Pareto solutions as limits of collective traps: an inexact multiobjective proximal point algorithm

Glaydston Carvalho Bento, João Xavier da Cruz Neto, L. Meireles, Antoine Soubeyran

Pareto solutions as limits of collective traps: an inexact multiobjective proximal point algorithm

G. C. Bento[1] · J. X. Cruz Neto[2] · L. V. Meireles[3] · A. Soubeyran[4]

Abstract
In this paper we introduce a definition of approximate Pareto efficient solution as well as a necessary condition for such solutions in the multiobjective setting on Riemannian manifolds. We also propose an inexact proximal point method for nonsmooth multiobjective optimization in the Riemannian context by using the notion of approximate solution. The main convergence result ensures that each cluster point (if any) of any sequence generated by the method is a Pareto critical point. Furthermore, when the problem is convex on a Hadamard manifold, full convergence of the method for a weak Pareto efficient solution is obtained. As an application, we show how a Pareto critical point can be reached as a limit of traps in the context of the variational rationality approach of stay and change human dynamics.

Keywords Multiobjective proximal method · Riemannian manifold · Approximate solution · Variational rationality · Worthwhile move · Trap

Mathematics Subject Classification 90C29 · 90C30 · 49M30

1 Introduction

In this paper, we present an inexact proximal point method for multiobjective optimization in the Riemannian setting based on the concept of approximate Pareto solution.

✉ G. C. Bento
glaydston@ufg.br

J. X. Cruz Neto
jxavier@ufpi.edu.br

L. V. Meireles
lucas.vidal@ifgoiano.edu.br

A. Soubeyran
antoine.soubeyran@gmail.com

1 IME, Universidade Federal de Goiás, Goiânia, GO 74001-970, Brazil

2 CCN, DM, Universidade Federal do Piauí, Teresina, PI 64049-550, Brazil

3 DM, Instituto Federal Goiano, Posse, GO 73900-000, Brazil

4 CNRS, EHESS, Aix-Marseille School of Economics, Aix-Marseille University, Aix-en-Provence, France

It should be noted that vector optimization models had found a lot of important applications in decision making problems such as in economics theory, management science and engineering design. Owing to these applications, a lot of literatures have been published to study optimality conditions, duality theories and topological properties of solutions of vector optimization problems (Chen et al., 2005; Jahn, 2004; Luc, 1989). It is worth noticing that Bonnel et al. (2005) constructed a vector-valued proximal point algorithm to investigate convex vector optimization problems in Hilbert spaces extending the convergence analysis from the scalar case [see, for instance, Rockafellar (1976)] to the vectorial case. For others references dealing with variants of the algorithms considered in Bonnel et al. (2005) for convex vector or multiobjective problems; see, for instance, (Ceng & Yao, 2007; Ceng et al., 2010; Choung et al., 2011; Gregório & Oliveira, 2011; Villacorta & Oliveira, 2011). An approach for the $\mathbb{R}^m_+$-quasi-convex case was discussed in Bento et al. (2014), Apolinário et al. (2016). More recently, in Bento et al. (2018) the authors present a new (scalarization-free) approach for convergence analysis of the exact proximal point method considered in Bento et al. (2005) where the first-order optimality condition of the scalarized problem is replaced by a necessary condition for weak Pareto points of a multiobjective problem. As a consequence, this allowed to consider the method without any assumption of convexity over the constraint sets that determine the vectorial improvement steps. In Bento et al. (2018) the authors extended the necessary condition for weak Pareto points of a multiobjective problem [(see Minami (1983) for a version in the linear setting] as well as the approach presented in Bento et al., (2018) to the Riemannian context, which allowed to authors to consider the proximal point method for locally Lipschitz functions in multiobjective programming on Hadamard manifolds. For an approach of the proximal method for convex vector problems on Hadamard manifolds via scalarization see Bento et al. (2018b); Tang and Huang (2017).

Motivated by the definition of approximate Pareto efficient solutions introduced by Loridan (1984, Definition 4.1), in this paper we present that definition in the Riemannian setting and establish a necessary condition for this type of Pareto efficient solution, see Theorem 2, which plays an important role in the course of this paper. Besides, we present an inexact version of the proximal algorithm given in Bento et al. (2018), where each term of the sequence generated need not be an exact solution of the vectorial proximal optimization subproblem, but just an approximate solution of it by considering a relative error criteria as introduced in Solodov and Svaiter (1999). Note that the inexact versions of the proximal methods developed in the Riemannian context for vector optimization, see Bento et al. (2018b), Tang and Huang (2017), were developed, as in Bento et al. (2005), via scalarization by considering the same error criteria whose measure of the relative error is constant or has summable square. In our approach, we show that the accumulation points, if any, of each generated sequence are Pareto critical points and, in the convex case, we show full convergence of any generated sequence to a weak Pareto point. As an application, we show how a Pareto critical point can be reached as a limit of traps in the context of the variational rationality approach of stay and change human dynamics; see Soubeyran (2009, 2010, 2021a, 2021b, 2021c, 2021d).

In the rest of the paper is organized as follows. Section 2 provides some basic definitions and auxiliary results. In Sect. 3, we formulate definition of approximate Pareto solution and necessary conditions for characterizing this solutions for a multiobjective optimization problem on Riemannian manifolds. Section 4 is devoted to presenting and to analysing an inexact version of the proximal point algorithm. Section 5 provides an application showing how a Pareto critical point can be reached as a limit of traps in the context of the variational rationality approach of stay and change human dynamics. In the last section we present perspective future of works.

2 Preliminaries

In this section, we present some pertinent concepts and results related to Riemannian geometry. For more details see, for example, (Do Carmo, 1992; Ledyaev & Zhu, 2007; Li et al., 2011; Sakai, 1996).

Assuming that M is a complete and connected Riemannian manifold, from Hopf-Rinow theorem it is known that any pair of points in M can be joined by a minimal geodesic. Moreover, (M, d) is a complete metric space, where d denotes the Riemannian distance, and bounded closed subsets are compact. Given $x \in M$, the exponential map $\exp_x : T_x M \to M$ is defined by $\exp_x v := \gamma_v(1, x)$, where $\gamma_v(\cdot, x)$ denotes the geodesic determined by its position x and velocity v at x.

A complete, simply connected Riemannian manifold of nonpositive sectional curvature is called a Hadamard manifold. The following result is well known [see, for example, (Sakai, (1996), Theorem 4.1, p. 221].

Proposition 1 *Let M be a Hadamard manifold and $x \in M$. Then, $\exp_x$ is a diffeomorphism, and for any two points $x, y \in M$ there exists an unique normalized geodesic joining x to y, which is, in fact, a minimal geodesic. In particular, $d(x, y) = \| \exp_x^{-1} y\|$.*

From now on, M is assumed to be a finite-dimensional Hadamard manifold. One important property is described in the following proposition, which is taken from [Sakai (1996), Proposition 4.5, p. 223] and will be useful in our study. Recall that a geodesic triangle $\triangle(xyz)$ of a Riemannian manifold is a set consisting of three points x, y and z and three minimal geodesics joining these points.

Proposition 2 (Law of cosines) *Given a geodesic triangle $\triangle(xyz)$ in a Hadamard manifold, it holds*

$$d^2(x, z) + d^2(z, y) - 2\langle \exp_z^{-1} x, \exp_z^{-1} y\rangle \le d^2(x, y),$$

where $\exp_z^{-1}(\cdot)$ denotes the inverse of $\exp_z(\cdot)$.

Now, we introduce some elements of nonsmooth analysis on Riemannian manifolds which will be useful throughout the paper.

We denote the extended real line by $\bar{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$. For an extended-valued function $f : M \to \bar{\mathbb{R}}$ the domain of f is defined by

$$\mathrm{dom}(f) := \{x \in M : f(x) < +\infty\}.$$

We recall that f is said to be proper if $\mathrm{dom}(f) \neq \emptyset$.

Definition 1 (Ledyaev & Zhu, 2007) Let $f : M \to \bar{\mathbb{R}}$ be a proper and lower semicontinuous function. The Fréchet subdifferential of f at $x \in \mathrm{dom}(f)$ is defined by

$$\partial_F f(x) := \{dg(x) : g \in C^1(M) \text{ and } f - g \text{ attains a local minimum at } x\},$$

where $dg(x)$ is the differential of g at x. The (limiting) subdifferential of f at $x \in M$ is defined by

$$\partial f(x) := \{v \in T_x M : \exists (x^k, v^k) \in \mathrm{Graph}(\partial_F f),\ (x^k, v^k) \to (x, v),\ f(x^k) \to f(x)\},$$

where $\mathrm{Graph}(\partial_F f) := \{(y, u) \in TM : u \in \partial_F f(y)\}$.

It follows directly from Definition 1 that $\partial_F f(x) \subset \partial f(x)$. Note further that $\partial_F f(x)$ may be empty, but if f attains a local minimum at x, then $0 \in \partial_F f(x)$. A necessary (but not sufficient) condition for $x \in M$ to be a minimizer of f is that

$$0 \in \partial f(x).$$

Two next propositions have appeared in Meireles (2019).

Proposition 3 *Let $f : M \to \bar{\mathbb{R}}$ be a locally Lipschitz function. Consider a sequence $(x^k)_k \subset \operatorname{dom}(\partial_F f)$ which is bounded. If the sequence $(v^k)_k$ is such that $v^k \in \partial_F f(x^k)$, for each $k \in \mathbb{N}$, then (v^k) is also bounded.*

Proposition 4 *If $f : M \to \bar{\mathbb{R}}$ is a locally Lipschitz function, then the limiting subdifferential is not empty.*

For further details and properties on (limiting) subdifferentials on Riemannian manifolds see, for example (Ledyaev & Zhu, 2007).

For a general closed subset $C \subset M$ and a point $x_0 \in C$, the Fréchet normal cone and (limiting) normal cone of C at x_0 were defined in Ledyaev and Zhu (2007) as

$$N_C^F(x_0) := \partial_F \delta_C(x_0); \qquad N_C(x_0) := \partial \delta_C(x_0),$$

respectively, where $\delta_C : M \to \bar{\mathbb{R}}$ is the indicator function of C. As remarked by Li et al. (2011), when C is convex, i.e., for any $x, y \in C$ there exists a minimizing geodesic of M connecting x, y is contained in C, then $N_C^F(x_0) = N_C(x_0)$.

Proposition 5 *Let $C \subset M$ be a not empty and closed set. If $x \in C$, then*

$$\partial_F d_C(x) \subset \mathbb{B}_x \cap N_C(x), \tag{2.1}$$

where $d_C(x) := \inf\{d(y, x) : y \in C\}$ and $\mathbb{B}_x := \{\xi \in T_x M : \|\xi\| \leq 1\}$.

In the particular case where M is a Hadamard manifold and C is a convex set, it follows from [(Li et al., (2011), Lemma 5.2 and Theorem 5.3], that the expression (2.1) of the last proposition takes the form

$$\partial d_C(x) = \mathbb{B}_x \cap N_C(x).$$

As noted previously, (M, d) is a metric space. Hence, the result below is valid in our context and they will be useful in next section.

Theorem 1 *Let (M, d) be a metric space. Take a set $C \subset M$ and a function $f : M \to \mathbb{R}$. Assume that f satisfies a Lipschitz condition on M with Lipschitz constant L. Let $\bar{x}$ be a minimizer for the constrained minimization problem,*

$$\min f(x), \qquad x \in C. \tag{2.2}$$

Choose any $\tau \geq L$. Then, $\bar{x}$ is also a minimizer for the unconstrained minimization problem,

$$\min\{f(x) + \tau d_C(x)\}, \qquad x \in M. \tag{2.3}$$

If $\tau > L$ and C is a closed set, then the converse assertion is also true: any minimizer $\bar{x}$ for the unconstrained problem (2.3) is also a minimizer for the constrained problem (2.2) and so, in particular, $\bar{x} \in C$.

Proof See [(Vinter (2000), Theorem 3.2.1]. □

3 ϵ-quasi-weakly Pareto optimality

In this section, we consider a multiobjective program and give necessary condition for ϵ-quasi-weakly Pareto solutions, which will play a very important role during the analysis of the our method later on.

Let h_i, $i \in \{1, 2, \ldots, m\}$, g_j, $j \in \{1, 2, \ldots, l\}$, be real-valued functions on a finite dimensional Riemannian manifold M, $C \subset M$ a not empty and closed set and $\Omega := \{x \in C \mid g_j(x) \leq 0, \ j = 1, 2, \ldots, l\}$. We consider the problem of finding a weakly Pareto solution of $H := (h_1, \ldots, h_m)$ in Ω, i.e., a point $x^* \in \Omega$ such that there does not exist another point $x \in \Omega$ with $h_i(x) < h_i(x^*)$ for all $i \in \{1, \ldots, m\}$. We denote this constrained problem as

$$\min_{\omega} H(x) \quad s.t. \quad x \in \Omega, \tag{3.1}$$

and the set of all its weakly Pareto solutions by

$$S^{\omega}(H, \Omega) := \operatorname{argmin}_{\omega}\{H(x) : x \in \Omega\}.$$

Definition 2 Let $\epsilon \in \mathbb{R}^m_+$. A point $x^* \in \Omega$ is an ϵ-quasi-weakly Pareto solution of the problem in (3.1) if there does not exist another point $x \in \Omega$ such that

$$h_i(x) + \epsilon_i d(x, x^*) < h_i(x^*), \quad i = 1, 2, \ldots, m,$$

where d is the Riemannian distance.

We denote the set of ϵ-quasi-weakly Pareto solutions H in Ω by $S^{\omega}_{\epsilon q}(H, \Omega)$.

Remark 1

(1) Note that Definition 2 is a natural extension of that introduced in the linear setting in Loridan (1984). See also Chuong and Kim (2016) and references therein;
(2) It is easy to see that, even in the linear context, in general, $S^{\omega}(H, \Omega) \subsetneq S^{\omega}_{\epsilon q}(H, \Omega)$ and $S^{\omega}(H, \Omega) = S^{\omega}_{\epsilon q}(H, \Omega)$ for $\epsilon = 0$.

Next, we present an illustrative example in the Riemannian context where it is possible to notice that an ϵ-quasi-weakly Pareto Solution is different from a classical weak Pareto Solution.

Example 1 Let $\mathcal{H} := \{p = (x, y) \in \mathbb{R}^2 : y > 0\}$, $H : \mathcal{H} \to \mathbb{R}^2$ given by $H(p) = (y, x^2 y^{-1} + 1)$ and let us consider $\mathcal{G}(p) := (g_{ij}(p))$, where

$$g_{11}(p) = g_{22}(p) = \frac{1}{y^2}, \qquad g_{12}(p) = g_{21}(p) = 0.$$

It is known that $\mathbb{H}^2 = (\mathcal{H}, \langle\langle \ , \ \rangle\rangle)$, with $\langle\langle u, v\rangle\rangle_p := v^T \mathcal{G}(p) u$, is a complete Riemannian manifold of sectional curvature -1, called the Poincaré plane. Note that the coordinate functions of H are locally Lipschitz, but not convex functions on $\mathbb{H}^2$. It can be shown that the geodesics in $\mathbb{H}^2$ are the semi-lines and the semicircles orthogonal to the line $v = 0$; see [(Udriste (1994), page 20]. The Riemannian metric $\mathcal{G}(\cdot)$ induces the hyperbolic distance between the points $p = (x, y)$ and $\tilde{p} = (z, w)$ given by:

$$d_p(\tilde{p}) := d(p, \tilde{p}) = \operatorname{arccosh}\left(1 + \frac{(z - x)^2 + (w - y)^2}{2yw}\right).$$

Let's denote $\theta := 1 + \left[(z-x)^2 + (w-y)^2\right]/2yw$. Thus,

$$\exp^{-1}_{(z,w)}(x, y) = \operatorname{arccosh}(\theta)\Big(\frac{wx - wz}{y\sqrt{\theta^2 - 1}}, \frac{y^2 - w^2 + (z-x)^2}{2y\sqrt{\theta^2 - 1}}\Big).$$

Take $\Omega = C := \{(x, y) \in \mathbb{H}^2 : x \geq 1, y \geq 1\}$. In this case,

$$S^{\omega}(H, \Omega) = \{(1, w) : w \geq 1\} \cup \{(z, 1) : z \geq 1\}.$$

Given $\epsilon_1 = \epsilon_2 = \epsilon < 1$, Definition 2 says us that $(x^*, y^*) \in S^{\omega}_{\epsilon q}(H, \Omega)$ iff there does not exist another point $(x, y) \in \Omega$ such that

$$\begin{cases} \epsilon.\operatorname{arccosh}\left(1 + \frac{(x^*-x)^2+(y^*-y)^2}{2yy^*}\right) < y^* - y, \\ \epsilon.\operatorname{arccosh}\left(1 + \frac{(x^*-x)^2+(y^*-y)^2}{2yy^*}\right) < (x^*)^2(y^*)^{-1} - x^2y^{-1} \end{cases}$$

For ϵ sufficiently small, it is not difficult to see that $P^* := \{(1 + \frac{\epsilon^2}{2}, y^*) : y^* \geq 1 + \frac{\epsilon^2}{2}\}$ is a set that is contained in $S^{\omega}_{\epsilon q}(H, \Omega) \setminus S^{\omega}(H, \Omega)$ and, as noted in item ii) of Remark 1, it follows that $S^{\omega}(H, \Omega) \subsetneq S^{\omega}_{\epsilon q}(H, \Omega)$.

From now on, we assume that h_i, $i \in \{1, 2, \ldots, m\}$, g_j, $j \in \{1, 2, \ldots, l\}$ are locally Lipschitz functions. Define

$$x \ni M \mapsto \phi(x) := \max_{\substack{i=1,\ldots,m \\ j=1,\ldots,l}} \{h_i(x) - h_i(x^*) + \epsilon_i d(x, x^*), g_j(x)\}. \tag{3.2}$$

The next result presents a Fritz-John type necessary condition for ϵ-quasi-(weakly) Pareto solutions associated to the problem in (3.1). For a version of this result in the linear setting, see Chuong and Kim (2016).

Theorem 2 *Let $x^* \in S^{\omega}_{\epsilon q}(H, \Omega)$. Then, there exist $\tau > 0$ and $\xi_i \geq 0$, $i = 1, \ldots, m$, $\mu_j \geq 0$, $j = 1, \ldots, l$ with $\sum_{i=1}^m \xi_i + \sum_{j=1}^l \mu_j = 1$, such that*

$$0 \in \sum_{i=1}^m \xi_i \partial h_i(x^*) + \sum_{j=1}^l \mu_j \partial g_j(x^*) + \sum_{i=1}^m \xi_i \epsilon_i \mathbb{B}_{x^*} + \tau \partial d_C(x^*).$$

Proof From the definition of ϕ in (3.2), it is easy to see that $\phi(x^*) = 0$ and $\phi(x) \geq 0$, for all $x \in C$ and, hence,

$$\phi(x^*) = \inf_{x \in C} \phi(x).$$

Since h_i, $i \in \{1, 2, \ldots, m\}$, g_j, $j \in \{1, 2, \ldots, l\}$ are locally Lipschitz functions, then $\phi(\cdot)$ is also a locally Lipschitz function. Let us suppose that L is a Lipschitz constant of $\phi(\cdot)$ at x^*. From Theorem 1 follows that x^* is a local minimal for $\phi(\cdot) + \tau d_C(\cdot)$, whenever $\tau \geq L$. From the first-order optimality condition and using the sum rule for the limiting subdifferential, see [Ledyaev & Zhu, (2007), Theorem 4.13(A.1)], we obtain

$$0 \in \partial\phi(x^*) + \tau \partial d_C(x^*). \tag{3.3}$$

By the definition of ϕ in (3.2), [Ledyaev & Zhu, (2007), Theorem 4.16(A2)] applied to ϕ implies that there exist non-negative real numbers ξ_i, μ_j with $i \in \{1, 2, \ldots, m\}$ and $j \in \{1, 2, \ldots, l\}$ such that $\sum_{i=1}^m \xi_i + \sum_{j=1}^l \mu_j = 1$ and

$$\partial\phi(x^*) \subset \sum_{\{i \mid \xi_i \neq 0\}} \xi_i \partial\left(h_i(\cdot) + \epsilon_i d(\cdot, x^*)\right)(x^*) + \sum_{\{j \mid \mu_j \neq 0\}} \mu_j \partial g_j(x^*).$$

Hence, using again [(Ledyaev & Zhu, (2007), Theorem 4.13(A.1)] and [(Ledyaev & Zhu, (2007), Theorem 5.2], we get

$$\partial\phi(x^*) \subset \sum_{\{i \mid \xi_i \neq 0\}} \xi_i \partial h_i(\cdot) + \sum_{\{i \mid \xi_i \neq 0\}} \xi_i \epsilon_i \mathbb{B}_{x^*} + \sum_{\{j \mid \mu_j \neq 0\}} \mu_j \partial g_j(x^*), \tag{3.4}$$

and the desired result follows by combining (3.3) with (3.4). □

Remark 2 From the last result it is possible retrieves the Fritz-John type necessary condition for weakly Pareto solutions which was introduced, in the Riemannian context, by Bento et al. in (2018).

Definition 3 Let $\Omega \subset M$ be a not empty, closed and convex set. A point $\bar{x} \in \Omega$ is said to be Pareto critical of H in Ω iff for any $y \in \Omega$, there are an index $i \in \{1, \ldots, m\}$ and $u \in \partial h_i(\bar{x})$ such that

$$\langle u, \exp_{\bar{x}}^{-1} y \rangle \geq 0.$$

Remark 3 (1) Note that, when M is a Hadamard manifold and $\Omega = M$, the last definition retrieves the Pareto critical point notion introduced in Bento and Cruz Neto (2013);

(2) If $\bar{x}$ is not a Pareto critical point of H in Ω, then there exists $y \in \Omega$ such that for all $i \in \{1, 2, \ldots, m\}$ and $u^i \in \partial h_i(\hat{x})$, $\langle u^i, \exp_{\bar{x}}^{-1} y \rangle < 0$.

4 Inexact proximal algorithm for multiobjective and convergence analysis

In this section, we introduce an inexact proximal point method for the multiobjective framework and we show that the accumulation points, if any, of each generated sequence are Pareto critical points. Moreover, in the convex case, we show full convergence of any generated sequence to a weak Pareto point.

Throughout this section, we assume that $C \subset M$ is a nonempty, convex, closed set and $F := (f_1, \ldots, f_m) : M \to \mathbb{R}^m$ such that each component function $f_i : M \to \mathbb{R}, i \in \{1, \ldots, m\}$, is a locally Lipschitz function. For $x, y \in \mathbb{R}^m_+$, $x \preceq y$ means that $y - x \in \mathbb{R}^m_+$ and $x \prec y$ means that $y - x \in \mathbb{R}^m_{++}$.

4.1 Inexact proximal point algorithm

In the following, we formally state an inexact proximal point algorithm for solving a constrained optimization problem as in (3.1).

Algorithm 1 (Inexact Proximal Algorithm)
Initialization: Take $(\lambda_k)_k, \subset \mathbb{R}_{++}$, $(\varsigma^k)_k \subset \mathbb{R}^m_{++}$ such that $\|\varsigma^k\| = 1$ for all $k \in \mathbb{N}$, and choose $x^0 \in C$.
Stopping rule: Given x^k, if x^k is a Pareto critical point, then set $x^{k+p} = x^k$, for all $p \in \mathbb{N}$.
Iterative step: Start with $x^0, x^1, \ldots, x^k$. Take x^{k+1} as any $x \in C$ such that there exists $\epsilon^k \in \mathbb{R}^m_+$ satisfying:

$$x \in S^{\omega}_{\epsilon^k q}(F_k, \Omega_k), \tag{4.1}$$

$$\epsilon^k \preceq \sigma_k \frac{\lambda_k}{2} d(x^k, x) \varsigma^k, \tag{4.2}$$

where $F_k(x) := F(x) + \frac{\lambda_k}{2} d^2(x^k, x)\varsigma^k$, $\Omega_k := \{x \in C \mid F(x) \preceq F(x^k)\}$, and $(\sigma_k)_k \subset [0, 1)$.

It is worth mentioning that for $\epsilon^k = 0$ we recover the proximal method proposed by Bento et al. in (2018). We note that some proposals for an inexact proximal point method for multiobjective optimization have also appeared, for example, in Souza (2018), Quiroz et al. (2020). In both references the authors consider a scalar iterative scheme and an approach fully based on scalarization to find possible weak Pareto of the original multiobjective problem.

We now make the following assumption on the map F which is common when dealing with non-convex problems:

$\mathcal{A}\mathbf{1}: \quad F \succ 0.$

There is no loss of generality in considering assumption $\mathcal{A}1$ since the vectorial functions $F(\cdot)$ and $e^{F(\cdot)} := (e^{f_1(\cdot)}, \ldots, e^{f_m(\cdot)})$ have the same set of Pareto critical points, where e^{α} denotes the exponential map valued at $\alpha \in \mathbb{R}$. This fact was first observed by Huang and Yang in (2004) and, in the Riemannian context, it can be verified in a similar way.

Proposition 6 *The Algorithm 1 is well defined.*

Proof From [Bento et al. (2018), Proposition 4.1], we have $S^{\omega}(F_k, \Omega_k) \neq \emptyset$. Therefore, the desired result follows from Remark 1 item ii). $\square$

Note that if Algorithm 1 terminates after a finite number of iterations, then it terminates at a critical Pareto point. From now on, we will assume that $(\lambda_k)_k$, $(\epsilon_k)_k$ and $(x^k)_k$ are infinite sequences generated by Algorithm 1.

4.2 Convergence analysis

The next result plays an important role in our subsequent considerations and our prove follows straight as application of Theorem 2.

Proposition 7 *For all $k \in \mathbb{N}$, there exist $\alpha^k, \beta^k \in \mathbb{R}^m_+$, $u_i^k, v^k, w^k \in T_{x^k}M$ and $\tau_k \in \mathbb{R}_{++}$ such that*

$$\sum_{i=1}^{m}(\alpha_i^k + \beta_i^k)u_i^k - \lambda_{k-1}\exp^{-1}_{x^k} x^{k-1}\sum_{i=1}^{m}\alpha_i^k\varsigma_i^{k-1} + v^k\sum_{i=1}^{m}\alpha_i^k\epsilon_i^{k-1} + \tau_k w^k = 0, \quad (4.3)$$

where $\sum_{i=1}^{m}(\alpha_i^k + \beta_i^k) = 1$, $v^k \in \mathbb{B}_{x^k}$, $u_i^k \in \partial f_i(x^k)$, and $w^k \in \partial d_C(x^k)$.

Proof Let us denote $F_{k-1}(x)$, $G_{k-1}(x)$ two functions defined by

$$F_{k-1}(x) := F(x) + \frac{\lambda_{k-1}}{2}d^2(x^{k-1}, x)\varsigma^{k-1} \quad \text{and} \quad G_{k-1}(x) := F(x) - F(x^{k-1}).$$

As F is a locally Lipschitz function, it follows that the coordinates functions

$$(F_{k-1})_i(\cdot) = f_i(\cdot) + \frac{\lambda_{k-1}}{2}d^2(x^{k-1}, \cdot)\varsigma_i^{k-1}, \quad \text{and} \quad (G_{k-1})_i(\cdot) = f_i(\cdot) - f_i(x^{k-1})$$

of F_{k-1} and G_{k-1}, respectively, are locally Lipschitz functions.

Therefore, since x^k is an ϵ^{k-1}-quasi-weak Pareto efficient for the problem min $F_{k-1}(x)$ s.t. $G_{k-1}(x) \preceq 0$, applying Theorem 2 with $H(\cdot) = F_{k-1}(\cdot)$ and $g_j(\cdot) = (G_{k-1})_j(\cdot)$, there exist real non-negative numbers α_i^k, β_i^k, $i = 1, 2, \ldots, m$ and a positive number τ_k, for every k, with

$$\sum_{i=1}^{m}(\alpha_i^k + \beta_i^k) = 1$$

and such that

$$0 \in \sum_{i=1}^{m}(\alpha_i^k + \beta_i^k)\partial f_i(x^k) - \lambda_{k-1}\exp_{x^k}^{-1} x^{k-1}\sum_{i=1}^{m}\alpha_i^k\varsigma_i^{k-1} + \sum_{i=1}^{m}\alpha_i^k\epsilon_i^{k-1}\mathbb{B}_{x^k} + \tau_k\partial d_C(x^k).$$

Thus, there exist $u_i^k \in \partial f_i(x^k)$, $v^k \in \mathbb{B}_{x^k}$ and $w^k \in \partial d_C(x^k)$ satisfying the desired result. □

As an immediate consequence of Proposition 7 we get the following stopping rule for Algorithm 1:

Corollary 1 *Let $k_0 \in \mathbb{N}$ be such that $\alpha^{k_0} = 0$. Then, x^{k_0} is a Pareto critical point of F.*

Proof If there exists $k_0 \in \mathbb{N}$ such that $\alpha^{k_0} = 0$ then, from (4.3), we have

$$\sum_{i=1}^{m}\beta_i^{k_0}u_i^{k_0} + \tau_{k_0}w^{k_0} = 0,$$

where $w^{k_0} \in \mathbb{B}_{x^{k_0}} \cap N_C(x^{k_0})$, $u_i^{k_0} \in \partial f_i(x^{k_0})$ and $\sum_{i=1}^{m}\beta_i^{k_0} = 1$ with $\beta^{k_0} \in \mathbb{R}_+^m$. Suppose that x^{k_0} is not a Pareto critical point of F. Then there exists $y \in C$ so that, for all $i \in \{1, 2, \ldots, m\}$ and $u_i^{k_0} \in \partial f_i(x^{k_0})$, $\langle u_i^{k_0}, \exp_{x^{k_0}}^{-1} y\rangle < 0$. Since, $\beta^{k_0} \in \mathbb{R}_+^m$, $\sum_{i=1}^{m}\beta_i^{k_0} = 1$ it holds

$$\sum_{i=1}^{m}\beta_i^{k_0}\langle u_i^{k_0}, \exp_{x^{k_0}}^{-1} y\rangle < 0,$$

which contradicts the fact that $-\sum_{i=1}^{m}\beta_i^{k_0}u_i^{k_0} \in N_C(x^{k_0})$. Hence, x^{k_0} is a Pareto critical point of F. □

Corollary 2 *If $x^{k+1} = x^k$, then x^k is a Pareto critical of F.*

Proof Suppose that $x^{k+1} = x^k$, then the error condition implies $\epsilon^k = 0$. Furthermore, from expression (4.3) it follows

$$\sum_{i=1}^{m}(\alpha_i^{k+1} + \beta_i^{k+1})u_i^{k+1} + \tau_{k+1}w^{k+1} = 0,$$

i.e., $-\sum_{i=1}^{m}(\alpha_i^{k+1} + \beta_i^{k+1})u_i^{k+1} \in N_C(x^{k+1})$, and the desired result follows by using arguments similar to those used in the proof of Corollary 1. □

Remark 4 Since the sequences $(\lambda_k)_k$, $(\epsilon_k)_k$ and $(x^k)_k$ were assumed to be infinite sequences, then $\alpha^k \neq 0$ and $x^{k+1} \neq x^k$ for all $k \in \mathbb{N}$, in view of Corollary 1 and 2, respectively.

In the sequel, we present and prove the main theorem of the section.

Theorem 3 *Assume that $\mathcal{A}\mathbf{1}$ holds and there exist scalars $a_1, a_2, a_3, a_4 \in \mathbb{R}_{++}$ such that $a_1 \leq \lambda_k \leq a_2$, $a_3 \leq \varsigma_i^k$ and $\sigma_k \leq a_4 < 1$, for all $k \in \mathbb{N}$ and $i = 1, \ldots, m$. Then, every cluster point of $(x^k)_k$, if any, is a Pareto critical point of F.*

Proof Since $x^{k+1} \in S_{\epsilon^k q}^{\omega}(F_k, \Omega_k)$ (this follows from (4.1)) there does not exist another decision point $x \in \Omega_k$ such that $(F_k)_i(x) + \epsilon_i^k d(x, x^{k+1}) < (F_k)_i(x^{k+1})$, for $i = 1, 2, \ldots, m$. Hence, for any k, there exists some index $i := i(k)$ such that $(F_k)_i(x^{k+1}) \leq (F_k)_i(x^k) + \epsilon_i^k d(x^k, x^{k+1})$ and, using the definition of F_k, the last inequality takes the following form

$$f_i(x^{k+1}) + \frac{\lambda_k}{2}d^2(x^k, x^{k+1})\varsigma_i^k \leq f_i(x^k) + \epsilon_i^k d(x^k, x^{k+1}).$$

Now, after some algebra, we obtain

$$\frac{\lambda_k}{2} d^2(x^k, x^{k+1}) \varsigma_i^k - \epsilon_i^k d(x^k, x^{k+1}) \leq \|F(x^k) - F(x^{k+1})\|,$$

which, combined with (4.2) yields

$$(1 - \sigma_k) \frac{\lambda_k}{2} d^2(x^k, x^{k+1}) \varsigma_i^k \leq \|F(x^k) - F(x^{k+1})\|.$$

From the iterative step (4.1) combined with the definition of Ω_k, it follows that $(F(x^k))_k$ is nonincreasing. Hence, by assumption $\mathcal{A}\mathbf{1}$ and taking into account that $0 < \sigma_k \leq a_4 < 1$ and $(\lambda_k)_k$, $(\varsigma^k)_k$ are bounded sequences, we conclude that $(d(x^k, x^{k+1}))_k$ converges to zero as k goes to infinity.

Take $\hat{x}$ a cluster point of $(x^k)_k$. Thus, there exists $\mathbb{K} \subset \mathbb{N}$ such that the sequence $(x^k)_{k \in \mathbb{K}}$ converges to $\hat{x}$ as k goes to infinity. Therefore, applying Proposition 7 for every $k \in \mathbb{K}$ we have existence of sequences

$$\begin{cases} (u_i^k)_k \subset \partial f_i(x^k), \quad i \in \{1, 2, \dots, m\}; \\ (\alpha^k)_k, \ (\beta^k)_k \subset \mathbb{R}_+^m; \quad (\tau_k)_k \subset \mathbb{R}_{++}; \\ (w^k)_k \subset T_{x^k} M \quad ; \quad (v^k)_k \in \mathbb{B}_{x^k}, \end{cases}$$

satisfying

$$w^k \in \mathbb{B}_{x^k} \cap N_C(x^k), \quad \sum_{i=1}^{m} (\alpha_i^k + \beta_i^k) = 1$$

and

$$\sum_{i=1}^{m} (\alpha_i^k + \beta_i^k) u_i^k - \lambda_{k-1} \exp_{x^k}^{-1} x^{k-1} \sum_{i=1}^{m} \alpha_i^k \varsigma_i^{k-1} + v^k \sum_{i=1}^{m} \alpha_i^k \epsilon_i^{k-1} + \tau_k w^k = 0. \quad (4.4)$$

From the convergence of $(x^k)_{k \in \mathbb{K}}$ we obtain boundedness. Using the locally Lipschitz continuity of F, Proposition 3 and previous conditions, it follows that sequences $(u_i^k)_k$, $(v^k)_k$, $(\alpha^k)_k$, $(\beta^k)_k$, $(w^k)_k$ are bounded. Thus, equality (4.4) implies that $(\tau_k)_k$ is also bounded. In this case, we can assume, without loss of generality, that these sequences converge to u_i, v, α, β, w and τ (take subsequences, if necessary), respectively. Note also that the sequence $(\lambda_k \sum_{i=1}^m \alpha_i^k \varsigma_i^k)_k$ is bounded and, letting k goes to infinity in (4.4), we obtain

$$\sum_{i=1}^{m} (\alpha_i + \beta_i) u_i + \tau w = 0, \quad (4.5)$$

because $(\epsilon^k)_k$ converges to zero due condition (4.2) combined with the fact that $(d(x^k, x^{k+1}))_k$ converges to zero as k goes to infinity. Since $w \in \mathbb{B}_{\hat{x}} \cap N_C(\hat{x})$, from (4.5) it follows that

$$-\sum_{i=1}^{m} (\alpha_i + \beta_i) u_i \in \mathbb{B}_{\hat{x}} \cap N_C(\hat{x}). \quad (4.6)$$

Assume by contradiction that $\hat{x}$ is not a Pareto critical point of F. Then, there is $y \in C$ such that for all $i \in \{1, 2, \cdots, m\}$ and $u_i \in \partial f_i(\hat{x})$ it holds $\langle u_i, \exp_{\hat{x}}^{-1} y \rangle < 0$. Therefore,

$$\sum_{i=1}^{m} (\alpha_i + \beta_i) \langle u_i, \exp_{\hat{x}}^{-1} y \rangle < 0,$$

because $\alpha, \beta \in \mathbb{R}^m_+$ and $\sum_{i=1}^m (\alpha_i + \beta_i) = 1$. But this contradicts (4.6) and the desired result is proved. □

4.3 Full convergence

In this section, under the convexity assumption on F and assuming that M is a Hadamard manifold, we establish the full convergence of the inexact proximal method point.

The function F is said to be convex on C iff for every $x, y \in C$

$$F(\gamma(t)) \preceq (1-t)F(x) + tF(y), \qquad t \in [0,1],$$

where $\gamma : [0,1] \to C$ is the geodesic segment joining x to y. In the particular case where $C = M$, this definition coincides with the notion of convexity introduced in Bento et al. (2012). In the particular case where F is convex, it is known that the limiting subdifferential coincides with the classical subdifferential for convex analysis and that criticality is equivalent to weak optimality; see Bento et al. (2018).

We now make the following assumption:

$\mathcal{A}\mathbf{2} : \bar{\Omega} := \cap_{k=0}^{+\infty} \Omega_k \neq \emptyset$.

The assumption $\mathcal{A}\mathbf{2}$ is naturally verified when assuming the known $\mathbb{R}^m_+$ - completeness on $(F(x^0) - \mathbb{R}^m_+) \cap F(C)$, for each $x^0 \in C$ [see (Luc, (1989), Section 19]), which was also made in various studies on proximal algorithms; see, for instance, (Bonnel et al., 2005; Bento et al., 2018; Ceng & Yao, 2007 for references in the linear setting and (Bento et al., 2018, 2018b) for references in the Riemannian context.

Theorem 4 *Let M be a Hadamard manifold. If $F : M \to \mathbb{R}^m$ is convex, $\sum_{k=0}^{\infty} \sigma_k < +\infty$ and assumption $\mathcal{A}\mathbf{2}$ holds, then, for all $x^* \in U$:*

(i) $d^2(x^*, x^{k+1}) \leq d^2(x^k, x^*) + \delta_k$ *where* $\delta_k > 0$ *and* $\sum_{k=1}^{\infty} \delta_k < +\infty$;
(ii) $(x^k)_k$ *converges to a weak Pareto efficient solution of* F.

Proof Take $x^* \in \bar{\Omega}$ (this is possible from the assumption $\mathcal{A}\mathbf{2}$). Consider the geodesic triangle $\triangle(x^k x^{k+1} x^*)$. By the law of cosines, we have

$$d^2(x^k, x^{k+1}) + d^2(x^*, x^{k+1}) - 2\langle \exp^{-1}_{x^{k+1}} x^k, \exp^{-1}_{x^{k+1}} x^* \rangle \leq d^2(x^k, x^*). \tag{4.7}$$

On the other hand, Proposition 7 tells us that for each $k \in \mathbb{N}$, there exist

$$\begin{cases} \alpha^{k+1}, \beta^{k+1} \in \mathbb{R}^m_+; \quad \tau_{k+1} \in \mathbb{R}_{++}; \quad u_i^{k+1} \in \partial f_i(x^{k+1}); \\ v^{k+1} \in \mathbb{B}_{x^{k+1}} \ \text{ and } \ w^{k+1} \in \mathbb{B}_{x^{k+1}} \cap N_C(x^{k+1}), \end{cases}$$

which verify the equality

$$\sum_{i=1}^m (\alpha_i^{k+1} + \beta_i^{k+1}) u_i^{k+1} - \lambda_k \exp^{-1}_{x^{k+1}} x^k \sum_{i=1}^m \alpha_i^{k+1} \varsigma_i^k + v^{k+1} \sum_{i=1}^m \alpha_i^{k+1} \epsilon_i^k + \tau_{k+1} w^{k+1} = 0.$$

Hence, this last identity combined with inequality (4.7) yields

$$\lambda_k \sum_{i=1}^m \alpha_i^{k+1} \varsigma_i^k \left[d^2(x^k, x^*) - d^2(x^k, x^{k+1}) - d^2(x^*, x^{k+1}) \right] \geq -2\tau_{k+1} \langle w^{k+1}, \exp^{-1}_{x^{k+1}} x^* \rangle$$
$$-2 \sum_{i=1}^m (\alpha_i^{k+1} + \beta_i^{k+1}) \langle u_i^{k+1}, \exp^{-1}_{x^{k+1}} x^* \rangle - 2 \sum_{i=1}^m \alpha_i^{k+1} \epsilon_i^k \langle v^{k+1}, \exp^{-1}_{x^{k+1}} x^* \rangle.$$

The characterization of normal cone, i.e., $\langle w^{k+1}, \exp^{-1}_{x^{k+1}} x^*\rangle \leq 0$, jointly with the convexity of F, $x^* \in U$ and $u_i^{k+1} \in \partial f_i(x^{k+1})$ imply

$$\lambda_k b_k \left[d^2(x^k, x^*) - d^2(x^k, x^{k+1}) - d^2(x^*, x^{k+1})\right] \geq -2\sum_{i=1}^{m} \alpha_i^{k+1}\epsilon_i^k \langle v^{k+1}, \exp^{-1}_{x^{k+1}} x^*\rangle,$$

where $b_k = \sum_{i=1}^m \alpha_i^{k+1}\varsigma_i^k$. Since $r + s \geq 2\sqrt{rs}$ holds for $r, s \geq 0$, taking $s := d^2(x^k, x^{k+1})$, $r := d^2(x, x^{k+1})$, after some algebraic manipulations and using (4.2), we get:

$$d^2(x^k, x^*) - d^2(x^k, x^{k+1}) - d^2(x^*, x^{k+1}) \geq -\frac{\sigma_k}{2}\left[d^2(x^k, x^{k+1}) + d^2(x^*, x^{k+1})\right].$$

Thus,

$$d^2(x^*, x^{k+1}) \leq \left(1 + \frac{\sigma_k}{2-\sigma_k}\right) d^2(x^k, x^*) - d^2(x^k, x^{k+1}). \tag{4.8}$$

Since $\sum_{k=0}^{\infty} \sigma_k < \infty$, it follows that

$$K_0 := \sum_{k=k_0}^{\infty} \frac{\sigma_k}{2-\sigma_k} < \infty \quad \text{and} \quad K_1 := \prod_{j=k_0}^{\infty}\left(1 + \frac{\sigma_j}{2-\sigma_j}\right) < \infty.$$

By (4.8), observe that for all $k \geq k_0$

$$\begin{aligned}
d^2(x^*, x^{k+1}) &\leq \left(1 + \frac{\sigma_k}{2-\sigma_k}\right) d^2(x^k, x^*) \\
&\leq \left(1 + \frac{\sigma_{k-1}}{2-\sigma_{k-1}}\right)\left(1 + \frac{\sigma_k}{2-\sigma_k}\right) d^2(, x^{k-1}, x^*) \\
&\vdots \\
&\leq \prod_{j=k_0}^{k}\left(1 + \frac{\sigma_j}{2-\sigma_j}\right) d^2(x^{k_0}, x^*) \\
&\leq \prod_{j=k_0}^{\infty}\left(1 + \frac{\sigma_j}{2-\sigma_j}\right) d^2(x^{k_0}, x^*)
\end{aligned}$$

This shows that (x^k) is bounded. Set $K = \sup_k d(x^k, x^*)$. Then again from (4.8), we obtain

$$d^2(x^*, x^{k+1}) \leq d^2(x^k, x^*) + \frac{\sigma_k}{2-\sigma_k}K^2, \tag{4.9}$$

and the item (i) follows for considering $\delta_k = \dfrac{\sigma_k}{2-\sigma_k}K^2$.

Let us now prove item (ii). Since $(x^k)_k$ is bounded, from the Hopf-Rinow's Theorem this sequence has some accumulation point $\bar{x} \in M$. It follows from the definition of the iterative step in (4.1) that $F(x^{k+1}) \preceq F(x^k)$ for all k and, consequently, we can conclude that $\bar{x} \in \bar{\Omega}$. Given $\eta > 0$, there exists $k_1 \in \mathbb{N}$ such that

$$\sum_{k=k_1}^{\infty} \delta_k < \frac{1}{2}\eta^2 \quad \text{and} \quad d^2(x^{k_1}, \bar{x}) < \frac{1}{2}\eta^2. \tag{4.10}$$

For $k > k_1$, using (4.9) it yields

$$d^2(x^k, \bar{x}) \leq d^2(x^{k_1}, \bar{x}) + \sum_{j=k_1}^{\infty} \delta_k.$$

Combining the inequality above with expressions in (4.10) we have $d(x^k, \bar{x}) < \eta$, for all $k > k_1$. Therefore, the item (ii) follows as consequence of Theorem 3 jointly with the assumption of convexity of the function F. $\square$

Next it is considered an ilustrative example for Algorithm 1 in Poincaré plane.

Example 2 Let us consider $\mathbb{H}^2 = (\mathcal{H}, \langle\langle\ ,\ \rangle\rangle)$, with $\langle\langle u, v\rangle\rangle_p := v^T \mathcal{G}(p)u$, $F := H : \mathcal{H} \to \mathbb{R}^2$ given by $F(p) = (y, x^2 y^{-1} + 1)$ and $C = \{(x, y) \in \mathbb{H}^2 : x \geq 1, y \geq 1\}$ as introduced in Example 1. Thus, $\mathbb{H}^2 = (\mathcal{H}, \langle\langle\ ,\ \rangle\rangle)$ is a complete Riemannian manifold of sectional curvature -1, called the Poincaré plane, the coordinate functions of F are locally Lipschitz, but not convex functions, on $\mathbb{H}^2$ and C is a convex set. Take $\varsigma^k = (1, 1)$, for all $k \in \{1, 2, \ldots\}$ and $(x^0, y^0) \in C$. If $(x^k, y^k)_k$ is a sequence generated from Algorithm 1 with $x^0, y^0 > 1$, then (x^{k+1}, y^{k+1}) is determined as follows:

$$(x^{k+1}, y^{k+1}) \in S^{\omega}_{\epsilon^k q}(F_k, \Omega_k),$$
$$\epsilon^k \preceq \sigma_k \frac{\lambda_k}{2} d((x^k, y^k), (x^{k+1}, y^{k+1}))\varsigma^k,$$

where $\Omega_k := \{x \in C \mid G_k(x) \preceq 0\}$, $G_k(x, y) := \Big(y - y^k, x^2 y^{-1} - (x^k)^2 (y^k)^{-1}\Big)$, $(\sigma_k)_k \subset [0, 1)$ and

$$F_k(x, y) := \Big(f_1(x, y) + \frac{\lambda_k}{2} d^2_{(x^k, y^k)}((x, y)),\ f_2(x, y) + \frac{\lambda_k}{2} d^2_{(x^k, y^k)}((x, y))\Big).$$

From Proposition 7, there exist $\alpha_i^{k+1}, \beta_i^{k+1} \in \mathbb{R}_+$, $i = 1, 2$, $w^{k+1} \in \partial d_C(x^{k+1}, y^{k+1})$ and $\tau_{k+1} > 0$ such that $\sum_{i=1}^2 (\alpha_i^{k+1} + \beta_i^{k+1}) = 1$ and

$$\sum_{i=1}^{2} \Big(\alpha_i^{k+1} + \beta_i^{k+1}\Big) \operatorname{grad} f_i(x^{k+1}, y^{k+1}) +$$
$$-\lambda_k \exp^{-1}_{(x^{k+1}, y^{k+1})}(x^k, y^k) \sum_{i=1}^{2} \alpha_i^{k+1} - v^{k+1} \sum_{i=1}^{m} \alpha_i^{k+1} \epsilon_i^k + \tau_{k+1} w^{k+1} = 0,$$

where $v^{k+1} \in \mathbb{B}_{(x^{k+1}, y^{k+1})}$. Note that

$$\partial d_C(x, y) = \begin{cases} \{(w, t) \in \mathbb{R}^2_- : w^2 + t^2 \leq 1\}, & x = y = 1, \\ \{(w, 0) : -1 \leq w \leq 0\}, & x = 1,\ y > 1, \\ \{(0, 0)\}, & x > 1,\ \ y \geq 1. \end{cases}$$

Due to the vector improvement steps, characterized by the constraint set Ω_k, it follows that $(F(x^k, y^k))_k$ is decreasing. For each $(x^0, y^0) \in C$ fixed Ω_k is bounded. Besides, $\mathcal{A}1$ holds and, hence, for $\sigma_k \leq \lambda_k := 1/2$ Theorem 3 implies that each cluster point of $(x^k, y^k)_k$ is a Pareto critical point of F in C.

5 Application: weak pareto points as limits of traps in behavioral sciences

At the application level the main message of this paper is that weak Pareto points of a multidimensional optimization problem can be reached as limits of traps of a perturbed multiobjective proximal point algorithm.

5.1 The variational rationality approach

The concept of trap is one of the building block of the recent (VR) Variational rationality approach of stay and change human dynamics; see Soubeyran (2009, 2010, 2021a, b, c, d). This recent theory provides, in the context of Variational analysis in mathematics, a general and formalized reformulation of the theory of motivation, emotion and behavior in psychology. The general problem of motivation theory is why, how and when individuals do what they do. Why, how and when, they stop, continue and start doing things each day of their lives. Why, how and when they disengage, reengage or engage in different goals and activities. This new approach is driven by only one concept. This is the concept of worthwhile move which provides a new and generalized formulation of sufficient descent conditions in a lot of different and recent optimizing algorithms including proximal algorithms. Thus, the VR approach is, both, able to illustrate and to generalize almost all of the main concepts of variational analysis in mathematics and, to provide a general and mathematical theory of motivation, emotion and behavior in psychology.

5.1.1 Making a worthwhile move

To save a lot of space, and to directly meet the concept of worthwhile move, we start our presentation starting from the proof of Theorem 3 which is one of the main result of this paper. For each k, it exists $i = i(k)$ such that

$$f_i(x^k) - f_i(x^{k+1}) \geq (\lambda_k/2)d^2(x^k, x^{k+1})\zeta_i^k - \epsilon_i^k d(x^k, x^{k+1}). \tag{5.1}$$

Let $\tilde{g}_i(x)$ be the utility of member $i \in \mathcal{I}$ of an organization or group of individuals $\mathcal{I} := \{1, 2, \ldots, m\}$. Given that $x \in X = M$ is a vector of collective activities done by all members of this group, let $\tilde{g}_i^* = \sup\{\tilde{g}_i(x), x \in X\} < +\infty$ be the aspiration level of member $i \in \mathcal{I}$. That is, the highest level of utility he can hope to get as a member of the group. Then, the difference $f_i(x) = \tilde{g}_i^* - \tilde{g}_i(x) \geq 0$ models the frustration, or unsatisfacton feeling, of member i, relative to the collective action x. Suppose that this organization moves from having done the collective action x^k in the previous period k to doing the collective action x^{k+1} in the current period $k + 1$. Then, given that $f_i(x) - f_i(y) = \tilde{g}_i(y) - \tilde{g}_i(x)$, inequality in (5.1) becomes:

$$\tilde{g}_i(x^{k+1}) - \tilde{g}_i(x^k) \geq (\lambda_k/2)d^2(x^k, x^{k+1})\zeta_i^k - \epsilon_i^k d(x^k, x^{k+1}). \tag{5.2}$$

Note that the inequality in (5.2) is a direct consequence of the condition (4.1), while condition (4.2) requires that

$$\epsilon_i^k \leq \sigma_k(\lambda_k/2)d(x^k, x^{k+1})\zeta_i^k, \qquad i \in \mathcal{I},$$

i.e., $\epsilon_i^k d(x^k, x^{k+1}) \leq \sigma_k(\lambda_k/2)d^2(x^k, x^{k+1})\zeta_i^k$. This implies the fundamental condition

$$f_i(x^k) - f_i(x^{k+1}) = \tilde{g}_i(x^{k+1}) - \tilde{g}_i(x^k) \geq (1 - \sigma_k)(\lambda_k/2)d^2(x^k, x^{k+1})\zeta_i^k. \tag{5.3}$$

5.1.2 A simple VR structure

Using the (VR) variational rationality approach, we are now in a good position to gives a clear interpretation of condition (5.3) in term of a worthwhile move. The VR approach compares, each current period $k+1$, two alternatives: to stay or to change, i.e.,

(1) to stay at the status quo x^k, making the stay $x^k \curvearrowright x^{k+1} = x^k$;
(2) to change, making the change $x^k \curvearrowright x^{k+1}$, $x^{k+1} \neq x^k$.

Then, the VR approach defines the following list of concepts relative to a move $m^k = x^k \curvearrowright x^{k+1}$:

(a) advantages to move (change rather than stay) are, for each member i of the group, $A_i(x^k, x^{k+1}) = \tilde{g}_i(x^{k+1}) - \tilde{g}_i(x^k) = f_i(x^k) - f_i(x^{k+1})$. These advantages refer to the difference between the utility to change from x^k to x^{k+1} and the utility to stay at x^k;
(b) inconveniences to move $I_i(x^k, x^{k+1}) = C_i(x^k, x^{k+1}) - C_i(x^k, x^k) \geq 0$ represent the difference between costs to change $C_i(x^k, x^{k+1})$ and costs to stay $C_i(x^k, x^k)$. The construction of such costs to move is complex and requires a lot of justifications; see Soubeyran (2021a, b, c, d). Costs to move are not symmetric. That is, $C_i(y, x) \neq C_i(x, y)$. In the present paper inconveniences to move $I_i(x^k, x^{k+1})$ are higher or proportional to the Riemannian distance between x^k and x^{k+1}, namely, $d(x^k, x^{k+1})$. Then, we will suppose that $I_i(x^k, x^{k+1}) \geq \zeta_i^k d(x^k, x^{k+1})$, for all $k \in \mathbb{N}$. This means that inconveniences to move must be high enough in the large, i.e., when the length $d(x^k, x^{k+1})$ of moves are high enough;
(c) motivation to move is $\mathcal{M}_i(x^k, x^{k+1}) = U_i\left[A_i(x^k, x^{k+1})\right]$, where $U_i\left[A_i\right]$ is the utility of advantages to move;
(d) resistance to move is $R_i(x^k, x^{k+1}) = D_i\left[I_i(x^k, x^{k+1})\right]$, where $D_i\left[I_i\right]$ is the disutility of inconveniences to move. Resistance to move is strong when $D_i\left[I_i\right] = I_i^{\beta}$, $0 < \beta \leq 1$. It is weak when $\beta > 1$.
(e) a worthwhile balance $B_i(x^k, x^{k+1}) = \mathcal{M}_i(x^k, x^{k+1}) - \xi_i R_i(x^k, x^{k+1})$ is a weighted difference between motivation and resistance to move. The importance that member i gives to resistance to move is the weight $\xi_i > 0$.

Then, for member i, a move is worthwhile if $B_i(x^k, x^{k+1}) \geq 0$. This condition means that motivation to move is high enough relative to resistance to move. Thus, condition in (5.3) defines a worthwhile move.

This paper supposes implicitly a linear -quadratic variational structure where $U_i\left[A_i\right] = A_i$ and $D_i\left[I_i\right] = I_i^2$ for all $A_i, I_i \in \mathbb{R}_+$.

5.1.3 Worthwhile stop and go group dynamics

Starting from the status quo x^0, these group dynamics can be:

(a) over one period (the current period),

- A collective stay at a weak desired collective end $x^* = x^0 \in X$ (collective desire) if there does not exist an other collective action $y \in X$ such that $A_i(x^*, y) > 0$ for all $i \in \mathcal{I}$;
- A collective stationary trap $x^* = x^0 \in$ if there does not exist an other collective action $y \in X$ such that $B_i(x^*, y) > 0$ for all $i \in \mathcal{I}$;
- A collective worthwhile move going from x^0 to y if $B_i(x^0, y) \geq 0$ for all $i \in \mathcal{I}$.

(b) over two periods,

- A collective variational trap $x^* \neq x^0 \in X$ if, starting from $x^0 \in X$,
  (1) $B_i(x^*, x^0) \geq 0$ for all $i \in I$ in the current period and, i.e., it is collectively worthwhile to change from x^0 to x^* for all members of the group;
  (2) x^* is a collective stationary trap in the future period, i.e., it is not worthwhile to move from x^*.

5.2 Interpretations of the main concepts and results

5.2.1 ϵ-quasi-weakly Pareto solutions as global traps

In the context of the VR approach, the concept of ϵ-quasi-weakly Pareto solution introduced in Definition 2 models a trap when resistance to move is strong ($\beta = 1$, see above the definition of resistance to move). That is, there is no way to make a collective worthwhile move $x^* \curvearrowright x$, $x \neq x^*$ such that $h_i(x^*) - h_i(x) > \epsilon_i d(x, x^*)$, for all $x \neq x^*$, $i \in \mathcal{I}$, where $\epsilon := (\epsilon_1, \ldots, \epsilon_m) \in \mathbb{R}^m_+$ is taken a priori.

5.2.2 Succession of traps of successive perturbation functions

Each period, the proximal point algorithm introduced in Sect. 4 considers functions $h_i(\cdot) = (F_k)_i(\cdot) = f_i(\cdot) + (\lambda_k/2)d^2(x^k, \cdot)\zeta_i^k$. They can be seen as perturbations of the initial functions $f_i(\cdot)$, $i \in \mathcal{I}$. This means that, each period, (4.1) models a succession of moves $m^k : x^k \curvearrowright x^{k+1}$ going from a trap x^k to a next trap relative the vectorial perturbation function $(F_k)(\cdot)$. This algorithm is a specific instance of the algorithm given by inequality in (5.3). This last algorithm defines a succession of worthwhile moves, for at least, each period, one member of the group $i = i(k)$. Condition $x^k \in \Omega_k$ requires that, each period, the payoff $f_i(\cdot)$ of any other member of the group $i \neq i(k)$ improves or be the same as before. Condition (4.2) requires that each worthwhile move must be large enough to accelerate convergence: $d(x^k, x^{k+1})$ large enough.

5.2.3 Succession of large enough worthwhile moves

The main result of Sect. 4, namely, Theorem 3, shows that every cluster point of a succession of large enough worthwhile moves (see (4.1) and (4.2)) is a Pareto critical point of the initial function $F(\cdot)$. As noted earlier, full convergence to a weak Pareto optimum of $F(\cdot)$ is obtained in Theorem 4 when the problem is assumed to be convex.

5.2.4 An illustration of Lewin's change management model

Moving from one trap to an other one refers to the famous Lewin's unfreezing, change, refreezing model of change; see Lewin (1952, 1959). Lewin believed a successful change project involved three steps:

1. Unfreezing. For Lewin, human behavior was based on a quasi-stationary equilibrium supported by a complex field of forces. Before old behavior can be discarded (unlearnt) and new behavior successfully adopted, the equilibrium needs to be destabilized (unfrozen).

This first stage of change involves preparing the organization to accept that change is

necessary, which involves breaking down the existing status quo before you can build up a new way of operating;
2. Change. After the uncertainty created in the unfreeze stage, the change stage is where people begin to resolve their uncertainty and look for new ways to do things. People start to believe and act in ways that support the new direction;
3. Refreezing. This seeks to stabilize the group at a new quasi-stationary equilibrium in order to ensure that the new behaviors are relatively safe from regression.

It is worth mentioning that the present paper and Bento et al. (2021) provide two variants of such Lewin's model when resistance to move is weak, with two different formulations of resistance to move. Then, both papers give an application to the theories of collective desires, showing how a group must escape to a succession of temporary traps to be able to reach, at the end, his desires.

6 Conclusions

We have given a definition of approximate Pareto efficient solution and a necessary condition for such solutions in multiobjective optimization on Riemannian manifolds. We also propose an inexact version of the algorithm given by Bento et al. (2018) using the notion of approximate solution. We have presented a convergence analysis, which proves that every accumulation point, if any, is a critical Pareto of the problem. Furthermore, under the assumption of convexity one has convergence to weakly Pareto of the problem. As an application, we show how a Pareto critical point can be reached as a limit of traps in the context of the variational rationality approach of stay and change human dynamics.

As future perspectives we intend to extend our convergence analysis by considering the subproblem of the iterative process presented in Bento et al. (2018) regularized by a Riemannian version of the proximal distance introduced in Auslender and Teboulle (2006). In the linear context an approach of the proximal method for multiobjective optimization with proximal distance was considered in Bento et al. (2018a).

Acknowledgements The authors was supported in part by CAPES, FAPEG/PRONEM- (grants 201710267000532) and CNPq (grants 308330/2018-8, 314106/2020-0).

References

Apolinário, H. C. F., Papa Quiroz, E. A., & Oliveira, P. R. (2016). A scalarization proximal point method for quasiconvex multiobjective minimization. *Journal of Global Optimization, 64,* 79–96.

Auslender, A., & Teboulle, M. (2006). Interior gradient and proximal methods for convex and conic optimization. *SIAM Journal on Optimization, 16*(3), 697–725.

Bento, G. C., Ferreira, O. P., & Oliveira, P. R. (2012). Unconstrained steepest descent method for multicriteria optimization on Riemannian manifolds. *Journal of Optimization Theory and Applications, 154*(1), 88–107.

Bento, G. C., & Cruz Neto, J. X. (2013). A subgradient method for multiobjective optimization on Riemannian manifolds. *Journal of Optimization Theory and Applications, 159*(1), 125–137.

Bento, G. C., Cruz Neto, J. X., & Soubeyran, A. (2014). A proximal point-type method for multicriteria optimization. *Set-Valued and Variational Analysis, 22,* 557–573.

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

Bento, G., Cruz Neto, J. X., & Meireles, L. V. (2018). Proximal point method for locally Lipschitz functions in multiobjective optimization of Hadamard manifolds. *Journal of Optimization Theory and Applications, 179*(1), 37–52.

Bento, G. C., Neto, J. C., Soares, P. A., & Soubeyran, A. (2021). A new regularization of equilibrium problems on Hadamard manifolds: Applications to theories of desires. *Annals of Operations Research*. https://doi.org/10.1007/s10479-021-04052-w.

Bento, G. C., Cruz Neto, J. X., Soubeyran, A., & Sousa Júnior, V. L. (2018). Inexact multi-objective local search proximal algorithms: Application to group dynamic and distributive justice problems. *Journal of Optimization Theory and Applications, 177*(1), 181–200.

Bento, G. C., Ferreira, O. P., & Pereira, Y. R. L. (2018). Proximal point method for vector optimization on Hadamard manifolds. *Operations Research Letters, 46*(1), 13–18.

Bonnel, H., Iusem, A. N., & Svaiter, B. F. (2005). Proximal methods in vector optimization. *SIAM Journal on Optimization, 15*(4), 953–970.

Ceng, L. C., & Yao, J. C. (2007). Approximate proximal methods in vector optimization. *European Journal of Operational Research, 183*(1), 1–19.

Ceng, L. C., Mordukhovich, B. S., & Yao, J. C. (2010). Hybrid approximate proximal method with auxiliary variational inequality for vector optimization. *Journal of Optimization Theory and Applications, 146,* 267–303.

Chen, G. Y., Huang, X., & Yang, X. (2005). *Vector optimization: Set-valued and variational analysis* Lecture notes in economics and mathematical systems (1st ed., Vol. 541). Springer-Verlag.

Choung, T. D., Mordukhovich, B. S., & Yao, J. C. (2011). Hybrid approximate proximal algorithms for efficient solutions in vector optimization. *Journal of Nonlinear and Convex Analysis, 12,* 257–286.

Chuong, T. D., & Kim, D. S. (2016). Approximate solutions of multiobjective optimization problems. *Positivity, 20*(1), 187–207.

Carmo, Do., & M.P.a. (1992). *Riemannian geometry*. Mathematics: Theory & applications. Birkhäuser Boston Inc.

Gregório, R., & Oliveira, P. R. (2011). A logarithmic-quadratic proximal point scalarization method for multiobjective programming. *Journal of Global Optimization, 49,* 281–291.

Huang, X. X., & Yang, X. Q. (2004). Duality for multiobjective optimization via nonlinear Lagrangian functions. *Journal of Optimization Theory and Applications, 120*(1), 111–127.

Jahn, J. (2004). *Vector optimization: Theory, applications, and extensions*. Springer.

Ledyaev, Y. S., & Zhu, Q. J. (2007). Nonsmooth analysis on smooth manifolds. *Transactions of the American Mathematical Society, 359*(8), 3687–3732.

Lewin, K. (1952). Frontiers in group dynamics. In D. Cartwright (Ed.), *(1947) and Field Theory in social Science*. Social Science Paperbacks.

Lewin, K. (1959). Group decisions and social change. In T. M. Newcomb & E. L. Hartley (Eds.), *Reading in social psychology*. Henry Holt.

Li, C., Mordukhovich, B. S., Wang, J., & Yao, J. C. (2011). Weak sharp minima on Riemannian manifolds. *SIAM Journal on Optimization, 21*(4), 1523–1560.

Loridan, P. (1984). ϵ-solutions in vector minimization problems. *Journal of Optimization Theory and Applications, 43*(2), 265–276.

Luc, D. T. (1989). *Theory of vector optimization lecture notes in economics and mathematical systems* (Vol. 319). Springer.

Meireles, L. V. (2019). Proximal point methods for multiobjective optimization in riemann- ian manifolds (PhD Thesis),

Minami, M. (1983). Weak Pareto-optimal necessary conditions in a nondifferentiable multiobjective program on a Banach space. *Journal of Optimization Theory and Applications, 41,* 451–461.

Quiroz, E. A. P., Cusihuallpa, N. B., & Maculan, N. (2020). Inexact proximal point methods for multiobjective quasiconvex minimization on hadamard manifolds. *Journal of Optimization Theory and Applications, 186*(3), 879–898.

Rockafellar, R. T. (1976). Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization, 14*(5), 877–898.

Sakai, T. (1996). *Riemannian geometry, translations of mathematical monographs* (Vol. 149). American Mathematical Society.

Solodov, M. V., & Svaiter, B. F. (1999). An inexact hybrid extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Analysis, 7,* 323–345.

Soubeyran, A. (2009). *Variational rationality, a theory of individual stability and change: worthwhile and ambidextry behaviors*. Preprint. GREQAM: Aix Marseille University.

Soubeyran, A. (2010). *Variational rationality and the "unsatisfied man": routines and the course pursuit between aspirations, capabilities and beliefs*. Preprint GREQAM: Aix Marseille University.

Soubeyran, A. (2021). *Variational rationality: towards a grand theory of motivation driven by worthwhile moves*. Preprint: GREQAM-AMSE, Aix Marseille University.
Soubeyran, A. (2021). *Variational rationality: The concepts of motivation and motivational force*. Aix-Marseille University.
Soubeyran, A., (2021). Variational rationality. The resolution of goal conflicts via stop and go approach-avoidance dynamics. Preprint. AMSE, Aix-Marseille University
Soubeyran, A., (2021). Variational rationality. A general theory of moving goals and intentions as satisficing worthwhile moves. Preprint. AMSE, Aix-Marseille University
Souza, J. C. O. (2018). Proximal point methods for Lipschitz functions on Hadamard manifolds: scalar and vectorial cases. *Journal of Optimization Theory and Applications, 179*(3), 745–760.
Tang, F. M., & Huang, P. L. (2017). On the convergence rate of a proximal point algorithm for vector function on Hadamard manifolds. *Journal of the Operations Research Society of China, 5,* 405–417.
Udriste, C. (1994). *Convex functions and optimization methods on riemannian manifolds, mathematics and its applications* (Vol. 297). Kluwer Academic Publishers Group.
Villacorta, K. D. V., & Oliveira, P. R. (2011). An interior proximal method in vector optimization. *European Journal of Operational Research, 214,* 485–492. Vinter, R. B. (2000). *Optimal control*. Birkhauser.