J Optim Theory Appl (2012) 154:88–107
DOI 10.1007/s10957-011-9984-2

Unconstrained Steepest Descent Method for Multicriteria Optimization on Riemannian Manifolds

G.C. Bento · O.P. Ferreira · P.R. Oliveira

Received: 21 June 2011 / Accepted: 28 December 2011 / Published online: 19 January 2012

Abstract In this paper, we present a steepest descent method with Armijo's rule for multicriteria optimization in the Riemannian context. The sequence generated by the method is guaranteed to be well defined. Under mild assumptions on the multicriteria function, we prove that each accumulation point (if any) satisfies first-order necessary conditions for Pareto optimality. Moreover, assuming quasiconvexity of the multicriteria function and nonnegative curvature of the Riemannian manifold, we prove full convergence of the sequence to a critical Pareto point.

Keywords Steepest descent · Pareto optimality · Vector optimization · Quasi-Fejér convergence · Quasiconvexity · Riemannian manifolds

1 Introduction

The steepest descent method with Armijo's rule for real continuously differentiable optimization problem (see, for instance, Burachik et al. [1]), generates a sequence such that any accumulation point of it, if any, is critical for the objective function. This fact was generalized for multicriteria optimization by Fliege and Svaiter [2], namely, whenever the objective function is a vectorial function. Full convergence for a real

Communicated by Dinh The Luc.

G.C. Bento (✉) · O.P. Ferreira
IME-Universidade Federal de Goiás, Goiânia, GO 74001-970, Brazil
e-mail: glaydston@mat.ufg.br

O.P. Ferreira
e-mail: orizon@mat.ufg.br

P.R. Oliveira
COPPE/Sistemas-Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ 21945-970, Brazil
e-mail: poliveir@cos.ufrj.br

optimization problem was assured under the assumption that the solution set of the problem be nonempty and the objective function is a convex function; see Burachik et al. [1] (or, more generally, a quasiconvex function; see Kiwiel and Murty [3]), which has been generalized for vector optimization by Graña Drummond and Svaiter [4] (see also Graña Drummond and Iusem [5]).

Extension of concepts and techniques, as well as methods from Euclidean spaces to Riemannian manifolds, is natural and, in general, nontrivial; see, for instance, [6]. In the last few years, such extension settings with practical and theoretical purpose have been the subject of much new research. Recent works dealing with this issue include [7–30]. The generalization of optimization methods from Euclidean space to Riemannian manifold have some important advantages. For example, constrained optimization problems can be seen as unconstrained ones from the Riemannian geometry viewpoint (the constrained set is a manifold), and in this case, we have an alternative possibility besides the projection idea for solving the problem. Moreover, nonconvex problems in the classical context may become convex through the introduction of an appropriate Riemannian metric (see, for example, [12]).

The steepest descent method for continuously differentiable Riemannian problems has been studied by Udriste [31], Smith [32], and Rapcsák [6], and partial convergence results were obtained. For the convex case, the full convergence using Armijo's rule has been generalized by da Cruz Neto et al. [33], in the particular case of the Riemannian manifold has nonnegative curvature. Using the same restrictive assumption on the manifold, Papa Quiroz et al. [18] generalized the full convergence result for quasiconvex objective function.

In this paper, following the ideas of Fliege and Svaiter [2], we generalize their convergence results for multicriteria optimization to the Riemannian context. Besides this, following the ideas of Graña Drummond and Svaiter [4], we generalize the full convergence result for multicriteria optimization, where the multicriteria function is quasiconvex and the Riemannian manifold has nonnegative curvature.

The organization of our paper is as follows. In Sect. 2, some notations and results of Riemannian geometry, used throughout of the paper, are defined. In Sect. 3, the multicriteria problem, the first-order optimality condition for it and some basic definitions are presented. In Sect. 4, the Riemannian steepest descent method for finding one solution of multicriteria problems is stated and the well-definedness of the sequence generated for it is established. In Sect. 5, a partial convergence result for continuous differentiability multicriteria optimization is presented without any additional assumption on the objective function. Moreover, assuming that the objective function be quasiconvex and the Riemannian manifold has nonnegative curvature, a full convergence result is presented. Finally, in Sect. 6, some examples of complete Riemannian manifolds with explicit geodesic curves and the steepest descent iteration of the sequence generated by the proposed method are presented.

2 Preliminaries on Riemannian Geometry

In this section, we introduce some fundamental properties and notations of Riemannian manifolds. These basic facts can be found in any introductory book on Riemannian geometry; see, for example, [34, 35].

Let M be an n-dimensional connected manifold. We denote by T_pM the n-dimensional *tangent space* of M at p, and by $TM = \bigcup_{p\in M} T_pM$ *tangent bundle* of M and by $\mathcal{X}(M)$ the space of smooth vector fields over M. Suppose that M be endowed with a Riemannian metric $\langle , \rangle$, with corresponding norm denoted by $\| \|$, that is, M is a Riemannian manifold. Recall that the metric can be used to define the length of piecewise smooth curves $\gamma : [a, b] \to M$ joining p to q, i.e., such that $\gamma(a) = p$ and $\gamma(b) = q$, by

$$l(\gamma) = \int_a^b \|\gamma'(t)\|\, dt,$$

and, moreover, by minimizing this functional length over the set of all such curves, we obtain a Riemannian distance $d(p, q)$, which induces the original topology on M. The metric induces a map $f \mapsto \operatorname{grad} f \in \mathcal{X}(M)$, which associates to each scalar function smooth over M its gradient via the rule $\langle \operatorname{grad} f, X\rangle = df(X),\ \ X \in \mathcal{X}(M)$. Let ∇ be the Levi–Civita connection associated to $(M, \langle , \rangle)$. A vector field V along γ is said to be *parallel* iff $\nabla_{\gamma'} V = 0$. If γ' itself is parallel we say that γ is a *geodesic*. Because the geodesic equation $\nabla_{\gamma'}\gamma' = 0$ is a second-order nonlinear ordinary differential equation, then the geodesic $\gamma = \gamma_v(., p)$ is determined by its position p and velocity v at p. It is easy to check that $\|\gamma'\|$ is constant. We say that γ is *normalized* if $\|\gamma'\| = 1$. The restriction of a geodesic to a closed bounded interval is called a *geodesic segment*. A geodesic segment joining p to q in M is said to be *minimal* iff its length is equals to $d(p, q)$, and in this case, the geodesic is called a *minimizing geodesic*.

A Riemannian manifold is *complete* iff geodesics are defined for any values of t. The Hopf–Rinow theorem asserts that, if this is the case then any pair of points, say p and q, in M can be joined by a (not necessarily unique) minimal geodesic segment. Moreover, (M, d) is a complete metric space and bounded and closed subsets are compact. If $p \in M$, then the *exponential map* $\exp_p : T_pM \to M$ is defined by $\exp_p v = \gamma_v(1, p)$.

We denote by R *the curvature tensor* defined by $R(X, Y)Z = \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z$, with $X, Y, Z \in \mathcal{X}(M)$, where $[X, Y] = YX - XY$. Then the *sectional curvature* with respect to X and Y is given by $K(X, Y) = \langle R(X, Y)Y, X\rangle / (\|X\|^2\|Y\|^2 - \langle X, Y\rangle^2)$, where $\|X\|^2 = \langle X, X\rangle$.

In Sect. 5.2 of this paper, we will be mainly interested in Riemannian manifolds for which $K(X, Y) \geq 0$ for any $X, Y \in \mathcal{X}(M)$. Such manifolds are referred to as *manifolds with nonnegative curvature*. A fundamental geometric property of this class of manifolds is that the distance between points on the geodesics issuing from one point is, at least locally, bounded from above by the distance between the points on the respective rays in the tangent space. A global formulation of this general principle is the *law of cosines* that we now pass to describe. A *geodesic hinge* in M is a pair of normalized geodesic segments γ_1 and γ_2 such that $\gamma_1(0) = \gamma_2(0)$ and at least one of them, say γ_1, is minimal. From now on $l_1 = l(\gamma_1)$, $l_2 = l(\gamma_2)$, $l_3 = d(\gamma_1(l_1), \gamma_2(l_2))$, and $\alpha = \sphericalangle(\gamma_1'(0), \gamma_2'(0))$.

Theorem 2.1 (Law of cosines) *In a complete Riemannian manifold with nonnegative curvature, with the notation introduced above, we have*

$$l_3^2 \leq l_1^2 + l_2^2 - 2l_1 l_2 \cos\alpha. \tag{1}$$

Proof See [34] and [35]. □

In this paper, M will denote a complete n-dimensional Riemannian manifold.

3 The Multicriteria Problem

In this section, we present the multicriteria problem, the first-order optimality condition for it and some basic definitions.

Let $I := \{1, \ldots, m\}$, $\mathbb{R}^m_+ = \{x \in \mathbb{R}^m : x_i \geq 0,\ j \in I\}$ and $\mathbb{R}^m_{++} = \{x \in \mathbb{R}^m : x_j > 0,\ j \in I\}$. 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 $y - x \in \mathbb{R}^m_{++}$.

Given a continuously differentiable vector function $F : M \to \mathbb{R}^m$, we consider the problem of finding a *optimum Pareto point* of F, i.e., a point $p^* \in M$ such that there exists no other $p \in M$ with $F(p) \preceq F(p^*)$ and $F(p) \neq F(p^*)$. We denote this unconstrained problem in the Riemannian context as

$$\min_{p \in M} F(p). \tag{2}$$

Let F be given by $F(p) := (f_1(p), \ldots, f_m(p))$. We denote the Riemannian Jacobian of F by

$$\operatorname{grad} F(p) := \big(\operatorname{grad} f_1(p), \ldots, \operatorname{grad} f_m(p)\big), \quad p \in M,$$

and the image of the Riemannian Jacobian of F at a point $p \in M$ by

$$\operatorname{Im}\big(\operatorname{grad} F(p)\big) := \big\{\operatorname{grad} F(p)v = \big(\langle \operatorname{grad} f_1(p), v\rangle, \ldots, \langle \operatorname{grad} f_m(p), v\rangle\big) : v \in T_pM\big\}, \quad p \in M.$$

Using the above equality, the first-order optimality condition for the problem (2) is stated as

$$p \in M, \quad \operatorname{Im}\big(\operatorname{grad} F(p)\big) \cap \big(-\mathbb{R}^m_{++}\big) = \emptyset. \tag{3}$$

Remark 3.1 Note that the condition in (3) generalizes to vector optimization the classical condition $\operatorname{grad} F(p) = 0$ for the scalar case, i.e., $m = 1$.

In general, (3) is necessary, but not sufficient, for optimality. So, a point $p \in M$ satisfying (3) is called *critical Pareto point*.

4 Steepest Descent Methods for Multicriteria Problems

In this section, we state the Riemannian steepest descent methods for solving multicriteria problems and establish a well-defined sequence generated for them.

Let $p \in M$ be a point which is not critical Pareto point. Then there exists a direction $v \in T_pM$ satisfying

$$\operatorname{grad} F(p)v \in -\mathbb{R}^m_{++},$$

that is, $\operatorname{grad} F(p)v \prec 0$. In this case, v is called a *descent direction* for F at p.

For each $p \in M$, we consider the following unconstrained optimization problem in the tangent plane T_pM:

$$\min_{v \in T_pM} \left\{ \max_{i \in I} \langle \operatorname{grad} f_i(p), v \rangle + (1/2)\|v\|^2 \right\}, \quad I = \{1, \ldots, m\}. \tag{4}$$

Lemma 4.1 *The unconstrained optimization problem in* (4) *has only one solution. Moreover, the vector v is the solution of the problem in* (4) *if and only if there exists $\alpha_i \geq 0$, $i \in I(p, v)$, such that*

$$v = -\sum_{i \in I(p,v)} \alpha_i \operatorname{grad} f_i(p), \qquad \sum_{i \in I(p,v)} \alpha_i = 1,$$

where $I(p, v) := \{i \in I : \langle \operatorname{grad} f_i(p), v \rangle = \max_{i \in I} \langle \operatorname{grad} f_i(p), v \rangle\}$.

Proof Since the function

$$T_pM \ni v \mapsto \max_{i \in I} \langle \operatorname{grad} f_i(p), v \rangle,$$

is the maximum of linear functions in the linear space T_pM, it is convex. So, it is easy to see that the function

$$T_pM \ni v \mapsto \max_{i \in I} \langle \operatorname{grad} f_i(p), v \rangle + (1/2)\|v\|^2, \tag{5}$$

is strongly convex, which implies that the problem in (4) has only one solution in T_pM and the first statement is proved.

From the convexity of the function in (5), it is well known that v is the solution of the problem in (4) if and only if

$$0 \in \partial \left(\max_{i \in I} \langle \operatorname{grad} f_i(p), . \rangle + (1/2)\|.\|^2 \right)(v),$$

or equivalently,

$$-v \in \partial \left(\max_{i \in I} \langle \operatorname{grad} f_i(p), . \rangle \right)(v).$$

Therefore, the second statement follows from the formula for the subdifferential of the maximum of convex functions (see [36], Vol. I, Corollary VI.4.3.2). □

Lemma 4.2 *If $p \in M$ is not a critical Pareto point of F and v is the solution of the problem in (4), then*

$$\max_{i\in I}\langle \operatorname{grad} f_i(p), v\rangle + (1/2)\|v\|^2 < 0.$$

In particular, v is a descent direction.

Proof Since p is not a critical Pareto point, there exists $0 \neq \hat{v} \in T_pM$ such that $\operatorname{grad} F(p)\hat{v} \prec 0$. In particular,

$$\beta = \max_{i\in I}\langle \operatorname{grad} f_i(p), \hat{v}\rangle < 0.$$

As $-\beta/\|\hat{v}\|^2 > 0$, letting $\bar{v} = (-\beta/\|\hat{v}\|^2)\hat{v}$, we obtain

$$\max_{i\in I}\langle \operatorname{grad} f_i(p), \bar{v}\rangle + (1/2)\|\bar{v}\|^2 = -\frac{\beta^2}{2\|\hat{v}\|^2} < 0,$$

Using v as the solution of the problem in (4), the first part of the lemma follows from the last inequality. The second part of the lemma is an immediate consequence of the first one. □

In view of the two previous lemmas and (4), we define the steepest descent direction function for F as follows.

Definition 4.1 The steepest descent direction function for F is defined as

$$M \ni p \mapsto v(p) := \operatorname*{argmin}_{v\in T_pM}\left\{\max_{i\in I}\langle \operatorname{grad} f_i(p), v\rangle + (1/2)\|v\|^2\right\} \in T_pM.$$

Remark 4.1 As an immediate consequence of Lemma 4.1, it follows that the steepest descent direction for vector functions becomes the steepest descent direction when $m = 1$. See, for example, [6, 31–33, 37]. When $M = \mathbb{R}^n$, we retrieve the steepest descent direction proposed in [2].

The *steepest descent method with the Armijo rule* for solving the unconstrained optimization problem (2) is as follows.

Method 4.1 (Steepest descent method with Armijo rule)

INITIALIZATION. Take $\beta \in (0, 1)$ and $p_0 \in M$. Set $k = 0$.
STOP CRITERION. If p^k is a critical Pareto point STOP. Otherwise.
ITERATIVE STEP. Compute the steepest descent direction v^k for F at p^k, i.e.,

$$v^k := v\big(p^k\big), \tag{6}$$

and the step-length $t_k \in]0, 1]$ is as follows:

$$t_k := \max\left\{2^{-j} : j \in \mathbb{N}, F\big(\exp_{p^k}\big(2^{-j}v^k\big)\big) \preceq F\big(p^k\big) + \beta 2^{-j} \operatorname{grad} F\big(p^k\big)v^k\right\}, \tag{7}$$

and set

$$p^{k+1} := \exp_{p^k}\left(t_k v^k\right), \tag{8}$$

and GOTO STOP CRITERION.

Remark 4.2 The steepest descent method for vector optimization in Riemannian manifolds becomes the classical steepest descent method when $m = 1$, which has appeared, for example, in [6, 31–33].

Proposition 4.1 *The sequence $\{p^k\}$ generated by the steepest descent method with Armijo rule is well defined.*

Proof Assume that p^k is not a critical Pareto point. From Definition 4.1 and Lemma 4.1, $v^k = v(p^k)$ is well defined. Thus, to prove that the method proposed is well defined, it is enough to prove that the step-length is well defined. For this, first note that from Definition 4.1 and Lemma 4.2

$$\operatorname{grad} F\left(p^k\right)v^k \prec 0.$$

Since $F : M \to \mathbb{R}^m$ is a continuously differentiable vector function, $\operatorname{grad} F(p^k)v^k \prec 0$ and $\beta \in (0, 1)$, we have

$$\lim_{t\to 0^+} \frac{F(\exp_{p^k}(tv^k)) - F(p^k)}{t} = \operatorname{grad} F\left(p^k\right)v^k \prec \beta \operatorname{grad} F\left(p^k\right)v^k \prec 0.$$

Therefore, it is straightforward to show that there exists $\delta \in (0, 1]$ such that

$$F\left(\exp_{p^k}\left(tv^k\right)\right) \prec F\left(p^k\right) + \beta t \operatorname{grad} F\left(p^k\right)v^k, \quad t \in (0, \delta).$$

As $\lim_{j\to\infty} 2^{-j} = 0$, last vector inequality implies that the step-length (7) is well defined. Hence, p^{k+1} is also well defined and the proposition is concluded. □

5 Convergence Analysis

In this section, following the ideas of [2] we prove a partial convergence result without any additional assumption on F besides the continuous differentiability. In the sequel, following [4], assuming quasiconvexity of F and nonnegative curvature for M, we extend the full convergence result presented in [33] and [18] to optimization of vector functions. It can be immediately seen that, if Method 4.1 terminates after a finite number of iterations, then it terminates at a critical Pareto point. From now on, we will assume that $\{p^k\}$, $\{v^k\}$ and $\{t_k\}$ are infinite sequences generated by Method 4.1.

5.1 Partial Convergence Result

In this section, we prove that every accumulation point of $\{p^k\}$ is a critical Pareto point. Before this, we prove the following preliminary fact that will be useful.

Lemma 5.1 *The steepest descent direction function for F, $M \ni p \mapsto v(p) \in T_pM$, is continuous.*

Proof Let $\{q^k\} \subset M$ be a sequence which converges to $\bar{q}$ as k goes to $+\infty$, and $U_{\bar{q}} \subset M$ a neighborhood of $\bar{q}$ such that $TU_{\bar{q}} \approx U_{\bar{q}} \times \mathbb{R}^n$. Since $\{q^k\}$ converges to $\bar{q}$ and $TU_{\bar{q}} \subset TM$ is an open set, we assume that the whole sequence $\{(q^k, v(q^k))\}$ is in $TU_{\bar{q}}$. Define $v^k := v(q^k)$. Combining Definition 4.1 with Lemma 4.2, it is easy to see that

$$\|v^k\| \leq 2 \max_{i \in I} \|\operatorname{grad} f_j(q^k)\|.$$

As F is continuously differentiable and $\{q^k\}$ is convergent, the above inequality implies that the sequence $\{v^k\}$ is bounded. Let $\bar{v}$ be an accumulation point of the sequence $\{v^k\}$. From Definition 4.1 and Lemma 4.1 we conclude that there exist $\alpha_i^k \geq 0$, $i \in I(q^k, v^k)$, such that

$$v^k = -\sum_{i \in I(q^k, v^k)} \alpha_i^k \operatorname{grad} f_i(q^k), \qquad \sum_{i \in I(q^k, v^k)} \alpha_i^k = 1, \quad k = 0, 1, \ldots, \tag{9}$$

where $I(q^k, v^k) := \{i \in I : \langle \operatorname{grad} f_i(q^k), v^k \rangle = \max_{i \in I} \langle \operatorname{grad} f_i(q^k), v^k \rangle\}$. Using the above constants and the associated indexes, define the sequence $\{\alpha^k\}$ as

$$\alpha^k := \left(\alpha_1^k, \ldots, \alpha_m^k\right), \quad \alpha_i^k = 0,\ i \in I \setminus I\left(q^k, v^k\right),\ k = 0, 1, \ldots.$$

Let $\|.\|_1$ be the sum norm in $\mathbb{R}^m$. Since $\sum_{i \in I(q^k, v^k)} \alpha_i^k = 1$, we have $\|\alpha^k\|_1 = 1$ for all k, which implies that the sequence $\{\alpha^k\}$ is bounded. Let $\bar{\alpha}$ be an accumulation point of the sequence $\{\alpha^k\}$. Let $\{v^{k_s}\}$ and $\{\alpha^{k_s}\}$ be subsequences of $\{v^k\}$ and $\{\alpha^k\}$, respectively, such that

$$\lim_{s \to +\infty} v^{k_s} = \bar{v}, \qquad \lim_{s \to +\infty} \alpha^{k_s} = \bar{\alpha}.$$

Now, as $I = \{1, \ldots, m\}$, the cardinality of the set of all subset of I is 2^m and $I(q^{k_s}, v^{k_s}) \subset I$ for all $s \in \mathbb{N}$, we can assume without loss of generality that

$$I\left(q^{k_1}, v^{k_1}\right) = I\left(q^{k_2}, v^{k_2}\right) = \cdots = \bar{I}. \tag{10}$$

Hence, we conclude from (9) and last equalities that

$$v^{k_s} = -\sum_{i \in \bar{I}} \alpha_i^{k_s} \operatorname{grad} f_i\left(q^{k_s}\right), \quad \sum_{i \in \bar{I}} \alpha_i^{k_s} = 1, \quad s = 0, 1, \ldots.$$

Letting s go to $+\infty$ in the above equalities, we obtain

$$\bar{v} = \sum_{i \in \bar{I}} \bar{\alpha}_i \operatorname{grad} f_i(\bar{q}), \quad \sum_{i \in \bar{I}} \bar{\alpha}_i = 1. \tag{11}$$

On the other hand, $I(q^{k_s}, v^{k_s}) = \{i \in I : \langle \operatorname{grad} f_i(q^{k_s}), v^{k_s} \rangle = \max_{i \in I} \langle \operatorname{grad} f_i(q^{k_s}), v^{k_s} \rangle\}$. So, (10) implies that

$$\left\langle \operatorname{grad} f_i\left(q^{k_s}\right), v^{k_s} \right\rangle = \max_{i \in I} \left\langle \operatorname{grad} f_i\left(q^{k_s}\right), v^{k_s} \right\rangle, \quad i \in \bar{I},\ s = 0, 1, \ldots.$$

By using continuity of grad F and last equality, we have

$$\langle \operatorname{grad} f_i(\bar{q}), \bar{v} \rangle = \max_{i \in I} \langle \operatorname{grad} f_i(\bar{q}), \bar{v} \rangle, \quad i \in \bar{I}.$$

From the definition of $I(\bar{q}, \bar{v})$, we obtain $\bar{I} \subset I(\bar{q}, \bar{v})$. Therefore, combining again Definition 4.1 with Lemma 4.1 and (11), we conclude that $\bar{v} = v(\bar{q})$ and the desired result is proved. □

In next result, we use F as continuously differentiable to assure that the sequence of the functional values of the sequence $\{p^k\}$, $\{F(p^k)\}$ is monotonously decreasing and that their accumulation points are critical Pareto points.

Theorem 5.1 *The following statements hold*:

(i) $\{F(p^k)\}$ *is decreasing*;
(ii) *Each accumulation point of the sequence* $\{p^k\}$ *is a critical Pareto point*.

Proof The iterative step in Method 4.1 implies that

$$F\left(p^{k+1}\right) \preceq F\left(p^k\right) + \beta t_k \operatorname{grad} F\left(p^k\right) v^k, \quad p^{k+1} = \exp_{p^k} t_k v^k, \quad k = 0, 1, \ldots. \tag{12}$$

Since $\{p^k\}$ is an infinite sequence, for all k, p^k is not a critical Pareto point of F. Thus, item i follows from the definition of v^k together with Definition 4.1, Lemma 4.2 and the last vector inequality.

Let $\bar{p} \in M$ be an accumulation point of the sequence $\{p^k\}$ and $\{p^{k_s}\}$ a subsequence of $\{p^k\}$ such that $\lim_{s \to +\infty} p^{k_s} = \bar{p}$. Since F is continuous and $\lim_{s \to +\infty} p^{k_s} = \bar{p}$ we have $\lim_{s \to +\infty} F(p^{k_s}) = F(\bar{p})$. So, taking into account that $\{F(p^k)\}$ is a decreasing sequence and has $F(\bar{p})$ as an accumulation point, it is easy to conclude that the whole sequence $\{F(p^k)\}$ converges to $F(\bar{p})$. Using (12), Definition 4.1 and Lemma 4.2, we conclude that

$$F\left(p^{k+1}\right) - F\left(p^k\right) \preceq \beta t_k \operatorname{grad} F\left(p^k\right) v^k \preceq 0, \quad k = 0, 1, \ldots.$$

Since $\lim_{s \to +\infty} F(p^k) = F(\bar{p})$, last inequality implies that

$$\lim_{k \to +\infty} \beta t_k \operatorname{grad} F\left(p^k\right) v^k = 0. \tag{13}$$

As $\{p^{k_s}\}$ converges to $\bar{p}$, we assume that $\{(p^{k_s}, v^{k_s})\} \subset TU_{\bar{p}}$, where $U_{\bar{p}}$ is a neighborhood of $\bar{p}$ such that $TU_{\bar{p}} \approx U_{\bar{p}} \times \mathbb{R}^n$. Moreover, as the sequence $\{t_k\} \subset]0, 1]$ has an accumulation point $\bar{t} \in [0, 1]$, we assume without loss of generality that $\{t_{k_s}\}$ converges to $\bar{t}$. We have two possibilities to consider:

(a) $\bar{t} > 0$;
(b) $\bar{t} = 0$.

Assume that item (a) holds. In this case, from (13), continuity of grad F, (6) and Lemma 5.1, we obtain

$$\operatorname{grad} F(\bar{p}) v(\bar{p}) = 0,$$

which implies that

$$\max_{i\in I}\langle \operatorname{grad} f_i(\bar{p}), v(\bar{p})\rangle = 0. \tag{14}$$

On the other hand, from Definition 4.1 together with Lemma 4.2,

$$\max_{i\in I}\langle \operatorname{grad} f_i\big(p^{k_s}\big), v^{k_s}\rangle + (1/2)\big\|v^{k_s}\big\|^2 < 0.$$

Letting s go to $+\infty$ in the above inequalities and using Lemma 5.1 combined with the continuity of grad F and equality (14), we conclude that

$$\max_{i\in I}\langle \operatorname{grad} f_i(\bar{p}), v(\bar{p})\rangle + (1/2)\big\|v(\bar{p})\big\|^2 = 0.$$

Hence, it follows from last equality, Definition 4.1 and Lemma 4.2 that $\bar{p}$ is a critical Pareto point.

Now, assume that item (b) holds true. Since p^{k_s} is not critical Pareto point, we have

$$\max_{i\in I}\langle \operatorname{grad} f_i\big(p^{k_s}\big), v^{k_s}\rangle \le \max_{i\in I}\langle \operatorname{grad} f_i\big(p^{k_s}\big), v^{k_s}\rangle + (1/2)\big\|v^{k_s}\big\|^2 < 0,$$

where the last inequality is a consequence from Definition 4.1 together with Lemma 4.2. Hence, letting s go to $+\infty$ in the last inequalities, using (6) and Lemma 5.1, we obtain

$$\max_{i\in I}\langle \operatorname{grad} f_i(\bar{p}), v(\bar{p})\rangle \le \max_{i\in I}\langle \operatorname{grad} f_i(\bar{p}), v(\bar{p})\rangle + (1/2)\big\|v(\bar{p})\big\|^2 \le 0. \tag{15}$$

Take $r\in\mathbb{N}$. Since $\{t_{k_s}\}$ converges to $\bar{t}=0$, we conclude that if s is large enough,

$$t_{k_s} < 2^{-r}.$$

From (7) this means that the Armijo condition (12) is not satisfied for $t=2^{-r}$, i.e.,

$$F\big(\exp_{p^k}\big(2^{-j}v^{k_s}\big)\big) \not\le F\big(p^{k_s}\big) + \beta 2^{-r}\operatorname{grad} F\big(p^{k_s}\big)v^{k_s},$$

which means that there exists at least one $i_0\in I$ such that

$$f_{i_0}\big(\exp_{p^{k_s}}\big(2^{-r}v^{k_s}\big)\big) > f_{i_0}\big(p^{k_s}\big) + \beta 2^{-r}\langle \operatorname{grad} f_{i_0}\big(p^{k_s}\big), v^{k_s}\rangle.$$

Letting s go to $+\infty$ in the above inequality, taking into account that grad F and exp are continuous and using Lemma 5.1, we obtain

$$f_{i_0}\big(\exp_{\bar{p}}\big(2^{-r}v(\bar{p})\big)\big) \ge f_{i_0}(\bar{p}) + \beta 2^{-r}\langle \operatorname{grad} f_{i_0}(\bar{p}), v(\bar{p})\rangle.$$

The last inequality is equivalent to

$$\frac{f_{i_0}(\exp_{\bar{p}}(2^{-r}v(\bar{p}))) - f_{i_0}(\bar{p})}{2^{-r}} \ge \beta\langle \operatorname{grad} f_{i_0}(\bar{p}), v(\bar{p})\rangle,$$

which letting r go to $+\infty$ and assuming that $0 < \beta < 1$, yields $\langle \operatorname{grad} f_{i_0}(\bar{p}), v(\bar{p})\rangle \geq 0$. Hence,

$$\max_{i\in I}\big\langle \operatorname{grad} f_i(\bar{p}), v(\bar{p})\big\rangle \geq 0.$$

Combining the last inequality with (15), we have

$$\max_{i\in I}\big\langle \operatorname{grad} f_i(\bar{p}), v(\bar{p})\big\rangle + (1/2)\big\|v(\bar{p})\big\|^2 = 0.$$

Therefore, again from Definition 4.1 and Lemma 4.2 it follows that $\bar{p}$ is a critical Pareto point and the proof is concluded. □

Remark 5.1 If the sequence $\{p^k\}$ begins in a bounded level set, for example, if

$$L_F\big(F(p_0)\big) := \big\{p \in M : F(p) \preceq F(p_0)\big\},$$

is a bounded set, then, since F is a continuous function, Hopf–Rinow theorem assures that $L_F(F(p_0))$ is a compact set. So, item (i) of Theorem 5.1 implies that $\{p^k\} \subset L_F(F(p_0))$, and consequently $\{p^k\}$ is bounded. In particular, $\{p^k\}$ has at least one accumulation point. Therefore, Theorem 5.1 extends to vector optimization the results of Theorem 5.1 of [33]. See also the Remark 4.5 of [38].

5.2 Full Convergence

In this section, under the quasiconvexity assumption on F and nonnegative curvature for M, full convergence of the steepest descent method is obtained.

Definition 5.1 Let $H : M \to \mathbb{R}^m$ be a vectorial function.

(i) H is called convex on M iff for every $p, q \in M$ and every geodesic segment $\gamma : [0, 1] \to M$ joining p to q (i.e., $\gamma(0) = p$ and $\gamma(1) = q$), the following holds:

$$H\big(\gamma(t)\big) \preceq (1-t)H(p) + tH(q), \quad t \in [0, 1].$$

(ii) H is called quasiconvex on M iff for every $p, q \in M$ and every geodesic segment $\gamma : [0, 1] \to M$ joining p to q, the following holds:

$$H\big(\gamma(t)\big) \preceq \max\big\{H(p), H(q)\big\}, \quad t \in [0, 1],$$

where the maximum is considered coordinate by coordinate.

Remark 5.2 The first definition above is a natural extension of the definition of convexity, while the second is an extension of a characterization of the definition of quasi-convexity, of the Euclidean space to the Riemannian context. See Definition 6.2 and Corollary 6.6 of [39], pages 29 and 31, respectively. Thus, when $m = 1$ these definitions merge to the scalar convexity and quasiconvexity defined in [31], respectively. Moreover, it is immediate from the above definitions that if H is convex then it is quasiconvex. In the case of H being differentiable, convexity of H implies that

for every $p, q \in M$ and every geodesic segment $\gamma : [0,1] \to M$ such that $\gamma(0) = p$ and $\gamma(1) = q$,

$$\operatorname{grad} H(p)\gamma'(0) \preceq H(q) - H(p).$$

Proposition 5.1 *Let $H : M \to \mathbb{R}^m$ be a differentiable quasi-convex function. Then, for every $p, q \in M$ and every geodesic segment $\gamma : [0,1] \to M$ joining p to q, it holds*

$$H(q) \preceq H(p) \quad \Rightarrow \quad \operatorname{grad} H(p)\gamma'(0) \preceq 0.$$

Proof Take $p, q \in M$ such that $H(q) \preceq H(p)$ and a geodesic segment $\gamma : [0,1] \to M$ such that $\gamma(0) = p$ and $\gamma(1) = q$. Since H is quasiconvex, we have

$$H\big(\gamma(t)\big) \preceq H(p), \quad t \in [0,1].$$

Using the last inequality, the result is an immediate consequence from the differentiability of H. □

We know that criticality is a necessary, but not sufficient, condition for optimality. However, under convexity of the vectorial function F, we will prove that criticality is equivalent to the weak optimality.

Definition 5.2 A point $p^* \in M$ is a weak optimal Pareto point of F iff there is no $p \in M$ with $F(p) \prec F(p^*)$.

Proposition 5.2 *Let $H : M \to \mathbb{R}^m$ be a continuously differentiable convex function. Then $p \in M$ is a critical Pareto point of H, i.e.,*

$$\operatorname{Im}\big(\operatorname{grad} H(p)\big) \cap \big(-\mathbb{R}^m_{++}\big) = \emptyset,$$

iff p is a weak optimal Pareto point of H.

Proof Let us suppose that p is a critical Pareto point of H. Assume by contradiction that p be not a weak optimal Pareto point of H. Since p is not a weak optimal Pareto point, there exists $\tilde{p} \in M$ such that

$$H(\tilde{p}) \prec H(p). \tag{16}$$

Let $\gamma : [0,1] \to M$ be a geodesic segment joining p to $\tilde{p}$ (i.e., $\gamma(0) = p$ and $\gamma(1) = \tilde{p}$). As H is differentiable and convex, the last part of Remark 5.2 and (16) imply that

$$\operatorname{grad} H(p)\gamma'(0) \preceq H(\tilde{p}) - H(p) \prec 0.$$

But this contradicts the fact of p being a critical Pareto point of H, and so the first part is concluded.

Now, let us suppose that p be a weak optimal Pareto point of H. Assume by contradiction that p be not critical Pareto point of H. Since p be not critical Pareto

point, then $\operatorname{Im}(\operatorname{grad} H(p)) \cap (-\mathbb{R}^m_{++}) \neq \emptyset$, that is, there exists $v \in T_pM$ a descent direction for F at p. Hence, from the differentiability of H, we have

$$\lim_{t\to 0^+} \frac{H(\exp_p(tv)) - H(p)}{t} = \operatorname{grad} H(p)v \prec 0,$$

which implies that there exists $\delta > 0$ such that

$$H\big(\exp_p(tv)\big) \prec H(p) + t \operatorname{grad} H(p)v, \quad t \in (0, \delta).$$

Since v is a descent direction for F at p and $t \in (0, \delta)$ we have $t \operatorname{grad} H(p)v \prec 0$. So, the last vector inequality yields

$$H\big(\exp_p(tv)\big) \prec H(p), \quad t \in (0, \delta),$$

contradicting the fact of p to be a weak optimal Pareto point of H, which concludes the proof. □

Definition 5.3 A sequence $\{q^k\} \subset M$ is quasi-Fejér convergent to a nonempty set U iff, for all $q \in U$, there exists a sequence $\{\epsilon_k\} \subset \mathbb{R}_+$ such that

$$\sum_{k=0}^{+\infty} \epsilon_k < +\infty, \quad d^2\big(q^{k+1}, q\big) \leq d^2\big(q^k, q\big) + \epsilon_k, \quad k = 0, 1, \ldots.$$

In next lemma, we recall the theorem known as quasi-Fejér convergence.

Lemma 5.2 *Let $U \subset M$ be a nonempty set and $\{q^k\} \subset M$ a quasi-Fejér convergent sequence. Then $\{q^k\}$ is bounded. Moreover, if an accumulation point $\bar{q}$ of $\{q^k\}$ belongs to U, then the whole sequence $\{q^k\}$ converges to $\bar{q}$ as k goes to $+\infty$.*

Proof Analogous to the proof of Theorem 1 in Burachik et al. [1], by replacing the Euclidean distance by the Riemannian distance d. □

Consider the following set:

$$U := \big\{p \in M : F(p) \preceq F\big(p^k\big),\ k = 0, 1, \ldots\big\}. \tag{17}$$

In general, the above set may be an empty set. To guarantee that U is nonempty, an additional assumption on the sequence $\{p^k\}$ is needed. In the next remark, we give such a condition.

Remark 5.3 If the sequence $\{p^k\}$ has an accumulation point, then U is nonempty. Indeed, let $\bar{p}$ be an accumulation point of the sequence $\{p^k\}$. Then there exists a subsequence $\{p^{k_j}\}$ of $\{p^k\}$ which converges to $\bar{p}$. Since F is continuous $\{F(p^k)\}$ has $F(\bar{p})$ as an accumulation point. Hence, using $\{F(p^k)\}$ as a decreasing sequence (see item (i) of Theorem 5.1) the usual arguments easily show that the whole sequence $\{F(p^k)\}$ converges to $F(\bar{p})$ and the following relation holds:

$$F(\bar{p}) \preceq F\big(p^k\big), \quad k = 0, 1, \ldots,$$

which implies that $\bar{p} \in U$, i.e., $U \neq \emptyset$.

In next lemma we present the main result of this section. It is fundamental to the proof of the global convergence result of the sequence $\{p^k\}$.

Lemma 5.3 *Suppose that F is quasiconvex, M has nonnegative curvature and U, defined in (17), is nonempty. Then, for all $\tilde{p} \in U$, the following inequality there holds*:

$$d^2\big(p^{k+1}, \tilde{p}\big) \leq d^2\big(p^k, \tilde{p}\big) + t_k^2 \big\|v^k\big\|^2.$$

Proof Consider the geodesic hinge $(\gamma_1, \gamma_2, \alpha)$, where γ_1 is a normalized minimal geodesic segment joining p^k to $\tilde{p}$; γ_2 is the geodesic segment joining p^k to p^{k+1} such that $\gamma_2'(0) = t_k v^k$ and $\alpha = \angle(\gamma_1'(0), v^k)$. By the law of cosines (Theorem 2.1), we have

$$d^2\big(p^{k+1}, \tilde{p}\big) \leq d^2\big(p^k, \tilde{p}\big) + t_k^2 \big\|v^k\big\|^2 - 2d\big(p^k, p\big) t_k \big\|v^k\big\| \cos\alpha, \quad k = 0, 1, \ldots.$$

Thus, taking into account that $\cos(\pi - \alpha) = -\cos\alpha$ and $\langle -v^k, \gamma_1'(0)\rangle = \|v^k\| \cos(\pi - \alpha)$, the above vector inequality becomes

$$d^2\big(p^{k+1}, \tilde{p}\big) \leq d^2\big(p^k, \tilde{p}\big) + t_k^2 \big\|v^k\big\|^2 + 2d\big(p^k, \tilde{p}\big) t_k \big\langle -v^k, \gamma_1'(0)\big\rangle, \quad k = 0, 1, \ldots.$$

On the other hand, from (6), Definition 4.1 and Lemma 4.1, there exists $\alpha_i^k \geq 0$, with $i \in I_k := I(p^k, v^k)$, such that

$$v^k = -\sum_{i \in I_k} \alpha_i \operatorname{grad} f_i\big(p^k\big), \quad \sum_{i \in I_k} \alpha_i^k = 1, \quad k = 0, 1, \ldots.$$

Hence, the last vector inequality yields

$$d^2\big(p^{k+1}, \tilde{p}\big) \leq d^2\big(p^k, \tilde{p}\big) + t_k^2 \big\|v^k\big\|^2 + 2d\big(p^k, \tilde{p}\big) t_k \sum_{i \in I_k} \alpha_i^k \big\langle \operatorname{grad} f_i\big(p^k\big), \gamma_1'(0)\big\rangle,$$

$$k = 0, 1, \ldots. \tag{18}$$

Since F is quasiconvex and $\tilde{p} \in U$, from Proposition 5.1 with $H = F$, $p = p^k$, $q = \tilde{p}$ and $\gamma = \gamma_1$, we have

$$\operatorname{grad} F\big(p^k\big)\gamma_1'(0) \preceq 0, \quad k = 0, 1, \ldots,$$

or equivalently,

$$\big\langle \operatorname{grad} f_i\big(p^k\big), \gamma_1'(0)\big\rangle \leq 0, \quad i = 1, \ldots, m, \ k = 0, 1, \ldots. \tag{19}$$

Therefore, by combining (18) with (19), the lemma proceeds. □

Proposition 5.3 *If F is quasiconvex, M has nonnegative curvature and U, defined in (17), is a nonempty set, then the sequence $\{p_k\}$ is quasi-Fejér convergent to U.*

Proof To simplify the notation, the scalar function $\varphi : \mathbb{R}^m \to \mathbb{R}$ is defined as follows:

$$\varphi(y) = \max_{i \in I} \langle y, e_i \rangle, \quad I = \{1, \ldots, m\},$$

where $\{e_i\} \subset \mathbb{R}^m$ is the canonical base of the space $\mathbb{R}^m$. It is easy to see that the following properties of the function φ hold:

$$\varphi(x + y) \leq \varphi(x) + \varphi(y), \quad \varphi(tx) = t\varphi(x), \quad x, y \in \mathbb{R}^m, \ t \geq 0. \tag{20}$$

$$x \preceq y \quad \Rightarrow \quad \varphi(x) \leq \varphi(y), \quad x, y \in \mathbb{R}^m. \tag{21}$$

From the definition of t_k in (7) and p^{k+1} in (8), we have

$$F\big(p^{k+1}\big) \preceq F\big(p^k\big) + \beta t_k \operatorname{grad} F\big(p^k\big) v^k, \quad k = 0, 1 \ldots.$$

Hence, using (20), (21), and the last inequality, we obtain

$$\varphi\big(F\big(p^{k+1}\big)\big) \leq \varphi\big(F\big(p^k\big)\big) + \beta t_k \varphi\big(\operatorname{grad} F\big(p^k\big) v^k\big), \quad k = 0, 1 \ldots. \tag{22}$$

On the other hand, combining the definition of v^k in (6), Definition 4.1, Lemma 4.2, and the definition of φ, we conclude that

$$\varphi\big(\operatorname{grad} F\big(p^k\big) v^k\big) + (1/2)\big\|v^k\big\|^2 < 0, \quad k = 0, 1 \ldots,$$

which together with (22) implies that

$$\varphi\big(F\big(p^{k+1}\big)\big) < \varphi\big(F\big(p^k\big)\big) - (\beta t_k/2)\big\|v^k\big\|^2, \quad k = 0, 1 \ldots.$$

But this tells us that

$$t_k\big\|v^k\big\|^2 < 2\big[\varphi\big(F\big(p^k\big)\big) - \varphi\big(F\big(p^{k+1}\big)\big)\big]/\beta, \quad k = 0, 1 \ldots.$$

As $t_k \in]0, 1]$, it follows that

$$t_k^2\big\|v^k\big\|^2 < 2\big[\varphi\big(F\big(p^k\big)\big) - \varphi\big(F\big(p^{k+1}\big)\big)\big]/\beta, \quad k = 0, 1 \ldots.$$

Thus, the latter inequality easily implies that

$$\sum_{k=0}^{n} t_k^2\big\|v^k\big\|^2 < 2\big[\varphi\big(F\big(p^0\big)\big) - \varphi\big(F\big(p^{n+1}\big)\big)\big]/\beta, \quad n > 0.$$

Take $\bar{p} \in U$. Then $F(\bar{p}) \preceq F(p^{n+1})$. So, from (21) $\varphi(F(\bar{p})) \leq \varphi(F(p^{n+1}))$ and the last inequality yields

$$\sum_{k=0}^{n} t_k^2\big\|v^k\big\|^2 < 2\big(\varphi\big(F\big(p^0\big)\big) - \varphi\big(F(\bar{p})\big)\big)/\beta,$$

which implies that $\{t_k^2\|v^k\|^2\}$ is a summable sequence. Therefore, Lemma 5.3 combined with Definition 5.3 achieves the desired result. □

Theorem 5.2 *Suppose that F is quasiconvex, M has nonnegative curvature and U, as defined in (17), is a nonempty set. Then, the sequence $\{p_k\}$ converges to a critical Pareto point of F.*

Proof From Proposition 5.3, $\{p^k\}$ is Fejér convergent to U. Thus, Lemma 5.2 guarantees that $\{p^k\}$ is bounded and from the Hopf–Rinow theorem, there exists $\{p^{k_s}\}$, subsequence of $\{p^k\}$, which converges to $\bar{p} \in M$ as s goes to $+\infty$. Since F is continuous and $\{F(p^k)\}$ is a decreasing sequence (see item (i) of Theorem 5.1), we conclude that $F(p^k)$ converges to $F(\bar{p})$ as k goes to $+\infty$, which implies that

$$F(\bar{p}) \preceq F\big(p^k\big), \quad k = 0, 1, \ldots,$$

i.e., $\bar{p} \in U$. Hence, from Lemma 5.2, we conclude that the whole sequence $\{p^k\}$ converges to $\bar{p}$ as k goes to $+\infty$, and the conclusion of the proof is a consequence of item (ii) of Theorem 5.1. □

Corollary 5.1 *If F is convex, M has nonnegative curvature and U, as defined in (17), is a nonempty set, then the sequence $\{p_k\}$ converges to a weak optimal Pareto point of F.*

Proof Since F is convex, and in particular quasiconvex (see Remark 5.2), the corollary is a consequence of the previous theorem and Proposition 5.2. □

6 Examples

In this section, we present some examples of complete Riemannian manifolds with explicit geodesic curves and the steepest descent iteration of the sequence generated by Method 4.1. We recall that the function $F : M \to \mathbb{R}^m$, $F(p) := (f_1(p), \ldots, f_m(p))$, is differentiable. If (M, G) is a Riemannian manifold then the Riemannian gradient of f_i is given by $\operatorname{grad} f_i(p) = G(p)^{-1} f_i'(p)$, $i \in I := \{1, \ldots, n\}$. Hence, if $v(p)$ is the steepest descent direction for F at p (see Definition 4.1) then, from Lemma 4.1, there exist constants $\alpha_i \geq 0$, $i \in I(p, v)$, such that

$$v = -\sum_{i \in I(p,v)} \alpha_i G(p)^{-1} f_i'(p), \qquad \sum_{i \in I(p,v)} \alpha_i = 1, \tag{23}$$

where $I(p, v) := \{i \in I : \langle G(p)^{-1} f_i'(p), v\rangle = \max_{i \in I} \langle G(p)^{-1} f_i'(p), v\rangle\}$.

6.1 A Steepest Descent Method for $\mathbb{R}^n_{++}$

Let M be the positive octant, $\mathbb{R}^n_{++}$, endowed with the Riemannian metric

$$M \ni v \mapsto G(p) = P^{-2} := \operatorname{diag}\big(p_1^{-2}, \ldots, p_n^{-2}\big),$$

(metric induced by the Hessian of the logarithmic barrier). Since (M, G) is isometric to the Euclidean space endowed with the usual metric (see, Da Cruz Neto et al. [12]),

it follows that M has constant curvature equal to zero. Besides, it is easy to see that the unique geodesic $p = p(t)$ such that $p(0) = p^0 = (p_1^0, \ldots, p_n^0)$ and that $p'(0) = v^0 = (v_1^0, \ldots, v_n^0)$ is given by $p(t) = (p_1(t), \ldots, p_n(t))$, where

$$p_j(t) = p_j^0 \mathrm{e}^{(v_j^0/p_j^0)t}, \quad j = 1, \ldots, n. \tag{24}$$

So, we conclude that (M, G) is also complete. In this case, from (24) and (23), there exists $\alpha_i^k \geq 0$ such that the steepest descent iteration of the sequence generated by Method 4.1 is given by

$$p_j^{k+1} = p_j^k \mathrm{e}^{(v_j^k/p_j^k)t_k}, \quad v_j^k = -\sum_{i \in I(p^k, v^k)} \alpha_i^k \big(p_j^k\big)^2 \frac{\partial f_i}{\partial p_j}\big(p^k\big),$$

$$\sum_{i \in I(p^k, v^k)} \alpha_i^k = 1, \quad j = 1, \ldots, n.$$

6.2 A Steepest Descent Method for the Hypercube

Let M be the hypercube $]0, 1[\times \cdots \times]0, 1[$ endowed with the Riemannian metric

$$M \ni v \mapsto G(p) = P^{-2}(I - P)^{-2} := \operatorname{diag}\big((p_1)^2(1 - p_1)^2, \ldots, (p_n)^2(1 - p_n)^2\big),$$

(metric induced by the Hessian of the barrier $b(p) = \sum_{i=1}^n (2p_i - 1)(\ln p_i - \ln(1 - p_i))$. The Riemannian manifold (M, G) is complete and the unique geodesic $p = p(t)$, satisfying $p(0) = p^0 = (p_1^0, \ldots, p_n^0)$ and $p'(0) = v^0 = (v_1^0, \ldots, v_n^0)$, is given by $p(t) = (p_1(t), \ldots, p_n(t))$,

$$p_j(t) = (1/2)\left[1 + \tanh\left((1/2)\frac{v_j}{p_j(1 - p_j)}t + (1/2)\ln\left(\frac{p_j}{1 - p_j}\right)\right)\right], \quad j = 1, \ldots, n, \tag{25}$$

where $\tanh(z) := (e^z - e^{-z})/(e^z + e^{-z})$. Moreover, (M, G) has constant curvature equal to zero, see Theorems 3.1 and 3.2 of [40]. In this case, from (25) and (23), there exists $\alpha_i^k \geq 0$ such that the steepest descent iteration of the sequence generated by Method 4.1 is given by

$$p_j^{k+1} = (1/2)\left[1 + \tanh\left((1/2)\frac{v_j^k}{p_j^k(1 - p_j^k)}t_k + (1/2)\ln\left(\frac{p_j^k}{1 - p_j^k}\right)\right)\right],$$

$$j = 1, \ldots, n,$$

with,

$$v_j^k = -\sum_{i \in I(p^k, v^k)} \alpha_i^k \big(p_j^k\big)^2 \big(1 - p_j^k\big)^2 \frac{\partial f_i}{\partial p_j}\big(p^k\big), \quad \sum_{i \in I(p^k, v^k)} \alpha_i^k = 1, \quad j = 1, \ldots, n.$$

6.3 Steepest Descent Method for the Cone of Positive Semidefinite Matrices

Let $\mathbb{S}^n$ be the set of the symmetric matrices $n \times n$, $\mathbb{S}^n_+$ the cone of the symmetric positive semidefinite matrices and $\mathbb{S}^n_{++}$ the cone of the symmetric positive definite matrices. Following Rothaus [41], let $M = \mathbb{S}^n_{++}$ be endowed with the Riemannian metric induced by the Euclidean Hessian of $\Psi(X) = -\ln\det X$, i.e., $G(X) := \Psi''(X)$. In this case, the unique geodesic segment connecting any $X, Y \in M$ is given by

$$X(t) = X^{1/2}\big(X^{-1/2}YX^{-1/2}\big)^t X^{1/2}, \quad t \in [0,1];$$

see [42]. More precisely, M is a Hadamard manifold (with curvature not identically zero); see, for example, [43], Theorem 1.2, p. 325. In particular, the unique geodesic $X = X(t)$ such that $X(0) = X$ and $X'(0) = V$ is given by

$$X(t) = X^{1/2}\mathrm{e}^{tX^{-1/2}VX^{-1/2}}X^{1/2}. \tag{26}$$

Thus, from (26) and (23), there exists $\alpha_j^k \geq 0$ such that the steepest descent iteration of the sequence generated by Method 4.1 is given by

$$X^{k+1} = \big(X^k\big)^{1/2}\mathrm{e}^{t_k(X^k)^{-1/2}V^k(X^k)^{-1/2}}\big(X^k\big)^{1/2},$$

with

$$V^k = -\sum_{i\in I(X^k,V^k)} \alpha_i^k X^k f_i'\big(X^k\big)X^k, \qquad \sum_{i\in I(X^k,V^k)} \alpha_i^k = 1.$$

Remark 6.1 Under the assumption of convexity on the vector function F, if (M, G) is the Riemannian manifold in the first or in the second example, then Corollary 5.1 assures the full convergence of the sequence generated by Method 4.1. This fact does not necessarily happen if (M, G) is the Riemannian manifold in the last example, since in this case (M, G) has nonpositive curvature, i.e., $K \leq 0$. However, Theorem 5.1 assures at least partial convergence.

7 Final Remarks

We have extended the steepest descent method with Armijo's rule for multicriteria optimization to the Riemannian context. Full convergence is obtained under the assumptions of quasiconvexity of the multicriteria function and nonnegative curvature of the Riemannian manifold. A subject in open is to obtain the same result without a restrictive assumption on the curvature of the manifold. Following the same structure of this paper, as a future work we propose the application of the proximal method (see Bonnel et al. [44]) and Newton method (see Fliege et al. [45]) to multiobjective optimization, within the Riemannian context.

Acknowledgements G.C. Bento was supported in part by CNPq Grant 473756/2009-9 and PROCAD/NF. O.P. Ferreira was supported in part by CNPq Grant 302618/2005-8, PRONEX–Optimization (FAPERJ/CNPq) and FUNAPE/UFG. P.R. Oliveira was supported in part by CNPq.

References

1. Burachik, R., Graña Drummond, L.M., Iusem, A.N., Svaiter, B.F.: Full convergence of the steepest descent method with inexact line searches. Optimization **32**(2), 137–146 (1995)
2. Fliege, J., Svaiter, B.F.: Steepest descent methods for multicriteria optimization. Math. Methods Oper. Res. **51**(3), 479–494 (2000)
3. Kiwiel, K.C., Murty, K.: Convergence of the steepest descent method for minimizing quasiconvex functions. J. Optim. Theory Appl. **89**(1), 221–226 (1996)
4. Graña Drummond, L.M., Svaiter, B.F.: A steepest descent method for vector optimization. J. Comput. Appl. Math. **175**(2), 395–414 (2005)
5. Graña Drummond, L.M., Iusem, A.N.: A projected gradient method for vector optimization problems. Comput. Optim. Appl. **28**(1), 5–29 (2004)
6. Rapcsák, T.: Smooth Nonlinear Optimization in R^n. Kluwer Academic, Dordrecht (1997)
7. Alvarez, F., Bolte, J., Munier, J.: A unifying local convergence result for Newton's method in Riemannian manifolds. Found. Comput. Math. **8**, 197–226 (2008)
8. Attouch, H., Bolte, J., Redont, P., Teboulle, M.: Singular Riemannian barrier methods and gradient-projection dynamical systems for constrained optimization. Optimization **53**(5–6), 435–454 (2004)
9. Azagra, D., Ferrera, J., López-Mesas, M.: Nonsmooth analysis and Hamilton–Jacobi equations on Riemannian manifolds. J. Funct. Anal. **220**, 304–361 (2005)
10. Barani, A., Pouryayevali, M.R.: Invariant monotone vector fields on Riemannian manifolds. Nonlinear Anal. **70**(5), 1850–1861 (2009)
11. Ferreira, O.P., Svaiter, B.F.: Kantorovich's theorem on Newton's method in Riemannian manifolds. J. Complex. **18**, 304–329 (2002)
12. Da Cruz Neto, J.X., Ferreira, O.P., Lucâmbio Pérez, L.R., Németh, S.Z.: Convex-and monotone-transformable mathematical programming problems and a proximal-like point method. J. Glob. Optim. **35**, 53–69 (2006)
13. Ferreira, O.P., Oliveira, P.R.: Subgradient algorithm on Riemannian manifolds. J. Optim. Theory Appl. **97**, 93–104 (1998)
14. Ferreira, O.P.: Proximal subgradient and a characterization of Lipschitz function on Riemannian manifolds. J. Math. Anal. Appl. **313**, 587–597 (2006)
15. Ledyaev, Yu.S., Zhu, Q.J.: Nonsmooth analysis on smooth manifolds. Trans. Am. Math. Soc. **359**, 3687–3732 (2007)
16. Li, C., López, G., Martín-Márquez, V.: Monotone vector fields and the proximal point algorithm on Hadamard manifolds. J. Lond. Math. Soc. **79**(2), 663–683 (2009)
17. Li, S.L., Li, C., Liou, Y.C., Yao, J.C.: Existence of solutions for variational inequalities on Riemannian manifolds. Nonlinear Anal. **71**, 5695–5706 (2009)
18. Papa Quiroz, E.A., Quispe, E.M., Oliveira, P.R.: Steepest descent method with a generalized Armijo search for quasiconvex functions on Riemannian manifolds. J. Math. Anal. Appl. **341**(1), 467–477 (2008)
19. Papa Quiroz, E.A., Oliveira, P.R.: Proximal point methods for quasiconvex and convex functions with Bregman distances on Hadamard manifolds. J. Convex Anal. **16**(1), 49–69 (2009)
20. Wang, J.H., Li, C.: Kantorovich's theorems of Newton's method for mappings and optimization problems on Lie groups. IMA J. Numer. Anal. **31**, 322–347 (2011)
21. Wang, J.H.: Convergence of Newton's method for sections on Riemannian manifolds. J. Optim. Theory Appl. **148**(1), 125–145 (2011)
22. Wang, J.H., Li, C.: Newton's method on Lie groups with applications to optimization. IMA J. Numer. Anal. **31**, 322–347 (2011)
23. Wang, J.H., Lopez, G., Martin-Marquez, V., Li, C.: Monotone and accretive vector fields on Riemannian manifolds. J. Optim. Theory Appl. **146**, 691–708 (2010)
24. Wang, J.H., Dedieu, J.P.: Newton's method on Lie groups: Smale's point estimate theory under the γ-condition. J. Complex. **25**, 128–151 (2009)
25. Wang, J.H., Huang, S.C., Li, C.: Extended Newton's algorithm for mappings on Riemannian manifolds with values in a cone. Taiwan. J. Math. **13**, 633–656 (2009)
26. Li, C., Wang, J.H.: Newton's method for sections on Riemannian manifolds: generalized covariant α-theory. J. Complex. **24**, 423–451 (2008)
27. Wang, J.H., Li, C.: Uniqueness of the singular points of vector fields on Riemannian manifolds under the γ-condition. J. Complex. **22**(4), 533–548 (2006)
28. Li, C., Wang, J.H.: Newton's method on Riemannian manifolds: Smale's point estimate theory under the γ-condition. IMA J. Numer. Anal. **26**(2), 228–251 (2006)

29. Bento, G.C., Melo, J.G.: A subgradient method for convex feasibility on Riemannian manifolds. J. Optim. Theory Appl. (2011). doi:10.1007/s10957-011-9921-4
30. Bento, G.C., Ferreira, O.P., Oliveira, P.R.: Local convergence of the proximal point method for a special class of nonconvex functions on Hadamard manifolds. Nonlinear Anal. **73**, 564–572 (2010)
31. Udriste, C.: Convex functions and optimization algorithms on Riemannian manifolds. In: Mathematics and Its Applications, vol. 297. Kluwer Academic, Norwell (1994)
32. Smith, S.T.: Optimization techniques on Riemannian Manifolds. Fields Institute Communications, vol. 3, pp. 113–146. Am. Math. Soc., Providence (1994)
33. da Cruz Neto, J.X., de Lima, L.L., Oliveira, P.R.: Geodesic algorithms in Riemannian geometry. Balk. J. Geom. Appl. **3**(2), 89–100 (1998)
34. Do Carmo, M.P.: Riemannian Geometry. Birkhauser, Boston (1992)
35. Sakai, T.: Riemannian Geometry. Translations of Mathematical Monographs, vol. 149. Am. Math. Soc., Providence (1996)
36. Hiriart-Urruty, J.-B., Lemaréchal, C.: Convex Analysis and Minimization Algorithms I and II. Springer, Berlin (1993)
37. Munier, J.: Steepest descent method on a Riemannian manifold: the convex case. Balk. J. Geom. Appl. **12**(2), 98–106 (2007)
38. Gabay, D.: Minimizing a differentiable function over a differentiable manifold. J. Optim. Theory Appl. **37**, 177–219 (1982)
39. Luc, D.T.: Theory of Vector Optimization. Lecture Notes in Economics and Mathematical Systems, vol. 319. Springer, Berlin (1989)
40. Papa Quiroz, E.A., Oliveira, P.R.: New self-concordant barrier for the hypercube. J. Optim. Theory Appl. **135**, 475–490 (2007)
41. Rothaus, O.S.: Domains of positivity. Abh. Math. Semin. Univ. Hamb. **24**, 189–235 (1960)
42. 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)
43. Lang, S.: Fundamentals of Differential Geometry. Springer, New York (1998)
44. Bonnel, H., Iusem, A.N., Svaiter, B.F.: Proximal methods in vector optimization. SIAM J. Optim. **15**(4), 953–970 (2005)
45. Fliege, J., Graña Drummond, L.M., Svaiter, B.F.: Newton's method for multiobjective optimization. SIAM J. Optim. **20**(2), 602–626 (2009)