An Inexact Steepest Descent Method for Multicriteria Optimization on Riemannian Manifolds

G. C. Bento, J. X. da Cruz Neto & P. S. M. Santos

Journal of Optimization Theory and Applications

ISSN 0022-3239

J Optim Theory Appl
DOI 10.1007/s10957-013-0305-9

J Optim Theory Appl
DOI 10.1007/s10957-013-0305-9

An Inexact Steepest Descent Method for Multicriteria Optimization on Riemannian Manifolds

G.C. Bento · J.X. da Cruz Neto · P.S.M. Santos

Received: 29 April 2012 / Accepted: 14 March 2013

Abstract In this paper, we present an inexact version of the steepest descent method with Armijo's rule for multicriteria optimization in the Riemannian context given in Bento et al. (J. Optim. Theory Appl., 154: 88–107, 2012). 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 that the multicriteria function is quasi-convex and the Riemannian manifold has nonnegative curvature, we show full convergence of any sequence generated by the method to a Pareto critical point.

Keywords Steepest descent · Pareto optimality · Multicriteria optimization · Quasi-Fejér convergence · Quasi-convexity · Riemannian manifolds

1 Introduction

In many applications, such as engineering, statistics, and design problems, several objective functions have to be minimized simultaneously; see, for instance, [1] and references therein. This characterizes the so-called multiobjective optimization problem. A well-known strategy for solving multiobjective optimization problems is the

Communicated by Alfredo Iusem.

G.C. Bento
Universidade Federal de Goiás, Goiania, Brazil
e-mail: glaydston@mat.ufg.br

J.X. da Cruz Neto · P.S.M. Santos (✉)
Universidade Federal Piauí, Teresina, Brazil
e-mail: paulosms@gmail.com

J.X. da Cruz Neto
e-mail: jxavier@ufpi.edu

Published online: 19 April 2013

scalarization approach. The most widely used scalarization technique is the weighting method; see [2]. The choice of the vector of "weights" is of capital importance because, even for very well behaved problems, this choice can lead to unbounded scalar minimization problems; see, for instance, [3, 4], where a disadvantage of this approach is characterized. For the sake of simplicity, we refer to classic scalarization methods, even if they are not the most efficient ones; but the approach can be extended to every other one.

We recall that a classical method for solving scalar minimization problems is the so-called gradient method. This method was proposed in the multiobjective context by Fliege and Svaiter [5]. Since then, it has been considered in more general settings, for instance, for vector optimization problems; see Graña-Drummond and Svaiter [3], and for constrained vector optimization, see Graña-Drummond and Iusem [4] and Fukuda and Graña Drummond [6, 7]. The convergence result presented in [5] is only partial, under mild assumptions on the multicriteria function. In [3, 4, 6, 7] the authors presented a global convergence result under the assumption of convexity of the multiobjective function. Particularly, in [3, 7] this result was restricted to inexact search directions of the s-compatible type. For full convergence of the exact gradient method for quasi-convex multicriteria optimization, see, for instance, Bello Cruz et al. [9]. We emphasize that the quasi-convex optimization problems have been receiving special attention from many researchers due to the broad range of applications as, for instance, in economic theory [10] and location theory [11]. For extensions of other scalar optimization methods to the vectorial setting, see, for instance, [12–14] and references therein. From the Euclidean viewpoint, following the ideas of [3], in the present paper we present the global convergence of any sequence generated for the inexact gradient method to a Pareto critical point (resp. weak Pareto optimal point) of the multiobjective optimization problem in the quasi-convex case (resp. pseudo-convex case).

Extension of concepts and techniques, as well as methods from Euclidean spaces to Riemannian manifolds, is natural and, in general, nontrivial; see, for instance, Udriste [15] and Rapcsák [16]. In the last few years such extensions have been the subject of many research papers with practical and theoretical purposes; see, for example, [17–20] and references therein. 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. Moreover, nonconvex problems in the classical context may become convex through the introduction of an appropriate Riemannian metric (see, for example, [21, 22]).

In the present paper, we propose an inexact version of the method presented in [8] by admitting relative errors on the search directions; more precisely, an approximation of the exact search direction is computed at each iteration, as considered by Fliege and Svaiter [5] in the Euclidean context (see also [3, 7]). Following the ideas of [5], we extend the partial convergence result presented in [8] for the case inexact. We point out that this result is different from its counterpart in [8] since here we deal with inexact search directions. In the sequel, we show that any sequence generated by this new method converges to a Pareto critical point when the objective function is quasi-convex and the Riemannian manifold has nonnegative curvature.

Moreover, we extend the definition of pseudo-convex function for the Riemannian context, and we observed that under this hypothesis, any sequence generated by this new method converges to a weak Pareto optimal point. We emphasize that such results extend the presented in [8], and, as noted in the second paragraph of this introduction, they are new even in the Euclidean context, where the existing full convergence results are obtained under the hypothesis of convexity of the multiobjective function.

The organization of our paper is as follows. In Sect. 2, some notation and results of Riemannian geometry are defined. In Sect. 3, the multicriteria problem and some basic definitions are presented. In Sect. 4, the Riemannian inexact steepest descent method is stated. 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 is quasi-convex and the Riemannian manifold has nonnegative curvature, a full convergence result is presented. Finally, in Sect. 6, examples of complete Riemannian manifold with explicit geodesic curve and the steepest descent iteration of the sequence generated by the proposed method are presented. In Sect. 7, we report two numerical experiments.

2 Preliminary Material on Riemannian Geometry

In this section, we introduce some fundamental properties and notation of Riemannian manifolds useful throughout the text as in [8, 23]. These basic facts can be found, for instance, in [24].

From now on, let M be an n-dimensional connected manifold. We denote by T_pM the n-dimensional *tangent space* of M at p, by $TM = \bigcup_{p\in M} T_pM$ the *tangent bundle* of M, and by $\mathscr{X}(M)$ the space of smooth vector fields over M. Suppose that M be endowed with a Riemannian metric $\langle\,,\,\rangle$ with the 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

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

moreover, by minimizing this functional length over the set of all such curves, we obtain a Riemannian distance $d(p,q)$ that induces the original topology on M. The metric induces a map

$$f \mapsto \operatorname{grad} f \in \mathscr{X}(M)$$

that associates to each scalar function smooth over M its gradient via the rule $\langle \operatorname{grad} f, X\rangle = df(X)$, $X \in \mathscr{X}(M)$. 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*. A geodesic $\gamma = \gamma_v(\cdot, p)$ is determined by its position p and velocity v at p. 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 equal 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 $X,Y,Z \in \mathscr{X}(M)$, 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, we will be interested mainly in Riemannian 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) *Let M be a complete Riemannian manifold with nonnegative curvature with the notation introduced above. The following inequality holds*: $l_3^2 \le l_1^2 + l_2^2 - 2l_1l_2\cos\alpha$.

Proof See, for example, [23]. □

3 The Multicriteria Problem

In this section, we present the multicriteria problem, the first order optimality condition for it, and some basic definitions that were presented in [8]. For completeness, here we also present some notation.

Let $I := \{1,\dots,m\}$, $\mathbb{R}^m_+ = \{x \in \mathbb{R}^m : x_i \ge 0, j \in I\}$, 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 does not exist any other $p \in M$ with $F(p) \preceq F(p^*)$ and $F(p) \ne F(p^*)$. We denote this unconstrained problem in the Riemannian context as

$$\min_{p\in M} F(p). \tag{1}$$

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

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

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

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

Using the above equality, the first-order optimality condition for problem (1) (see, for instance, [8]) is stated as

$$x \in M, \quad \operatorname{Im}\big(JF(x)\big) \cap \big(-\mathbb{R}^m_{++}\big) = \emptyset. \tag{2}$$

In general, (2) is necessary but not sufficient for optimality. A point of M satisfying (2) is called a *Pareto critical point*.

4 Inexact Steepest Descent Methods for Multicriteria Problems

In this section, we state the inexact steepest descent methods for solving multicriteria problems admitting relative errors in the search directions; more precisely, an approximation of the exact search direction is computed at each iteration, as considered, for example, in [3, 5, 7] in the Euclidean context.

Let $p \in M$ be a point that is not Pareto critical point. Then there exists a direction $v \in T_pM$ satisfying $JF(p)v \prec 0$. 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:

$$\min_{v \in T_pM} \Big\{\max_{i \in I} \langle \operatorname{grad} f_i(p), v\rangle + (1/2)\|v\|^2\Big\}, \qquad I := \{1, \dots, m\}. \tag{3}$$

Lemma 4.1 *The following statements hold*:

(i) *The unconstrained optimization problem in* (3) *has only one solution. Moreover, the vector v is the solution of problem* (3) *if and only if there exist $\alpha_i \geq 0$, $i \in I(p, v)$, 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 $I(p, v) := \{i \in I : \langle \operatorname{grad} f_i(p), v\rangle = \max_{i \in I} \langle \operatorname{grad} f_i(p), v\rangle\}$;

(ii) *If p is a Pareto critical point of F and v denotes the solution of problem* (3), *then $v = 0$, and the optimal value associated to v is equal to zero*;

(iii) *If $p \in M$ is not a Pareto critical point of F and v is the solution of problem* (3), *then $v \neq 0$, and*

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

In particular, v is a descent direction for F at p.

Proof The proof of items (i) and (iii) may be found in [8]. To prove item (ii), let us suppose that p is a Pareto critical point of F. Then, $\max_{1\le i\le m}\langle \operatorname{grad} f_i(p), u\rangle \ge 0$ for all $u \in T_pM$, and, hence,

$$\max_{1\le i\le m}\big\langle \operatorname{grad} f_i(p), u\big\rangle + 1/2\|u\|^2 \ge 0, \qquad u \in T_pM.$$

But this tells us that $v = 0$ and, in particular, that the optimal value of problem (3) is equal to zero. □

Remark 4.1 From item (i) of Lemma 4.1 we note that the solution of the minimization problem (3) is of the form

$$v = -JF(p)^t w, \qquad w = (\alpha_1, \dots, \alpha_m) \in \mathbb{R}^m_+, \qquad \|w\|_1 = 1 \quad \big(\text{sum norm in } \mathbb{R}^m\big),$$

with $\alpha_i = 0$ for $i \in I \setminus I(p, v)$. In other words, if $S := \{e_i \in \mathbb{R}^m : i \in I\}$ (the set of the elements of the canonical base of Euclidean space $\mathbb{R}^m$), then w is an element of the convex hull of $S(p, v)$, where

$$S(p, v) := \Big\{\bar{u} \in S : \big\langle \bar{u}, JF(p)v\big\rangle = \max_{u\in S}\big\langle u, JF(p)v\big\rangle\Big\}. \tag{4}$$

Note that the minimization problem (3) may be rewritten as follows:

$$\min_{v\in T_pM}\Big\{\max_{u\in S}\big\langle u, JF(p)v\big\rangle + (1/2)\|v\|^2\Big\} = \min_{v\in T_pM}\Big\{\max_{u\in S}\big\langle JF(p)^t u, v\big\rangle + (1/2)\|v\|^2\Big\}.$$

In view of the previous lemma and (3), we define the steepest descent direction function for F as follows.

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

$$p \in M,\ p \longmapsto v(p) := \arg\min_{v\in T_pM}\Big\{\max_{i\in I}\big\langle \operatorname{grad} f_i(p), v\big\rangle + (1/2)\|v\|^2\Big\} \in T_pM.$$

Remark 4.2 This definition was considered in the Riemannian context in [8]. When $M = \mathbb{R}^n$, we are with the steepest descent direction proposed in [5].

The optimal valued associated to $v(p)$ will be denoted by $\alpha(p)$. Note that the function

$$p \in M, \qquad p \longmapsto \max_{i\in I}\big\langle \operatorname{grad} f_i(p), v\big\rangle + (1/2)\|v\|^2 \in \mathbb{R},$$

is strongly convex with modulus $1/2$ and $0 \in \partial\ (\max_{i\in I}\langle \operatorname{grad} f_i(p),\ .\ \rangle + 1/2\|.\|^2) \times (v(p))$.

So, for all $v \in T_pM$,

$$\max_{i\in I}\big\langle \operatorname{grad} f_i(p), v\big\rangle + (1/2)\|v\|^2 - \alpha(p) \ge 1/2\big\|v - v(p)\big\|^2. \tag{5}$$

Lemma 4.2 *The steepest descent direction function for F, $p \in M$, $p \mapsto v(p) \in T_pM$, is continuous. In particular, the function $p \in M$, $p \mapsto \alpha(p) \in \mathbb{R}$, is also continuous.*

Proof See [8] for the proof of the first part. The second part is an immediate consequence of the first. □

Definition 4.2 Let $\sigma \in [0, 1[$. A vector $v \in T_pM$ is said to be a σ-approximate steepest descent direction at p for F iff

$$\max_{1 \le i \le m} \langle \operatorname{grad} f_i(p), v \rangle + 1/2\|v\|^2 \le (1-\sigma)\alpha(p).$$

Note that the exact steepest descent direction at p is a σ-approximate steepest descent direction for F with $\sigma = 0$. As an immediate consequence of Lemma 4.1 together with last definition, it is possible to prove the following:

Lemma 4.3 *Given $p \in M$,*

(i) *$v = 0$ is a σ-approximate steepest descent direction at p if and only if p is a Pareto critical point*;
(ii) *if p is not a Pareto critical point and v is a σ-approximate steepest descent direction at p, then v is a descent direction for F.*

Next lemma establishes the degree of proximity between an approximate direction v and the exact direction $v(p)$ in terms of the optimal value $\alpha(p)$.

Lemma 4.4 *Let $\sigma \in [0, 1[$. If $v \in T_pM$ is a σ-approximate steepest descent direction at p, then*

$$\left\| v - v(p) \right\|^2 \le 2\left|\alpha(p)\right|.$$

Proof The proof follows from (5) combined with Definition 4.2. □

A particular class of σ-approximate steepest descent directions for F at p is given for the directions $v \in T_pM$ that are *compatible scalarization*, i.e., such that there exists $\tilde{w} \in \operatorname{conv} S$ with

$$v = -JF(p)^t \tilde{w}. \tag{6}$$

From Remark 4.1 we observe that, for each $p \in M$, the steepest descent direction for F at p, $v(p)$, is compatible scalarization. Note that $\tilde{w}$ determines a scalar function $g(p) := \langle \tilde{w}, F(p) \rangle$ whose steepest descent direction coincides with v, which justifies the name previously attributed to the direction v; see [3] for a good discussion.

The *inexact steepest descent method with the Armijo rule* for solving the unconstrained optimization problem (1) is as follows.

Method 4.1 (Inexact 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 Pareto critical point, STOP. Otherwise,*
ITERATIVE STEP *Compute a* σ*-approximate steepest descent direction* v^k *for* F *at* p^k, *and the steplength* $t_k \in]0,1]$ *is as follows*:

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

Set

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

and GOTO STOP CRITERION.

Remark 4.3 The Method 4.1 is a natural extension of the method proposed by Fliege and Svaiter [5] in the Riemannian context. Moreover, it becomes the steepest descent method for vector optimization in Riemannian manifolds proposed in [8].

Next proposition ensures that the sequence generated by Method 4.1 is well defined.

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

Proof The proof follows from item ii of Lemma 4.3 combined with the fact that F is continuously differentiable. See [8] for more details. □

5 Convergence Analysis

In this section, following the ideas of [5], we extend the partial convergence result presented in [8] for the case inexact. In the sequel, following the ideas of [3] and assuming that F is quasi-convex and M has nonnegative curvature, we extend the full convergence result presented in [23] and [25] to multicriteria optimization, as well as the full convergence result presented in [8] for the inexact case.

If Method 4.1 terminates after a finite number of iterations, then it terminates at a Pareto critical point. From now on, we will assume that Method 4.1 generates infinite sequences $\{p^k\}$, $\{v^k\}$, and $\{t_k\}$.

To simplify the notation, in what follows, we will use the scalar function $\varphi : \mathbb{R}^m \to \mathbb{R}$ defined as

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

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

$$\varphi(x+y) \le \varphi(x) + \varphi(y), \qquad \varphi(tx) = t\varphi(x), \quad x, y \in \mathbb{R}^m,\ t \ge 0. \quad (9)$$

$$x \preceq y \quad \Rightarrow \quad \varphi(x) \le \varphi(y), \quad x, y \in \mathbb{R}^m. \quad (10)$$

5.1 Partial Convergence Result

In this section, we prove that every accumulation point of $\{p^k\}$ is a Pareto critical point. Although the proof of this result is similar to that presented in [8, Theorem 5.1] (here we deal with inexact search directions), we chose to present its proof here for reasons of completeness.

Theorem 5.1 *The sequence $\{F(x^k)\}$ is decreasing, and the following statements hold*:

(i) *If $\{p^k\}$ has accumulation point, then $\{t_k^2\|v^k\|^2\}$ is a summable sequence, and*

$$\lim_{k\to+\infty} t_k\left\|v^k\right\|^2 = 0; \tag{11}$$

(ii) *Each accumulation point of the sequence $\{p^k\}$, if any, is a Pareto critical point.*

Proof The iterative step in Method 4.1 implies that

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

Since $\{p^k\}$ is an infinite sequence, for all k, p^k is not a Pareto critical point of F. Thus, the first part of the theorem follows from item ii of Lemma 4.3 combined with the last vector inequality.

Suppose now that $\{p^k\}$ has an accumulation point $\bar{p} \in M$. Taking into account that $\{F(p^k)\}$ is a decreasing sequence, it is easy to conclude that the whole sequence $\{F(p^k)\}$ converges to $F(\bar{p})$. So, from the definition of the function φ we conclude that $\{\varphi(F(x^k))\}$ converges to $\varphi(F(\bar{x}))$ and, in particular,

$$\varphi\left(F(\bar{x})\right) \le \varphi\left(F\left(x^k\right)\right), \quad k = 0, 1, \ldots. \tag{13}$$

From (12), (9), (10) and the definition of v^k we obtain

$$\varphi\left(F\left(p^{k+1}\right)\right) - \varphi\left(F\left(p^k\right)\right) \le \beta\left((1-\sigma)t_k\alpha\left(x^k\right) - (1/2)t_k\left\|v^k\right\|^2\right), \quad k = 0, 1 \ldots. \tag{14}$$

Summing the last inequality from $k = 0$ to n and taking into account that $|\alpha(x^k)| = -\alpha(x^k)$, $\beta \in]0, 1[$, and $\varphi(F(\bar{p})) \le \varphi(F(p^{n+1}))$ (see (13)), we get

$$\sum_{k=0}^{n}\left[(1-\sigma)t_k\left|\alpha\left(p^k\right)\right| + (1/2)t_k\left\|v^k\right\|^2\right] \le \frac{\varphi(F(p^0)) - \varphi(F(\bar{p}))}{\beta}, \quad n \ge 0.$$

But this tells us that (recall that $\sigma \in [0, 1[$)

$$\sum_{k=0}^{+\infty} t_k\left|\alpha\left(p^k\right)\right| < +\infty \quad \text{and} \quad \sum_{k=0}^{+\infty} t_k\left\|v^k\right\|^2 < +\infty, \tag{15}$$

from which the second part of item i follows. The first part of item i follows from last inequality in (15) together with the fact that $t_k \in]0, 1]$.

We assume initially that $\bar{p}$ is an accumulation point of the sequence $\{p^k\}$ and that $\{p^{k_s}\}$ is a subsequence of $\{p^k\}$ converging to $\bar{p}$. As $\{p^{k_s}\}$ converges to $\bar{p}$, we assume that $\{(p^{k_s}, v(p^{k_s}))\} \subset TU_{\bar{p}}$, where $U_{\bar{p}}$ is a neighborhood of $\bar{p}$ such that $TU_{\bar{p}} \approx U_{\bar{p}} \times \mathbb{R}^n$. From Lemma 4.2 we may to conclude that $\{v(p^{k_s})\}$ and $\{\alpha_{p^{k_s}}\}$ converge, respectively, to $v(\bar{p})$ and $\alpha_{\bar{p}}$. In particular, from Lemma 4.4 it follows that $\{v^{k_s}\}$ is bounded and, hence, has a convergent subsequence. Moreover, the sequence $\{t_k\} \subset]0,1]$ also has an accumulation point $\bar{t} \in [0,1]$. We assume, without loss of generality, that $\{t_{k_s}\}$ converges to $\bar{t}$ and $\{v^{k_s}\}$ converges to some $\bar{v}$. From Eq. (11) it follows that

$$\lim_{s\to+\infty} t_{k_s}\left\|v^{k_s}\right\|^2 = 0. \tag{16}$$

We have two possibilities to consider: (a) $\bar{t} > 0$ and (b) $\bar{t} = 0$. Assume that item (a) holds. Then, from (16) it follows that $\bar{v} = 0$. So, using the definition of v^k, it is easy to see that $\bar{v} = 0$ is a σ-approximation steepest descent direction for F at $\bar{p}$, and from item i of Lemma 4.3 we conclude that $\bar{p}$ is a Pareto critical point of F.

Now, assume that item (b) holds. Since v^{k_s} is a σ-approximation steepest descent method for F at p^{k_s} and $\{p^{k_s}\}$ is not a Pareto critical point, we have

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

where the last inequality is a consequence of item (iii) of Lemma 4.1. Hence, letting s to $+\infty$ in the last inequalities and using that $\{v^{k_s}\}$ converges to $\bar{v}$, we obtain

$$\max_{i\in I}\big\langle \operatorname{grad} f_i(\bar{p}), v(\bar{p})\big\rangle \le (1-\sigma)\alpha(\bar{p}) \le 0. \tag{17}$$

Take $r \in \mathbb{N}$. Since $\{t_{k_s}\}$ converges to $\bar{t} = 0$, we conclude that if s is large enough, then $t_{k_s} < 2^{-r}$. From (7) this means that the Armijo condition (12) is not satisfied for $t = 2^{-r}$, i.e.,

$$F\big(\exp_{p^k}\big(2^{-j}v^{k_s}\big)\big) \npreceq F\big(p^{k_s}\big) + \beta 2^{-r} JF\big(p^{k_s}\big)v^{k_s},$$

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

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

Letting s to $+\infty$ in the above inequality, taking into account that $\operatorname{grad} f_{i_0}$ and the exponential mapping are continuous, and using that $\{v^{k_s}\}$ converges to $\bar{v}$, we obtain

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

So, letting r to $+\infty$ and assuming that $0 < \beta < 1$, we obtain $\langle \operatorname{grad} f_{i_0}(\bar{p}), v(\bar{p})\rangle \ge 0$. Hence,

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

which, combined with (17) and taking into account that $\sigma \in [0,1[$, implies $\alpha(\bar{p}) = 0$. Therefore, from item (iii) of Lemma 4.1 it follows that $\bar{p}$ is a Pareto critical point of F, and the proof is concluded. □

5.2 Full Convergence

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

Definition 5.1 A function $H\colon M \to \mathbb{R}^m$ is called pseudo-convex on M iff H is differentiable and, for every $p, q \in M$ and every geodesic segment $\gamma : [0,1] \to M$ joining p to q, the following holds:

$$JH(p)\gamma'(0) \not\prec 0 \quad \Rightarrow \quad H(q) \not\prec H(p).$$

Remark 5.1 The definitions of convex and quasi-convex functions were presented in [8]. In the particular case that H is differentiable, convexity implies pseudo-convexity.

The next proposition provides a characterization for differentiable quasi-convex functions. From this characterization it follows that pseudo-convex functions are quasi-convex.

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

$$H(q) \prec H(p) \quad \Rightarrow \quad JH(p)\gamma'(0) \preceq 0. \tag{18}$$

Proof Let us assume that, for every pair of points $p, q \in M$ and every geodesic segment $\gamma : [0,1] \to M$ joining p to q, (18) holds. Take $\tilde{p}, \tilde{q} \in M$ and assume that, for every geodesic segment $\gamma : [0,1] \to M$ with $\gamma(0) = \tilde{p}$ and $\gamma(1) = \tilde{q}$, it follows that $H(\tilde{q}) \prec H(\gamma(t))$ for $t \in [0,1[$. So, using (18) with $q = \tilde{q}$ and $p = \gamma(t)$, we obtain

$$\begin{aligned} &JH\big(\gamma(t)\big)\gamma'(t) \preceq 0 \\ \Rightarrow \quad &\frac{d}{dt} h_i\big(\gamma(t)\big) = \big\langle \operatorname{grad} h_i\big(\gamma(t)\big), \gamma'(t)\big\rangle \le 0, \quad i \in \{1, \ldots, m\}, \end{aligned}$$

where $H = (h_1, \ldots, h_m)$. But this implies that $h_i(\gamma(t)) \le h_i(\gamma(0)) = h_i(\tilde{p})$ for $i \in \{1, \ldots, m\}$ and, hence, that $H(\gamma(t)) \preceq H(\tilde{p}) = \max\{H(\tilde{p}), H(\tilde{q})\}$, which proves the first part of the proposition. The proof of the second part follows immediately from the definition of quasi-convexity combined with differentiability of H; see [8] for more details. □

We know that criticality is a necessary, but not sufficient, condition for optimality. In [8] the authors proved that, under the convexity of the vectorial function F, criticality is equivalent to the weak optimality. Next we prove that the equivalence still happens if F is just pseudo-convex.

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 pseudo-convex function. Then, $p \in M$ is a Pareto critical point of H iff p is a weak optimal Pareto point of H.*

Proof Let us suppose that p is a Pareto critical point of H. Assume by contradiction that p is not a weak Pareto optimal point of H, i.e., that there exists $\tilde{p} \in M$ such that $H(\tilde{p}) \prec H(p)$. 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 pseudo-convex, then the last inequality implies that $JH(p)\gamma'(0) \prec 0$. But this contradicts the fact of p being a Pareto critical point of H, and so the first part is concluded. The second part is a simple consequence of the fact that F is differentiable with the definitions of Pareto critical point and weak Pareto optimal point. For more details, see [8]. □

Consider the following set

$$U := \{p \in M : F(p) \preceq F(p^k), k = 0, 1, \dots\}. \tag{19}$$

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. See [8, Remark 5.3].

Assumption 5.1 Each v^k of the sequence $\{v^k\}$ is a compatible scalarization, i.e., there exists a sequence $\{w^k\} \subset \operatorname{conv} S$ such that

$$v^k = -JF(p^k)^t w^k, \quad k = 0, 1, \dots.$$

As was observed in Sect. 4, this assumption holds if $v^k = v(x^k)$, i.e., if v^k is the exact steepest descent direction at x^k. We observe that Assumption 5.1 also was used in [3] for proving the full convergence of the sequence generated for the Algorithm in the case that M is the Euclidean space and F is convex. From now on, we will assume that Assumption 5.1 holds.

Next lemma generalizes [8, Lemma 5.3] for directions satisfying Assumption 5.1. It is the main result of this section that is fundamental for the proof of the global convergence result of the sequence $\{p^k\}$.

Lemma 5.1 *Suppose that F is quasi-convex, M has nonnegative curvature, and U, defined in (19), is nonempty. Then, for all $\tilde{p} \in U$, the following inequality holds:*

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

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 $\alpha = \angle(\gamma_1'(0), v^k)$. Taking into account that $\cos(\pi - \alpha) = -\cos\alpha$ and $\langle -v^k, \gamma_1'(0)\rangle = \|v^k\| \cos(\pi - \alpha)$, from the law of cosines (Theorem 2.1) we have

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

For each $k \in \mathbb{N}$, Assumption 5.1 implies that there exists $w^k \in \operatorname{conv} S$ such that $v^k = -JF(p^k)^t w^k$. So, the last vector inequality yields

$$d^2(p^{k+1}, \tilde{p}) \le d^2(p^k, \tilde{p}) + t_k^2 \|v^k\|^2 + 2d(p^k, \tilde{p}) t_k \langle w^k, JF(p^k)\gamma_1'(0)\rangle,$$
$$k = 0, 1, \dots . \tag{20}$$

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

$$JF(p^k)\gamma_1'(0) \preceq 0, \quad k = 0, 1, \dots .$$

Now, because $w^k \in \operatorname{conv} S$, we get

$$\langle w^k, JF(p^k)\gamma_1'(0)\rangle \le 0, \quad k = 0, 1, \dots . \tag{21}$$

Therefore, the lemma follows by combining (20) with (21). □

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

$$\sum_{k=0}^{+\infty} \epsilon_k < +\infty, \qquad d^2(q^{k+1}, q) \le d^2(q^k, q) + \epsilon_k, \quad k = 0, 1, \dots .$$

Proposition 5.3 *If F is quasi-convex, M has nonnegative curvature, and U, defined in (19), is a nonempty set, then the sequence $\{p^k\}$ is quasi-Fejér convergent to U.*

Proof The resulted follows from item (ii) of Theorem 5.1 and Lemma 5.1 combined with Definition 5.3. □

Theorem 5.2 *Suppose that F is quasi-convex, M has nonnegative curvature, and U, as defined in (19), is a nonempty set. Then, the sequence $\{p^k\}$ converges to a Pareto critical point of F.*

Proof From Proposition 5.3, $\{p^k\}$ is quasi-Fejér convergent to U. Thus, [23, Lemma 5.2] guarantees that $\{p^k\}$ is bounded and, by the Hopf–Rinow theorem, has an accumulation point $\bar{p} \in M$. Since $\{F(p^k)\}$ is a decreasing sequence, we conclude that $\bar{p} \in U$ and, hence, that the whole sequence $\{p^k\}$ converges to $\bar{p}$ (see [23, Lemma 5.2]). The conclusion of the proof is a consequence of item (iii) of Theorem 5.1. □

Corollary 5.1 *If F is pseudo-convex, M has nonnegative curvature, and U, as defined in (19), is a nonempty set, then the sequence $\{p^k\}$ converges to a weak Pareto optimal point of F.*

Proof Since F is pseudo-convex and in particular quasi-convex, the corollary is a consequence of the previous theorem and Proposition 5.2. □

6 Examples

For examples of complete Riemannian manifolds with explicit geodesic curves and the steepest descent iteration of the sequence generated by Method 4.1, see [8, Examplea 6.1, 6.2, and 6.3]. The manifolds of Examples 6.1 and 6.2 have sectional curvature identically null. The manifold of Example 6.3 has sectional curvature $K \leq 0$. In this section we present another example of complete Riemannian manifold with explicit geodesic curves whose sectional curvature is $K \geq 0$, not identically null. We recall that the function $F : M \to \mathbb{R}^m$, $F(p) := (f_1(p), \dots, f_m(p))$, is differentiable. If (M, G) is a Riemannian manifold, then the Riemannian gradient of f_i is given by $\operatorname{grad} f_i(p) = G(p)^{-1} f_i'(p)$, $i \in I := \{1, \dots, n\}$. Hence, if $v(p)$ is the steepest descent direction for F at p (see Definition 4.1), it takes the form given by Lemma 4.1.

Example 6.1 (Steepest descent method for the revolution surfaces) Let $f, g : \mathbb{R} \to \mathbb{R}$ be continuously differentiable functions such that $f(u) > 0$, and let M be a revolution surface, obtained from a regular plane curve $\beta(u) := (f(u), g(u))$ (known as generating curve of M), endowed with the induced metric from $\mathbb{R}^3$. Note that M is connected and by [24, Corollary 2.10], M is also a complete Riemannian manifold. Consider the following parameterization of M:

$$X(u, v) = \big(f(u)\cos v, f(u)\sin v, g(u)\big), \quad 0 \leq v < 2\pi. \tag{22}$$

In this case, when the curve β is parameterized by arc length, the sectional curvature (see [26, p. 162]) is given by $K = -f''/f$. Hence, a necessary and sufficient condition to have $K \geq 0$ is the concavity of f.

In the particular case where the generating curve $\beta(u) = (u, u^2)$ is parameterized by arc length, we have that $K > 0$ and M is a connected and complete Riemannian manifold. At each iteration k, we minimize the function F along the geodesics on M by solving the following second-order differential equation (with ODE/Scilab):

$$\begin{cases} (1 + 4u^2)u'' + 4u\big(u'\big)^2 - u\big(v'\big)^2 = 0, \\ uv'' + 2u'v' = 0, \end{cases}$$

with the initial conditions $u(0) = u^k$, $v(0) = v^k$, and $(u'(0), v'(0)) = d^k$, where d^k is a steepest descent direction at $p^k = X(u^k, v^k)$ for F.

7 Numerical Experiments

In this section, we present some numerical experiments. The examples presented illustrate the performance of Method 4.1 when the manifold is a hypercube and when the manifold is an elliptic paraboloid.

The algorithm was coded in SCILAB 5.3 on a 2-GB RAM Atom notebook. We denote by Iter(k) the number of iterations and by *Call.Armijo* the number of steps in Armijo search. The stop condition is $\alpha(p^k) \in \,]-\epsilon, 0[$, where $\epsilon = 10^{-4}$.

7.1 Quasi-Convex Minimization on a Hypercube

We consider problem (1), where $M = (]0,1[^n, P^{-2}(I-P)^{-2})$, and $F(p) = (f_1(p), f_2(p), f_3(p))$ is given by $f_1(p) = \sqrt{-log(p_1(1-p_1)p_2(1-p_2))}$, $f_2(p) = \log(1-\log(p_1(1-p_1)p_2(1-p_2)))$, and

$$f_3(p) = \arctan\bigl(-\log\bigl(p_1(1-p_1)p_2(1-p_2)\bigr)\bigr).$$

It is known that M is a connected and complete Riemannian manifold and each f_i, $i = 1, 2, 3$, is quasi-convex in M; see [25]. Note that problem (1) has a unique minimal point $p^* = (0.5, 0.5)$ and $f_1(p^*) = 2\sqrt{\log 2}$, $f_2(p^*) = \log(1 + 4\log 2)$, and $f_3(p^*) = \arctan(4\log 2)$. We show the ISDM behavior in Table 1.

7.2 A Convex Problem on an Elliptic Paraboloid

Now, let us consider problem (1), where $M = \{(p_1, p_2, p_3) \in \mathbb{R}^3 \colon p_3^2 = p_1^2 + p_2^2\}$ with the parameterization (22), and $F(p) = (f_1(p), f_2(p))$ is given by $f_1(p) = \sqrt{p_3}e^{\sqrt{p_3}}$ and $f_2(p) = \frac{\sqrt{p_3}\arctan(\frac{p_2}{p_1})^2}{16\pi^2}$, whose optimal value is $F^\star = 0$. In Fig. 1, we show the evolution of $\alpha(p^k)$ along of the 500 Iterations(k).

Table 1 Behavior of ISDM with different starting points

p^0	Iter(k)	Call.Armijo	p^k	$\alpha(p^k)$
(0.10, 0.90)	115	114	(0.4552, 0.4561)	$-9.8038D$-05
(0.20, 0.80)	80	79	(0.4533, 0.4554)	$-9.9238D$-05
(0.60, 0.40)	23	22	(0.5540, 0.5280)	$-9.7997D$-05
(0.85, 0.15)	98	97	(0.5437, 0.5436)	$-9.8880D$-05

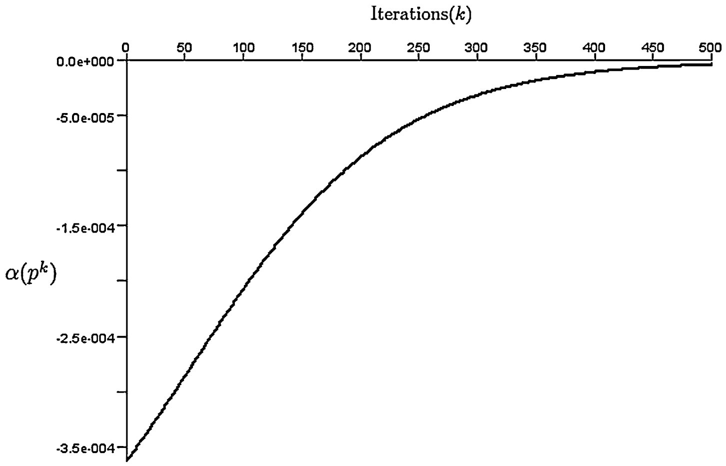

Fig. 1 Minimization on the elliptic paraboloid

8 Final Remarks

The convergence analysis presented in this paper uses the principle of quasi-Fejér convergence. It is known, since the preliminary studies on the convergence analysis in the Riemannian context, that such an approach determines a restriction on the sign of the sectional curvature of the manifold and on the objective function (in our case multiobjective function). In particular, this approach is limited to manifolds that have infinite volume; see Yamaguchi [27]. As future work, we intend to extend our analysis to compact Riemannian manifolds whose sign of the sectional curvature is not necessarily constant; see Rapsáck [28] for examples of such manifolds. We mention that the analysis proposed in this present paper do not extend trivially to this situation. For this purpose, a different approach is necessary; see [29].

Acknowledgements The authors would like to extend their gratitude toward anonymous referees whose suggestions helped us to improve the presentation of this paper. The first author was partially supported by CNPq Grant 471815/2012-8, Project CAPES-MES-CUBA 226/2012, PROCAD-nf-UFG/UnB/IMPA, and FAPEG/CNPq. The second author was partially supported by CNPq GRANT 301625-2008 and PRONEX-Optimization (FAPERJ/CNPq).

References

1. Janh, J.: Vector Optimization. Springer, Berlin (2011)
2. Graña Drummond, L.M., Maculan, N., Svaiter, B.F.: On the choice of parameters for the weighting method in vector optimization. Math. Program. **111**, 201–216 (2008)
3. Graña Drummond, L.M., Svaiter, B.F.: A steepest descent method for vector optimization. J. Comput. Appl. Math. **175**(2), 395–414 (2005)
4. Graña Drummond, L.M., Iusem, A.N.: A projected gradient method for vector optimization problems. Comput. Optim. Appl. **28**(1), 5–29 (2004)
5. Fliege, J., Svaiter, B.F.: Steepest descent methods for multicriteria optimization. Math. Methods Oper. Res. **51**(3), 479–494 (2000)
6. Fukuda, E.H., Graña Drummond, L.M.: On the convergence of the projected gradient method for vector optimization. Optimization **60**, 1009–1021 (2011)
7. Fukuda, E.H., Graña Drummond, L.M.: Inexact projected gradient method for vector optimization. Comput. Optim. Appl. **54**(3), 473–493 (2013)
8. Bento, G.C., Ferreira, O.P., Oliveira, P.R.: Unconstrained steepest descent method for multicriteria optimization on Riemannian manifolds. J. Optim. Theory Appl. **154**, 88–107 (2012)
9. Bello Cruz, J.Y., Lucambio Pérez, L.R., Melo, J.G.: Convergence of the projected gradient method for quasiconvex multiobjective optimization. Nonlinear Anal. **74**, 5268–5273 (2011)
10. Takayama, A.: Mathematical Economics. Cambridge Univ. Press, Cambridge (1997)
11. Gromicho, J.: Quasiconvex Optimization and Location Theory. Kluwer Academic Publishers, Dordrecht (1998)
12. Bonnel, H., Iusem, A.N., Svaiter, B.F.: Proximal methods in vector optimization. SIAM J. Optim. **15**(4), 953–970 (2005)
13. Fliege, J., Graña Drummond, L.M., Svaiter, B.F.: Newton's method for multiobjective optimization. SIAM J. Optim. **20**(2), 602–626 (2009)
14. Ceng, L.C., Mordukhovich, B.S., Yao, J.C.: Hybrid approximate proximal method with auxiliary variational inequality for vector optimization. J. Optim. Theory Appl. **146**, 267–303 (2010)
15. Udriste, C.: Convex Functions and Optimization Algorithms on Riemannian Manifolds. Mathematics and Its Applications, vol. 297. Kluwer Academic Publishers, Dordrecht (1994)
16. Rapcsák, T.: Smooth Nonlinear Optimization in R^n. Kluwer Academic Publishers, Dordrecht (1997)
17. Absil, P.-A., Mahony, R., Sepulchre, R.: Optimization Algorithms on Matrix Manifolds. Princeton University Press, Princeton (2008)
18. 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)

19. 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)
20. Li, C., Mordukhovich, B.S., Wang, J.H., Yao, J.C.: Weak sharp minima on Riemannian manifolds. SIAM J. Optim. **21**, 15–23 (2011)
21. 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)
22. Bento, G.C., Melo, J.G.: A subgradient method for convex feasibility on Riemannian manifolds. J. Optim. Theory Appl. **152**, 773–785 (2012)
23. 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)
24. Do Carmo, M.P.: Riemannian Geometry. Birkhauser, Boston (1992)
25. 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)
26. do Carmo, M.P.: Differential Geometry of Curves and Surfaces. Prentice Hall, New Jersey (1976)
27. Yamaguchi, T.: Locally geodesically quasiconvex functions on complete Riemannian manifolds. Trans. Am. Math. Soc. **298**(1), 307–330 (1986)
28. Rapsáck, T.: Sectional curvature in nonlinear optimization. J. Glob. Optim. **40**(1–3), 375–388 (2008)
29. Bento, C.G., Cruz Neto, J.X., Oliveira, P.R.: Convergence of inexact descent methods for nonconvex optimization on Riemannian manifolds (2011). arXiv:1103.4828