The proximal point method for locally Lipschitz functions in multiobjective optimization

G. C. Bento · J. X. Cruz Neto · G.
López · A. Soubeyran · J. C. O. Souza

Abstract This paper studies the constrained multiobjective optimization problem of finding Pareto critical points of vector-valued functions. The proximal point method considered by Bonnel et al. (SIAM J. Optim., 4 (2005), pp. 953-970) is extended to locally Lipschitz functions in the finite dimensional multiobjective setting. To this end, a new approach for convergence analysis of the method is proposed where the first-order optimality condition of the "scalarized" problem is replaced by a necessary condition for weakly Pareto points of a multiobjective problem.

Keywords Proximal method · multiobjective optimization · locally Lipschitz function · Pareto critical point · compromise problem · variational rationality

G. C. Bento
IME, Universidade Federal de Goiás, Goiânia, 74001-970, Brazil
E-mail: glaydston@ufg.br

J. X. Cruz Neto
CCN, DM, Universidade Federal do Piauí, Teresina, 64049-550, Brazil
E-mail: jxavier@ufpi.edu.br

G. López
Departamento de Análisis Matemático, Universidad de Sevilla, 41080, Sevilla, Spain
E-mail: glopez@us.es

A. Soubeyran
Aix-Marseille University (Aix-Marseille School of Economics), CNRS and EHESS, France
E-mail: antoine.soubeyran@gmail.com

J. C. O. Souza
PESC, Universidade Federal do Rio de Janeiro and Universidade Federal do Piauí, Brazil
E-mail: joaocos.mat@ufpi.edu.br

1 Introduction

We discuss a method for solving vector-valued optimization problems in the following multiobjective (or multicriteria) context. Let $\mathbb{R}^m$ be the m-dimensional Euclidean space with the partial order " $\preceq$ " in $\mathbb{R}^m$ induced by the Paretian cone $\mathbb{R}_+^m$, given by $y \preceq z$ (or $z \succeq y$) if and only if $z - y \in \mathbb{R}_+^m$ with its associate relation " $\prec$ ", given by $y \prec z$ (or $z \succ y$) if and only if $z - y \in \mathbb{R}_{++}^m$, where

$$\mathbb{R}_+^m := \{x \in \mathbb{R}^m \ : \ x_j \geq 0, \ j \in \mathcal{I}\}, \quad \mathbb{R}_{++}^m := \{x \in \mathbb{R}^m \ : \ x_j > 0, \ j \in \mathcal{I}\},$$

and $\mathcal{I} := \{1, \ldots, m\}$. Given the vector-valued function $F : \mathbb{R}^n \to \mathbb{R}^m$, we analyze the proximal point method for finding a Pareto critical point of $F := (f_1, \ldots, f_m)$. A point $x \in \mathbb{R}^n$ is a Pareto critical point of F if there exists a component function f_i of F for which the Clarke directional derivative of f_i at x in the direction of $y - x$ is nonnegative for all $y \in \mathbb{R}^n$, with $f_i : \mathbb{R}^n \to \mathbb{R}$ and $i \in \mathcal{I}$; see the details of this concept in Section 2.

One of the main strategies for solving a multiobjective optimization problem is the scalarization approach; see for instance [8,13,15,16,18,25,28]. Multiobjective optimization algorithms that do not scalarize have recently been developed; see [7] for an overview on the subject. Some of these techniques are extensions of scalar optimization algorithms, e.g., steepest descent method [17, 24], projected gradient method [19,23], subgradient method [2] and Newton's method [18], while others borrow heavily from ideas developed in heuristic optimization; see e.g. [29,39] and references therein. For the latter, no convergence proofs are known.

There is a wider research program consisting of the extension to vector-valued setting of several iterative methods for scalar-valued functions. In 2005, Bonnel et al. [6] proposed an extension of the proximal point method to vector optimization, i.e., when other underlying ordering cones are used instead of the nonnegative orthant $\mathbb{R}_+^m$. They actually perform a similar approach for the case of the proximal point method for scalar-valued optimization. The (scalar) proximal point method was introduced in the literature by Moreau [37], Martinet [35] and later popularized by Rockafellar [43] who performs the proximal point method for the problem of finding zeros of operators. A brief description of this method can be found in Bonnel et al. [6]. We also refer to Lemaire [30] who surveys the literature on proximal point algorithm for real-valued functions up to 1989.

In this paper, we introduce and analyze a new approach for convergence of the proximal point method in (finite dimensional) multiobjective optimization which allows us to apply this method to a large class of functions, namely, locally Lipschitz vector-valued functions. The importance of the generalization to Lipschitz objective functions comes from applications. We will consider the

famous compromise solution problem, where a group of agents tries to minimize the distances of their current positions to the ideal point of the group (Gearhart [21]). In this case, distances are locally Lipschitz vector functions; see for instance [40, 41]. In a broad range of application (see location theory, utility theory, consumer theory, ...) such distance functions are used as objectives.

Next, we briefly describe how the (exact) proximal method analyzed in Bonnel et al. [6] provides convergence for solving a vector optimization problem. They use this method in order to find a weakly efficient minimizer (or weakly Pareto minimizer) of a map $F : X \to Y$ from a real Hilbert space X to a real Banach space Y containing a closed, convex and pointed cone $\mathcal{C}$ with nonempty interior, where "pointed" means that $\mathcal{C} \cap (-\mathcal{C}) = \{0\}$, with respect to the partial order "$\preceq_{\mathcal{C}}$" induced by the cone $\mathcal{C}$. In this context, weakly efficient minimizer means a point $x \in X$ such that there exists no $y \in X$ satisfying $F(y) \prec_{\mathcal{C}} F(x)$. For orders induced by non-Paretian cones, the problem of finding efficient points (weakly or not) is certainly not as frequent as the one concerning the point-wise partial order, but, nevertheless, it is not just an extension of the Paretian case and has its own importance. The (exact) method analyzed in [6] takes as the $(k + 1)$th iteration a weakly efficient solution of $F_k : X \to Y$ defined as

$$F_k(x) = F(x) + \lambda_k \|x - x^k\|^2 \varepsilon^k$$

subject to the constrained set $\Omega_k = \{x \in X \ : \ F(x) \preceq_{\mathcal{C}} F(x^k)\}$, where $\{\lambda_k\}$ is a bounded sequence of positive scalars and ε^k is an exogenously selected vector belonging to the interior of $\mathcal{C}$ such that $\|\varepsilon^k\| = 1$ for each $k \geq 0$. The idea underlying the convergence results is based on the first-order optimality condition of the scalar problem

$$\min_{x \in \Omega_k} \eta_k(x), \tag{1}$$

where $\eta_k(x) = \langle F(x), z^k \rangle + \frac{\lambda_k}{2} \langle \varepsilon^k, z^k \rangle \|x - x^k\|^2$ and $\{z^k\}$ is an exogenous sequence belonging to the positive polar cone $\mathcal{C}^+ \subset Y^*$ given by $\mathcal{C}^+ = \{z \in Y^* \ : \ \langle y, z \rangle \geq 0, \ \forall y \in \mathcal{C}\}$ such that $\|z^k\| = 1$, for all $k \geq 0$, and Y^* is the topological dual space of Y with $\langle \cdot, \cdot \rangle : Y \times Y^* \to \mathbb{R}$ the duality pairing. Thus, (1) implies that

$$0 \in \partial \psi_k(x^{k+1}) + \lambda_k \langle \varepsilon^k, z^k \rangle (x^{k+1} - x^k), \tag{2}$$

where $\psi_k(x) = \langle F(x), z^k \rangle + \delta_{\Omega_k}(x)$, with $\partial \psi_k$ denoting the subdifferential of ψ_k in the sense of convex analysis and $\delta_{\Omega_k}(\cdot)$ is the indicator function, that is, $\delta_{\Omega_k}(x) = 0$, if $x \in \Omega_k$, and $\delta_{\Omega_k}(x) = +\infty$, otherwise. Bonnel et al. [6] establish that any sequence generated by this algorithm converges (in the weak topology of X) to a weakly efficient minimizer of F under the following two assumptions:

(A1) (Convexity and lower semicontinuity) F is $\mathcal{C}$-convex with respect to the order "$\preceq_{\mathcal{C}}$", i.e., $F((1 - t)x + ty) \preceq_{\mathcal{C}} (1 - t)F(x) + tF(y)$ for all $x, y \in X$ and $t \in [0, 1]$, and F is positively lower semicontinuous which means that for every $z \in \mathcal{C}^+$, the scalar function $x \mapsto \langle F(x), z \rangle$ is lower semicontinuous;

(A2) (Completeness) The set $(F(x^0) - \mathcal{C}) \cap F(X)$ is $\mathcal{C}$-complete, i.e., for every
sequence $\{a^k\} \subset X$, with $a^0 = x^0$, such that $F(a^{k+1}) \preceq_{\mathcal{C}} F(a^k)$ for all
$k \in \mathbb{N}$, there exits $a \in X$ such that $F(a) \preceq_{\mathcal{C}} F(a^k)$ for all $k \in \mathbb{N}$.

Assumption (A1) guarantees that the constrained set Ω_k is closed and
convex for all $k \in \mathbb{N}$. Thus, (2) can be viewed as

$$\alpha_k(x^k - x^{k+1}) \in \partial(\langle F(\cdot), z^k \rangle)(x^{k+1}) + N_{\Omega_k}(x^{k+1}), \tag{3}$$

where $\alpha_k = \lambda_k \langle \varepsilon^k, z^k \rangle$ and $N_{\Omega_k}(x^{k+1})$ stands for the normal cone to Ω_k at
$x^{k+1} \in \Omega_k$ in the classical sense of convex analysis. In this approach, convexity
of each set Ω_k plays an important role. The set Ω_k forces the algorithm to
be a descent process. A motivation, in a dynamic context, to consider the
constrained set Ω_k is given in Bento et al. [4]. They mention that the set Ω_k
characterizes a vector improving process where a vectorial minimizing solution
x^k of the current proximal problem moves to a next one such that it improves
the current solution, which is essential to justify the process at a behavioral
level where a risk averse agent accepts to change only if the change is improving
on all aspects (all components of the vector). If we consider a group of agents,
as we do here, this constraint set is even more important. It imposes that the
payoff of each agent of the group does not decrease.

Other authors have proposed variants of the algorithm considered by Bon-
nel et al. [6] for convex vector or multiobjective problems; see for instance Ceng
and Yao [10], Choung et al. [11], Gregório and Oliveira [26] and Villacorta and
Oliveira [49]. Recently, the $\mathbb{R}_+^m$-quasiconvex case was discussed in Bento et
al. [4] and Apolinario et al. [1]; see the definition of $\mathbb{R}_+^m$-quasiconvexity on Sec-
tion 2. In these works, their corresponding algorithms, at $(k+1)$th iteration,
compute a point x^{k+1} satisfying

$$0 \in \partial g(F(x^{k+1})) + \alpha_k(x^{k+1} - x^k) + N_{\Omega_k}(x^{k+1}), \tag{4}$$

where $g : \mathbb{R}^m \to \mathbb{R}$ is a kind of "scalarization" function, ∂g denotes some sub-
differential of g and $\{\alpha_k\}$ is a sequence of positive real numbers; see Section 4
for more details about these algorithms. In both [4] and [1], convexity of Ω_k
comes from the $\mathbb{R}_+^m$-quasiconvexity of F.

The aim of this paper is two-fold. First, we present a new approach for con-
vergence of the proximal point algorithm in (finite dimensional) multiobjective
optimization. In [6], the authors use an optimality condition for the "scalar-
ized" problem (1), while here we establish convergence results combining the
fact that each iteration of the algorithm is a weakly efficient solution for a con-
strained multiobjective problem with a necessary condition for a point to be a
weakly efficient solution of a constrained multiobjective problem. We mention
that our approach does not use a convexity assumption of the constrained
set Ω_k as the previously mentioned works do. As a second contribution, we
expand the application of proximal methods in (finite dimensional) multiob-
jective optimization for locally Lipschitz vector-valued functions. We mention
that the $\mathcal{C}$-convex case analyzed by Bonnel et al. [6], restricted to the finite
dimensional multiobjective framework, is indeed a particular instance of our

locally Lipschitz case. In light of our approach, the $\mathbb{R}^m_+$-quasiconvex case is also analyzed and a convergence result as given by [1] and [4] is presented.

As a motivation, we give a dynamic formulation of the well known static group compromise problem to modelize, in a crude way, how, starting from an initial situation, a group of agents with interrelated payoffs can succeed to approach and reach, following an acceptable transition, a desired end, defined as a compromise solution. This is a very important problem related to cooperative dynamical games.

This paper is organized as follows. Section 2 introduces some notations as well as some basic concepts and results in multiobjective optimization. In Section 3, we define the proximal method and, state and prove some of its properties. Section 4 is devoted to the convergence analysis of the algorithm. Finally, some remarks and future works are discussed in Section 5.

2 Multiobjective Optimization

In this section we recall some basic definitions and properties of multiobjective optimization which can be found, for instance, in T. D. Luc [33]. We state and prove some results that will allow to define the algorithm and study its convergence properties.

Given a nonempty set $\Omega \subset \mathbb{R}^n$ and a vector-valued function $F = (f_1, \ldots, f_m) : \mathbb{R}^n \to \mathbb{R}^m$, the associate vector optimization problem (VP) consists of finding a Pareto optimal point (or Pareto point) of F in Ω, i.e., a point $x^* \in \Omega$ such that there exists no other $x \in \Omega$ with $F(x) \preceq F(x^*)$ and $F(x) \neq F(x^*)$. This problem will be denoted as

$$\min\{F(x) \ : \ x \in \Omega\}. \tag{5}$$

The problem of finding a weakly Pareto optimal point (or weakly Pareto point) of F in Ω consists of finding a point $x^* \in \Omega$ such that there exists no $x \in \Omega$ with $F(x) \prec F(x^*)$. We denote this problem by

$$\min{}_w\{F(x) \ : \ x \in \Omega\}, \tag{6}$$

and the set of all weakly Pareto points of F in Ω is denoted by $\arg\min_w\{F(x) \ : \ x \in \Omega\}$.

Remark 1 As mentioned in Huang and Yang [27], the vector functions

$$F(\cdot) \quad \text{and} \quad e^{F(\cdot)} := (e^{f_1(\cdot)}, \ldots, e^{f_m(\cdot)})$$

have the same set of weakly Pareto points, where e^α denotes the exponential map valued at $\alpha \in \mathbb{R}$. This result can be easily extended to the Pareto critical setting. Hence, concerning Pareto critical points, we can assume without loss of generality that $F \succ 0$.

Given a scalar-valued function $g : \mathbb{R}^m \to \mathbb{R}$, here called a scalarization function, one can define a scalar optimization problem, denoted by (SP), corresponding to (VP) as follows

$$\min\{g(F(x)) \ : \ x \in \Omega\}. \tag{7}$$

We say that (SP) is a scalar weak representation of (VP), if $F(y) \prec F(x)$ implies $g(F(y)) < g(F(x))$.

Example 1 Let us consider the scalar function $g : \mathbb{R}^m \to \mathbb{R}$ given by

$$g(y) = \max_{1 \leq i \leq m} \langle y, e_i \rangle, \tag{8}$$

where $\{e_i\}$ is the canonical base of the space $\mathbb{R}^m$. In this case, it is clear that (SP) is a scalar weak representation of (VP). The scalarization function (8) was used in [4] to study the convergence analysis of a proximal point-type method for multiobjective optimization problems.

Proposition 1 *For the vector and scalar problems (VP) and (SP), the following inclusion holds:*

$$\arg\min\{g(F(x)) \ : \ x \in \Omega\} \subseteq \arg\min_w\{F(x) \ : \ x \in \Omega\},$$

if (SP) is scalar weak representation of (VP).

Proof [33, page 87].   □

Let $\mathcal{G}$ be a family of functions from $\mathbb{R}^m$ to $\mathbb{R}$. We say that this family is a complete weak scalarization for (VP) if, for every weakly Pareto optimal solution x^* of (VP), there exists $g \in \mathcal{G}$ such that x^* is an optimal solution of (SP) corresponding to g and (VP), and

$$\arg\min\{g(F(x)) \ : \ x \in \Omega\} \subseteq \arg\min_w\{F(x) \ : \ x \in \Omega\}.$$

For a vector function $F : \mathbb{R}^n \to \mathbb{R}^m$, we say that:

i) F is $\mathbb{R}_+^m$-convex if, for every $x, y \in \mathbb{R}^n$, the following holds:

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

ii) F is $\mathbb{R}_+^m$-quasiconvex if, for every $x, y \in \mathbb{R}^n$, the following holds:

$$F((1 - t)x + ty) \preceq \max\{F(x), F(y)\}, \qquad \forall t \in [0, 1],$$

where the maximum is taken component-wise.

Example 2 Consider the family of scalar-valued functions $h_z : \mathbb{R}^m \to \mathbb{R}$, for each $z \in \mathbb{R}^m \setminus \{0\}$, given by

$$h_z(y) = \langle y, z \rangle. \tag{9}$$

If the vector problem (5) is convex, it follows from [6, Theorem 2.1] that this family is a complete weak scalarization for the vector problem (5).

The scalarization function (9) was used in the convergence analysis of the proximal methods proposed in [6]. However, the authors show that the solution set of the scalarized problem may be empty. They actually provide an example where the solution set of the scalarized problem is nonempty only if z belongs to a set of measure zero in $\mathbb{R}^2$; see [6, Remark 1]. This means that the convexity assumption of the vector problem does not guarantee that the scalarized problem is nonempty but it is a sufficient condition for such a scalarization to be a complete weak scalarization for a vector problem (5). Next proposition gives us a complete weak scalarization for a vector problem which does not depend of the convexity of a vector function.

Proposition 2 *Let $\Omega \subset \mathbb{R}^n$ be a nonempty set and let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a vector function such that $F(x) \succ 0$, for all $x \in \Omega$. Then, the family of scalar-valued functions $\{g_z\}_{z \in \mathbb{R}^m_{++}}$ given by*

$$g_z(y) = \max_{1 \leq i \leq m} \frac{\langle y, e_i \rangle}{\langle z, e_i \rangle} \tag{10}$$

is a complete weak scalarization for (VP), where $\{e_i\}$ is the canonical basis of the space $\mathbb{R}^m$.

Proof Given a vector problem (VP), it is easy to verify that the related scalar problem (SP), with the scalarization given by (10), is a scalar weak representation of (VP), for each $z \in \mathbb{R}^m_{++}$. Hence, using Proposition 1, we get

$$\arg\min\{g_z(F(x)) \; : \; x \in \Omega\} \subseteq \arg\min_w\{F(x) \; : \; x \in \Omega\},$$

for all $z \in \mathbb{R}^m_{++}$. Now, for each $x^* \in \arg\min_w\{F(x) \; : \; x \in \Omega\}$ take $z = F(x^*) \succ 0$. From the definition of a weakly Pareto point, we have

$$g_z(F(y)) = \max_{1 \leq i \leq m} \frac{\langle F(y), e_i \rangle}{\langle z, e_i \rangle} \geq 1,$$

and $g_z(F(x^*)) = 1$. Then, $x^* \in \arg\min\{g_z(F(x)) \; : \; x \in \Omega\}$, and the proof is completed. $\square$

We recall now that a scalar-valued function $f : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz at a point $x \in \mathbb{R}^n$ if there exists a neighborhood U of this point and some real number $L > 0$ such that

$$|f(y) - f(y')| \leq L||y - y'||, \quad \forall y, y' \in U.$$

A function f is locally Lipschitz when it is locally Lipschitz at all points of its domain.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz function at $x \in \mathbb{R}^n$ and let $d \in \mathbb{R}^n$. The Clarke directional derivative of f at x in the direction of d, denoted by $f^\circ(x, d)$, is defined as follows

$$f^\circ(x, d) := \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y + td) - f(y)}{t},$$

and the Clarke subdifferential of f at x, denoted by $\partial f(x)$, as follows

$$\partial f(x) := \{w \in \mathbb{R}^n \ : \ \langle w, d \rangle \leq f^\circ(x, d), \quad \forall \, d \in \mathbb{R}^n\},$$

see Clarke [12]. Given a locally Lipschitz vector-valued function $F : \mathbb{R}^n \to \mathbb{R}^m$, i.e., all component functions $f_i : \mathbb{R}^n \to \mathbb{R}$ are locally Lipschitz functions, the Clarke subdifferential of F at $x \in \mathbb{R}^n$, denoted by $\partial F(x)$, is defined as

$$\partial F(x) := \{U \in \mathbb{R}^{m \times n} \ : \ U^\top d \preceq F^\circ(x; d), \quad \forall d \in \mathbb{R}^n\}, \tag{11}$$

where $F^\circ(x; d) := (f_1^\circ(x; d), \ldots, f_m^\circ(x; d))$. It is worth to point out that an equivalent definition has appeared, in a more general context, in Thibault [48]. If F is $\mathcal{C}$-convex, for some ordering cone $\mathcal{C}$, a similar definition can be found in Luc [34].

Let Ω be a closed and convex set. In this context, we say that a point $x \in \Omega$ is a Pareto-Clarke critical point (or Pareto critical point) of F in Ω if, for any $y \in \Omega$, there exists $i \in \mathcal{I}$ such that

$$f_i^\circ(x, y - x) \geq 0. \tag{12}$$

Remark 2 Note that if $m = 1$ in the previous definition, we retrieve the (classical) definition of critical points for nonsmooth functions: $0 \in \partial f(x)$. It is worth to notice that, combining (12) with Clarke [12, Proposition 1.4], we have the following alternative definition: a point $x \in \mathbb{R}^n$ is a Pareto critical point of F in Ω if, for any $y \in \Omega$, there exist $i \in \mathcal{I}$ and $\xi \in \partial f_i(x)$ such that $\langle \xi, y - x \rangle \geq 0$. Thus, if x is not a Pareto critical point of F in Ω, then there exists $y \in \Omega$ such that

$$U(y - x) \prec 0, \quad \forall U \in \partial F(x).$$

The next result gives a necessary condition for a point to be a Pareto critical point of a vector-valued function.

Lemma 1 *Let $w \in \mathbb{R}^m_+ \backslash \{0\}$ and assume that Ω is a nonempty, closed and convex set. If $-U^\top w \in N_\Omega(x)$, for some $U \in \partial F(x)$, then x is a Pareto critical point of F.*

Proof Take $x \in \Omega$ such that $-U^\top w \in N_\Omega(x)$ and let us suppose, by contradiction, that x is not a Pareto critical point of F. From Remark 2, there exists $y \in \Omega$ such that

$$U(y - x) \prec 0.$$

Since $w \in \mathbb{R}^m_+ \backslash \{0\}$, we have $\langle w, U(y - x) \rangle < 0$, but this contradicts the fact that $-U^\top w \in N_\Omega(x)$ and $\langle U^\top w, y - x \rangle = \langle w, U(y - x) \rangle$. Hence, the desired result is proved. $\square$

We denote the distance function $d : \mathbb{R}^n \to \mathbb{R}$ of a point $x \in \mathbb{R}^n$ to a set $C \subset \mathbb{R}^n$ as

$$d_C(x) := \inf\{\|x - c\| \ : \ c \in C\}. \tag{13}$$

Consider the problem (6) of finding a weakly Pareto point of a vector-valued function F subject to the following constrained set

$$\Omega = \{x \in \mathcal{D} \; : \; g_j(x) \leq 0, \quad j \in \mathcal{I}\}, \tag{14}$$

where $\mathcal{D} \subset \mathbb{R}^n$ is a nonempty, closed and convex set, and $g_j : \mathbb{R}^n \to \mathbb{R}$ is a locally Lipschitz function for each $j \in \mathcal{I}$. The next result presents a necessary condition for a point $x^* \in \Omega$ to be a weakly Pareto solution of (6).

Theorem 1 *Let $\mathcal{D} \subset \mathbb{R}^n$ be a nonempty, closed and convex set. Assume that the set Ω in (6) is given as in (14), and the functions $f_j, g_j : \mathbb{R}^n \to \mathbb{R}$, $j \in \mathcal{I}$, are locally Lipschitz. If $x^* \in \Omega$ is a weakly Pareto solution of (6), then there exist real numbers $u_j \geq 0$, $v_j \geq 0$, with $j \in \mathcal{I}$, and $\tau > 0$ such that*

$$\sum_{j \in \mathcal{I}} u_j \partial f_j(x^*) + \sum_{j \in \mathcal{I}} v_j \partial g_j(x^*) + \tau \partial d_{\mathcal{D}}(x^*) = 0,$$

$$with \quad \sum_{j \in J}(u_j + v_j) = 1 \quad and \quad v_j g_j(x^*) = 0, \quad j \in \mathcal{I}.$$

Proof The proof follows from Minami [36, Theorem 3.1]. □

As remarked by Minami [36, Remark 3.1], since $\mathcal{D}$ is a closed and convex set, the cone

$$\{w \; : \; w \in \tau \partial d_{\mathcal{D}}(x^*), \; \tau > 0\}$$

is the normal cone $N_{\mathcal{D}}(x^*)$.

The next result provides an explicit formula for the Clarke subdifferential of the distance function (13), which can be found in a more general context in Burke et al. [9].

Theorem 2 *Let C be a nonempty, closed and convex subset of $\mathbb{R}^m$. If $x \in C$, then $\partial d_C(x) = B[0,1] \cap N_C(x)$, where $B[0,1]$ denotes the closed unit ball in $\mathbb{R}^m$.*

3 The proximal point method

In this section we prove some facts related to our new approach for convergence of the proximal method for vector-valued functions. As a motivation, we show how this method can be a nice tool to solve the famous compromise solution problem.

3.1 Compromise problem

Let us consider a group of producers $i \in \mathcal{I} = \{1, \ldots, m\}$. The decision variables of the group is the vector $x \in \mathbb{R}^n$ which must satisfy some constraints $x = (x_1, \ldots, x_n) \in \mathcal{D} \subset \mathbb{R}^n$. The objective of each of them is a "to be increased" payoff (profit, utility), $h_i(x) \in \mathbb{R}_+$. The vectorial objective of the group is $H(x) \in \mathbb{R}^m$, where $H(x) = (h_1(x), \ldots, h_m(x))$. Thus, the subset of feasible vectorial payoffs of the group, i.e., the payoff sub-space of the group is $H(\mathcal{D}) = \{H(x) : x \in \mathcal{D}\} \subset \mathbb{R}^m$. Each agent wants a payoff as high as possible.

Suppose that the maximum payoff of each agent of the group is bounded above, i.e., $\overline{h}_i = \sup\{h_i(x) : x \in \mathcal{D}\} < +\infty$, $i \in \mathcal{I}$. Then, the vectorial payoff $\overline{H} = (\overline{h}_1, \ldots, \overline{h}_m)$ is the ideal (or utopia) vectorial payoff of this group. Usually, the ideal vectorial payoff is not feasible, which means, $\overline{H} \notin H(\mathcal{D})$. Let us consider "to be decreased" payoffs

$$f_i(x) = \overline{h}_i - h_i(x) \geq 0, \quad i \in \mathcal{I},$$

which refer, in Psychology, to unsatisfaction gap functions f_i. They measure how much each individual payoff $h_i(x)$, with $x \in \mathcal{D}$, fails to reach its maximum (ideal or utopia) value $\overline{h}_i$. These vectorial unsatisfaction gaps

$$F(x) = (f_1(x), \ldots, f_m(x)) = \overline{H} - H(x) \succeq 0$$

generate individual regrets or unsatisfactions with respect to ideal payoffs. The compromise solution (with respect to a norm) is some feasible alternative $x^* \in \mathcal{D}$ which minimizes the whole unsatisfaction of the group, in other words, it minimizes the distance between the ideal vectorial payoff $\overline{H}$ and the payoff sub-space $H(\mathcal{D})$. For the "compromise" problem in Multicriteria decision making see the well known references Gearhart [21] and Goetzmann et al. [22].

Using this (static) compromise model, let us consider a simple group dynamic model. It includes a starting point, an acceptable transition and some desired ends. This simple group dynamics model considers that transitions are acceptable if, each period, all members of the group improve their payoffs. In the opposite case, some agents will quit the group or resist to change. The desired end of the group $\mathcal{I}$ is to approach and reach an end point, which itself approaches as much as possible the ideal point. In a dynamic cooperative setting, all agents of the group will accept to change from the last position $x = x^k$ to the next, $y = x^{k+1}$ only if their payoff do not decrease, i.e., if

$$h_i(x^k) \leq h_i(x^{k+1}), \; \forall i \in \mathcal{I} \quad \Longleftrightarrow \quad H(x^k) \preceq H(x^{k+1}).$$

This defines a cooperative improving dynamic $x^{k+1} \in \Omega(x^k)$, where $\Omega(x^k) = \{x \in \mathcal{D} : H(x^k) \preceq H(x)\}$. The cooperative group dynamic problem is to find a cooperative improving dynamic $x^{k+1} \in \Omega(x^k)$ which approaches and reaches (converges to) a desired end position close enough to the ideal point; see Lewin [31, 32] for the details of "Group dynamics" and "Organizational change" management problems in Psychology and Management Sciences.

3.2 The algorithm

Throughout this paper, we consider $\mathcal{D} \subset \mathbb{R}^n$ a nonempty, closed and convex set, and $F = (f_1, \ldots, f_m) : \mathbb{R}^n \to \mathbb{R}^m$ such that each component function $f_i : \mathbb{R}^n \to \mathbb{R}$, $i \in \mathcal{I}$, is a locally Lipschitz function. From Remark 1, we can assume without loss of generality that $F \succ 0$.

Next, we consider the proximal point algorithm for finding a Pareto critical point of F in $\mathcal{D}$. Let $\{\lambda_k\}$ be a sequence of positive real numbers and let $\{\varepsilon^k\} \subset \mathbb{R}^m_{++}$ be a sequence such that $||\varepsilon^k|| = 1$ for all $k \geq 0$. The method generates a sequence $\{x^k\} \subset \mathcal{D}$ as follows:

Algorithm 1
Initialization: Choose $x^0 \in \mathcal{D}$.
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: Take, as next iterate, $x^{k+1} \in \mathcal{D}$ such that

$$x^{k+1} \in \operatorname{argmin}_w \left\{ F(x) + \frac{\lambda_k}{2} ||x - x^k||^2 \varepsilon^k \ : \ x \in \Omega_k \right\}, \qquad (15)$$

where $\Omega_k = \{x \in \mathcal{D} \ : \ F(x) \preceq F(x^k)\}$.

We would like to mention that this method finds separate solutions at time and not the whole solution set. It has been noticed by Fukuda and Graña Drummond [20], and Fliege et al. [18] that we can expect to somehow approximate the solution set by just performing this method for different initial points. In the well known weighting method, this kind of idea also appears; see Burachik et al. [8]. More precisely, the method can be performed for different weights in order to find the solution set, or a reasonable approximation of this set. However, in some cases, arbitrary choices of the weighting vectors may lead the weighting method to unbounded problems. The Pareto front, i.e., the objective values of these solutions, is in general an infinite set. Thus, in practice, only an approximation of the Pareto front is obtained.

Proposition 3 *The Algorithm 1 is well-defined.*

Proof The starting point $x^0 \in \mathcal{D}$ is chosen in the initialization step. Assuming that the algorithm reached iteration k, we show next that the $(k+1)$th iteration exists. Take $z \in \mathbb{R}^m_{++}$, and define the scalar-valued function

$$g_z(y) = \max_{1 \leq i \leq m} \frac{\langle y, e_i \rangle}{\langle z, e_i \rangle},$$

where $\{e_i\}$ is the canonical basis of the space $\mathbb{R}^m$. It is easy to check that $\min\{g_z(F_k(x)) \ : \ x \in \Omega_k\}$ is a scalar weak representation of $\min_w\{F_k(x) \ : \ x \in \Omega_k\}$ for all $z \in \mathbb{R}^m_{++}$, where $F_k(x) = F(x) + \frac{\lambda_k}{2}||x - x^k||^2 \varepsilon^k$. Therefore, by Proposition 1, we obtain

$$\arg\min\{g_z(F_k(x)) \ : \ x \in \Omega_k\} \subseteq \arg\min_w\{F_k(x) \ : \ x \in \Omega_k\}. \qquad (16)$$

We will prove next that $\arg\min\{g_z(F_k(x)) : x \in \Omega_k\}$ is nonempty. Indeed,

$$g_z(F_k(x)) = \max_{1 \le i \le m} \frac{\langle F(x) + \frac{\lambda_k}{2}||x - x^k||^2 \varepsilon^k, e_i \rangle}{\langle z, e_i \rangle}$$

$$\ge \frac{\langle F(x), e_i \rangle}{\langle z, e_i \rangle} + \frac{\lambda_k \langle \varepsilon^k, e_i \rangle}{2\langle z, e_i \rangle}||x - x^k||^2, \tag{17}$$

for $i = 1, \ldots, m$. Note that $\langle z, e_i \rangle > 0$ and $\lambda_k \langle \varepsilon^k, e_i \rangle > 0$ for each $i = 1, \ldots, m$ and $k \in \mathbb{N}$. By virtue of $F \succ 0$, we also have that $\langle F(x), e_i \rangle > 0$ for $i = 1, \ldots, m$. These facts combined with (17) imply that $g_z(F_k(\cdot))$ is coercive, i.e., $\lim_{||x|| \to +\infty} g_z(F_k(x)) = +\infty$. By using Rockafellar and Wets [44, Theorem 1.9], from the continuity of $g_z(F_k(\cdot))$ together with closeness of Ω_k, we have that there exists a point $\tilde{x} \in \Omega_k$ such that

$$\tilde{x} \in \arg\min\{g_z(F_k(x)) : x \in \Omega_k\}.$$

Therefore, from (16), we can take $x^{k+1} = \tilde{x}$ as the $(k+1)$th iteration, which ends the proof. □

From now on, $\{x^k\}$, $\{\lambda_k\}$ and $\{\varepsilon^k\}$ denote the sequences considered in Algorithm 1. Next, we explore deeply the structure of the vector problem by using the necessary condition for a weakly Pareto optimal point of a multiobjective problem given by Theorem 1. The following result will be used in our main convergence results.

Proposition 4 *For all $k \in \mathbb{N}$, there exist $A_k \in \mathbb{R}^{m \times n}$, $u^k, v^k \in \mathbb{R}_+^m$, $w^k \in \mathbb{R}^m$ and $\tau_k \in \mathbb{R}_{++}$ such that*

$$A_k^\top(u^k + v^k) + \lambda_{k-1}\langle \varepsilon^{k-1}, u^k \rangle(x^k - x^{k-1}) + \tau_k w^k = 0, \tag{18}$$

where

$$w^k \in B[0,1] \cap N_{\mathcal{D}}(x^k) \quad and \quad ||u^k + v^k||_1 = 1, \quad \forall k \in \mathbb{N}. \tag{19}$$

Proof It follows from definition of the algorithm that x^k is a weakly Pareto solution of the problem

$$\min_w\{F_{k-1}(x) : x \in \Omega_{k-1}\},$$

where $F_{k-1}(x) = F(x) + \frac{\lambda_{k-1}}{2}||x - x^{-1}k||^2\varepsilon^{k-1}$. Denoting $G_{k-1}(x) = F(x) - F(x^{k-1})$, it is easy to verify, from the locally Lipschitz continuity of F, that all component functions

$$(g_{k-1})_j(\cdot) = f_j(\cdot) - f_j(x^{k-1}), \text{ with } j \in \mathcal{I}, \tag{20}$$

and

$$(f_{k-1})_j(\cdot) = f_j(\cdot) + \frac{\lambda_{k-1}}{2}|| \cdot -x^{k-1}||^2\varepsilon_j^{k-1}, \text{ with } j \in \mathcal{I}, \tag{21}$$

are locally Lipschitz functions. Hence, the desired result follows by applying Theorem 1, for each $k \in \mathbb{N}$ fixed, with g_j and f_j given in (20) and (21), respectively, and taking into account that, from Theorem 2, we have

$$\partial d_{\mathcal{D}}(x^k) = B[0,1] \cap N_{\mathcal{D}}(x^k), \quad \forall k \in \mathbb{N}.$$

In this case, $A_k^\top = [a_1^k \ \ldots \ a_m^k]^\top$, where $a_j^k \in \partial f_j(x^k)$, with $j \in \mathcal{I}$, $u^k = (u_1^k, \ldots, u_m^k)^\top$ and $v^k = (v_1^k, \ldots, v_m^k)^\top$. $\qquad\square$

Remark 3 Note that from (19), $\{u^k\}$, $\{v^k\}$ and $\{w^k\}$ are bounded sequences. From Bolte et al. [5, Remark 1] ∂f_j is bounded on compact sets. So, we have that $\{A_k\}$ is bounded as long as $\{x^k\}$ is bounded because $a_j^k \in \partial f_j(x^k)$, $j \in \mathcal{I}$. Therefore, if $\{\lambda_k\}$ and $\{x^k\}$ are bounded sequences, it follows from (18) that $\{\tau_k\}$ is also bounded.

As a consequence of the previous proposition, we obtain the following stopping rule for Algorithm 1.

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

Proof If there exists $k_0 \in \mathbb{N}$ such that $u^{k_0} = 0$, then from (18), we have

$$A_{k_0}^\top v^{k_0} + \tau_{k_0} w^{k_0} = 0.$$

As $\tau_{k_0} > 0$, and $w^{k_0} \in N_{\mathcal{D}}(x^{k_0})$, last equality is equivalent to

$$-A_{k_0}^\top v^{k_0} \in N_{\mathcal{D}}(x^{k_0}).$$

On the other hand, from the second assertion in (19), we can say that $v^{k_0} \in \mathbb{R}_+^m \backslash \{0\}$. Since $A_{k_0} \in \partial F(x^{k_0})$, the desired result follows by using Lemma 1 with $U = A_{k_0}$, $w = v^{k_0}$ and $x = x^{k_0}$. $\qquad\square$

As in [6], the stopping rule in Algorithm 1 can be changed by the following rule, which is easier to check: after computing x^{k+1} the algorithm stops if $x^{k+1} = x^k$, i.e., we set $x^{k+p} = x^k$ for all $p \geq 1$. Proposition 4 combined with Lemma 1 allows us to see that this condition is sufficient to get the stopping rule given in Algorithm 1. However, even in the convex case, it is possible to note that this rule might fail to recognize weakly Pareto solutions; see [6, Proposition 3.2].

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

Remark 4 Note that, if Algorithm 1 terminates after a finite number of iterations, it terminates at a Pareto critical point. This leads us to suppose that $\{x^k\}$ is an infinite sequence, and hence, $u^k \neq 0$ and $x^{k+1} \neq x^k$ for all $k \in \mathbb{N}$, in view of Corollary 1 and 2, respectively.

4 Convergence Analysis

In this section, we will consider the following $\mathbb{R}^m_+$-completeness assumption on the vector function F and the initial point x^0:

   ($\mathbb{R}^m_+$-**Completeness assumption**) For all sequences $\{a^k\} \subset \mathbb{R}^n$, with $a^0 = x^0$, such that $F(a^{k+1}) \preceq F(a^k)$ for all $k \in \mathbb{N}$, there exists $a \in \mathbb{R}^n$ such that

$$F(a) \preceq F(a^k) \quad \forall k \in \mathbb{N}.$$

   In the classical unconstrained (convex) optimization case, this condition is equivalent to existence of solutions of the optimization problem. The $\mathbb{R}^m_+$-completeness assumption is standard for ensuring existence of efficient points for vector optimization problems; see [33, page 46]. An interesting discussion on existence conditions of efficient points for vector optimization problems can also be found in [33, Chapter 2].

4.1 Locally Lipschitz case

Let us go back to our group compromise motivation. The nature of the "group dynamic" problem depends heavily on the nature of the objective functions which determine the properties of the improving sets and other relevant constraints. Objective functions can be convex or concave, quasiconvex or quasiconcave, difference of convex or concave functions, and more generally, Lipschitz or locally Lipschitz functions. In the recent (VR) variational rationality approach of human behaviors, see Soubeyran [45–47]), Lipschitz and locally Lipschitz payoffs are very interesting for two reasons: they mean that when inconveniences to change are low, you cannot expect large advantages to change; there is no free lunch, which is a reasonable hypothesis. Furthermore, these functions are easy to estimate locally. This helps badly informed agents, who know their payoffs functions only at some given points, to be able to find, each step, some improving changes. Lipschitz functions f have concave underestimating functions

$$y \in \mathcal{D} \longmapsto u_0(y) = f(x_0) - L||y - x_0||,$$

for each $x_0 \in \mathcal{D}$. This allows us to analyze convergence properties of an algorithm for locally Lipschitz vector-valued functions.

   Next we prove our convergence result for the locally Lipschitz case.

Theorem 3 *Suppose that there exist scalars $a, b, c \in \mathbb{R}_{++}$ such that $0 < a \leq \lambda_k \leq b$ and $0 < c \leq \varepsilon^k_j$, for all $k \in \mathbb{N}$ and $j = 1, \ldots, m$. Then, every cluster point of $\{x^k\}$, if any, is a Pareto critical point of F.*

Proof Let $\hat{x}$ be a cluster point of $\{x^k\}$, and let $\{x^{k_l}\}$ be a subsequence of $\{x^k\}$ converging to $\hat{x}$. According to Proposition 2, the scalarization given by (10) is

a complete weak scalarization for a vector problem. It follows from definition of the algorithm that x^{k_l}, for each $l \in \mathbb{N}$, is an optimal solution of the problem

$$\min_w \left\{ F(x) + \frac{\lambda_{k_l-1}}{2} ||x - x^{k_l-1}||^2 \varepsilon^{k_l-1} \ : \ x \in \Omega_{k_l-1} \right\}.$$

Using these facts combined with definition of complete weak scalarization, we have that there exists a sequence $\{z^{k_l}\} \subset \mathbb{R}^m_{++}$, such that

$$\max_{1 \leq j \leq m} \left\{ \frac{\langle F(x^{k_l}) + \frac{\lambda_{k_l-1}}{2} ||x^{k_l} - x^{k_l-1}||^2 \varepsilon^{k_l-1}, e_j \rangle}{\langle z^{k_l}, e_j \rangle} \right\} \leq \max_{1 \leq j \leq m} \frac{\langle F(x^{k_l-1}), e_j \rangle}{\langle z^{k_l}, e_j \rangle}, \tag{22}$$

for all $l \in \mathbb{N}$. Since (22) is not altered through multiplication by a positive scalar, we can assume, without any loss of generality, that $||z^{k_l}|| = 1$, for all $l \in \mathbb{N}$, and hence, we can assume that $z^{k_l} \to \hat{z}$ as $l \to +\infty$ (extracting other subsequence if necessary). From (22), applying the Cauchy-Schwarz inequality, we have

$$\frac{\langle F(x^{k_l}), e_i \rangle}{\langle z^{k_l}, e_j \rangle} + \frac{\lambda_{k_l-1}}{2} ||x^{k_l} - x^{k_l-1}||^2 \langle \varepsilon^{k_l-1}, e_j \rangle \leq \max_{1 \leq j \leq m} \frac{\langle F(x^{k_l-1}), e_j \rangle}{\langle z^{k_l}, e_j \rangle},$$

for each $j = 1, \ldots, m$. In particular, it holds

$$\frac{\langle F(x^{k_l}), e_j \rangle}{\langle z^{k_l}, e_j \rangle} + \frac{\lambda_{k_l-1}}{2} ||x^{k_l} - x^{k_l-1}||^2 \min_{1 \leq j \leq m} \langle \varepsilon^{k_l-1}, e_j \rangle \leq \max_{1 \leq j \leq m} \frac{\langle F(x^{k_l-1}), e_j \rangle}{\langle z^{k_l}, e_j \rangle},$$

for each $j = 1, \ldots, m$. Once again, last inequality holds, in particular, for the index where the maximum of the first term in the left-hand side is attained. Thus,

$$\frac{\lambda_{k_l-1}}{2} ||x^{k_l} - x^{k_l-1}||^2 \min_{1 \leq j \leq m} \langle \varepsilon^{k_l-1}, e_j \rangle \leq \max_{1 \leq j \leq m} \frac{\langle F(x^{k_l-1}), e_j \rangle}{\langle z^{k_l}, e_j \rangle}$$
$$- \max_{1 \leq j \leq m} \frac{\langle F(x^{k_l}), e_j \rangle}{\langle z^{k_l}, e_j \rangle}. \tag{23}$$

Since $z^{k_l} \to \hat{z}$ as $l \to +\infty$ and, $\{F(x^k)\}$ is nonincreasing and $F \succ 0$, we obtain that the right-hand side of (23) converges to 0 as $l \to +\infty$. Using this fact combined with the lower bound assumption $0 < a \leq \lambda_k$ and $0 < c \leq \varepsilon_j^k$, for all $k \in \mathbb{N}$ and $j = 1, \ldots, m$, we must have

$$(x^{k_l} - x^{k_l-1}) \to 0 \text{ as } l \to +\infty. \tag{24}$$

Now, applying Proposition 4 for the sequence $\{x^{k_l}\}$, we have that there exist sequences $\{A_{k_l}\} \subset \mathbb{R}^{m \times n}$, $\{u^{k_l}\}$, $\{v^{k_l}\} \subset \mathbb{R}^m_+$, $\{w^{k_l}\} \subset \mathbb{R}^m$ and $\{\tau_{k_l}\} \subset \mathbb{R}_{++}$ satisfying

$$A_{k_l}^\top (u^{k_l} + v^{k_l}) + \lambda_{k_l-1} \langle \varepsilon^{k_l-1}, u^{k_l} \rangle (x^{k_l} - x^{k_l-1}) + \tau_{k_l} w^{k_l} = 0. \tag{25}$$

Note that $\{\lambda_{k_l}\}$ is bounded and $\{x^{k_l}\}$ converges to $\hat{x}$, and hence $\{x^{k_l}\}$ is bounded. Thus, from Remark 3, we can assume that the sequences $\{A_{k_l}\}$,

$\{u^{k_l}\}$, $\{v^{k_l}\}$, $\{w^{k_l}\}$ and $\{\tau_{k_l}\}$ are bounded. Without loss of generality, we may assume that $A_{k_l} \to \hat{A}$, $u^{k_l} \to \hat{u}$, $v^{k_l} \to \hat{v}$ and $\tau_{k_l} \to \hat{\tau}$ as $l \to +\infty$ (we will use the same notation for the index even if we need to extract other subsequences). Since $\{\lambda_{k_l-1}\langle\varepsilon^{k_l-1}, u^{k_l}\rangle\}$ is bounded, it follows from (24) that $\lambda_{k_l-1}\langle\varepsilon^{k_l-1}, u^{k_l}\rangle(x^{k_l} - x^{k_l-1})$ vanishes as $l \to +\infty$. Therefore, we get, taking the limit in (25) as $l \to +\infty$, that

$$\hat{A}^\top \hat{y} + \hat{\tau}\hat{w} = 0, \qquad (26)$$

where $\mathbb{R}^m_+ \setminus \{0\} \ni \hat{y} := \hat{u} + \hat{v}$, $\hat{A} \in \partial F(\hat{x})$ and $\hat{w} \in N_{\mathcal{D}}(\hat{x})$, because $\partial F(\cdot)$ and $N_{\mathcal{D}}(\cdot)$ are closed. Thus, from (26), we obtain

$$-\hat{A}^\top \hat{y} \in N_{\mathcal{D}}(\hat{x}),$$

and this together with Lemma 1, enables us to say that $\hat{x}$ is a Pareto critical point of F. This completes the proof. $\qquad \square$

4.2 Quasiconvex case

In this section, we consider Algorithm 1 with the additional assumption that $F : \mathbb{R}^n \to \mathbb{R}^m$ is a $\mathbb{R}^m_+$-quasiconvex function. This case was analyzed by Apolinario et al. [1] in the (finite dimensional) multiobjective framework. They compute the $(k + 1)$th iteration as follows

$$0 \in \partial\left(\langle F(\cdot), z^k\rangle + \frac{\lambda_k}{2}\langle\varepsilon^k, z^k\rangle || \cdot -x^k ||^2\right)(x^{k+1}) + N_{\Omega_k}(x^{k+1}). \qquad (27)$$

The $\mathbb{R}^m_+$-quasiconvex case for multiobjective optimization was also studied by Bento et al. [4]. They consider the following iterative procedure

$$x^{k+1} \in \arg\min_{x\in\mathbb{R}^n} f\left(F(x) + \delta_{\Omega_k} + \frac{\lambda_k}{2}||x - x^k||^2 e\right), \qquad (28)$$

where $e = (1, \ldots, 1) \in \mathbb{R}^m$, the scalarization function $f : \mathbb{R}^m \to \mathbb{R}$ is given by $f(y) = \max_{1 \leq i \leq m}\langle y, e_i\rangle$ and $\{e_i\}$ is the canonical base of the space $\mathbb{R}^m$.

The convergence analysis of both algorithms (27) and (28) are based on the Fejér monotonicity (see definition below), using the same approach proposed by Bonnel et al. [6]. In these works, the scalarization plays an important role in the proof, since the vectorial subproblems are replaced by scalar optimality conditions by using a scalarization function; see [6, Theorem 3.1], [1, Proposition 3.4.1] and [4, Theorem 4.1].

We emphasize that, as in the previously mentioned works, our convergence analysis for the $\mathbb{R}^m_+$-quasiconvex case is also based on the Fejér monotonicity of the sequence generated by the algorithm. However, it does not depend of the scalarization.

Before we give the main result of this section, let us recall that a sequence $\{y^k\}$ is said to be Fejér convergent (or Fejér monotone) to a nonempty set $U \subset \mathbb{R}^n$ if, for all $k \in \mathbb{N}$

$$||y^{k+1} - y|| \leq ||y^k - y||, \quad \forall y \in U.$$

The following result is well known and its proof is elementary.

Proposition 5 *Let $U \subset \mathbb{R}^n$ be a nonempty set and $\{y^k\}$ be a Fejér convergent sequence to U. Then, $\{y^k\}$ is bounded. Moreover, if a cluster point y of $\{y^k\}$ belongs to U, then $\{y^k\}$ converges to y.*

The next theorem shows that for the $\mathbb{R}^m_+$-quasiconvex case we have convergence of the (whole) sequence to a Pareto critical point.

Theorem 4 *The sequence $\{x^k\}$ converges to a Pareto critical point of F.*

Proof We divide the proof into five steps.
Step1 (Fejér convergence) Define $E \subset \mathcal{D}$ as

$$E = \{x \in \mathcal{D} \ : \ F(x) \preceq F(x^k), \quad \forall k \in \mathbb{N}\}.$$

From the $\mathbb{R}^m_+$-completeness assumption of F at x^0, E is nonempty. Now, take an arbitrary point $x^* \in E$, which means that $x^* \in \Omega_k$, for all $k \in \mathbb{N}$. Denote by $\gamma_{k+1} = \lambda_k \langle \varepsilon^k, u^{k+1} \rangle$. Note that $\gamma_{k+1} > 0$, for each $k \in \mathbb{N}$, because $\lambda_k > 0$, $\varepsilon^k \in \mathbb{R}^m_{++}$ and $u^k \in \mathbb{R}^m_+ \backslash \{0\}$ for all $k \in \mathbb{N}$. Since

$$||x^k - x^*||^2 = ||x^k - x^{k+1}||^2 + ||x^{k+1} - x^*||^2 + 2\langle x^k - x^{k+1}, x^{k+1} - x^* \rangle,$$

we conclude from (18) that

$$\begin{aligned}
||x^k - x^*||^2 &= ||x^k - x^{k+1}||^2 + ||x^{k+1} - x^*||^2 \\
&\quad + \frac{2}{\gamma_{k+1}} \langle A^\top_{k+1}(u^{k+1} + v^{k+1}) + \tau_{k+1} w^{k+1}, x^{k+1} - x^* \rangle \\
&= ||x^k - x^{k+1}||^2 + ||x^{k+1} - x^*||^2 \\
&\quad + \frac{2}{\gamma_{k+1}} \sum_{i=1}^m (u_i^{k+1} + v_i^{k+1}) \langle a_i^{k+1}, x^{k+1} - x^* \rangle + \tau_{k+1} \langle w^{k+1}, x^{k+1} - x^* \rangle,
\end{aligned}$$
$$(29)$$

where $a_i^{k+1} \in \partial f_i(x^{k+1})$, for all k and $i = 1, \ldots, m$. On the other hand, since F is $\mathbb{R}^m_+$-quasiconvex and, $x^* \in \Omega_k$ and $\gamma_k > 0$, for all k, we obtain

$$\frac{2}{\gamma_{k+1}} \sum_{i=1}^m (u_i^{k+1} + v_i^{k+1}) \langle a_i^{k+1}, x^{k+1} - x^* \rangle \geq 0. \tag{30}$$

Moreover, $w^{k+1} \in N_{\mathcal{D}}(x^{k+1})$, together with $\tau_k > 0$, leads to

$$\tau_{k+1} \langle w^{k+1}, x^{k+1} - x^* \rangle \geq 0. \tag{31}$$

Thus, using (30) and (31) in (29), we have

$$||x^{k+1} - x^k||^2 \leq ||x^k - x^*||^2 - ||x^{k+1} - x^*||^2, \quad \forall k \in \mathbb{N} \qquad (32)$$

which means that, $||x^{k+1} - x^*|| \leq ||x^k - x^*||$ for any $x^* \in E$. In other words, $\{x^k\}$ is Fejér convergent to E.

Step 2 (The cluster points of $\{x^k\}$ belongs to E) Since $\{x^k\}$ is Fejér convergent to E, it follows from Proposition 5 that $\{x^k\}$ is bounded. Let x^* be a cluster point of $\{x^k\}$. It follows from definition of the algorithm that $F(x^{k+1}) \preceq F(x^k)$ for all k. Thus, from the continuity of F, we can easily conclude that $F(x^*) \preceq F(x^k)$ for all k, which means that $x^* \in E$.

Step 3 (Convergence of the sequence) This step directly follows from Proposition 5 combined with Step 1 and 2.

Step 4 (Proximity of consecutive iterates) Assume that $\{x^k\}$ converges to $\hat{x}$. From the triangular inequality, we have

$$||x^{k+1} - x^k|| \leq ||x^{k+1} - \hat{x}|| + ||x^k - \hat{x}||, \quad \forall k \in \mathbb{N}. \qquad (33)$$

Noting that the right-hand side of (33) vanishes as $k \to +\infty$ because $x^k \to \hat{x}$ as $k \to +\infty$, we conclude

$$\lim_{k \to +\infty} ||x^{k+1} - x^k|| = 0. \qquad (34)$$

Step 5 (Pareto criticality of the limit point) The proof of this step uses the same argument as in the proof of Theorem 3 from (25) on. This establishes the result.   □

Remark 5 It is well known that, under the $\mathbb{R}_+^m$-convexity assumption, the concepts of weakly Pareto and Pareto critical are equivalent. In this context, from last theorem, one can say that Algorithm 1 converges to a weakly Pareto point of F.

5 Final remarks

In this paper, we have proposed a new approach for convergence of the proximal point algorithm in finite dimensional multiobjective optimization. It is proved that our approach can be successfully applied to obtain convergence properties of the proximal method for locally Lipschitz vector-valued maps. Although this approach can be applied to $\mathbb{R}_+^m$-quasiconvex vector functions (and, in particular, $\mathbb{R}_+^m$-convex vector functions), this new technique seems to be particularly useful for vector functions which make Ω_k in (15) a nonconvex set. To the best of our knowledge, it was the first time that a possible nonconvex Ω_k was considered in the proximal method (15).

It is well known that the classical proximal point method is indeed a conceptual scheme for optimization which has been the starting point for other methods, e.g., Augmented Lagrangians, both classical or generalized. As such, its convergence properties have been analyzed in a large list of papers, written

by many distinguished authors. We see our paper as part of this substantial body of the literature, within the realm of theoretical optimization. On the other hand, as observed in [6], "the performance of the method depends essentially on the algorithm used to solve the subproblems. In this situation, it makes little sense to compare the proximal method with other methods for finding efficient points (weakly or not) in terms of computational efficiency, unless a specific procedure is chosen for solving the subproblems". In this paper, we refrain from discussing algorithms for solve the subproblems, and hence we skip a discussion of implementation issues or comparison with an alternative approach.

The next steps as future works would be to propose inexact versions of Algorithm 1, as for instance in Ceng and Yao [10], following our new approach of convergence as well as a dynamic formulation of the well known static group compromise problem using the recent (VR) variational rationality approach of human behaviors (Soubeyran [45–47]). This allows us to show, how, starting from an initial situation, a group of agents with interrelated payoffs can succeed to approach and reach, following an acceptable transition, a desired end, defined as a compromise solution. This improving group dynamic, $x = x^k \curvearrowright y = x^{k+1}$ such that $H(x^{k+1}) \succeq H(x^k)$, for all $k \in \mathbb{N}$, must have two drawbacks:

i) It can fail to be acceptable for some members of the group, because, each period, it can be too costly for some agent to improve their payoffs;
ii) It does not converge in a non compact space of alternatives $\mathcal{D}$.

To remedy to these drawbacks, we can follow the (VR) approach. In fact, to improve, each period, generates both advantages and inconveniences to change for each agent $i \in \mathcal{I}$, i.e., vectorial advantages to change $A(x, y) \in \mathbb{R}^m$ and vectorial inconveniences to change $I(x, y) \in \mathbb{R}^m$ to be defined later. In turn vectorial advantages to change generate vectorial motivation to change $M(x, y) = U[A(x, y)] \in \mathbb{R}^m_+$ defined as the utilities $U_i[A_i]$ of advantages to change A_i, $i \in \mathcal{I}$. Similarly, vectorial inconveniences to change generate vectorial resistance to change $R(x, y) = D[A(x, y)] \in \mathbb{R}^m_+$ defined as the disutilities $D_i[I_i]$ of inconveniences to change I_i, $i \in \mathcal{I}$. Then, following again the (VR) approach, to be acceptable, a transition $x = x^k \curvearrowright y = x^{k+1}$, $k \in \mathbb{N}$, must generate, each period, vectorial motivation to change high enough ($\lambda = \lambda_{k+1} \in \mathbb{R}_{++}$ high enough) with respect to vectorial resistance to change. This means that each current change $x = x^k \curvearrowright y = x^{k+1}$ must be worthwhile, i.e.,

$$M(x, y) \geq \lambda R(x, y), \quad \lambda \in \mathbb{R}_{++}.$$

Thus, to be acceptable, a group dynamic must be worthwhile. Given the lessons of the (VR) approach and several of its applications, we can expect that such worthwhile group dynamic will converge.

This paper examines the case where inconveniences to change can be identified to a distance, i.e., costs to be able to change $C(x, y) = C(y, x)$ and costs to be able to stay $C(x, x) = 0$. In the general case $C(x, y) \neq C(y, x)$ and $C(x, x) \succ 0$ is possible. Future research will also examine the case where

Euclidean norm in (15) is replaced by a "like-distance" as done for instance in Bento and Soubeyran [3] and Moreno et al. [38] for scalar-valued functions. This is better adapted to applications in behavioral sciences.

References

1. H. C. F. APOLINÁRIO, E. A. PAPA QUIROZ, AND P. R. OLIVEIRA, A Scalarization Proximal Point Method for Quasiconvex Multiobjective Minimization. J. Glob. Optim., 64 (2016), pp. 79-96.
2. J. Y. BELLO CRUZ, A subgradient method for vector optimization problems. SIAM J. Optim., 23 (2013), pp. 2169-2182.
3. G. C. BENTO, AND A. SOUBEYRAN, Generalized Inexact Proximal Algorithms: Routine's Formation with Resistance to Change, Following Worthwhile Changes. J. Optim. Theory Appl., 166 (2014), pp. 172-187.
4. G. C. BENTO, J. X. CRUZ NETO, AND A. SOUBEYRAN, A proximal point-type method for multicriteria optimization. Set-Valued Var. Anal., 22 (2014), pp. 557-573.
5. J. BOLTE, A. DANIILIDIS, A. LEWIS, AND M. SHIOTA, Clarke critical values of subanalytic Lipschitz continuous functions. Ann. Polon. Math., 87 (2005), pp. 13-25.
6. H. BONNEL, A. N. IUSEM, AND B. F. SVAITER, Proximal methods in vector optimization. SIAM J. Optim., 15 (2005), pp. 953-970.
7. J. BRANKE, K. DEB, K. MIETTINEN, AND R. SLOWINSKI, EDITORS, Practical Approaches to Multi-Objective Optimization. Dagstuhl Seminar, 06501, Germany, 2007.
8. R. S. BURACHIK, C. Y. KAYA, AND M. M. RIZVI, A new scalarization technique to approximate Pareto fronts of problems with disconnected feasible sets. J. Optim. Theory Appl., 162 (2014), pp. 428-446.
9. J. V. BURKE, M. C. FERRIS, AND M. QIAN, On the Clarke subdifferential of the distance function of a closed set. J. Math. Anal. Appl., 166 (1992), pp. 199-213.
10. L. C. CENG, AND J. C. YAO, Approximate proximal methods in vector optimization. Eur. J. Oper. Res., 183 (2007), pp. 1-19.
11. T. D. CHOUNG, T. D., B. S. MORDUKHOVICH, AND J. C. YAO, Hybrid approximate proximal algorithms for efficient solutions in vector optimization. J. Nonlinear Convex Anal., 12 (2011), pp. 257-286.
12. F. H. CLARKE, Generalized gradients and applications. Trans. Amer. Math. Soc., 205 (1975), pp. 247-262.
13. J. X. CRUZ NETO, G. J. P. SILVA, O. P. FERREIRA, AND J. O. LOPES, A subgradient method for multiobjective optimization. Comput. Optim. Appl., 54 (2013), pp. 461-472.
14. A. L. CUSTÓDIO, J. F. A. MADEIRA, A. I. F. VAZ, AND L. N VICENTE. Direct multisearch for multiobjective optimization. SIAM J. Optim., 21 (2011), pp. 1109-1140.
15. I. DAS, AND J. E. DENNIS, Normal-boundary intersection: a new method for generating the pareto surface in nonlinear multicriteria optimization problems. SIAM J. Optim., 8 (1998), pp. 631-657.
16. G. EICHFELDER, An adaptive scalarization method in multiobjective optimization. SIAM J. Optim., 19 (2009), pp. 1694-1718.
17. J. FLIEGE, AND B. F. SVAITER, Steepest descent methods for multicriteria optimization. Math. Methods Oper. Res., 51 (2000), pp. 479-494.
18. J. FLIEGE, L. M. GRAÑA DRUMMOND, AND B. F. SVAITER, Newton's method for multiobjective optimization. SIAM J. Optim. 20 (2009), pp. 602-626.
19. E. H. FUKUDA, AND L. M. GRAÑA DRUMMOND, On the convergence of the projected gradient method for vector optimization. Optimization, 60 (2011), pp. 1009-1021.

20. E. H. FUKUDA, AND L. M. GRAÑA DRUMMOND, A survey on multiobjective decent methods. Pesqui. Oper., 34 (2014), pp. 585-620.
21. W. B. GEARHART, Compromise solutions and estimation of the non inferior set. J. Optim. Theory Appl., 47 (1979), pp. 29-47.
22. K. S. GOETZMANN, B. BUSING, AND J. MATUSCHKE, Multicriteria optimization and compromise solutions. MDS Colloquium, (2011).
23. L. M. GRAÑA DRUMMOND, AND A. N. IUSEM, A projected gradient method for vector optimization problems. Comput. Optim. Appl. 28 (2004), pp. 5-29.
24. L. M. GRAÑA DRUMMOND, AND B. F. SVAITER, A steepest descent method for vector optimization. J. Comput. Appl. Math. 175 (2005), pp. 395-414.
25. L. M. GRAÑA DRUMMOND, N. MACULAN, AND B. F. SVAITER, On the choice of parameters for the weighting method in vector optimization. Math. Program. Ser. B, 111 (2008), pp. 201-216.
26. R. GREGÓRIO, AND P. R. OLIVEIRA, A logarithmic-quadratic proximal point scalarization method for multiobjective programming. J. Glob. Optim., 49 (2011), pp. 281-291.
27. X. X. HUANG, AND X. Q. YANG, Duality for multiobjective optimization via nonlinear Lagrangian functions. J. Optim. Theory Appl., 120 (2004), pp. 111-12.
28. J. JAHN, Scalarization in vector optimization. Math. Program., 29 (1984), pp. 203-218.
29. M. LAUMANNS, L. THIELE, K. DEB, AND E. ZITZLER, Combining convergence and diversity in evolutionary multiobjective optimization. Evolutionary Computation, 10 (2002), pp. 263-282.
30. B. LEMAIRE, The proximal algorithm. In: New methods in Optimization and Their Industrial Uses, (J. P. Penot, Ed.), International Series of Numerical Mathematics, 87 (1989), pp. 73-87, Birkhäuser, Boston, MA.
31. K. LEWIN, Frontiers in group dynamics: Concept, method and reality in social science, social equilibria and social change. Human Relations, 1 (1947), pp. 5-41.
32. K. LEWIN, Field Theory in Social Science. Harper and Row, New York (1951).
33. D. T. LUC, , Theory of Vector Optimization, Lecture Notes in Econom. and Math. Syst. Springer-Verlag, New York, (1989).
34. D. T. LUC, N. X. TAN, AND P. N. TINH, Convex Vector Functions and their Subdifferential. Acta Math Vietnam, 23, (1998), pp. 107-127.
35. B. MARTINET, Régularisation d'inéquations variationelles par approximations successives. Rev. Française d'Inform. Recherche Oper. 4 (1970), pp. 154-159.
36. M. MINAMI, Weak Pareto-Optimal Necessary Conditions in a Nondifferentiable Multiobjective Program on a Banach Space. J. Optim. Theory Appl., 41 (1983), pp. 451-461.
37. J. J. MOREAU, Proximité et dualité dans un espace Hilbertien. Bull. Soc. Math. France, 93 (1965) pp. 273-299.
38. F. G. MORENO, P. R. OLIVEIRA, AND A. SOUBEYRAN, A proximal algorithm with quasidistance. Application to Habit's Formation. Optimization, 61 (2011), pp. 1383-1403.
39. S. MOSTAGHIM, J. BRANKE, AND H. SCHMECK, Multi-objective particle swarm optimization on computer grids. Technical report. 502 (2006), Institut AIFB, University of Karlsruhe.
40. S. OPRICOVIC, AND G-H. TZENG, Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS. Eur. J. Oper. Res., 156 (2004), pp. 445-455.
41. S. OPRICOVIC, AND G-H. TZENG, Extended VIKOR method in comparison with outranking methods. Eur. J. Oper. Res., 178 (2007), pp. 514-529.
42. M. S POOLE, AND A. H. VAN DE VEN, Handbook of Organizational Change and Innovation, Oxford University Press., New York, 2004.
43. R. T. ROCKAFELLAR, Monotone operators and the proximal point algorithm. SIAM J. Control. Optim., 14 (1976), pp. 877-898.
44. R. T. ROCKAFELLAR, AND R.J-B. WETS, Variational Analysis. Springer, Berlin (1998).
45. A. SOUBEYRAN, Variational rationality, a theory of individual stability and change: worthwhile and ambidextry behaviors. Preprint, GREQAM, Aix Marseillle University, France (2009).
46. A. SOUBEYRAN, Variational rationality and the "unsatisfied man": routines and the course pursuit between aspirations, capabilities and beliefs. Preprint, GREQAM, Aix Marseillle University, France (2010).

47. A. SOUBEYRAN, Variational rationality, Part 1. A theory of worthwhile stay and change approach-avoidance transitions ending in traps. Preprint, GREQAM-AMSE, Aix Marseille University, France (2016).
48. L. THIBAULT, Subdifferentials of Nonconvex Vector-Valued Functions. J. Math. Anal. Appl., 86, (1982), pp. 319-344.
49. K. D. V. VILLACORTA, AND P. R. OLIVEIRA, An interior proximal method in vector optimization. Eur. J. Oper. Res., 214 (2011), pp. 485-492.