

The Proximal Point Method for Locally Lipschitz Functions in Multiobjective Optimization with Application to the Compromise Problem

G. Bento, J. Cruz Neto, G. López, Antoine Soubeyran, J. Souza

▶ **To cite this version:**

G. Bento, J. Cruz Neto, G. López, Antoine Soubeyran, J. Souza. The Proximal Point Method for Locally Lipschitz Functions in Multiobjective Optimization with Application to the Compromise Problem. SIAM Journal on Optimization, 2018, 28 (2), pp.1104-1120. 10.1137/16M107534X . hal-01985333

HAL Id: hal-01985333
https://amu.hal.science/hal-01985333

Submitted on 1 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SIAM J. OPTIM.
Vol. 28, No. 2, pp. 1104–1120
© 2018 Society for Industrial and Applied Mathematics

THE PROXIMAL POINT METHOD FOR LOCALLY LIPSCHITZ FUNCTIONS IN MULTIOBJECTIVE OPTIMIZATION WITH APPLICATION TO THE COMPROMISE PROBLEM*

G. C. BENTO†, J. X. CRUZ NETO‡, G. LÓPEZ§, A. SOUBEYRAN¶, AND 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, Iusem, and Svaiter [*SIAM J. Optim.*, 15 (2005), pp. 953–970] is extended to locally Lipschitz functions in the finite dimensional multiobjective setting. To this end, a new (scalarization-free) 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 weak Pareto points of a multiobjective problem. As a consequence, this has allowed us to consider the method without any assumption of convexity over the constraint sets that determine the vectorial improvement steps. This is very important for applications; for example, to extend to a dynamic setting the famous compromise problem in management sciences and game theory.

Key words. proximal point method, multiobjective optimization, locally Lipschitz function, Pareto critical point, compromise problem, variational rationality

AMS subject classifications. 49J52, 65K05, 90C26, 90C29, 91E99

DOI. 10.1137/16M107534X

1. Introduction.

In this paper, we extend the applicability of the exact proximal point method introduced by Bonnel, Iusem, and Svaiter [7] for solving (finite dimensional) multiobjective optimization problems with nonconvex and closed constraints. Our approach solves iteratively multiobjective problems by using a vectorial optimality condition rather than a first-order optimality condition to scalarized problems. This has allowed us to increase the range of application of the method to locally Lipschitz vector-valued functions without any assumption of convexity over the constraint sets that determine the vectorial improvement steps. The importance of the generalization to locally Lipschitz objective functions comes from applications. Our generalization offers two main advantages. First, it allows us to model the famous static 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 [25]). In

*Received by the editors May 16, 2016; accepted for publication (in revised form) November 30, 2017; published electronically April 10, 2018.

http://www.siam.org/journals/siopt/28-2/M107534.html

Funding: The research of first author was supported in part by FAPEG 201210267000909 - 05/2012, CNPq grants 458479/2014-4, 471815/2012-8, 312077/2014-9, and CAPES (88881.117595/2016-01). The research of the second author was supported in part by CNPq grant 305462/2014-8. The research of the third author has been supported by grant MTM2015-65242-C2-1-P. The research of the fourth author was supported by ANR GREEN-Econ research project (ANR-16-CE03-0005). The research of the fifth author was supported in part by CNPq-Ciências sem Fronteiras grant 203360/2014-1.

†IME, Universidade Federal de Goiás, Goiânia, 74001-970, Brazil (glaydston@ufg.br).

‡CCN, DM, Universidade Federal do Piauí, Teresina, 64049-550, Brazil (jxavier@ufpi.edu.br).

§Departamento de Análisis Matemático, Universidad de Sevilla, 41080, Sevilla, Spain (glopez@us.es).

¶Aix-Marseille University (Aix-Marseille School of Economics), CNRS and EHESS, France (antoine.soubeyran@gmail.com).

‖PESC, Universidade Federal do Rio de Janeiro and Universidade Federal do Piauí, Brazil (joaocos.mat@ufpi.edu.br).

this case, distances are locally Lipschitz vector functions; see, for instance, [45, 46]. In a broad range of applications (see location theory, utility theory, consumer theory, ...) such distance functions are used as objectives. Second, taking advantages of a nice specificity of the algorithm proposed in Bonnel, Iusem, and Svaiter [7], it allows us to examine a dynamic version of the compromise problem in the context of human dynamics in behavioral sciences. This comes from the fact that the algorithm in [7] is a descent process. Thus, it can model in a crude but precise way a transition that a group of agents accepts to follow before reaching a compromise solution as an ascent process, each agent required not to decrease his payoff from one period to the other one.

The beginning of the story starts with the (scalar) proximal point method introduced in the literature by Moreau [42], Martinet [40], and later popularized by Rockafellar [47] 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, Iusem, and Svaiter [7]. We also refer to Lemaire [35] who surveys the literature on proximal point algorithms for real-valued functions up to 1989.

We consider multiobjective (or multicriteria) optimization in the following 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 $\mathcal{I} := \{1, \ldots, m\}$. We recall that the cone $\mathbb{R}^m_+$ is Daniell, i.e., any decreasing sequence having a lower bound converges to its infimum, and $\mathbb{R}^m_+$ is correct in the sense that $(\mathrm{cl}\,\mathbb{R}^m_+) + \mathbb{R}^m_+ \backslash l(\mathbb{R}^m_+) \subset \mathbb{R}^m_+$, where cl denotes the closure and $l(\mathbb{R}^m_+)$ stands for the set $\mathbb{R}^m_+ \cap -\mathbb{R}^m_+$; see Luc [38]. Given a 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.

Two different strategies have been used for solving multiobjective optimization problems: scalarization techniques (see, for instance, [9, 16, 18, 19, 20, 22, 30, 33]) and nonscalarization approaches (see [8] for an overview on this subject). In the first case, for finding a (weak or not) Pareto optimal point of F (see these concepts in section 2), the scalarization approach finds a minimizer of the scalar function $\zeta(F(\cdot))$, for some scalarization functional ζ which has the property that a solution of the scalar problem is a Pareto solution of the vectorial problem. In the second case, multiobjective optimization algorithms that do not scalarize have been developed and some of these techniques are extensions of scalar optimization algorithms, e.g., steepest descent method [21, 29], projected gradient method [23, 28], subgradient method [2], and Newton's method [22], while others borrow heavily from ideas developed in heuristic optimization; see, e.g., [34, 44] and references therein. For the latter, no convergence proofs are known.

Our paper extends the work of Bonnel, Iusem, and Svaiter [7] to a nonconvex setting without using a scalarization method. Thus, it will be useful for the reader to describe the exact method analyzed in Bonnel, Iusem, and Svaiter [7] who 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 functions. They use this method in order to find a weak Pareto optimal (or weak Pareto) 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, weak Pareto point 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 Pareto points (weak or not) is certainly not as frequent as the one concerning the pointwise partial order, but, nevertheless, it is not just an extension of the Paretian case and has its own importance. The (exact) method analyzed in [7] takes as the $(k+1)$th iteration a weak Pareto 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 $\mathcal{C}^+ = \{z \in Y^* : \langle y, z \rangle \geq 0$, for all $y \in \mathcal{C}\}$ such that $||z^k|| = 1$ for all $k \geq 0$, and Y^* is the topological dual space of Y, where $\langle \cdot, \cdot \rangle : Y \times Y^* \to \mathbb{R}$ is the duality pairing. Thus, x^{k+1} is a solution of (1) and, hence, it satisfies the following inclusion:

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

where $\psi_k(x) = \langle F(x), z^k \rangle + \delta_{\Omega_k}(x)$, $\partial \psi_k$ denotes 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, Iusem, and Svaiter [7] establish that any sequence generated by this algorithm converges (in the weak topology of X) to a weak Pareto point 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 exists $a \in X$ 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

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

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, Cruz Neto, and Soubeyran [4]. They mention that the set Ω_k characterizes a

vector improvement 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 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; see section 3.1.

Other authors have proposed variants of the algorithm considered by Bonnel, Iusem, and Svaiter [7] for convex vector or multiobjective problems; see, for instance, Ceng and Yao [11], Ceng, Mordukhovich, and Yao [12], Choung, Mordukhovich, and Yao [13], Gregório and Oliveira [31], and Villacorta and Oliveira [52]. Recently, the $\mathbb{R}^m_+$-quasi-convex case was discussed in Bento, Cruz Neto, and Soubeyran [4] and Apolinário, Papa Quiroz, and Oliveira [1]; see the definition of $\mathbb{R}^m_+$-quasi-convexity on section 2. In these works, their corresponding algorithms, at the $(k+1)$th iteration, compute a point x^{k+1} satisfying

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

where $\zeta : \mathbb{R}^m \to \mathbb{R}$ is a scalarization function, $\partial\zeta$ denotes some subdifferential of ζ, and $\{\alpha_k\}$ is a sequence of positive real numbers; see section 4 for more details about these algorithms. In both [1] and [4], convexity of Ω_k comes from the $\mathbb{R}^m_+$-quasi-convexity of F.

The aim of this paper is twofold. First, we present an approach (which does not use a scalarization), for convergence of the proximal point algorithm in (finite dimensional) multiobjective optimization. In [7], the authors use an optimality condition for the scalarized problem (1), while here we establish convergence results without using any scalarization method combining the fact that each iteration of the algorithm is a weak Pareto solution for a constrained multiobjective problem with a necessary condition for a point to be a weak Pareto solution of a constrained multiobjective problem. We mention that our approach does not use convexity assumption of the constraint sets as the previously mentioned works do. As a second contribution, we expand the application of proximal methods in (finite dimensional) multiobjective optimization for locally Lipschitz vector-valued functions with nonconvex constraints. We mention that the $\mathcal{C}$-convex case analyzed by Bonnel, Iusem, and Svaiter [7], restricted to the finite dimensional multiobjective framework, is indeed a particular instance of our locally Lipschitz case. In light of our approach, the $\mathbb{R}^m_+$-quasi-convex case is also analyzed and convergence results as proved in [1] and [4] are presented.

As an application, we give a dynamic formulation of the well-known static group compromise problem to model, in a crude way, how, starting from an initial situation, a group of agents with interrelated payoffs are able 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 discuss some basic definitions and properties of multiobjective optimization which can be found, for instance, in Luc [38].

Given a nonempty set $\Omega \subset \mathbb{R}^n$ and $F = (f_1, \ldots, f_m) : \mathbb{R}^n \to \mathbb{R}^m$ a vector-valued function, a point $x^* \in \Omega$ is called a Pareto optimal point (or Pareto point) of F in Ω if there exists no other $x \in \Omega$ with $F(x) \preceq F(x^*)$ and $F(x) \neq F(x^*)$. A point $x^* \in \Omega$ is called weak Pareto optimal (or weak Pareto) of F in Ω if there exists no $x \in \Omega$ with $F(x) \prec F(x^*)$. This means that for all $x \in \Omega$ there exists an index $j(x) = j \in \{1, \ldots, m\}$ such that $F_j(x) - F_j(x^*) \geq 0$. We call (VP) the problem of finding a weak Pareto optimal point which we denote by

$$\min_w \{F(x) \ : \ x \in \Omega\}. \tag{3}$$

The set of all weak 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 [32], 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 weak 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 $F \succ 0$.

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 $\mathbb{R}^m_+$-quasi-convex 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 componentwise.

We recall now some concepts involving locally Lipschitz functions and nonconvex constrained sets. The definitions and notation are taken from [17]. 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:

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

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

$$d_C(x) := \inf\{||x - c|| \ : \ c \in C\}. \tag{4}$$

Let $\Omega \subset \mathbb{R}^n$ be a nonempty and closed set. As in [17], 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 $v \in T_\Omega(x)$, there exists $i \in \mathcal{I}$ such that

$$f_i^\circ(x, v) \geq 0, \tag{5}$$

where $T_\Omega(x) := \{v \in \mathbb{R}^n \ : \ d^\circ_\Omega(x, v) = 0\}$ denotes the set of all tangent vectors to Ω at x. As mentioned in [15, page 11], a vector v belongs to $T_\Omega(x)$ if and only if it satisfies the following property: for every sequence $\{x^k\}$ in Ω converging to x

and every sequence t_n in $(0,\infty)$ converging to 0, there is a sequence v^n converging to v such that $x^n + t_n v^n$ belongs to Ω for all n. Having defined a tangent cone, the likely candidate for the normal cone is the one obtained from $T_\Omega(x)$ by polarity. Accordingly, we define $N_\Omega(x)$, the normal cone to Ω at x, as follows:

$$N_\Omega(x) := \{\xi \in \mathbb{R}^n \,:\, \langle \xi, v\rangle \le 0 \quad \forall v \in T_\Omega(x)\}.$$

If Ω is convex, $N_\Omega(x)$ coincides with the cone of normals in the sense of convex analysis; see [15, Proposition 2.4.4].

The Clarke subdifferential of f at x, denoted by $\partial f(x)$, is defined as follows:

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

see Clarke [14]. 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

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

where $F^\circ(x;d) := (f_1^\circ(x;d), \ldots, f_m^\circ(x;d))$. It is worth pointing out that an equivalent definition has appeared, in a more general context, in Thibault [51]. If F is $\mathcal{C}$-convex for some ordering cone $\mathcal{C}$, a similar definition can be found in Luc, Tan, and Tinh [39].

Remark 2. Note that if $m = 1$ in the previous definition of a Pareto critical point, we retrieve the (classical) definition of critical points for nonsmooth functions: $0 \in \partial f(x)$. It is worth noticing that, combining (5) with Clarke [14, 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 $v \in T_\Omega(x)$, there exist $i \in \mathcal{I}$ and $\xi \in \partial f_i(x)$ such that $\langle \xi, v\rangle \ge 0$. Thus, if x is not a Pareto critical point of F in Ω, then there exists $v \in T_\Omega(x)$ such that

$$Uv \prec 0 \quad \forall U \in \partial F(x).$$

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 and closed 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 $v \in T_\Omega(x)$ such that

$$Uv \prec 0.$$

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

Consider the problem (3) of finding a weak Pareto point of a vector-valued function F subject to the following constrained set

$$\Omega := \{x \in \mathcal{D} \,:\, g_s(x) \le 0, \quad s \in \mathcal{J}\}, \tag{6}$$

where $\mathcal{D} \subset \mathbb{R}^n$ is a nonempty and closed set, and $g_s : \mathbb{R}^n \to \mathbb{R}$ is a locally Lipschitz function for each $s \in \mathcal{J} := \{1, \ldots, p\}$. The next result presents a necessary condition for a point $x^* \in \Omega$ to be a weak Pareto solution of (3).

THEOREM 1. *Let $\mathcal{D} \subset \mathbb{R}^n$ be a nonempty and closed set. Assume that the set Ω in (3) is given as in (6), and the functions $f_j, g_s : \mathbb{R}^n \to \mathbb{R}$, $j \in \mathcal{I}$, and $s \in \mathcal{J}$, are locally Lipschitz. If $x^* \in \Omega$ is a weak Pareto solution of (3), then there exist real numbers $u_j \geq 0$, $v_s \geq 0$, with $j \in \mathcal{I}$ and $s \in \mathcal{J}$, and $\tau > 0$ such that*

$$\sum_{j\in\mathcal{I}} u_j \partial f_j(x^*) + \sum_{s\in\mathcal{J}} v_s \partial g_s(x^*) + \tau \partial d_{\mathcal{D}}(x^*) = 0$$

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

Proof. The proof follows from Minami [41, Theorem 3.1]. ∎

As remarked by Minami [41, Remark 3.1], if $\mathcal{D}$ is additionally convex (which we do not assume), then the cone

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

is the normal cone in the classical sense of convex analysis. For the nonconvex case, a formula for the Clarke subdifferential of the distance function (4) can be found in Burke, Ferris, and Qian [10]. We present a proof just for the sake of clarity.

THEOREM 2. *Let C be a nonempty and closed subset of $\mathbb{R}^m$. If $x \in C$, then*

$$\partial d_C(x) \subset B[0,1] \cap N_C(x),$$

where $B[0,1]$ denotes the closed unit ball in $\mathbb{R}^m$.

Proof. It is known that the distance function, $d_C(x)$, is globally Lipschitz with constant $L = 1$; see, for instance, [15, Proposition 2.4.1]. From [15, Proposition 2.1.2], we have that a Lipschitz function f of rank L near to x satisfies $||\xi|| \leq L$ for every $\xi \in \partial f(x)$. Thus, $\partial d_C(x) \subset B[0,1]$. On the other hand, it follows from [15, Proposition 2.4.2] that $N_C(x) = \text{cl}\,\{\cup_{\lambda\geq 0} \lambda \partial d_C(x)\}$, where cl denotes closure. This implies that $\partial d_C(x) \subset N_C(x)$ and the proof is completed. ∎

3. The proximal point method. In this section, we prove some facts related to our approach for convergence of the proximal method for vector-valued functions. As an application, 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 form 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 subspace 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 utopian) 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 utopian) 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 subspace $H(\mathcal{D})$. For the "compromise" problem in multicriteria decision making, see the well known references Gearhart [25] and Goetzmann, Busing, and Matuschke [26].

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 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 change from the last position $x = x^k$ to the next, $y = x^{k+1}$ only if their payoff does not decrease, i.e., if

$$h_i(x^k) \leq h_i(x^{k+1}) \; \forall i \in \mathcal{I} \quad \Longleftrightarrow \quad H(x^k) \preceq H(x^{k+1}). \tag{7}$$

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 [36, 37] 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 and closed 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

$$x^{k+1} \in \operatorname{argmin}_w \left\{ F(x) + \frac{\lambda_k}{2} ||x - x^k||^2 \varepsilon^k \; : \; x \in \Omega_k \right\}, \tag{8}$$

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 [24], and

Fliege, Graña Drummond, and Svaiter [22] 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, Kaya, and Rizvi [9]. 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.

Next, we prove that Algorithm 1 is well defined. To this end, we consider the concept of completeness as in (A2).

PROPOSITION 1. *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. Denote by $F_k(x) := F(x) + \frac{\lambda_k}{2}||x - x^k||^2\varepsilon^k$. Note that $x^k \in \Omega_k$ which implies that $F_k(\Omega_k)$ is nonempty. It is straightforward to check that $F_k(x) \succ 0$ and $F_k(\Omega_k)$ is closed. Since the cone $\mathbb{R}^m_+$ is Daniell, it follows from [38, Lemma 3.5] that $F_k(\Omega_k)$ is $\mathbb{R}^m_+$-complete. Thus, from [38, Theorem 3.3] the set

$$\arg\min_w\{F_k(x) \,:\, x \in \Omega_k\}$$

is nonempty. ∎

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 weak Pareto optimal point of a multiobjective problem given by Theorem 1. The following result will be used in our main convergence results.

PROPOSITION 2. *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*

$$A_k^\top(u^k + v^k) + \lambda_{k-1}\langle\varepsilon^{k-1}, u^k\rangle(x^k - x^{k-1}) + \tau_k w^k = 0, \tag{9}$$

where

$$w^k \in B[0,1] \cap N_{\mathcal{D}}(x^k) \quad \textit{and} \quad ||u^k + v^k||_1 = 1 \quad \forall k \in \mathbb{N}. \tag{10}$$

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

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

where $F_{k-1}(x) = F(x) + \frac{\lambda_{k-1}}{2}||x - x^{k-1}||^2\varepsilon^{k-1}$. 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

$$(g_{k-1})_j(\cdot) = f_j(\cdot) - f_j(x^{k-1}) \ \text{ with } j \in \mathcal{I}, \tag{11}$$

and

$$(f_{k-1})_j(\cdot) = f_j(\cdot) + \frac{\lambda_{k-1}}{2}||\cdot - x^{k-1}||^2\varepsilon_j^{k-1} \ \text{ with } j \in \mathcal{I}, \tag{12}$$

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 by (11) and (12), respectively, and taking into account that, from Theorem 2, we have

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

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

Remark 3. Note that from (10), $\{u^k\}$, $\{v^k\}$, and $\{w^k\}$ are bounded sequences. From Bolte et al. [6, 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 (9) 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 (9), we have

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

As $\tau_{k_0} > 0$ and $w^{k_0} \in N_{\mathcal{D}}(x^{k_0})$, the 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 (10), 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 $x = x^{k_0}$. ∎

As in [7], 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 2 combined with Lemma 1 allows us to see that this condition is sufficient getting 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 weak Pareto solutions; see [7, 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 Algorithm 1 generates an infinite sequence $\{x^k\}$ which remains constant from a Pareto critical point on, i.e., if x^k is a Pareto critical point, then $x^{k+p} = x^k$ for all $p \in \mathbb{N}$. In view of Corollaries 1 and 2, we can suppose without loss of generality that $u^k \neq 0$ and $x^{k+1} \neq x^k$ for all $k \in \mathbb{N}$, respectively.

4. Convergence analysis. As mentioned before, the idea underlying the results of the scalarization approach is to solve a scalar problem in order to obtain a solution of the related vectorial problem. In our approach, the main result does not solve a scalar problem to obtain convergence of the sequence but it uses a necessary condition for weak Pareto points of a multiobjective problem instead. Hence, it allows us to analyze convergence properties of proximal methods for both locally Lipschitz vector and scalar functions with nonconvex constraints.

4.1. Locally Lipschitz case. Before presenting our convergence results 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, quasi-convex or quasi-concave, difference of convex or concave functions, and, more generally, Lipschitz or locally Lipschitz functions. In the recent variational rationality (VR) approach of human behaviors (see Soubeyran [48, 49, 50]), 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 payoff functions only at some given points, to be able to find, at 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 $x_0 \in \mathcal{D}$.

Next, we prove our main 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 $0 < c \leq \varepsilon_j^k$ 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. It follows from the definition of the algorithm that x^k for each $k \in \mathbb{N}$, is an optimal solution of the problem

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

This implies

$$\max_{1 \leq j \leq m} \left\{ F_j(x^{k-1}) - F_j(x^k) - \frac{\lambda_{k-1}}{2}||x^k - x^{k-1}||^2 \varepsilon_j^{k-1} \right\} \geq 0.$$

Take $j_0(k) = j_0 \in \{1, \ldots, m\}$ as the index where the maximum in the last inequality is attained. Then, from the lower boundedness assumption of $\{\lambda_k\}$ and $\{\varepsilon^k\}$, we have

$$\frac{ac}{2}||x^k - x^{k-1}||^2 \leq F_{j_0}(x^{k-1}) - F_{j_0}(x^k). \tag{13}$$

Since $\{F(x^k)\}$ is nonincreasing and $F \succ 0$, we obtain that the right-hand side of (13) converges to 0 as $k \to +\infty$. Hence,

$$(x^k - x^{k-1}) \to 0 \text{ as } k \to +\infty. \tag{14}$$

Now, let $\hat{x}$ be a cluster point of $\{x^k\}$, and let $\{x^{k_l}\}$ be a subsequence of $\{x^k\}$ converging to $\hat{x}$. Applying Proposition 2 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

$$A_{k_l}^\top (u^{k_l} + v^{k_l}) + \lambda_{k_l - 1} \langle \varepsilon^{k_l - 1}, u^{k_l} \rangle (x^{k_l} - x^{k_l - 1}) + \tau_{k_l} w^{k_l} = 0. \tag{15}$$

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 (14) 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 (15) as $l \to +\infty$, that

$$\hat{A}^\top \hat{y} + \hat{\tau} \hat{w} = 0, \tag{16}$$

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 (16), 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. ∎

Remark 5. Note that, if $\Omega_0 = \{x \in \mathbb{R}^n : F(x) \preceq F(x^0)\}$ is bounded, then $\{x^k\}$ is bounded. It is worthwhile to mention that, if we take $m = 1$ and $\mathcal{D} = \mathbb{R}^n$ throughout this work, our method coincides with the one studied in [5] and the exact version analyzed in [19] for the finite dimensional Euclidean setting with the square Euclidean norm as the regularization term.

4.2. Quasi-convex case. In this section, we consider Algorithm 1 with the additional assumptions that $F : \mathbb{R}^n \to \mathbb{R}^m$ is a $\mathbb{R}^m_+$-quasi-convex function, $\mathcal{D}$ is convex, and the well-known $\mathbb{R}^m_+$-completeness assumption on the set $(F(x^0) - \mathbb{R}^m_+) \cap F(\mathcal{D})$:

$(\mathcal{H})$: For every sequence $\{a^k\} \subset \mathcal{D}$, with $a^0 = x^0$, such that $F(a^{k+1}) \preceq F(a^k)$, for all $k \in \mathbb{N}$, there exists $a \in \mathcal{D}$ such that

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

The $\mathbb{R}^m_+$-quasi-convex case was analyzed by Apolinário, Papa Quiroz, and Oliveira [1] in the unconstrained framework, i.e., $\mathcal{D} = \mathbb{R}^n$. They compute the $(k+1)$th iteration as follows:

$$0 \in \partial \left(\langle F(\cdot), z^k \rangle + \frac{\lambda_k}{2} \langle \varepsilon^k, z^k \rangle || \cdot - x^k||^2 \right) (x^{k+1}) + N_{\Omega_k}(x^{k+1}). \tag{17}$$

This case was also studied by Bento, Cruz Neto, and Soubeyran [4]. They consider the following iterative procedure

$$x^{k+1} \in \arg\min_{x \in \mathbb{R}^n} \vartheta \left(F(x) + \delta_{\Omega_k} + \frac{\lambda_k}{2} ||x - x^k||^2 e \right), \tag{18}$$

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

The convergence analyses of both algorithms (17) and (18) are based on Fejér monotonicity (see definition below), using the same approach proposed by Bonnel, Iusem, and Svaiter [7]. In these works, the scalarization plays an important role in their proofs because the vectorial subproblems are replaced by scalar optimality conditions using a scalarization function; see [7, 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 $\mathbb{R}^m_+$-quasi-convex case is also based on the Fejér monotonicity of the sequence generated by the algorithm. However, it does not depend on any scalarization functional.

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 $k \in \mathbb{N}$

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

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

PROPOSITION 3. *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 $\mathbb{R}^m_+$-quasi-convex 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.

Step 1 *(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 assumption $(\mathcal{H})$, the set 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 $\gamma_{k+1} = \lambda_k \langle \varepsilon^k, u^{k+1} \rangle$. 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 (9) that

$$\begin{aligned} ||x^k - x^*||^2 &= ||x^k - x^{k+1}||^2 + ||x^{k+1} - x^*||^2 \\ &\quad + \frac{2}{\gamma_{k+1}} \langle A^\top_{k+1}(u^{k+1} + v^{k+1}) + \tau_{k+1} w^{k+1}, x^{k+1} - x^* \rangle \\ &= ||x^k - x^{k+1}||^2 + ||x^{k+1} - x^*||^2 \\ &\quad + \frac{2}{\gamma_{k+1}} \sum_{i=1}^m (u_i^{k+1} + v_i^{k+1}) \langle a_i^{k+1}, x^{k+1} - x^* \rangle + \tau_{k+1} \langle w^{k+1}, x^{k+1} - x^* \rangle, \end{aligned} \tag{19}$$

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 $\mathbb{R}^m_+$-quasi-convex and $x^* \in \Omega_k$ and $\gamma_k > 0$ for all k, we obtain

$$\frac{2}{\gamma_{k+1}} \sum_{i=1}^m (u_i^{k+1} + v_i^{k+1}) \langle a_i^{k+1}, x^{k+1} - x^* \rangle \geq 0. \tag{20}$$

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

$$\tau_{k+1} \langle w^{k+1}, x^{k+1} - x^* \rangle \geq 0. \tag{21}$$

Thus, using (20) and (21) in (19), we have

$$||x^{k+1} - x^k||^2 \leq ||x^k - x^*||^2 - ||x^{k+1} - x^*||^2 \quad \forall k \in \mathbb{N}$$

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 $\{x^k\}$ belong to E).* Since $\{x^k\}$ is Fejér convergent to E, it follows from Proposition 3 that $\{x^k\}$ is bounded. Let x^* be a cluster point of $\{x^k\}$. It follows from the 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 3 combined with Steps 1 and 2.

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

$$||x^{k+1} - x^k|| \leq ||x^{k+1} - \hat{x}|| + ||x^k - \hat{x}|| \quad \forall k \in \mathbb{N}. \tag{22}$$

Noting that the right-hand side of (22) vanishes as $k \to +\infty$ because $x^k \to \hat{x}$ as $k \to +\infty$, we conclude

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

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 (15) on. This establishes the result. □

Remark 6. It is well known that, under the $\mathbb{R}^m_+$-convexity assumption, the concepts of weak Pareto and Pareto critical are equivalent. In this case, under the assumption of $\mathbb{R}^m_+$-convexity, the last theorem ensures that any sequence generated from Algorithm 1 converges to a weak Pareto point of F as in [7]. Now, regarding assumption $(\mathcal{H})$, it is standard for ensuring existence of Pareto points for vector optimization problems and an interesting discussion on existence conditions of such points can be found in [38, Chapters 2 and 3]. Taking into account that $F \succ 0$, if we suppose that $F(D)$ is a closed set, then assumption $(\mathcal{H})$ holds. Note that this occurs naturally when $F(D)$ is a compact set; for more details, see [38, Lemma 3.5, page 47]. It is worth noting that assumption $(\mathcal{H})$ is used just to ensure that the set E, defined in step 1 of the proof of Theorem 4, is a nonempty set. As observed in [7], in the absence of $(\mathcal{H})$, this could be obtained by considering, for example, that the sequence $\{x^k\}$ has a cluster point.

5. Final remarks. In this paper, we have proposed a new approach for the convergence of the proximal point algorithm in finite dimensional multiobjective optimization. We 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 $\mathbb{R}^m_+$-quasi-convex 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 (8) 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 (8).

The next steps as future works would be to propose inexact versions of Algorithm 1, as for instance in Ceng and Yao [11] and Durea and Strugariu [19], following our approach of convergence as well as a dynamic formulation of the well-known static group compromise problem using the recent VR approach of human behaviors (Soubeyran [48, 49, 50]). In the VR context, an extension of this paper would examine the case where inconveniences to change can be identified at a distance, i.e., costs able to change $C(x, y) = C(y, x)$ and costs 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. Hence, future research will examine the case where the Euclidean norm in (8) is replaced by a "like-distance" as done for instance in Bento and Soubeyran [3] and Moreno, Oliveira, and Soubeyran [43] for scalar-valued functions. This is better adapted to applications in behavioral sciences.

Finally let us emphasize two advantages of our approach compared to the one in [7] and [19].

Remark 7. Given a nonempty set Ω and a vector-valued function F. Let $\zeta : \mathbb{R}^m \to \mathbb{R}$ be a scalar function, here called a scalarization function. One can define a

scalar optimization problem, corresponding to the vector problem VP as follows:

$$\min\{\zeta(F(x)) \ : \ x \in \Omega\}. \tag{23}$$

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

$$\varrho_z(y) := \langle y, z\rangle. \tag{24}$$

If the corresponding vector problem (VP) is convex, it follows from [38, Proposition 3.2] that this family is a complete weak scalarization for the corresponding vector problem, i.e.,

$$\arg\min_w\{F(x) \ : \ x \in \Omega\} = \bigcup_{z\in\mathbb{R}^m_+\setminus\{0\}} \arg\min\{\varrho_z(F(x)) \ : \ x \in \Omega\}.$$

The scalarization function (24) was used in the convergence analysis of the proximal methods proposed in [7] for a convex context. 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 [7, Remark 1]. This means that even the convexity assumption of the vector problem being sufficient to ensure the family given by (24) is a complete weak scalarization for the vector problem, it does not guarantee that the scalarized problem is nonempty.

Remark 8. Let $\varrho_z : \mathbb{R}^m \to \mathbb{R}$ be a scalarization given by

$$\varrho_z(y) := \inf\{t \in \mathbb{R} \ : \ y \in tz - \mathbb{R}^m_+\},$$

where $z \in \mathbb{R}^m_{++}$. Such a scalarization is a well-known functional in the vectorial optimization literature; see [27, section 3]. A particular instance of this functional was used in [4] to study the convergence of a proximal-point-type method for multiobjective optimization problems in the quasi-convex setting. A more general version of this scalarization functional was used in [19] for finding a weak Pareto point of an F $\mathcal{K}$-convex, for an ordering cone $\mathcal{K}$ by means of the following (exact) proximal method:

$$x^{k+1} \in \arg\min_{x\in X}\{\varrho_e(F(x)) + \frac{\lambda_k}{2}||x - x^k||^2\},$$

where $e \in \operatorname{int}\mathcal{K}$ and X is a Hilbert space. However, it is worth mentioning that in this kind of method, we have the following descent property

$$\varrho_e(F(x^{k+1})) \le \varrho_e(F(x^k)) \quad \forall k \in \mathbb{N}$$

which clearly does not imply the descent property $F(x^{k+1}) \preceq F(x^k)$ as proposed by [7] whereas the converse implication holds. As mentioned in the introduction the vectorial improvement $F(x^{k+1}) \preceq F(x^k)$ plays an important role in applications.

Acknowledgments. G. Bento, J.X. Cruz Neto, and J.C.O. Souza wish to express their gratitude to Professor Genaro López and Professor Antoine Soubeyran for their hospitality during the authors' visit to Universidad de Sevilla and Aix-Marseille University (Aix-Marseille School of Economics). The authors wish to express their gratitude to the anonymous referees for their helpful comments.

REFERENCES

[1] H. C. F. Apolinário, E. A. Papa Quiroz, and P. R. Oliveira, *A scalarization proximal point method for quasiconvex multiobjective minimization*, J. Global 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] S. Chrétien and A. O. Hero, *Generalized Proximal Point Algorithms and Bundle Implementations*, Technical Report 316, Department of EECS, University of Michigan, Ann Arbor, MI, 1998.

[6] 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.

[7] H. Bonnel, A. N. Iusem, and B. F. Svaiter, *Proximal methods in vector optimization*, SIAM J. Optim., 15 (2005), pp. 953–970.

[8] J. Branke, K. Deb, K. Miettinen, and R. Slowinski, eds., *Practical approaches to multiobjective optimization*, Dagstuhl Seminar, Dagstuhl, Wadern, Germany, 2007, 06501.

[9] 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.

[10] 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.

[11] L. C. Ceng and J. C. Yao, *Approximate proximal methods in vector optimization*, European J. Oper. Res., 183 (2007), pp. 1–19.

[12] L. C. Ceng, B. S. Mordukhovich, and J. C. Yao, *Hybrid approximate proximal method with auxiliary variational inequality for vector optimization*, J. Optim. Theory Appl., 146 (2010), pp. 267–303.

[13] T. D. Choung, 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.

[14] F. H. Clarke, *Generalized gradients and applications*, Trans. Amer. Math. Soc., 205 (1975), pp. 247–262.

[15] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Classics Appl. Math. 5, SIAM, Philadelphia, 1990.

[16] 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.

[17] 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.

[18] 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.

[19] M. Durea and R. Strugariu, *Some remarks on proximal point algorithm in scalar and vectorial cases*, Nonlinear Funct. Anal. Appl., 15 (2010), pp. 307–319.

[20] G. Eichfelder, *An adaptive scalarization method in multiobjective optimization*, SIAM J. Optim., 19 (2009), pp. 1694–1718.

[21] J. Fliege and B. F. Svaiter, *Steepest descent methods for multicriteria optimization*, Math. Methods Oper. Res., 51 (2000), pp. 479–494.

[22] J. Fliege, L. M. Graña Drummond, and B. F. Svaiter, *Newton's method for multiobjective optimization*, SIAM J. Optim., 20 (2009), pp. 602–626.

[23] 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.

[24] E. H. Fukuda and L. M. Graña Drummond, *A survey on multiobjective descent methods*, Pesquisa Oper., 34 (2014), pp. 585–620.

[25] W. B. Gearhart, *Compromise solutions and estimation of the non inferior set*, J. Optim. Theory Appl., 47 (1979), pp. 29–47.

[26] K. S. Goetzmann, *The Power of Compromise. Reference Point Methods and Approximation in Multicriteria Optimization*, Doctorial thesis, TU Berlin, 2013.

[27] A. Göpfert, H. Riahi, C. Tammer, and C. Zălinescu, *Variational Methods in Partially Ordered Spaces*, Springer, Berlin, 2003.
[28] L. M. Graña Drummond and A. N. Iusem, *A projected gradient method for vector optimization problems*, Comput. Optim. Appl., 28 (2004), pp. 5–29.
[29] L. M. Graña Drummond and B. F. Svaiter, *A steepest descent method for vector optimization*, J. Comput. Appl. Math., 175 (2005), pp. 395–414.
[30] 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.
[31] R. Gregório and P. R. Oliveira, *A logarithmic-quadratic proximal point scalarization method for multiobjective programming*, J. Global Optim., 49 (2011), pp. 281–291.
[32] X. X. Huang and X. Q. Yang, *Duality for multiobjective optimization via nonlinear Lagrangian functions*, J. Optim. Theory Appl., 120 (2004), pp. 111–127.
[33] J. Jahn, *Scalarization in vector optimization*, Math. Program., 29 (1984), pp. 203–218.
[34] M. Laumanns, L. Thiele, K. Deb, and E. Zitzler, *Combining convergence and diversity in evolutionary multiobjective optimization*, Evol. Comput., 10 (2002), pp. 263–282.
[35] B. Lemaire, *The Proximal Algorithm*, in New Methods in Optimization and Their Industrial Uses, J. P. Penot, ed., Internat. Ser. Numer. Math. 87, Birkhäuser, Boston, 1989, pp. 73–87.
[36] 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.
[37] K. Lewin, *Field Theory in Social Science*, Harper and Row, New York, 1951.
[38] D. T. Luc, *Theory of Vector Optimization*, Lecture Notes in Econom. and Math. Systems, Springer, New York, 1989.
[39] D. T. Luc, N. X. Tan, and P. N. Tinh, *Convex vector functions and their subdifferential*, Acta Math. Vietnam., 23 (1998), pp. 107–127.
[40] B. Martinet, *Régularisation d'inéquations variationelles par approximations successives*, RAIRO Oper. Res., 4 (1970), pp. 154–159.
[41] 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.
[42] J. J. Moreau, *Proximité et dualité dans un espace Hilbertien*, Bull. Soc. Math. France, 93 (1965) pp. 273–299.
[43] 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.
[44] S. Mostaghim, J. Branke, and H. Schmeck, *Multi-Objective Particle Swarm Optimization on Computer Grids*, Technical report 502, Institut AIFB, University of Karlsruhe, Karlsruhe, 2006.
[45] S. Opricovic and G-H. Tzeng, *Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS*, European J. Oper. Res., 156 (2004), pp. 445–455.
[46] S. Opricovic and G-H. Tzeng, *Extended VIKOR method in comparison with outranking methods*, European J. Oper. Res., 178 (2007), pp. 514–529.
[47] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control. Optim., 14 (1976), pp. 877–898.
[48] A. Soubeyran, *Variational Rationality, a Theory of Individual Stability and Change: Worthwhile and Ambidextry Behaviors*, preprint, GREQAM, Aix Marseillle University, Marseille, 2009.
[49] A. Soubeyran, *Variational Rationality, and the "Unsatisfied Man": Routines and the Course Pursuit Between Aspirations, Capabilities and Beliefs*, preprint, GREQAM, Aix Marseille University, Marseille, 2010.
[50] A. Soubeyran, *Variational Rationality. A Theory of Worthwhile Stay and Change Approach-Avoidance Transitions Ending in Traps*, preprint, GREQAM-AMSE, Aix Marseille University, Marseille, 2016.
[51] L. Thibault, *Subdifferentials of nonconvex vector-valued functions*, J. Math. Anal. Appl., 86 (1982), pp. 319–344.
[52] K. D. V. Villacorta and P. R. Oliveira, *An interior proximal method in vector optimization*, European J. Oper. Res., 214 (2011), pp. 485–492.