

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

O. P. Ferreira ^{*} M. S. Louzeiro[†] L. F. Prudente ^{*}

June 17, 2019

Abstract

The steepest descent method for multiobjective optimization on Riemannian manifolds with lower bounded sectional curvature is analyzed in this paper. The aim of the paper is twofold. Firstly, an asymptotic analysis of the method is presented with three different finite procedures for determining the stepsize, namely, Lipschitz stepsize, adaptive stepsize and Armijo-type stepsize. The second aim is to present, by assuming that the Jacobian of the objective function is componentwise Lipschitz continuous, iteration-complexity bounds for the method with these three stepsizes strategies. In addition, some examples are presented to emphasize the importance of working in this new context. Numerical experiments are provided to illustrate the effectiveness of the method in this new setting and certify the obtained theoretical results.

Keywords: Steepest descent method, multiobjective optimization problem, Riemannian manifold, lower bounded curvature, iteration-complexity bound.

AMS subject classification: 90C33, 49K05, 47J25.

1 Introduction

A constrained multiobjective optimization problem with constraint set \mathcal{M} , consists of m objective functions f_1, \dots, 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 [1–7]. Here, among the methods designed for solving multiobjective optimization problems, we are interested in the steepest descent method. This method, was proposed in [8] and since of then several variants have been considered, including but not limited to [9–14]. Recently some iteration-complexity results to gradient method for unconstrained multi-objective optimization problem were

^{*}IME/UFG, Avenida Esperança, s/n, Campus Samambaia, Goiânia, GO, 74690-900, Brazil (e-mails: orizon@ufg.br, lfprudente@ufg.br).

[†]TU Chemnitz, Fakultät für Mathematik, D-09107, Chemnitz, Germany (e-mail: mauricio.silva-louzeiro@mathematik.tu-chemnitz.de).

presented in [15]. These results have been shown to be the same global rates as for steepest descent method in scalar objective optimization.

Constrained optimization problems, where the constraint set \mathcal{M} can be endowed with Riemannian manifold structure, have been studied extensively in the last few years. Some aspects about the use of Riemannian geometry tools to study these class of problems arises from the following interesting fact. Endowing \mathcal{M} with a suitable Riemannian metric, an Euclidean non-convex constrained problem with constraint set \mathcal{M} can be seen as a Riemannian convex unconstrained problem. In addition to this property, for differentiable functions, its gradient can also become *Riemannian Lipschitz continuous*; see [16]. 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 well known that the iteration-complexity of several optimization methods for convex optimization problems such that objective functions have Lipschitz continuous gradient is much lower than nonconvex optimization problems; see for example [17–21] and references therein. Furthermore, many optimization problems are naturally posed on the Riemannian context; see [18, 20, 22, 23]. Then, to take advantage of the intrinsic 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 [23–25]. In this sense constrained optimization problems can be seen as unconstrained from the point of view of Riemannian geometry. Moreover, intrinsic Riemannian structures can also opens up new research directions that aid in developing competitive optimization algorithms; see [18, 20, 22, 23, 26, 27]. More about concepts and techniques of optimization on Riemannian context can be found in [21, 25, 28–34] and the bibliographies therein.

In this paper we will study the steepest descent method for multiobjective optimization on Riemannian manifolds. The aim is twofold. First, asymptotic analysis will be done for quasi-convex and convex vectorial functions. In fact, in [35] asymptotic analysis of this method has already been done in Riemannian context; see also [36]. However, the analysis asymptotic presented in these previous works is just to stepsize given by Armijo rule and it demand that the Riemannian manifolds have nonnegative sectional curvature. The asymptotic analysis presented in the present paper increase the previous ones in two different aspects. *It is provided an analysis with three different finite procedures for determining the stepsize, namely, Lipschitz stepsize, adaptive stepsize and Armijo-type stepsize 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 [15]. 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. Numerical experiments are provided to illustrate the effectiveness of the method in this new setting and certify the obtained theoretical results.

The organization of this paper is as follows. In Section 2, some notations and auxiliary results, used throughout of the paper, are placed. In Section 3, we present the

algorithm and the stepsizes that will be used. In Section 3.1, the asymptotic convergence analysis of the sequence generated by the steepest descent method is made. In Section 3.2, we present iteration-complexity bounds related to the steepest descent method. In Section 4, we present examples of vectorial convex functions with componentwise Lipschitz continuous Jacobian. Numerical experiments are present in Section 5. Finally, some conclusions are given in Section 6.

2 Notations and Auxiliary Concepts

In this section, we recall some concepts, notations, and basics results about Riemannian manifolds and vector optimization. For more details we refer the reader to [19, 25, 37, 38].

We denote by $T_p\mathcal{M}$ the *tangent space* of a finite dimensional Riemannian manifold \mathcal{M} at p , and by $T\mathcal{M} = \cup_{p \in \mathcal{M}} T_p\mathcal{M}$ *tangent bundle* of \mathcal{M} . 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 $\alpha : [a, b] \rightarrow \mathcal{M}$. The Riemannian distance between p and q in \mathcal{M} is denoted by $d(p, q)$. Denote by $\mathcal{X}(\mathcal{M})$, the space of smooth vector fields on \mathcal{M} . Let ∇ be the Levi-Civita connection associated to $(\mathcal{M}, \langle \cdot, \cdot \rangle)$. For each $t \in [a, b]$ and a piecewise smooth curve $\alpha : [a, b] \rightarrow \mathcal{M}$, the covariant derivative ∇ induces an isometry, relative to $\langle \cdot, \cdot \rangle$, $P_{\alpha, a, t} : T_{\alpha(a)}\mathcal{M} \rightarrow T_{\alpha(t)}\mathcal{M}$ defined by $P_{\alpha, a, t} v = V(t)$, where V is the unique vector field on the curve α such that $\nabla_{\alpha'(t)} V(t) = 0$ and $V(a) = v$, the so-called *parallel transport* along of α joining $\alpha(a)$ to $\alpha(t)$. When there is no confusion, $P_{\alpha, p, q}$ denotes the parallel transport along the segment α joining p to q . Given that the geodesic equation $\nabla_{\gamma'} \gamma' = 0$ is a second order nonlinear ordinary differential equation, then the geodesic $\gamma = \gamma_v(\cdot, p)$ is determined by its position p and velocity v at p . The restriction of a geodesic to a closed bounded interval is called a *geodesic segment*. For any two points $p, q \in \mathcal{M}$, Γ_{pq} denotes the set of all geodesic segments $\gamma : [0, 1] \rightarrow \mathcal{M}$ with $\gamma(0) = p$ and $\gamma(1) = q$. A geodesic segment joining p to q in \mathcal{M} is said to be *minimal* if its length is equal to $d(p, q)$. *In this paper, all manifolds are assumed to be connected, finite dimensional, and complete.* 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} , the *exponential map* $\exp_p : T_p\mathcal{M} \rightarrow \mathcal{M}$ is given by $\exp_p v = \gamma_v(1, p)$, for each $p \in \mathcal{M}$. For $f : \mathcal{M} \rightarrow \mathbb{R}$ a differentiable function on \mathcal{M} , the Riemannian metric induces the mapping $f \mapsto \text{grad } f$ which associates its *gradient* via the following rule $\langle \text{grad } f(p), V(p) \rangle := df(p)V(p)$, for all $p \in \mathcal{M}$ and $V \in \mathcal{X}(\mathcal{M})$. 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 $V \in \mathcal{X}(\mathcal{M})$, where the last equalities imply that $\text{hess } f V = \nabla_V \text{grad } f$, for all $V \in \mathcal{X}(\mathcal{M})$. Let us to introduce some concepts of vector optimization on a Riemannian manifold \mathcal{M} . Letting $\mathcal{I} := \{1, \dots, m\}$ define $\mathbb{R}_+^m := \{x \in \mathbb{R}^m : x_i \geq 0, i \in \mathcal{I}\}$ and $\mathbb{R}_{++}^m := \{x \in \mathbb{R}^m : x_i > 0, i \in \mathcal{I}\}$. 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 $F := (f_1, \dots, f_m) : \mathcal{M} \rightarrow \mathbb{R}^m$ be a differentiable function. We denote the *Riemannian jacobian* of F at a point $p \in \mathcal{M}$ by $\nabla F(p)v := (\langle \text{grad } f_1(p), v \rangle, \dots, \langle \text{grad } f_m(p), v \rangle)$,

where $v \in T_p\mathcal{M}$, and the image of the Riemannian jacobian of F at p by $\text{Im}(\nabla F(p)) := \{\nabla F(p)v : v \in T_p\mathcal{M}\}$. A vectorial function $F : \mathcal{M} \rightarrow \mathbb{R}^m$ is said to be *convex* on \mathcal{M} if for any $p, q \in \mathcal{M}$ and $\gamma \in \Gamma_{pq}$ the composition $F \circ \gamma : [0, 1] \rightarrow \mathbb{R}^m$ satisfies $F \circ \gamma(t) \preceq (1-t)F(p) + tF(q)$, for all $t \in [0, 1]$. By convexity of F , it follows that $\nabla F(p)\gamma'(0) \preceq F(q) - F(p)$. A vectorial function F is called *quasi-convex* on \mathcal{M} if, for every $p, q \in \mathcal{M}$ and $\gamma \in \Gamma_{pq}$, it 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, then $F(q) \preceq F(p)$ implies $\nabla F(p)\gamma'(0) \preceq 0$.

The next result plays an important role in next sections. Its proof, which will be omitted here, follows the same ideas as those presented in the proof of [30, Lemma 3.2], with some minor technical adjustments needed to settle it to our goals. *For simplifying our notations throughout the paper*, we define

$$\kappa < 0, \quad \hat{\kappa} := \sqrt{|\kappa|}. \quad (1)$$

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

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

and, consequently, the following inequality holds

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

Next we present the definition of Lipschitz continuous gradient vector field; see [39].

Definition 2. *Let f be a differentiable function on the set \mathcal{M} . The gradient vector field of f is said to be Lipschitz continuous on \mathcal{M} with constant $L \geq 0$ if, for any $p, q \in \mathcal{M}$ and $\gamma \in \Gamma_{pq}$, it holds that $\|P_{\gamma,p,q} \text{grad } f(p) - \text{grad } f(q)\| \leq L\ell(\gamma)$.*

The norm of the hessian $\text{hess } f$ at $p \in \mathcal{M}$ is given by

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

The next result has similar proof to its Euclidean version and it will be omitted.

Lemma 3. *Let $f : \mathcal{M} \rightarrow \mathbb{R}$ be a twice continuously differentiable function. The gradient vector field of f is Lipschitz continuous with constant $L \geq 0$ if, and only if, there exists $L \geq 0$ such that $\|\text{hess } f(p)\| \leq L$, for all $p \in \mathcal{M}$.*

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

Definition 4. Let $F := (f_1, \dots, f_m) : \mathcal{M} \rightarrow \mathbb{R}^m$ be a differentiable function. If for each $f_i : \mathcal{M} \rightarrow \mathbb{R}$ there exists a $L_i \geq 0$ such that $\|P_{\gamma,p,q} \text{grad } f_i(p) - \text{grad } f_i(q)\| \leq L_i \ell(\gamma)$, for any $p, q \in \mathcal{M}$ and $\gamma \in \Gamma_{pq}$, then we say that ∇F is componentwise Lipschitz continuous on \mathcal{M} with constant $L := \max_{i=1, \dots, m} L_i$.

The proof of the next lemma follows, with appropriate adjustments, the same idea of proof of the scalar version presented in [17, Corollary 2.1]. Throughout of the paper we will use the following notation

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

Lemma 5. Let $F := (f_1, \dots, f_m) : \mathcal{M} \rightarrow \mathbb{R}^m$ be a differentiable function. Assume that ∇F is componentwise Lipschitz continuous on \mathcal{M} with constant $L \geq 0$ and $p \in \mathcal{M}$. Then there holds

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

Next we introduce the concept of quasi-Fejér convergence, which played an important role in the analysis of the gradient method.

Definition 6. 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 exist 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 $k = 0, 1, \dots$.

In the following we state the main property of the quasi-Fejér concept, its proof follows the same path as its Euclidean counterpart proved in [40], by replacing the Euclidean distance by the Riemannian one.

Theorem 7. 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 $\lim_{k \rightarrow \infty} y_k = \bar{y}$.

Hereafter, we assume that \mathcal{M} is a complete Riemannian manifolds with sectional curvature $K \geq \kappa$, where $\kappa < 0$. We point out that for Riemannian manifold with nonnegative sectional curvature, the convergence analysis of the steepest descent method for convex and quasi-convex vector functions is well understood; see for example [35, 36].

3 Steepest Descent for Multiobjective Optimization

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

$$\min\{F(p) : p \in \mathcal{M}\}. \quad (2)$$

A point $p \in \mathcal{M}$ satisfying $\text{Im}(\nabla F(p)) \cap (-\mathbb{R}_{++}^m) = \emptyset$ is called *critical Pareto*. An *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

$$\min_{v \in T_p \mathcal{M}} \left\{ \max_{i \in \mathcal{I}} \langle \text{grad } f_i(p), v \rangle + \frac{1}{2} \|v\|^2 \right\}, \quad \mathcal{I} = \{1, \dots, m\}. \quad (3)$$

Whenever $p \in \mathcal{M}$ is not critical Pareto, the optimization problem (3) has only one solution, which is called *steepest descent direction* for F in p and it is denoted by

$$v_p := \operatorname{argmin}_{v \in T_p \mathcal{M}} \left\{ \max_{i \in \mathcal{I}} \langle \text{grad } f_i(p), v \rangle + \frac{1}{2} \|v\|^2 \right\}. \quad (4)$$

In the next lemma we state an important property of the steepest descent direction. Its proof can be found in [35, Lemma 5.1].

Lemma 8. *The steepest descent direction mapping $\mathcal{M} \ni p \mapsto v_p \in T_p \mathcal{M}$, is a continuous vector field.*

Moreover, the vector v_p is the solution of the problem (3) if and only if there exist $\mu_j \geq 0$, for $j \in \mathcal{I}(v_p) := \{j \in \mathcal{I} : \langle \text{grad } f_j(p), v_p \rangle = \max_{i \in \mathcal{I}} \langle \text{grad } f_i(p), v_p \rangle\}$, such that

$$v_p = - \sum_{j \in \mathcal{I}(v_p)} \mu_j \text{grad } f_j(p), \quad \sum_{j \in \mathcal{I}(v_p)} \mu_j = 1, \quad (5)$$

see [35, 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 9. *Let $p \in \mathcal{M}$ and v_p as defined (4). Then,*

$$\max_{i \in \mathcal{I}} \langle \text{grad } f_i(p), v_p \rangle = - \|v_p\|^2. \quad (6)$$

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 in (4). Thus, from the first equality in (5) we have

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

Hence, by the definition of $\mathcal{I}(v_p)$ and the second equality in (5), it is easy to verify that (6) holds. The second statement follows by using the definitions of $\nabla F(p)v_p$ and $\mathcal{I}(v_p)$. 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 exists $i \in \mathcal{I}$ such that $\langle \text{grad } f_i(p), v_p \rangle \geq 0$. Then, the by first part of lemma we have $\|v_p\| = 0$. The converse follows from [35, Lemma 4.2] and the proof is concluded. \square

The proof of the next lemma is a straight combination of Lemma 5 with first part of Lemma 9 and will be omitted.

Lemma 10. Assume that ∇F is componentwise Lipschitz continuous on \mathcal{M} with constant $L \geq 0$. Let $p \in \mathcal{M}$ and v_p as defined in (4). Then, there holds

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

Next we state the steepest descent algorithm in Riemannian manifold to solve (2).

Algorithm 1: Steepest descent algorithm in a Riemannian 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). If $v_{p_k} = 0$, then **stop**; otherwise, choose a stepsize $t_k > 0$ and compute

$$p_{k+1} := \exp_{p_k}(t_k v_k). \quad (7)$$

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

Our goal is to analyze Algorithm 1 with three different strategies for choosing the stepsize $t_k > 0$. An analogous analysis done in the scalar case can be found in [16]. In the first strategy we assume that ∇F is componentwise Lipschitz continuous and in the last two without any Lipschitz condition. The statements of the strategies are as follows:

Strategy 1 (Lipschitz stepsize). Assume that ∇F is componentwise Lipschitz continuous on \mathcal{M} with constant $L \geq 0$. Let $\varepsilon > 0$ and take

$$\varepsilon < t_k \leq \frac{1}{L}. \quad (8)$$

Despite knowing that ∇F is componentwise Lipschitz continuous, in general the Lipschitz constant is not computable. Then, the next strategy can be used to compute the stepsize without any Lipschitz condition. However, as we shall show, if ∇F is componentwise Lipschitz continuous with constant $L > 0$ the stepsize computed is an approximation to $1/L$; see the scalar case in [16, 41].

Strategy 2 (adaptive stepsize). Take $\zeta \in (0, 1/2]$, $L_0 > 0$, $t_0 := L_0^{-1}$, and $0 < \eta < 1$. Consider v_k is defined as in (4). Set $t_k := \eta^{i_k} t_{k-1}$, where

$$i_k := \min \{ i : F(\exp_{p_k}(\eta^i t_{k-1} v_k)) \preceq F(p_k) - \zeta \eta^i t_{k-1} \|v_k\|^2 e, \ i = 0, 1, \dots \}. \quad (9)$$

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

Remark 11. Suppose that ∇F is componentwise Lipschitz continuous on \mathcal{M} with constant $L > 0$. Let $L_0 > 0$ be an estimate for L and $v_k = v_{p_k}$ be defined as in (4). Taking $t = 1/L$, using Lemma 10 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 $t_k = 1/L$ is always accepted for Strategies 2 with $i_k = 0$. Therefore, if $L_0 \geq L$ then we have $t_k = 1/L_0$, i.e., the step-size is constant. On the other hand, if $L_0 \leq L$ then owing to $\eta < 1$ we conclude that t_k in Strategies 2 satisfies

$$\frac{\eta}{L} \leq t_k \leq \frac{1}{L_0}. \quad (10)$$

In the following strategy a stepsize satisfying an Armijo-type sufficient descent condition is chosen using a backtracking approach.

Strategy 3 (Armijo-type stepsize). Let $t_{\max} > t_{\min} > 0$, $0 < \omega_1 < \omega_2 < 1$ and $\delta \in (0, 1)$. Let $v_k = v_{p_k}$ be defined as in (4). The stepsize t_k is chosen according the following algorithm:

STEP 0. Set $\ell = 0$ and take $\hat{t}_{k_0} \in [t_{\min}, t_{\max}]$.

STEP 1. If

$$F(\exp_{p_k}(\hat{t}_{k_\ell} v_k)) \preceq F(p_k) - \delta \hat{t}_{k_\ell} \|v_k\|^2 e, \quad (11)$$

then set $t_k := \hat{t}_{k_\ell}$ and **stop**.

STEP 2. Choose a stepsize $\hat{t}_{k_{\ell+1}} \in [\omega_1 \hat{t}_{k_\ell}, \omega_2 \hat{t}_{k_\ell}]$, set $\ell \leftarrow \ell + 1$ and proceed to STEP 1.

In the next remark we show that, for ∇F componentwise Lipschitz continuous on \mathcal{M} , the stepsizes in Strategy 3 are bounded below by a positive constant.

Remark 12. Assume that ∇F is componentwise Lipschitz continuous on \mathcal{M} with constant $L \geq 0$, $t_{\max} > 2[1-\delta]/L$ and $t_{\min} < 2\omega_1(1-\delta)/L$. Hence, for any $t \in (0, 2[1-\delta]/L]$, from Lemma 10 we have

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

Therefore, t_k in Strategies 3 satisfies the inequality $t_k > t_{\min}$, for all $k = 0, 1, \dots$.

Since well-definedness of Strategies 2 and 3 follows by using ordinary arguments, we will omitted its proof here. Hence, the sequence $\{p_k\}$ generated by Algorithm 1 with Strategies 1, 2 or 3 is well-defined. Finally we remind that, p is a critical Pareto if, and only if, $\|v_p\| = 0$. Therefore, from now on we assume that $\|v_k\| \neq 0$, for all k . Moreover, let us denote by $\{p_k\}$ the infinity sequence generated by Algorithm 1.

3.1 Asymptotic Convergence Analysis

In this section, we analyze asymptotic convergence of the sequence $\{p_k\}$ generated by Algorithm 1 with Strategies 1, 2 and 3. Let us define

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

To proceed with our analysis, from now on, we will assume that the set \mathcal{A} is non-empty. A condition guaranteeing this assumption is the existence of accumulation point for the sequence $\{p_k\}$.

Lemma 13. *Let $\{p_k\}$ be generated with any of Strategies 1, 2 or 3. Then,*

$$F(p_{k+1}) \preceq F(p_k) - \nu t_k \|v_k\|^2 e, \quad k = 0, 1, \dots, \quad (12)$$

where $\nu = 1/2$ for Strategy 1, $\nu = \zeta$ for Strategy 2 and $\nu = \delta$ for Strategy 3. As a consequence, there holds $\lim_{k \rightarrow +\infty} t_k \|v_k\|^2 = 0$.

Proof. The inequality (12) for Strategies 2 and 3 follows from (7), (9) and (11), respectively. Now, assume that $\{p_k\}$ is generated by using Strategies 1. In this case, combining (7) with Lemma 10 and taking into account that (8) implies $(Lt_k/2 - 1) \leq -1/2$, (12) follows with $\nu = 1/2$. To proceed with the proof of the last statement, take $q \in \mathcal{A}$ and an integer number $\ell > 0$. Thus, (12) yields

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

with implies the desired result, and the proof of the lemma is concluded. \square

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 (12) together with (8), (10) and (11) define the first constant $\rho > 0$ as follows

$$\sum_{k=0}^{\infty} t_k^2 \|v_k\|^2 \leq \rho := \begin{cases} \min_{i \in \mathcal{I}} \{2[f_i(p_0) - f_i(q)]/L : q \in \mathcal{A}\}, & \text{for Strategy 1;} \\ \min_{i \in \mathcal{I}} \{[f_i(p_0) - f_i(q)]/(\zeta L_0) : q \in \mathcal{A}\}, & \text{for Strategy 2;} \\ \min_{i \in \mathcal{I}} \{t_{\max}[f_i(p_0) - f_i(q)]/\delta : q \in \mathcal{A}\}, & \text{for Strategy 3.} \end{cases} \quad (13)$$

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

$$\mathcal{C}_{\rho, \kappa}^q := \cosh^{-1} \left(\cosh(\hat{\kappa} d(p_0, q)) e^{\frac{1}{2}(\hat{\kappa} \sqrt{\rho}) \sinh(\hat{\kappa} \sqrt{\rho})} \right), \quad (14)$$

$$\mathcal{K}_{\rho, \kappa}^q := \frac{\sinh(\hat{\kappa} \sqrt{\rho})}{\hat{\kappa} \sqrt{\rho}} \frac{\mathcal{C}_{\rho, \kappa}^q}{\tanh \mathcal{C}_{\rho, \kappa}^q}, \quad (15)$$

where the constants $\hat{\kappa}$ and ρ , are defined in (1) and (13), respectively.

Lemma 14. *Let $\{p_k\}$ be generated with any of Strategies 1, 2 or 3 and $q \in \mathcal{A}$. Assume that the function F is quasi-convex on \mathcal{M} . Then,*

$$d(p_{k+1}, q) \leq \frac{1}{\hat{\kappa}} \mathcal{C}_{\rho, \kappa}^q, \quad k = 0, 1, \dots \quad (16)$$

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

$$d^2(p_{k+1}, q) \preceq d^2(p_k, q) + \mathcal{K}_{\rho, \kappa}^q t_k^2 \|v_k\|^2, \quad k = 0, 1, \dots \quad (17)$$

Proof. For each k , let $\gamma_k : [0, \infty) \rightarrow \mathbb{R}$ be defined by $\gamma_k(t) = \exp_{p_k}(tv_k)$. Let $\beta_k : [0, 1] \rightarrow \mathcal{M}$ be a minimizing geodesic with $\beta_k(0) = p_k$ and $\beta_k(1) = q$. By using (5), the definition of v_k , the quasi-convexity of F , and taking into account that $q \in \mathcal{A}$, we have

$$\langle v_k, \beta'(0) \rangle = - \sum_{j \in \mathcal{I}(v_k)} \mu_j \langle \text{grad } f_j(p_k), \beta'(0) \rangle \geq 0, \quad \sum_{j \in \mathcal{I}(v_k)} \mu_j = 1. \quad (18)$$

Thus, applying the first inequality of Lemma 1, with $t = t_k$, $\gamma = \gamma_k$, $\beta = \beta_k$ and $p = p_k$, and using (7) and (18), we obtain

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

Since (13) implies $t_k \|v_k\| \leq \sqrt{\rho}$, and the map $(0, +\infty) \ni t \mapsto \sinh(t)/t$ is increasing, we conclude that

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

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 (13), 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 σ and (14), yields (16). The boundedness of $\{p_k\}$ is immediate from (16). We proceed with the proof of (17). Now, we apply the second inequality of Lemma 1 and again we take into account (7) and (18) to conclude that

$$d^2(p_{k+1}, q) \leq d^2(p_k, q) + \frac{\sinh(\hat{\kappa}t_k \|v_k\|)}{\hat{\kappa}t_k \|v_k\|} \frac{\hat{\kappa}d(p_k, q)}{\tanh(\hat{\kappa}d(p_k, q))} t_k^2 \|v_k\|^2. \quad (19)$$

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 (16) and that $t_k \|v_k\| \leq \sqrt{\rho}$, the inequality (19) becomes

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

Therefore, by using (15) we have the desired inequality. \square

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

Theorem 15. *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 14 and (13) imply that $\{p_k\}$ is bounded and quasi-Fejér convergent to set \mathcal{A} . Taking into account Lemma 13 we conclude that $\{f_s(p_k)\}$ is non-increasing, for all $s = 1, \dots, m$. Thus, we conclude that all cluster points of $\{p_k\}$ belongs to \mathcal{A} . Hence, Theorem 7 implies that $\{p_k\}$ converges to a point $\bar{p} \in \mathcal{A}$. Hence, remains to prove that \bar{p} is a critical Pareto point of F . We know that, for any of the three strategies 1, 2 or 3, 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 $\lim_{j \rightarrow \infty} t_{k_j} = \bar{t}$. First we suppose that $\bar{t} > 0$. Since $\lim_{j \rightarrow \infty} p_{k_j} = \bar{p}$ and $\lim_{j \rightarrow \infty} t_{k_j} = \bar{t}$, (13) and Lemma 8 imply that $0 = \lim_{j \rightarrow \infty} t_{k_j} \|v_{k_j}\| = \bar{t} \|v_{\bar{p}}\|$. Thus, considering that we are under the assumption $\bar{t} > 0$, we obtain $v_{\bar{p}} = 0$. Therefore, Lemma 9 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 2 and 3, due to Strategy 1 we have $\epsilon \leq \bar{t}$. First assume that Strategy 2 is used and take $r \in \mathbb{N}$. Since $\lim_{j \rightarrow \infty} t_{k_j} = 0$ we conclude that if j is large enough, $t_{k_j} < \eta^r t_0 =: C_r$. Thus, for each j large enough, from (9) 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 $s_j \in \{1, \dots, m\}$. Since the set $\{1, \dots, m\}$ 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 $\lim_{j \rightarrow \infty} p_{k_j} = \bar{p}$ and $\lim_{j \rightarrow \infty} t_{k_j} = \bar{t}$, letting j goes to $+\infty$ and taking into account that v_p and the exponential map are continuous, we obtain

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

Thus, letting r goes to $+\infty$, yields $\langle \text{grad } f_{\hat{s}}(\bar{p}), v_{\bar{p}} \rangle \geq -\zeta \|v_{\bar{p}}\|^2$. Hence, from Lemma 9 we conclude 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, using again Lemma 9 we have \bar{p} is a critical Pareto of F . Finally, assume that Strategy 3 is used. Since $\lim_{j \rightarrow \infty} t_{k_j} = 0$ 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 $s_j \in \{1, \dots, m\}$. Since the set $\{1, \dots, m\}$ 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, the mean value theorem implies that there exists $\bar{t}_j \in (0, \hat{t}_j)$ such that

$$\left\langle \text{grad } f_{\hat{s}}(\gamma_j(\bar{t}_j)), P_{\gamma_j, 0, \bar{t}_j} v_{k_j} \right\rangle > -\delta \|v_{k_j}\|^2. \quad (20)$$

On the other hand, let $B_\epsilon(\bar{p}) \subset \mathcal{M}$ be a totally normal ball. Hence, considering that $\lim_{j \rightarrow +\infty} p_{k_j} = \bar{p}$, Lemma 8 implies that $\lim_{j \rightarrow +\infty} v_{k_j} = \bar{v}_{\bar{p}}$. Moreover, $0 < \omega_1 \hat{t}_j \leq t_{k_j}$ implies that $\lim_{j \rightarrow +\infty} \hat{t}_j = 0$. Owing to $0 < \bar{t}_j \leq \hat{t}_j$ we obtain that $\lim_{j \rightarrow +\infty} \bar{t}_j = 0$. Hence, for all j large enough we have $\{\bar{t}_j\} \subset (0, 1)$ and $\gamma_j(\bar{t}_j) \in B_\epsilon(\bar{p})$, which implies

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

Thus, letting j goes to $+\infty$ and using [42, Lemma 1.1], we conclude that $\lim_{j \rightarrow +\infty} P_{\gamma_j, 0, \bar{t}_j} v_{k_j} = \bar{v}_{\bar{p}}$ (a general version for this equality, see [42, Lemma 1.2]). Then, letting j goes to $+\infty$ in (20) and taking into account Lemma 8, that $\text{grad } f_{\hat{s}}$ and the exponential map are continuous, we obtain $\langle \text{grad } f_{\hat{s}}(\bar{p}), v_{\bar{p}} \rangle \geq -\delta \|v_{\bar{p}}\|^2$. Hence, Lemma 9 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, using again Lemma 9 we conclude that \bar{p} is a critical Pareto of F . Therefore, for all Strategies 1, 2 or 3, \bar{p} is a critical Pareto point of F , which concludes the proof. \square

Corollary 16. *Let $\{p_k\}$ be generated with any of Strategies 1, 2 or 3. 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 [35, Proposition 5.2]. Considering that convex functions are also quasi-convex the result follows from Theorem 15. \square

3.2 Iteration-Complexity Analysis

In this section we present iteration-complexity bounds related to the steepest descent method with Strategies 1, 2 and 3, for F having ∇F with componentwise Lipschitz continuous constant $L > 0$. For this purpose, by using (8), (10) and Remark 12, define

$$\xi := \begin{cases} \epsilon, & \text{for Strategy 1;} \\ \eta/L, & \text{for Strategy 2;} \\ t_{\min}, & \text{for Strategy 3.} \end{cases} \quad (21)$$

The following result extends the scalar result [17, Theorem 3.1] to multiobjective settings. Moreover, it also extends to Riemannian context [15, Theorem 3.1].

Theorem 17. *Let $\{p_k\}$ be generated with any of Strategies 1, 2 or 3, 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*

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

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

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

where $\nu = 1/2$ for Strategy 1, $\nu = \zeta$ for Strategy 2 and $\nu = \delta$ for Strategy 3.

Proof. It follows from Lemma 13 that $\nu t_k \|v_k\|^2 e \preceq F(p_k) - F(p_{k+1})$, for all $k = 0, 1, \dots$. By summing both sides of this inequality for $k = 0, 1, \dots, N-1$ and using (21), 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

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

which implies the statement of the theorem. \square

Remark 18. *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 17 can be improved. We begin by presenting an auxiliary inequality.

Lemma 19. *Let $\{p_k\}$ be generated with any of Strategies 1, 2 or 3. 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*

$$d^2(p_{k+1}, q) \leq d^2(p_k, q) + \mathcal{K}_{\rho, \kappa}^q t_k^2 \|v_k\|^2 + 2t_k \sum_{j \in \mathcal{I}(v_k)} \mu_j^k [f_j(q) - f_j(p_k)], \quad (22)$$

where ρ is defined in (13).

Proof. For each k , let $\gamma_k : [0, \infty) \rightarrow \mathbb{R}$ be defined by $\gamma_k(t) = \exp_{p_k}(tv_k)$ and $\beta_k : [0, 1] \rightarrow \mathcal{M}$ with $\beta_k(0) = p_k$ and $\beta_k(1) = q$ be a minimizing geodesic. Using (5) and the convexity of F we conclude that exist $\mu_{j,s}^k \geq 0$ satisfying $\sum_{j \in \mathcal{I}(v_k)} \mu_j^k = 1$ such that

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

Applying the second inequality of Lemma 1 with $\beta = \beta_k$, $\gamma = \gamma_k$ and $t = t_k$ and using the last inequality we obtain

$$d^2(p_{k+1}, q) \leq d^2(p_k, q) + \frac{\sinh(\hat{\kappa} t_k \|v_k\|)}{\hat{\kappa} t_k \|v_k\|} \left(\frac{\hat{\kappa} d(p_k, q)}{\tanh(\hat{\kappa} d(p_k, q))} t_k^2 \|v_k\|^2 + 2t_k \sum_{j \in \mathcal{I}(v_k)} \mu_j^k (f_j(q) - f_j(p_k)) \right). \quad (23)$$

Since $(0, +\infty) \ni t \mapsto t/\tanh(t)$ and $(0, +\infty) \ni t \mapsto \psi(t) := \sinh(t)/t$ are increasing, taking into account that (13) implies $t_k \|v_k\| \leq \sqrt{\rho}$, and using (16), the inequality (23) becomes

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

Therefore, due to $f_j(q) - f_j(p_k) \leq 0$ and ψ be bounded from below by 1, the inequality (22) follows by using (15), which concludes the proof. \square

The next result, with minor adjustments, is a generalization of [15, Theorem 4.1] to Riemannian setting, when the Armijo's type strategy is used.

Proposition 20. *Let $\{p_k\}$ be generated with any of Strategies 1, 2 or 3. 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 $\lambda_1, \dots, \lambda_m$ with $\sum_{i=1}^m \lambda_i = 1$, satisfying*

$$\sum_{i=1}^m \lambda_i [f_i(p_N) - f_i(q)] \leq \frac{d^2(p_0, q) + \mathcal{K}_{\rho, \kappa}^q \rho}{2\xi N}, \quad (24)$$

where ρ is defined in (13).

Proof. Since $f_i(p_k) - f_i(q) \geq 0$ for all i , Lemma 19 and (21) implies there exist $\mu_{i's}^k \geq 0$ such that

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

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 $k = 0, 1, \dots, N-1$, and using (13) follows

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

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

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

Defining $\lambda_i := \sum_{k=0}^{N-1} \mu_i^k / N$ we obtain the inequality in (24). 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 . \square

Finally we are ready to present the main result of this section, namely, the improvement of Theorem 17. We remark that this result is new, even in Euclidean context.

Theorem 21. *Let $\{p_k\}$ be generated with any of Strategies 1, 2 or 3. Assume that F is a convex function on \mathcal{M} and $q \in \mathcal{A}$. Then, for every $N \in \mathbb{N}$, there holds*

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

where ρ is defined in (13) and $\nu = 1/2$ for Strategy 1, $\nu = \beta$ for Strategy 2 and $\nu = \delta$ for Strategy 3.

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 13 that $\nu t_k \|v_k\|^2 e \preceq F(p_k) - F(p_{k+1})$, for all $k = 0, 1, \dots$. Thus, by summing both sides of this inequality for $k = \lceil N/2 \rceil, \dots, N$ and using (21), we obtain

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

Hence, taking non-negative numbers $\lambda_1, \dots, \lambda_m$ as in the Proposition 20 and considering that $q \in \mathcal{A}$, we conclude from the last inequality that

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

Thus, from Proposition 20 and considering that $N/2 \leq \lceil N/2 \rceil$ it follows that

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

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

4 Examples

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

Example 22. *Let \mathbb{P}_{++}^n be the cone of symmetric positive definite matrices. Define the vectorial function $F(X) = (f_1(X), \dots, f_m(X))$, where $f_i : \mathbb{P}_{++}^n \rightarrow \mathbb{R}$ is given by*

$$f_i(X) = a_i \ln(\det(X)^{b_i} + c_i) - d_i \ln(\det(X)), \quad (25)$$

$a_i, b_i, c_i, d_i \in \mathbb{R}_{++}$ with $d_i < a_i b_i$ for all $i = 1, \dots, m$. Endowing \mathbb{P}_{++}^n with the Riemannian metric given by

$$\langle U, V \rangle := \text{tr}(V X^{-1} U X^{-1}), \quad X \in \mathbb{P}_{++}^n, \quad U, V \in T_X \mathbb{P}_{++}^n,$$

where $\text{tr}(X)$ denotes the trace of $X \in \mathbb{P}^n$, we obtain a Riemannian manifolds $\mathcal{M} := (\mathbb{P}_{++}^n, \langle \cdot, \cdot \rangle)$ with nonpositive sectional curvature, see [43, Theorem 1.2. p. 325]. In \mathcal{M} , f_i is convex and has Lipschitz gradient with constant $L_i \leq a_i b_i^2 n$, for each $i = 1, \dots, m$, see [16, example 4.5]. Hence, from Definition 4 the Jacobian ∇F is componentwise Lipschitz continuous with constant $L \leq n \max\{a_1 b_1^2, \dots, a_m b_m^2\}$. In \mathcal{M} , the exponential mapping $\exp_X : T_X \mathcal{M} \rightarrow \mathcal{M}$, is given by

$$\exp_X(V) = X^{1/2} e^{X^{-1/2} V X^{-1/2}} X^{1/2}, \quad V \in \mathbb{P}^n, \quad X \in \mathbb{P}_{++}^n. \quad (26)$$

Therefore, from Corollary 16 we can apply Algorithm 1 with Strategies 1, 2 or 3 to find weak optimal Pareto of F .

In the following we present, without giving the details, one more example of convex vectorial function with Lipschitz gradients in the Riemannian manifolds $\mathcal{M} := (\mathbb{P}_{++}^n, \langle \cdot, \cdot \rangle)$.

Example 23. Let $F(X) = (f_1(X), \dots, f_m(X))$ be a vectorial function, where $f_i : \mathbb{P}_{++}^n \rightarrow \mathbb{R}$ is defined by

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

$a_i, b_i \in \mathbb{R}_{++}$ for all $i = 1, \dots, m$. In $\mathcal{M} := (\mathbb{P}_{++}^n, \langle \cdot, \cdot \rangle)$, f_i is convex and has Lipschitz gradient with constant $L_i \leq 2a_i \sqrt{n}$, for each $i = 1, \dots, m$, [16, example 4.4]. The Jacobian ∇F is componentwise Lipschitz continuous with constant $L \leq 2\sqrt{n} \max\{a_1, \dots, a_m\}$.

Now, we present some preliminaries results to study examples of convex vectorial functions with componentwise Lipschitz continuous Riemannian Jacobians. We begin with a result that, with some adjustments in the notation, can be found in [44, Lemma 2].

Lemma 24. Let $\bar{\mathcal{M}}$ and \mathcal{M} be Riemannian manifold, $\bar{\nabla}$ be the Levi-Civita connection associated to $\bar{\mathcal{M}}$ and $\varphi : \bar{\mathcal{M}} \rightarrow \mathcal{M}$ be an isometry. Then, $\nabla : \mathcal{X}(\mathcal{M}) \times \mathcal{X}(\mathcal{M}) \rightarrow \mathcal{X}(\mathcal{M})$ defined by

$$\nabla_V U := d\varphi(\bar{\nabla}_{\bar{V}} \bar{U}), \quad \forall V, U \in \mathcal{X}(\mathcal{M}). \quad (27)$$

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 φ is a diffeomorphism, $f \circ \varphi$ is continuously differentiable, $\bar{V} = d\varphi^{-1}V$ and $\bar{U} = d\varphi^{-1}U$ are vector fields in $\bar{\mathcal{M}}$. Thus, we can prove that (27) satisfies [38, 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} . \square

The next result is the main tool used in the following examples.

Theorem 25. Let \mathcal{M} and $\bar{\mathcal{M}}$ be Riemannian manifolds, $f : \mathcal{M} \rightarrow \mathbb{R}$ be a twice-differentiable function and $\varphi : \bar{\mathcal{M}} \rightarrow \mathcal{M}$ be an isometry. Then, f has gradient vector field Lipschitz continuous with constant $L \geq 0$ if, and only if, $g : \bar{\mathcal{M}} \rightarrow \mathbb{R}$ 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 $V(\varphi(q)) = d\varphi(q)\bar{V}(q)$. Thus, by using the definition of the gradient vector field and the chain rule, we have

$$\begin{aligned}\langle \text{grad } g(q), \bar{V}(q) \rangle &= dg(q)\bar{V}(q) \\ &= df(\varphi(q))d\varphi(q)\bar{V}(q) \\ &= df(\varphi(q))V(\varphi(q)) \\ &= \langle \text{grad } f(\varphi(q)), V(\varphi(q)) \rangle.\end{aligned}$$

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

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

Hence, combining the two above equality we conclude that $\text{grad } f(\varphi(q)) = d\varphi(q) \text{grad } g(q)$. Moreover, the definition of the hessian of f together with Lemma 24 yield

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

which implies that $\text{hess } f(\varphi(q))d\varphi(q) = d\varphi(q)\text{hess } g(q)$. Then, using again that φ is an isometry, we have $\|\text{hess } f(\varphi(q))\| = \|\text{hess } g(q)\|$. Therefore, by using Lemma 3 the results follows. \square

The next result is an important property of isometries, its prove is in [45, Proposition 5.6.1, p. 196].

Proposition 26. *Let \mathcal{M} and $\bar{\mathcal{M}}$ be complete Riemannian manifolds. If $\varphi : \bar{\mathcal{M}} \rightarrow \mathcal{M}$ is a isometry and γ 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 26.

Theorem 27. *Let $\mathcal{M}, \bar{\mathcal{M}}$ be Riemannian manifold and $\varphi : \bar{\mathcal{M}} \rightarrow \mathcal{M}$ an isometry. The function $g : \mathcal{M} \rightarrow \mathbb{R}$ is convex if and only if $f : \bar{\mathcal{M}} \rightarrow \mathbb{R}$, 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 gradient 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 [25, p. 83].

Example 28 (Rosenbrock's banana function class). Let $f_j : \mathbb{R}^{2n} \rightarrow \mathbb{R}$ be a variant of the Rosenbrock's banana function, defined by

$$f_j(x_1, \dots, x_{2n}) := \sum_{i=1}^n a_{ij} (x_{2i-1}^2 - x_{2i})^2 + (x_{2i-1} - b_{ij})^2, \quad a_{ij} \in \mathbb{R}_{++}, \quad b_{ij} \in \mathbb{R}, \quad (28)$$

for $j = 1, \dots, m$. 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 $G(x) = \text{diag}(G_1(x), \dots, G_n(x))$, where the blocks are given by

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

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

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

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

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

is a quadratics function, which is convex with gradient vector field Lipschitz in $\bar{\mathcal{M}}$ with constant $L_j := \max\{2, 2a_{1j}, \dots, 2a_{nj}\}$. Therefore, Theorem 27 and Theorem 25 imply, respectively, that f_j is also convex and has gradient vector field Lipschitz continuous, with constant L_j , in \mathcal{M} . Let $F = (f_1, \dots, f_m)$ be the Rosenbrock's banana vectorial function. Hence, F is convex and Definition 4 implies that ∇F is componentwise Lipschitz continuous with constant $L = \max\{2, 2a_{11}, \dots, 2a_{nm}\}$. The gradient of f_j is given by $\text{grad } f_j(x) = G(x)^{-1} f'_j(x)$, where f'_j is the usual gradient of f_j . Given $z \in \bar{\mathcal{M}}$ the exponential map in $\bar{\mathcal{M}}$, $\bar{\text{exp}}_z : T_z \bar{\mathcal{M}} \rightarrow \bar{\mathcal{M}}$, is given by $\bar{\text{exp}}_z(\bar{v}) = z + \bar{v}$. Since φ is an isometry, Proposition 26 implies that the exponential map in \mathcal{M} , $\text{exp}_x : T_x \mathcal{M} \rightarrow \mathcal{M}$, is given by $\text{exp}_x(v) = \varphi(\varphi^{-1}(x) + d\varphi^{-1}(x)v)$. Thus, due to $\varphi^{-1}(x) = (x_1, x_1^2 - x_2, \dots, x_{2n-1}, x_{2n-1}^2 - x_{2n})$ and $d\varphi^{-1}(x)v = (v_1, 2x_1 v_1 - v_2, \dots, v_{2n-1}, 2x_{2n-1} v_{2n-1} - v_{2n})$, we obtain that

$$\text{exp}_x(v) = (x_1 + v_1, v_1^2 + x_2 + v_2, \dots, x_{2n-1} + v_{2n-1}, v_{2n-1}^2 + x_{2n} + v_{2n})$$

where $x := (x_1, \dots, x_{2n})$ and $v := (v_1, \dots, v_{2n})$.

We end this section by presenting, in particular, a family of vectorial functions in positive orthant \mathbb{R}_+^n that are not convex and their gradients are not componentwise Lipschitz continuous. However, by a suitable change of the metric of \mathbb{R}_+^n the functions of that family are convex and have componentwise Lipschitz continuous gradients on this new Riemannian manifold.

Example 29. Let $f_j : \mathbb{R}_{++}^n \rightarrow \mathbb{R}$ be defined by

$$f_j(x) := a_j \ln \left(\prod_{i=1}^n x_i^{u_{ij}} + b_i \right) - \sum_{i=1}^n w_{ij} \ln(x_i) + c_j \sum_{i=1}^n \ln^2(x_i), \quad (29)$$

where $x := (x_1, \dots, x_n) \in \mathbb{R}_{++}^n$, $u_j := (u_{1j}, \dots, u_{nj})^T \in \mathbb{R}_+^n$, $w_j := (w_{1j}, \dots, w_{nj})^T \in \mathbb{R}_+^n$ and $a_j, b_j, c_j \in \mathbb{R}_{++}$, for all $j = 1, \dots, m$. 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

$$G(x) := \text{diag}(x_1^{-2}, x_2^{-2}, \dots, x_n^{-2}),$$

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

$$\varphi(z_1, \dots, z_n) = (e^{z_1}, \dots, e^{z_n}), \quad (30)$$

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

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

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 27 and Theorem 25 imply, respectively, that f_j is also convex and has gradient Lipschitz in \mathcal{M} with constant L_j . Therefore, the vectorial function $F(x) = (f_1(x), \dots, f_m(x))$ is convex and Definition 4 implies that ∇F is componentwise Lipschitz continuous with constant $L = \max\{L_1, \dots, L_m\}$. The gradient of f_j is given by

$$\text{grad } f_j(x) = \text{diag}(x)^2 f'_j(x), \quad x \in \mathcal{M}$$

where $\text{diag}(x) := \text{diag}(x_1, \dots, x_n)$ and f'_j is the usual derivative. Using the isometry (30) Proposition 26 implies that the exponential map in \mathcal{M} , $\exp_x : T_x \mathcal{M} \rightarrow \mathcal{M}$, is given by $\exp_x(v) = \varphi(\varphi^{-1}(x) + d\varphi^{-1}(x)v)$. Since $\varphi^{-1}(x) = (\ln x_1, \dots, \ln x_n)$ and $d\varphi^{-1}(x)v = (x_1^{-1}v_1, \dots, x_n^{-1}v_n)$, where $v = (v_1, \dots, v_n)$, we have

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

5 Numerical experiments

In order to illustrate the applicability of our proposal, we implemented Algorithm 1 with the Armijo-type stepsize and tested it in the functions of the examples in Section 4. Without attempting to go into details, we mention that the Armijo-type line search sketched out in Strategy 3 was coded based on (quadratic) polynomial interpolations of the coordinate functions. We refer the reader to [46] for a careful discussion about line search strategies for vector optimization problems. We set $\delta = 10^{-4}$, $t_{\min} = 10^{-2}$,

$t_{\max} = 10^2$, $\omega_1 = 0.05$, and $\omega_2 = 0.95$. Given a Riemannian manifold \mathcal{M} , the steepest descent direction v_p at a non-critical point $p \in \mathcal{M}$ as in (4) can be calculated by solving for $\lambda \in \mathbb{R}$ and $u \in T_p\mathcal{M}$ the following differentiable problem

$$\begin{aligned} \text{Minimize} \quad & \lambda + \frac{1}{2} \langle u, u \rangle \\ \text{subject to} \quad & \langle \text{grad } f_i(p), u \rangle \leq \lambda, \quad i = 1, \dots, m, \end{aligned} \quad (31)$$

which is a convex quadratic problem with linear inequality constraints, see [8]. In our implementation, for calculating v_p , we solve problem (31) using Algencan [47], an augmented Lagrangian code for general nonlinear programming.

We stopped the execution of the algorithm at p_k declaring convergence if

$$\max_{i \in \mathcal{I}} \langle \text{grad } f_i(p_k), v_k \rangle + \frac{1}{2} \|v_k\|^2 \geq -5 \times \text{eps}^{1/2},$$

where $\mathcal{I} = \{1, \dots, m\}$, and eps denotes the machine precision given. In our experiments we used $\text{eps} = 2^{-52} \approx 2.22 \times 10^{-16}$. We point out that this convergence criterion was proposed in the numerical tests of [48] and also used in [1, 16]. The maximum number of allowed iterations was set to 10000. Codes are written in double precision Fortran 90 and are freely available at <https://orizon.ime.ufg.br/>.

5.1 Rosenbrock's Problem

We start the numerical experiments by verifying the practical behavior of Algorithm 1 in a small instance of the Rosenbrock's problem given by the functions in Example 28. We considered $n = 1$, $m = 2$ in (28), and set $F(x) = (f_1(x), f_2(x))$ where

$$f_1(x_1, x_2) = 100 (x_1^2 - x_2)^2 + (x_1 - 1)^2, \quad (32)$$

$$f_2(x_1, x_2) = 100 (x_1^2 - x_2)^2 + (x_1 - 2)^2. \quad (33)$$

Functions f_1 and f_2 have global minimizers at $x^* = (1, 1)$ and $\hat{x} = (2, 4)$, respectively. Note that $f_1(x^*) = f_2(\hat{x}) = 0$ and $f_1(\hat{x}) = f_2(x^*) = 1$. Figure 1(a) shows a representation of the image set of $F(x)$ around the Pareto front, obtained by discretizing the square $[-5, 5] \times [-5, 5]$ by a fine grid and plotting all the image points. We run the algorithm 1000 times using starting points from a uniform random distribution belonging to $(-5, 5) \times (-5, 5)$. In all instances, the Algorithm 1 stopped at a point satisfying the convergence criterion. Figure 1(b) shows the image set of all final iterates. Thus, given a reasonable number of starting points, Algorithm 1 was able to estimate the Pareto front of the considered Rosenbrock's problem. The value space generated by the Riemannian gradient method using others 200 random starting points with image belonging to the box $(0, 4) \times (0, 4)$ can be seen in Figure 2(a). A full point represents a final iterate whereas the beginning of a straight segment represents the corresponding starting point.

For comparative purposes, we implemented and tested the Euclidean gradient method for minimizing (32)–(33). In summary, the Euclidean method corresponds to Algorithm 1 with the usual inner product and the exponential map given by $\exp_x(v) = x + v$.

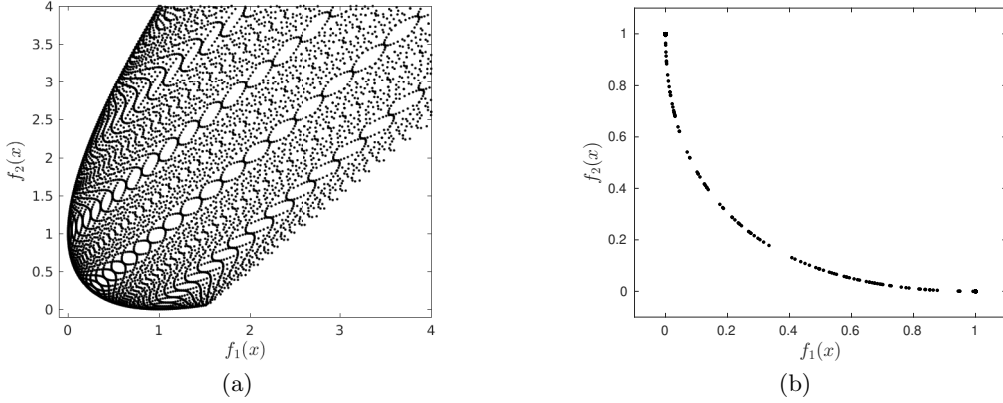


Figure 1: (a): Image set of the Rosenbrock's problem *around* its Pareto front; (b) value space of Rosenbrock's problem for 1000 random starting points belonging to $(-5, 5) \times (-5, 5)$.

We point out that an equivalent Armijo-type line search employed in the Riemannian case was coded in the Euclidean algorithm. We also run the Euclidean algorithm using the same 1000 starting points belonging to $(-5, 5) \times (-5, 5)$ considered for the Riemannian algorithm. For each method, Table 1 reports the percentages of runs that has reached a critical point (%) and, for the successful runs, the median of number of iterations (it), the median of functions evaluations (evalf), and the median of gradient evaluations (evalg). Thus, the reported data in Table 1 represents a typical run of the Riemannian and the Euclidean algorithms. It is worth noting that we considered each evaluation of a coordinate function (resp. gradient) in the calculation of evalf (resp. evalg). Note that the number of steepest descent direction calculations is equal to the number of iterations.

	%	it	evalf	evalg
Riemannian method	100.0	5.0	49.0	12.0
Euclidean method	95.1	1629.0	5721.0	3260.0

Table 1: Performance of the Riemannian and Euclidean gradient methods in the Rosenbrock's problem.

As can be seen in Table 1, in the considered Rosenbrock's problem, the Riemannian algorithm is much superior to the Euclidean one. The introduction of a suitable metric that makes F convex with componentwise Lipschitz continuous Jacobian enabled a huge reduction in computational cost to solve the problem. Figure 2(b) shows a typical behavior of the methods on the Rosenbrock's problem (32)–(33). For each method, we plotted the image set of the generated sequence for the particular case where the starting point is $(0.5, 0.2)$. The convergence criterion was satisfied with 25 and 1585 iterations for the Riemannian and Euclidean gradient methods, respectively. Due to the small

steps sizes performed by the Euclidean method (typically of the order of 10^{-3}), the corresponding path illustrated in the Figure 2(b) appears to be a continuous segment. In its turn, the Riemannian method quickly approaches the Pareto front.

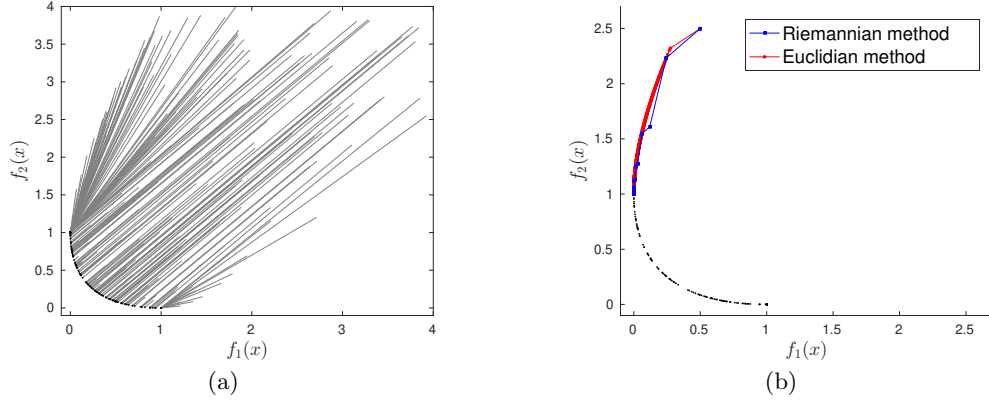


Figure 2: (a) Value space of Rosenbrock's problem for 200 starting points with image belonging to the box $(0, 4) \times (0, 4)$; (b) a typical behavior of the Riemannian and the Euclidean gradient methods on the Rosenbrock's problem.

5.2 Example in the Positive Orthant

Now we consider the application of Algorithm 1 for minimizing the vector function $F(x) = (f_1(x), \dots, f_m(x))$ where $f_j(x)$ is given by (29). Note that for the Riemannian manifold $\mathcal{M} = (\mathbb{R}_{++}^n, G)$ and $x \in \mathbb{R}_{++}^n$, the tangent space $T_x \mathcal{M}$ corresponds to \mathbb{R}^n . Thus, problem (31) to calculate v_x is directly posed as a quadratic programming problem.

Since in the previous section we solved only a small Rosenbrock's problem, we now consider larger instances of the problem related to Example 29. First, we kept the number of objectives equal to two and varied the dimension of the space assigning the following values: $n = 10, 100, 400$, and 1000 . In the second set of tests, we set $n = 100$ and varied the number of objectives taking $m = 10, 20, 100$, and 200 . All the parameters of each function f_j in (29) were random generated belonging to $(0, 1)$. Each problem instance was solved 20 times using starting points from a uniform random distribution inside the box $(0, 10)^n$. The results in Table 2 are given in the same form as Table 1.

The highlight of Table 2 is that Algorithm 1 was robust with respect to the dimension and to the number of objectives, which is consistent with the theoretical results. The results of the present section suggest that Algorithm 1 is potentially able to solve large problems. Surprisingly, for the first set of problems, a fewer number of function/gradient evaluations were required for the case where $n = 1000$ compared to smaller instances of the problem.

n	m	%	it	evalf	evalg
10	2	100.0	26.5	117.5	55.0
100	2	100.0	71.5	220.0	145.0
400	2	100.0	273.0	622.0	548.0
1000	2	100.0	17.0	104.0	36.0

(a)

n	m	%	it	evalf	evalg
100	10	100.0	33.5	592.5	345.0
100	20	100.0	34.0	1039.5	700.0
100	100	100.0	24.5	3251.5	2550.0
100	200	100.0	36.5	8864.0	7500.0

(b)

Table 2: Performance of the Riemannian gradient method related to Example 29 varying: (a) the dimension of the space; (b) the number of objectives.

5.3 Example in the Cone of Symmetric Positive Definite Matrices

Let \mathcal{M} be the Riemannian manifold $(\mathbb{P}_{++}^n, \langle \cdot, \cdot \rangle)$, where the inner product is defined as in Example 22. For $X \in \mathbb{P}_{++}^n$, the tangent space $T_X \mathcal{M}$ corresponds to the set of the symmetric matrices \mathbb{P}^n . In our implementation, in order to compute the steepest descent direction, in addition to λ , the unknowns of problem (31) are the $(n^2 + n)/2$ entries of the lower triangular part of the symmetric matrix u .

Given $X \in \mathbb{P}_{++}^n$ and $V \in \mathbb{P}^n$, direct calculations shows that the exponential map in (26) can be rewritten as $\exp_X(V) = X e^{X^{-1}V}$. For computing the inverse of matrix X , we used the LAPACK routine `dpotri` which uses the Cholesky factorization of X . For computing matrix exponentials, we used `dgpadm` routine of EXPOKIT package [49]. It should be noted that `dpotri` and `dgpadm` are dense routines.

We considered bicriteria and three-criteria problem instances related to Example 22. The parameters of function (25) were randomly generated belonging to $(0, 1)$. For each instance, we run the Riemannian gradient method 20 times using random starting points with eigenvalues belonging to the interval $(0, 100)$. The results in Table 3 show that Algorithm 1 solved all the instances with a moderate computational effort. It is worth mentioning that in a typical iteration, the first trial step size of Strategy 3 defined by

$$\hat{t}_{k_0} = \max\{t_{\min}, \min\{\bar{t}_{k_0}, t_{\max}\}\}, \quad \bar{t}_{k_0} = \begin{cases} 1/\|v_0\|, & \text{if } k = 0, \\ t_{k-1} \|v_{k-1}\|^2 / \|v_k\|^2, & \text{if } k \geq 1, \end{cases} \quad (34)$$

satisfies the sufficient descent condition (11). Indeed, as it can be seen Table 3, the values reported in `evalf` columns are slightly greater than the corresponding number of iterations times the number of objectives m . We observe that the choice (34) corresponds to the safeguarded Shanno and Phua [50] recommendation and was first proposed in the multiobjective optimization setting in [1].

Finally, we report that Algorithm 1 converges with a single iteration when applied to instances of Example 23. The considered metric makes it possible to explore the structure of the problem turning it into a trivial problem from the Riemannian perspective.

6 Conclusions

In this paper, the behavior of the steepest descent method for multiobjective optimization on Riemannian manifolds with lower bounded sectional curvature is analyzed. It

n	m	%	it	evalf	evalg
5	2	100.0	8.0	26.5	18.0
10	2	100.0	13.0	38.5	28.0
20	2	100.0	18.0	49.0	38.0
50	2	100.0	27.0	64.0	56.0

(a)

n	m	%	it	evalf	evalg
5	3	100.0	7.0	28.5	24.0
10	3	100.0	12.0	45.5	39.0
20	3	100.0	18.0	68.0	57.0
50	3	100.0	28.0	91.0	87.0

(b)

Table 3: Performance of the Riemannian gradient method related to Example 22 for: (a) bicriteria problems; (b) three-criteria problems.

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.

References

1. Lucambio Pérez, L.R., Prudente, L.F.: Nonlinear conjugate gradient methods for vector optimization. *SIAM J. Optim.* **28**(3), 2690–2720 (2018).
2. Gonçalves, M.L.N., Prudente, L.F.: On the extension of the Hager-Zhang conjugate gradient method for vector optimization. Technical report pp. 1–19 (2018).
3. Bento, G.C., Cruz Neto, J.X., López, G., Soubeyran, A., Souza, J.C.O.: 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).
4. Montonen, O., Karmitsa, N., Mäkelä, M.M.: Multiple subgradient descent bundle method for convex nonsmooth multiobjective optimization. *Optimization* **67**(1), 139–158 (2018).
5. Carrizo, G.A., Lotito, P.A., Maciel, M.C.: Trust region globalization strategy for the nonconvex unconstrained multiobjective optimization problem. *Math. Program.* **159**(1-2, Ser. A), 339–369 (2016).
6. Fliege, J., Vaz, A.I.F.: A method for constrained multiobjective optimization based on SQP techniques. *SIAM J. Optim.* **26**(4), 2091–2119 (2016).
7. Morovati, V., Pourkarimi, L., Basirzadeh, H.: Barzilai and Borwein’s method for multiobjective optimization problems. *Numer. Algorithms* **72**(3), 539–604 (2016).
8. Fliege, J., Svaiter, B.F.: Steepest descent methods for multicriteria optimization. *Math. Methods Oper. Res.* **51**(3), 479–494 (2000).
9. Fukuda, E.H., Graña Drummond, L.M.: Inexact projected gradient method for vector optimization. *Comput. Optim. Appl.* **54**(3), 473–493 (2013).
10. Fukuda, E.H., Graña Drummond, L.M.: On the convergence of the projected gradient method for vector optimization. *Optimization* **60**(8-9), 1009–1021 (2011).

11. Graña Drummond, L.M., Svaiter, B.F.: A steepest descent method for vector optimization. *J. Comput. Appl. Math.* **175**(2), 395–414 (2005).
12. Graña Drummond, L.M., Iusem, A.N.: A projected gradient method for vector optimization problems. *Comput. Optim. Appl.* **28**(1), 5–29 (2004).
13. Bello Cruz, J.Y., Bouza Allende, G.: A steepest descent-like method for variable order vector optimization problems. *J. Optim. Theory Appl.* **162**(2), 371–391 (2014).
14. Bello Cruz, J.Y.: A subgradient method for vector optimization problems. *SIAM J. Optim.* **23**(4), 2169–2182 (2013).
15. Fliege, J., Vaz, A.I.F., Vicente, L.N.: Complexity of gradient descent for multiobjective optimization. *Optimization Methods and Software* **0**, 1–11 (2018).
16. Ferreira, O.P., Louzeiro, M.S., Prudente, L.F.: Gradient Method for Optimization on Riemannian Manifolds with Lower Bounded Curvature. To appear in *SIAM J. Optim.* ArXiv e-prints. (2018).
17. Bento, G.C., Ferreira, O.P., Melo, J.G.: Iteration-Complexity of Gradient, Subgradient and Proximal Point Methods on Riemannian Manifolds. *J. Optim. Theory Appl.* **173**(2), 548–562 (2017).
18. Jeuris, B., Vandebril, R., Vandereycken, B.: A survey and comparison of contemporary algorithms for computing the matrix geometric mean. *Electron. Trans. Numer. Anal.* **39**, 379–402 (2012)
19. Rapcsák, T.: Smooth nonlinear optimization in \mathbf{R}^n , *Nonconvex Optimization and its Applications*, vol. 19. Kluwer Academic Publishers, Dordrecht (1997).
20. Sra, S., Hosseini, R.: Conic geometric optimization on the manifold of positive definite matrices. *SIAM J. Optim.* **25**(1), 713–739 (2015).
21. Zhang, H., Reddi, S.J., Sra, S.: Fast stochastic optimization on Riemannian manifolds. ArXiv e-prints pp. 1–17 (2016).
22. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.* **20**(2), 303–353 (1999).
23. Smith, S.T.: Optimization techniques on Riemannian manifolds. In: Hamiltonian and gradient flows, algorithms and control, *Fields Inst. Commun.*, vol. 3, pp. 113–136. Amer. Math. Soc., Providence, RI (1994)
24. Luenberger, D.G.: The gradient projection method along geodesics. *Management Sci.* **18**, 620–631 (1972)
25. Udriște, C.: Convex functions and optimization methods on Riemannian manifolds, *Mathematics and its Applications*, vol. 297. Kluwer Academic Publishers Group, Dordrecht (1994).

26. Absil, P.A., Mahony, R., Sepulchre, R.: Optimization algorithms on matrix manifolds. Princeton University Press, Princeton, NJ (2008). <https://doi.org/10.1515/9781400830244>. With a foreword by Paul Van Dooren
27. Nesterov, Y.E., Todd, M.J.: On the Riemannian geometry defined by self-concordant barriers and interior-point methods. *Found. Comput. Math.* **2**(4), 333–361 (2002).
28. Li, C., Mordukhovich, B.S., Wang, J., Yao, J.C.: Weak sharp minima on Riemannian manifolds. *SIAM J. Optim.* **21**(4), 1523–1560 (2011).
29. Li, C., Yao, J.C.: 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).
30. Wang, X., Li, C., Wang, J., Yao, J.C.: Linear convergence of subgradient algorithm for convex feasibility on Riemannian manifolds. *SIAM J. Optim.* **25**(4), 2334–2358 (2015).
31. Wang, X.M., Li, C., Yao, J.C.: Subgradient projection algorithms for convex feasibility on Riemannian manifolds with lower bounded curvatures. *J. Optim. Theory Appl.* **164**(1), 202–217 (2015).
32. Manton, J.H.: A framework for generalising the Newton method and other iterative methods from Euclidean space to manifolds. *Numer. Math.* **129**(1), 91–125 (2015).
33. Zhang, H., Sra, S.: First-order methods for geodesically convex optimization. *JMLR: Workshop and Conference Proceedings* **49**(1), 1–21 (2016). <https://arxiv.org/abs/1602.06053>
34. Wang, J., Li, C., Lopez, G., Yao, J.C.: Proximal point algorithms on Hadamard manifolds: linear convergence and finite termination. *SIAM J. Optim.* **26**(4), 2696–2729 (2016).
35. Bento, G.C., Ferreira, O.P., Oliveira, P.R.: Unconstrained steepest descent method for multicriteria optimization on Riemannian manifolds. *J. Optim. Theory Appl.* **154**(1), 88–107 (2012).
36. Bento, G.C., da Cruz Neto, J.X., Santos, P.S.M.: An inexact steepest descent method for multicriteria optimization on Riemannian manifolds. *J. Optim. Theory Appl.* **159**(1), 108–124 (2013).
37. do Carmo, M.P.: Riemannian geometry. *Mathematics: Theory & Applications*. Birkhäuser Boston, Inc., Boston, MA (1992). doi:10.1007/978-1-4757-2201-7.
38. Sakai, T.: Riemannian geometry, *Translations of Mathematical Monographs*, vol. 149. American Mathematical Society, Providence, RI (1996).

39. da Cruz Neto, J., De Lima, L., Oliveira, P.: Geodesic algorithms in riemannian geometry. *Balkan J. Geom. Appl* **3**(2), 89–100 (1998)
40. Burachik, R., Drummond, L.M.G., Iusem, A.N., Svaiter, B.F.: Full convergence of the steepest descent method with inexact line searches. *Optimization* **32**(2), 137–146 (1995).
41. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2**(1), 183–202 (2009).
42. Batista, E., Bento, G., Ferreira, O.: An extragradient-type algorithm for variational inequality on hadamard manifolds. *arXiv preprint arXiv:1804.09292* (2018)
43. Lang, S.: Fundamentals of differential geometry, *Graduate Texts in Mathematics*, vol. 191. Springer-Verlag, New York (1999).
44. Lin, B., He, X., Zhang, C., Ji, M.: Parallel vector field embedding. *J. Mach. Learn. Res.* **14**, 2945–2977 (2013)
45. Petersen, P.: Riemannian geometry, *Graduate Texts in Mathematics*, vol. 171, third edn. Springer, Cham (2016).
46. Lucambio Pérez, L.R., Prudente, L.F.: A Wolfe line search algorithm for vector optimization. Technical report pp. 1–21 (2018). http://www.optimization-online.org/DB_HTML/2018/07/6744.html
47. Birgin, E.G., Martínez, J.M.: Practical augmented Lagrangian methods for constrained optimization. SIAM (2014)
48. Fliege, J., Graña Drummond, L.M., Svaiter, B.F.: Newton’s Method for Multiobjective Optimization. *SIAM Journal on Optimization* **20**(2), 602–626 (2009)
49. Sidje, R.B.: Expokit: A software package for computing matrix exponentials. *ACM Transactions on Mathematical Software (TOMS)* **24**(1), 130–156 (1998)
50. Shanno, D.F., Phua, K.H.: Remark on algorithm 500: Minimization of unconstrained multivariate functions. *ACM Transactions on Mathematical Software* **6**(4), 618–622 (1980)