993–94)*, volume 30 of *Math. Sci. Res. Inst. Publ.*, pages 221–262. Cambridge Univ. Press, Cambridge, 1997.