

Inexact Multi-Objective Local Search Proximal Algorithms: Application to Group Dynamic and Distributive Justice Problems

Glaydston de Carvalho Bento¹ ·
Orizon Pereira Ferreira¹ · Antoine Soubeyran² ·
Valdinês Leite de Sousa Júnior¹

Abstract We introduce and examine an inexact multi-objective proximal method with a proximal distance as the perturbation term. Our algorithm utilizes a local search descent process that eventually reaches a weak Pareto optimum of a multi-objective function, whose components are the maxima of continuously differentiable functions. Our algorithm gives a new formulation and resolution of the following important distributive justice problem in the context of group dynamics: In each period, if a group creates a cake, the problem is, for each member, to get a high enough share of this cake; if this is not possible, then it is better to quit, breaking the stability of the group.

Keywords Multi-objective · Inexact proximal · Group dynamic · Distributive justice · Behavioral sciences · Variational rationality

Mathematics Subject Classification 90C29 · 90C30 · 49M30

✉ Valdinês Leite de Sousa Júnior
valdineslds@gmail.com
Glaydston de Carvalho Bento
glaydstonc@gmail.com
Orizon Pereira Ferreira
orizon@ufg.br
Antoine Soubeyran
antoine.soubeyran@gmail.com

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

² CNRS & EHESS, Aix-Marseille School of Economics, Aix-Marseille University, Marseille, France

1 Introduction

We propose an inexact multi-objective proximal algorithm with a strong (semi) local search aspect that eventually reaches a Pareto–Clarke critical point of a nonnecessary differentiable multi-objective function, whose regularization term is a proximal distance. The motivation for this paper comes from behavioral sciences. Our algorithm gives a new formulation and resolution of the following famous distributive justice problem in the context of group dynamics: In each period, if a group creates a cake, the problem is, for each member, to get a high enough share of this cake; if this is not possible, then it is better to quit, breaking the stability of the group. In a dynamic context, this problem is an instance of the very general and famous “Value creation, value appropriation problem,” which plays a central role in management science, economics, psychology, sociology, game theory, and political science, where value creation problems cannot be determined independently of value appropriation problems, such as distributive justice and bargaining problems. Our approach considers the Fréchet subdifferentials introduced in [1], which have, in the convex case, an important relationship with the directional derivative, which will be useful in the convergence analysis of the method.

Methods such as the proximal point algorithm are descent processes. In the scalar case, this property is an immediate consequence of the definition of the method. However, in higher instances, to reach this property, additional assumptions must be made. To solve this issue, Bonnel et al. [2] introduced an omega improving set as in (9), to analyze inexact (and exact) proximal algorithms in vector optimization. In that approach, the convexity of each omega improving set plays an important role, since, for convex sets, the normal cone has a well-known structure, allowing a clearer analysis of the method. However, since our objective function is not a convex function, the set in (9) is not convex. Recently, Bento et al. [3] presented the exact version of the scheme defined in (9), under the convexity of the omega set. However, this assumption seems to be rather restrictive (see Examples 4.1 and 4.2). To solve this issue, we analyze the tangent cone of an omega set at the current iterate, which allows us to reach a structure similar to that in [2]. We emphasize that our analysis holds without requiring the convexity of the set in (9).

The algorithm we choose mixes three different kinds of well-known methods from the literature: local search, adaptive, and multi-objective proximal algorithms. That is, we consider three points:

- Local search algorithm: There is an enormous amount of literature on this topic, but the multi-objective case seems to be largely underdeveloped; see Aarts and Lenstra [4]. A local search multi-objective subproblem is a problem is to find a subset of neighbors of the current iterate. For example, the neighborhood of a vertex cover is another vertex cover only differing by one node. At the previous stage, a local search method starts from a candidate solution and then iteratively moves to a neighbor solution in an adaptive neighborhood structure that depends on each stage. The justification of the term local search is the following: Every candidate solution has more than one neighbor solution; the choice of the next move is taken using only information about the solutions in the neighborhood

of the current one. When the choice of the neighbor solution is made by taking the one locally maximizing the criterion, the meta-heuristic takes the name hill climbing. When no improving configurations are present in the neighborhood, the local search is stuck at a locally optimal point. This local optima problem can be cured by using restarts (repeating the local search with different initial conditions) or more complex schemes based on iterations (e.g., an iterated local search), on memory (e.g., reactive search optimization), or on memoryless stochastic modifications (e.g., simulated annealing). Local search methods include hill-climbing search, simulated annealing search, local beam search, genetic methods, ant colony optimization, and tabu search.

- **Multi-objective proximal algorithms:** This method adds a perturbation term to the initial vectorial function. The term is chosen in such a way that the proximal vectorial function becomes simple enough, that is, linear or convex in all its components. Hence, the perturbed multi-objective program must be rather easy to optimize. In our case, we get convex vectorial proximal payoffs. This is a very important point. It entirely justifies our method, which replaces a very complicated initial multi-objective optimization problem by a succession of much simpler convex multi-objective perturbed optimization problems. Furthermore, the multi-objective proximal method considers a sequence of intermediate steps and a succession of intermediate subproblems. The algorithm requires that the current solution of each subproblem exists, be unique (if possible), and be rather easy to calculate. In this case, each successive subproblem is well defined and implementable.
- **Semi-local search multi-objective proximal algorithms:** In the case of our specific context, the neighborhood structure is semi-local, that is, not restricted to neighbors. In the scalar case, a class of semi-local search proximal algorithms has been examined in a metric space by Attouch and Soubeyran [5]. The motivation to study these semi-local search proximal methods came directly from a recent variational VR approach (Soubeyran, [6–8]). Hence, our process is a semi-local search inexact multi-objective proximal method with a nonconvex neighborhood and adaptive structure. The motivation to study such an algorithm comes from behavioral sciences (e.g., economics, psychology, and sociology), where the problem of group dynamics and distributive justice is crucial. The static Rawls problem of distributive justice is to maximize the weighted sum of the different objectives of a group of agents (the efficiency aspect, to get the larger pie for the group) as well as to protect the shares of the poor agents of the group (the distributive justice aspect).

The organization of the paper is as follows. In Sect. 2, some notations and basic results used throughout the paper are presented. In Sect. 3, we introduce the definition of proximal distance and its induced proximal distance, study the main properties of our objective function, and present some results of multi-objective optimization. Our inexact multi-objective proximal method is presented in Sect. 4, and the main results are stated and proved. We will show, in Sect. 5, that our multi-objective program, where each objective is the maximum of a subset of subobjectives, models the famous static Rawls distributive justice problem. In Sect. 6, we consider the also famous Lewin group dynamic problem, adding a dynamic distributive justice component, where the problem is to know how to reach, step by step, a static distributive justice solution as

a Pareto point. Hence, we show how our algorithm models, in a crude but compelling way, a dynamic group process that converges to a solution of the static Rawlsian max-min problem of distributive justice. Some final remarks are made in Sect. 7.

2 Basic Approach and Notation

In this section, we present several concepts of nonsmooth analysis that will be useful throughout this presentation. Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be finite at a point \bar{x} and $\varepsilon \geq 0$. The set

$$\hat{\partial}_\varepsilon f(\bar{x}) := \left\{ w \in \mathbb{R}^n : \liminf_{x \rightarrow \bar{x}} \frac{f(x) - f(\bar{x}) - \langle w, x - \bar{x} \rangle}{\|x - \bar{x}\|} \geq -\varepsilon \right\}$$

is the *analytic ε -subdifferential* [1, Definition 1.83, page 87] of f at \bar{x} with the elements referred to as *analytic ε -subgradients* of f at \bar{x} . If $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is a convex function, the analytic ε -subdifferential can be described in more detail, i.e., if $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is convex and finite at a point \bar{x} , then

$$\hat{\partial}_\varepsilon f(\bar{x}) = \{w \in \mathbb{R}^n : \langle w, x - \bar{x} \rangle \leq f(x) - f(\bar{x}) + \varepsilon \|x - \bar{x}\|, x \in \mathbb{R}^n\}; \quad (1)$$

see [1, Theorem 1.93, page 95]. Note that the set on the right-hand side of (1) is the subdifferential of the convex function $f(x) + \varepsilon \|x - \bar{x}\|$ at \bar{x} . By the classical Moreau-Rockafellar theorem, this set is equal to $\partial f(\bar{x}) + \varepsilon \mathbb{B}$ for any proper convex function $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$, where \mathbb{B} denotes the closed unit ball in \mathbb{R}^n . Observe that for $\varepsilon > 0$, the latter set is different from the classical ε -subdifferential of convex analysis defined as the collection of $w \in \mathbb{R}^n$ satisfying $\langle w, x - \bar{x} \rangle \leq f(x) - f(\bar{x}) + \varepsilon$, $x \in \mathbb{R}^n$; see Hiriart-Urruty and Lemaréchal [9]. Based on (1), it is straightforward to check that the analytic ε -subdifferential for convex functions has the following representation:

$$\hat{\partial}_\varepsilon f(\bar{x}) = \{w \in \mathbb{R}^n : f'(\bar{x}; v) \geq \langle w, v \rangle - \varepsilon \|v\|, v \in \mathbb{R}^n\}, \quad (2)$$

where $f'(\bar{x}; v)$ denotes the usual directional derivative.

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a locally Lipschitz function at $\bar{x} \in \mathbb{R}^n$ with constant $L > 0$ and $v \in \mathbb{R}^n$. The *Clarke's directional derivative* [10, page 25] of f at \bar{x} in the direction v , denoted by $f^\circ(\bar{x}; v)$, is defined as

$$f^\circ(\bar{x}; v) := \limsup_{t \downarrow 0, y \rightarrow \bar{x}} \frac{f(y + tv) - f(y)}{t},$$

and *Clarke's subdifferential* [10, page 27] of f at \bar{x} , denoted by $\partial^\circ f(\bar{x})$, is defined as

$$\partial^\circ f(\bar{x}) := \{w \in \mathbb{R}^n : f^\circ(\bar{x}; v) \geq \langle w, v \rangle, v \in \mathbb{R}^n\}.$$

The function $v \mapsto f^\circ(\bar{x}; v)$ is finite, positive homogeneous, i. e., $f^\circ(\bar{x}; \lambda v) = \lambda f^\circ(\bar{x}; v)$, and subadditive, i. e., $f^\circ(\bar{x}; v+w) \leq f^\circ(\bar{x}; v) + f^\circ(\bar{x}; w)$, for all $\lambda > 0$ and

$v, w \in \mathbb{R}^n$. Moreover, $f^\circ(\bar{x}; v)$ is upper semi-continuous as a function of (\bar{x}, v) , i.e., if $\{(x^k, v^k)\}$ is a sequence in $\mathbb{R}^n \times \mathbb{R}^n$ that converges to some point $(x, v) \in \mathbb{R}^n \times \mathbb{R}^n$, then

$$\limsup_{k \rightarrow +\infty} f^\circ(x^k; v^k) \leq f^\circ(x; v); \quad (3)$$

see [10, page 25, Proposition 2.1.1]. Another well-known property of Clarke's directional derivative is the following: given two locally Lipschitz functions $f_1, f_2 : \mathbb{R}^n \rightarrow \mathbb{R}$ at $\bar{x} \in \mathbb{R}^n$, we have

$$(f_1 + f_2)^\circ(\bar{x}; v) \leq f_1^\circ(\bar{x}; v) + f_2^\circ(\bar{x}; v) \quad (4)$$

for all $v \in \mathbb{R}^n$; see [10, page 38, Proposition 2.3.3].

Next, we recall the definition of the tangent and the normal cones presented in [10]. Given a nonempty closed set $\Omega \subset \mathbb{R}^n$, the *tangent cone* [10, page 53, Theorem 2.4.5] $T_\Omega(\bar{x})$ to Ω at a point $\bar{x} \in \Omega$ is defined as follows:

$$v \in T_\Omega(\bar{x}) \iff \begin{cases} \text{for every sequence } \{x^k\} \subset \Omega \text{ converging to } \bar{x} \text{ and every sequence} \\ \{t_k\} \subset]0, +\infty[\text{ converging to } 0, \text{ there is a sequence } \{v^k\}, \\ \text{converging to } v \text{ such that, } x^k + t_k v^k \text{ belongs to } \Omega \text{ for all } k. \end{cases}$$

An immediate consequence of the definition is that $T_\Omega(\bar{x})$ is a closed convex cone in \mathbb{R}^n . Having defined a tangent cone, the likely candidate for the normal cone is the one obtained from $T_\Omega(\bar{x})$ by polarity. We define $N_\Omega(\bar{x})$, the *normal cone* [10, page 51] to Ω at \bar{x} , as $N_\Omega(\bar{x}) := \{w \in \mathbb{R}^n : \langle w, v \rangle \leq 0, v \in T_\Omega(\bar{x})\}$. For any nonempty set $\Omega \subset \mathbb{R}^n$, δ_Ω denotes the *indicator function* of Ω . Note that δ_Ω is a convex function iff Ω is a convex set, and δ_Ω is l.s.c. iff Ω is closed.

3 Basic Framework for a Distributive Justice Multi-objective Program

In our approach, we choose a proximal distance $d : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ as the regularization term. Such a well-known distance allows us to analyze the convergence of the algorithm under various settings. Following [11], let us recall the definition of the proximal and induced proximal distances.

Definition 3.1 A function $d : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ is called a proximal distance with respect to an open nonempty convex set $C \subset \mathbb{R}^n$ if for each $y \in C$ it satisfies the following properties:

- (d1) $d(\cdot, y)$ is proper, lsc, convex, and C^1 on C ;
- (d2) $\text{dom } d(\cdot, y) \subset \bar{C}$ and $\text{dom } \partial_1 d(\cdot, y) = C$, where $\partial_1 d(\cdot, y)$ denotes the subgradient map of the function $d(\cdot, y)$ with respect to the first variable;
- (d3) $d(\cdot, y)$ is level-bounded on \mathbb{R}^n , i.e., $\lim_{\|u\| \rightarrow +\infty} d(u, y) = +\infty$;
- (d4) $d(y, y) = 0$.

For each $y \in C$, let $\nabla_1 d(\cdot, y)$ denote the gradient map of the function $d(\cdot, y)$ with respect to the first variable. Besides the above conditions, we consider that such a d satisfies the following property:

(d5) there exists $\sigma > 0$ such that for all $\tilde{x} \in C$, $d(\cdot, \tilde{x})$ is σ -strongly convex over C , i.e.,

$$\langle \nabla_1 d(x, \tilde{x}) - \nabla_1 d(y, \tilde{x}), x - y \rangle \geq \sigma \|x - y\|^2, \quad x, y \in C,$$

for some norm $\|\cdot\|$ in \mathbb{R}^n .

We denote by $\mathcal{D}(C)$ the family of functions d satisfying (d1)-(d5). Next, following the approach presented in [11], we associate to a given $d \in \mathcal{D}(C)$ a corresponding induced distance H that satisfies some desirable properties.

Definition 3.2 Given $C \subset \mathbb{R}^n$, open and convex, and $d \in \mathcal{D}(C)$, a function $H : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ is called the induced proximal distance to d if H is finite-valued on $C \times C$ and for each $x, y \in C$ satisfies the following properties:

(H1) $H(x, x) = 0$;

(H2) $\langle z - y, \nabla_1 d(y, x) \rangle \leq H(z, x) - H(z, y)$, $z \in C$.

Hereafter, C is an open nonempty convex set $C \subset \mathbb{R}^n$. We write $(d, H) \in \Phi(C)$ to quantify the triple $[C, d, H]$ that satisfies the premises of Definition 3.2. Similarly, we write $(d, H) \in \Phi(\bar{C})$ for the triple $[\bar{C}, d, H]$ whenever there exists H , which is finite-valued on $\bar{C} \times C$, satisfies (H1)-(H2) for any $z \in C$, and is such that $z \in \bar{C}$ has $H(z, \cdot)$ level-bounded on C . Clearly, one has $\Phi(\bar{C}) \subseteq \Phi(C)$. For examples and a thorough discussion about proximal and induced proximal distances, see, for instance, [11, 12]. We adopt the standard notation in multi-objective optimization. We denote $I := \{1, \dots, m\}$. Further, the sets of n -vectors with nonnegative (positive) real components are denoted by \mathbb{R}_+^n (\mathbb{R}_{++}^n). For each $x, y \in \mathbb{R}^n$, we have $x \preceq y \iff y - x \in \mathbb{R}_+^n$ and $x < y \iff y - x \in \mathbb{R}_{++}^n$. Similarly, we can define this relation with the reverse inequality. Before we present our most significant results, let us introduce the notion of Pareto–Clarke critical point of a multi-objective function. Let D be a closed and convex set and $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be locally Lipschitz on \mathbb{R}^n . We say that a point $x^* \in D$ is a Pareto–Clarke critical point of F in D if, for any $y \in D$, there exists $j_0 \in I$ such that $f_{j_0}^\circ(x^*; y - x^*) \geq 0$. We denote by $\mathcal{S}_D^*(F)$, the set of the Pareto–Clarke critical points of F in D ; see [13]. In this context, we have the following result, which is very useful in our convergence analysis.

Proposition 3.1 Let D be a closed and convex set and $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be locally Lipschitz on \mathbb{R}^n . If Ω is defined as $\{y \in \mathbb{R}^n : F(y) \preceq F(x)\}$, then for all $\bar{z} \in \Omega \setminus \mathcal{S}_D^*(F)$,

$$\{v \in \mathbb{R}^n : F^\circ(\bar{z}; v) \leq 0\} \subset T_\Omega(\bar{z}), \quad (5)$$

where $F^\circ(\bar{z}; v) := (f_1^\circ(\bar{z}; v), \dots, f_m^\circ(\bar{z}; v))$.

Proof First, let us prove that

$$\{v \in \mathbb{R}^n : F^\circ(\bar{z}; v) < 0\} \subset T_\Omega(\bar{z}). \quad (6)$$

holds. Consider any $v \in \mathbb{R}^n$, such that $F^\circ(\bar{z}; v) < 0$. Hence, $f_j^\circ(\bar{z}; v) < 0$, $j \in I$. It follows from the definition of $f_j^\circ(\bar{z}; v)$ that there are δ and $\varepsilon > 0$ such that, for

all $y \in B(\bar{z}, \delta)$ and $t \in]0, \varepsilon[$, $f_j(y + tv) - f_j(y) \leq -\varepsilon t$, $j \in I$. Now, let $\{x^k\}$ be any sequence in Ω converging to \bar{z} , and $\{t_k\}$ be any sequence decreasing to 0. By the definition of Ω , $f_j(x^k) \leq f_j(x)$ for all $j \in I$, and for all k sufficiently large,

$$f_j(x^k + t_k v) \leq f_j(x^k) - \varepsilon t \leq f_j(x) - \varepsilon t, \quad j \in I.$$

It follows that $\{x^k + t_k v\}$ (for a large k) belongs to Ω , and this establishes that $v \in T_\Omega(\bar{z})$. Now, let us prove (5). Since \bar{z} is not a Pareto–Clarke critical point of F , there exists $v_0 \in D$ such that $f_j^\circ(\bar{z}; v_0 - \bar{z}) < 0$, for all $j \in I$. If v belongs to the left-hand side of (5), then for all $k \in \mathbb{N}$, $f_j^\circ(\bar{z}; v + (1/k)(v_0 - \bar{z})) < 0$, since $f_j^\circ(\bar{z}; \cdot)$ is subadditive. Based on (6), $v + (1/k)(v_0 - \bar{z}) \in T_\Omega(\bar{z})$. As $T_\Omega(\bar{z})$ is a closed set and $\{v + (1/k)(v_0 - \bar{z})\}$ converges to v , we have $v \in T_\Omega(\bar{z})$. \square

The objective of this paper is to study the convergence of an inexact multi-objective proximal method for a particular class of multi-objective functions that are not necessarily differentiable. Let us introduce such a special class of multi-objective functions, which represent distributive justice payoffs in behavioral sciences (see Sect. 5), and study their main properties. Set $I_j := \{1, \dots, \ell_j\}$, with $\ell_j \in \mathbb{Z}_+$, for $j \in I$. Let $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a multi-objective function $F(x) := (f_1(x), \dots, f_m(x))$, where

$$f_j(x) := \max_{i \in I_j} f_{ij}(x), \quad j \in I, \quad (7)$$

$f_{ij} : \mathbb{R}^n \rightarrow \mathbb{R}$ is a continuously differentiable function on C and continuous on \bar{C} (closure of C), for all $i \in I_j$. We can easily see from (7) that each $f_j : \mathbb{R}^n \rightarrow \mathbb{R}$ is locally Lipschitz on C . Throughout this paper, we make the following assumptions:

- A1.** For all $j \in I$, $-\infty < \inf_{x \in \mathbb{R}^n} f_j(x)$;
- A2.** For all $j \in I$, ∇f_{ij} is Lipschitz continuous on C with constant $L_{ij} \geq 0$ for each $i \in I_j$.

Functions with the structure defined in (7) are very flexible when we are working with algorithms such as proximal point type. In general, such functions are not convex functions. However, the next result shows us that, under certain conditions, its proximal regularization is a convex function.

Lemma 3.1 *Consider any $d \in \mathcal{D}(C)$, $\tilde{x} \in C$, $\lambda \in \mathbb{R}_{++}$ and $\omega := (\omega_1, \dots, \omega_m) \in \mathbb{R}_{++}^m$, satisfying*

$$\sup_{i \in I_j} L_{ij} < \lambda \sigma \omega_j, \quad j \in I. \quad (8)$$

If $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a multi-objective function as in (7), then the functions $P_{ij} := f_{ij} + \lambda \omega_j d(\cdot, \tilde{x})$ and $P_j := f_j + \lambda \omega_j d(\cdot, \tilde{x})$ are strongly convex on C . Moreover, the function $\langle F(\cdot), z \rangle + \lambda \langle \omega, z \rangle d(\cdot, \tilde{x})$ is strongly convex on C for each $z \in \mathbb{R}_+^m \setminus \{0\}$.

Proof Consider $j \in I$, $i \in I_j$, and $x, y \in C$. Since $\nabla P_{ij}(x) = \nabla f_{ij}(x) + \lambda \omega_j \nabla_1 d(x, \tilde{x})$, we have

$$\begin{aligned} \langle \nabla P_{ij}(x) - \nabla P_{ij}(y), x - y \rangle &= \langle \nabla f_{ij}(x) - \nabla f_{ij}(y), x - y \rangle \\ &\quad + \lambda \omega_j \langle \nabla_1 d(x, \tilde{x}) - \nabla_1 d(y, \tilde{x}), x - y \rangle. \end{aligned}$$

Using property **(d5)** and Cauchy inequality, the last equality becomes

$$\langle \nabla P_{ij}(x) - \nabla P_{ij}(y), x - y \rangle \geq -\|\nabla f_{ij}(x) - \nabla f_{ij}(y)\| \|x - y\| + \lambda \sigma \omega_j \|x - y\|^2.$$

As ∇f_{ij} is Lipschitz on C with constant L_{ij} , the last inequality gives us

$$\langle \nabla P_{ij}(x) - \nabla P_{ij}(y), x - y \rangle \geq (\lambda \sigma \omega_j - L_{ij}) \|x - y\|^2.$$

Combining last inequality with (8), we conclude that P_{ij} is strongly convex. Since $\max_{i \in I_j} P_{ij} = P_j$, we deduce the strong convexity of the function P_j . It is straightforward to check the strong convexity of the function $\langle F(\cdot), z \rangle + \lambda \langle \omega, z \rangle d(\cdot, \tilde{x})$. \square

4 An Inexact Multi-objective Proximal Method

In this section, we present an inexact multi-objective proximal method. Hereafter, $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is the multi-objective function as defined in (7) and it satisfies **A1** and **A2**. Let $d \in \mathcal{D}(C)$, $\{\lambda_k\}, \{\varepsilon_k\} \subset \mathbb{R}_{++}$, $\{e^k\} \subset \mathbb{R}_{++}^m$, and $\{z^k\} \subset \mathbb{R}_+^m \setminus \{0\}$. Choose $x^0 \in C$. Given $x^k \in C$, if $x^k \in \mathcal{S}_C^*(F)$, then set $x^{k+p} = x^k$ for all $p \geq 1$. If $x^k \notin \mathcal{S}_C^*(F)$, the *inexact multi-objective proximal point algorithm* formally generates a sequence $\{x^k\}$ as follows

$$0 \in \hat{\partial}_{\varepsilon_k} \left(\left\langle F(\cdot), z^k \right\rangle + \gamma_k d(\cdot, x^k) \right) (x^{k+1}) + \gamma_k \nabla_1 d(x^{k+1}, x^k) + N_{\Omega_k}(x^{k+1}), \quad (9)$$

$k = 0, 1, \dots$, where $\Omega_k := \{x \in \mathbb{R}^n : F(x) \leq F(x^k)\}$ and $\gamma_k := \lambda_k \langle e^k, z^k \rangle$.

Remark 4.1 If $\varepsilon_k = 0$, and $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a convex multi-objective function, then inclusion (9) implies that

$$0 \in \partial \left(\left\langle F(\cdot), z^k \right\rangle + 2\gamma_k d(\cdot, x^k) \right) (x^{k+1}) + N_{\Omega_k}(x^{k+1}), \quad k = 0, 1, \dots$$

Using standard arguments of analysis, we can conclude that

$$x^{k+1} \in \operatorname{argmin}_w \left\{ F(x) + 2\lambda_k d(x, x^k) e^k : x \in \Omega_k \right\},$$

which is the finite dimensional version, with the respective choices of d , of the exact proximal point method for multi-objective optimizations presented in [2, 14], where

$$\operatorname{argmin}_w G(x) := \{x^* \in \mathbb{R}^n : \text{there is no } \bar{y} \in \mathbb{R}^n \text{ such that } G(\bar{y}) < G(x^*)\}.$$

Before we introduce the main theorem of the present paper, we recall the following well-known results of nonnegative sequences.

Lemma 4.1 (see [15]) Let $\{u^k\}$, $\{\alpha_k\}$, and $\{\beta_k\}$ be nonnegative sequences of real numbers satisfying $u^{k+1} \leq (1 + \alpha_k)u^k + \beta_k$ such that $\sum_k \alpha_k < \infty$, $\sum_k \beta_k < \infty$. Then, the sequence $\{u^k\}$ converges.

Lemma 4.2 (see [15]) Let $\{\lambda_k\}$ be a sequence of positive numbers, $\{a_k\}$ a sequence of real numbers, and $b_n := \sigma_n^{-1} \sum_{k=1}^n \lambda_k a_k$, where $\sigma_n := \sum_{k=1}^n \lambda_k$. If $\sigma_n \rightarrow \infty$, $\liminf a_n \leq \liminf b_n \leq \limsup b_n \leq \limsup a_n$.

Theorem 4.1 Let $\lambda_+ > 0$, $\mu > 0$, $(d, H) \in \Phi(\bar{C})$, and $\sigma > 0$ be the constant that appears in the strong convexity of d . Assume that $\lambda_k \in \mathbb{R}_{++}$, $\{e^k\} \subset \mathbb{R}_{++}^m$, and $\{z^k\} \subset \mathbb{R}_+^m \setminus \{0\}$ satisfy

$$\|z^k\|_1 = 1, \quad \|e^k\| = 1, \quad \mu < e_j^k, \quad (1/\sigma\mu) \max_{i \in I_j} L_{ij} < \lambda_k \leq \lambda_+, \quad (10)$$

$j \in I$, $k = 0, 1, \dots$. Then, the sequence $\{x^k\}$ generated by (9) is well defined and $x^{k+1} \in C \cap \Omega_k$, for $k = 0, 1, \dots$. Moreover, if $\{x^k\}$ is bounded and $\{\varepsilon_k\} \subset \mathbb{R}_{++}$ is such that $\sum_k \varepsilon_k < \infty$, then each accumulation point of $\{x^{k+1}\}$ is a Pareto–Clarke critical point of F in \bar{C} .

Proof We prove the well-definition by induction over k . Let $k = 0$ and $x^0 \in C$. Define $\varphi_0 : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ by $\varphi_0(x) = \psi(x) + \delta_{\Omega_0}(x)$, where

$$\psi(x) = \begin{cases} \langle F(\cdot), z^0 \rangle + \gamma_0 d(\cdot, x^0), & x \in C, \\ +\infty, & x \notin C. \end{cases}$$

Owing to $-\infty < \inf_{x \in \mathbb{R}^n} f_j(x)$, for all $j \in I$, $z^0 \in \mathbb{R}_+^m \setminus \{0\}$, the function $x \mapsto \langle F(x), z^0 \rangle$ is bounded below, and taking into account that $\gamma_0 > 0$ and $d(\cdot, x^0)$ satisfies (d3) of Definition 3.1, it follows that φ_0 is coercive. Then, as Ω_0 is closed, we can apply [16, Theorem 1.9, page 11] to ensure that there exists $\tilde{x} \in \Omega_0$ such that $\tilde{x} \in \operatorname{argmin} \{\psi(x) + \gamma_0 d(x, x^0) : x \in \Omega_0\}$. From [10, Corollary, page 52], we have $0 \in \partial^\circ (\psi(x) + \gamma_0 d(x, x^0))(\tilde{x}) + N_{\Omega_0}(\tilde{x})$. As (10) holds, Lemma 3.1 implies that $\langle F(\cdot), z^0 \rangle + \gamma_0 d(\cdot, x^0)$ is a convex function on C and consequently ψ is convex. Since $\operatorname{dom} \psi \cap \operatorname{dom}(\gamma_0 d(\cdot, x^0)) = C \neq \emptyset$, we can apply [17, Theorem 23.8, page 223] to obtain $0 \in \partial \psi(\tilde{x}) + \gamma_0 \partial_1 d(\tilde{x}, x^0) + N_{\Omega_0}(\tilde{x})$. Taking into account that $\operatorname{dom} \partial_1 d(\cdot, y) = C$, it follows from the last inclusion that $\tilde{x} \in C$. Since $x \mapsto \langle F(x), z^0 \rangle + \gamma_k d(x, x^0)$ is a convex function on C , using (2) we have

$$\partial \left(\langle F(x), z^0 \rangle + \gamma_0 d(x, x^0) \right) (\tilde{x}) \subset \hat{\partial}_{\varepsilon_0} \left(\langle F(x), z^0 \rangle + \gamma_0 d(x, x^0) \right) (\tilde{x}),$$

with $\varepsilon_0 > 0$. Therefore, we can consider $x^1 := \tilde{x} \in C \cap \Omega_0$. Thus, the method is well defined for $k = 1$. The proof of the induction step proceeds similarly. Therefore, method (9) is well defined and $x^{k+1} \in C \cap \Omega_k$, for $k = 0, 1, \dots$. Now, let us prove the second part of the theorem. Based on (9), there exists $v^{k+1} \in N_{\Omega_k}(x^{k+1})$ satisfying

$$-v^{k+1} - \gamma_k \nabla_1 d(x^{k+1}, x^k) \in \hat{\partial}_{\varepsilon_k} \left(\langle F(\cdot), z^k \rangle + \gamma_k d(\cdot, x^k) \right) (x^{k+1}), \quad k \geq 0. \quad (11)$$

Consider any $y_0 \in \bar{C}$ as fixed. We claim that

$$\liminf_{k \rightarrow +\infty} \left(- \sum_{j=1}^m z_j^k f_j^\circ \left(x^{k+1}; y_0 - x^{k+1} \right) - \left\langle v^{k+1}, y_0 - x^{k+1} \right\rangle \right) \leq 0. \quad (12)$$

Indeed, since (10) holds, we can apply Lemma 3.1 to conclude that $x \mapsto G_k(x) := \langle F(x), z^k \rangle + \gamma_k d(x, x^k)$ is a convex function on C . Then, we can combine (11) with (2) to obtain

$$G'_k(x^{k+1}; y_0 - x^{k+1}) \geq \left\langle -v^{k+1} - \gamma_k \nabla_1 d(x^{k+1}, x^k), y_0 - x^{k+1} \right\rangle - \varepsilon_k \|y_0 - x^{k+1}\|, \quad k \geq 0.$$

From (H2), with $z = y_0$, $y = x^{k+1}$, $x = x^k$, we have

$$\langle y_0 - x^{k+1}, \nabla_1 d(x^{k+1}, x^k) \rangle \leq H(y_0, x^k) - H(y_0, x^{k+1}), \quad k \geq 0. \quad (13)$$

Since $(d(\cdot, x^k))^\circ(x^{k+1}, v) = \langle \nabla_1 d(x^{k+1}, x^k), v \rangle$ holds for all $v \in \mathbb{R}^n$, it follows from (4) that

$$G'_k(x^{k+1}; y_0 - x^{k+1}) \leq \sum_{j=1}^m z_j^k f_j^\circ \left(x^{k+1}; y_0 - x^{k+1} \right) + \gamma_k \left\langle y_0 - x^{k+1}, \nabla_1 d(x^{k+1}, x^k) \right\rangle, \quad k \geq 0.$$

Combining the last inequality with (13) and (13), for all $k \geq 0$ we obtain

$$\begin{aligned} & H(y_0, x^k) - H(y_0, x^{k+1}) + \|y_0 - x^{k+1}\| \frac{\varepsilon_k}{2\gamma_k} \\ & \geq \frac{1}{2\gamma_k} \left(- \sum_{j=1}^m z_j^k f_j^\circ \left(x^{k+1}; y_0 - x^{k+1} \right) - \left\langle v^{k+1}, y_0 - x^{k+1} \right\rangle \right). \end{aligned}$$

As $\{x^k\}$ is bounded, there exists $\alpha > 0$ such that $\|y_0 - x^{k+1}\| \leq \alpha$, $k \geq 0$. Since $\|z^k\|_1 = 1$ and $\{e^k\}$ and $\{\lambda_k\}$ satisfy (10), we have

$$\gamma_k = \lambda_k \langle e^k, z^k \rangle = \lambda_k \sum_{j=1}^m e_j^k z_j^k \geq \mu(1/\sigma\mu) \max_{i \in I_j} L_{ij} \sum_{j=1}^m z_j^k = 1/\sigma \max_{i \in I_j} L_{ij}.$$

Hence, $1/\gamma_k \leq (1/\sigma \max_{i \in I_j} L_{ij})^{-1}$, $k \geq 0$. Then, there exists $M > 0$ satisfying

$$\begin{aligned} & H(y_0, x^k) - H(y_0, x^{k+1}) + M\varepsilon_k \\ & \geq \frac{1}{2\gamma_k} \left(- \sum_{j=1}^m z_j^k f_j^\circ(x^{k+1}; y_0 - x^{k+1}) - \langle v^{k+1}, y_0 - x^{k+1} \rangle \right), \quad k \geq 0. \end{aligned}$$

Summing the last inequality over $k = 1, \dots, n$, for all $k \geq 0$, we have

$$\begin{aligned} & H(y_0, x^1) - H(y_0, x^{n+1}) + M \sum_{k=1}^n \varepsilon_k \\ & \geq \sum_{k=1}^n \frac{1}{2\gamma_k} \left(- \sum_{j=1}^m z_j^k f_j^\circ(x^{k+1}; y_0 - x^{k+1}) - \langle v^{k+1}, y_0 - x^{k+1} \rangle \right). \end{aligned}$$

Since $H(\cdot, \cdot) \geq 0$, for all $k \geq 0$ we obtain

$$\begin{aligned} & \sigma_n^{-1} H(y_0, x^1) + M \sigma_n^{-1} \sum_{k=1}^n \varepsilon_k \\ & \geq \sigma_n^{-1} \sum_{k=1}^n \frac{1}{2\gamma_k} \left(- \sum_{j=1}^m z_j^k f_j^\circ(x^{k+1}; y_0 - x^{k+1}) - \langle v^{k+1}, y_0 - x^{k+1} \rangle \right), \end{aligned}$$

where $\sigma_n := \sum_{k=1}^n (1/2\gamma_k)$. As $1/\gamma_k \geq 1/\lambda_+$, then $\sigma_n \rightarrow \infty$, and considering that $\sum_{k=1}^{\infty} \varepsilon_k < \infty$, we can use Lemma 4.2 to obtain (12).

Now, consider any point x^* such that $\lim_{\ell \rightarrow +\infty} x^{k_\ell+1} = x^*$. Suppose, by contradiction, that x^* is not a Pareto–Clarke critical point of F in \bar{C} . Then, there exists $y_0 \in \bar{C}$ such that

$$f_j^\circ(x^*; y_0 - x^*) < 0, \quad j \in I. \quad (14)$$

From (14) and (3), there exists $\ell_0 \in \mathbb{N}$ such that for all $\ell \geq \ell_0$,

$$f_j^\circ(x^{k_\ell+1}; y_0 - x^{k_\ell+1}) < 0, \quad j \in I.$$

Since $x^{k_\ell+1} \in \Omega_{k_\ell} \setminus \mathcal{S}_{\bar{C}}^*(F)$, the last inequality with Proposition 3.1 implies that $y_0 - x^{k_\ell+1} \in T_{\Omega_{k_\ell}}(x^{k_\ell+1})$, $\ell \geq \ell_0$. As $v^{k_\ell+1} \in N_{\Omega_{k_\ell}}(x^{k_\ell+1})$ and $y_0 - x^{k_\ell+1} \in T_{\Omega_{k_\ell}}(x^{k_\ell+1})$, $\ell \geq \ell_0$, we have $\langle v^{k_\ell+1}, y_0 - x^{k_\ell+1} \rangle \leq 0$, for $\ell \geq \ell_0$. Thus, the last inequality along with (12) implies that

$$\limsup_{\ell \rightarrow +\infty} \sum_{j=1}^m z_j^{k_\ell} f_j^\circ(x^{k_\ell+1}; y_0 - x^{k_\ell+1}) \geq 0.$$

Combining a standard property of \limsup with the homogeneity of f° , we obtain

$$\sum_{j=1}^m \limsup_{\ell \rightarrow +\infty} f_j^\circ \left(x^{k\ell+1}; z_j^{k\ell} (y_0 - x^{k\ell+1}) \right) \geq 0.$$

Without loss of generality, suppose that $\lim_{\ell \rightarrow +\infty} z_j^{k\ell} = \bar{z}_j$. Using the last inequality with (3), we conclude that

$$\sum_{j=1}^m f_j^\circ (x^*; \bar{z}_j (y_0 - x^*)) \geq \sum_{j=1}^m \limsup_{\ell \rightarrow +\infty} f_j^\circ \left(x^{k\ell+1}; z_j^{k\ell} (y_0 - x^{k\ell+1}) \right) \geq 0.$$

Further, from the homogeneity of f° , $\sum_{j=1}^m \bar{z}_j f_j^\circ (x^*; y_0 - x^*) \geq 0$. Hence, there exists $j_0 \in I$ such that $f_{j_0}^\circ (x^*; y_0 - x^*) \geq 0$. This is not possible since (14) holds. Therefore, x^* is a Pareto–Clarke critical point of F in \bar{C} . \square

Let us illustrate Theorem 4.1 with some examples.

Example 4.1 (Nonnegative constraints) To solve that problem $\min_w \{F(x, y): (x, y) \in \mathbb{R}_+^2\}$, where $F: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is defined as $F(x, y) = (\max\{x^2 + y^2, -(x^2 + y^2) + 1\}, x^2 + y^2)$, we will take advantage of the flexibility of the proximal distances. To find a Pareto–Clarke critical point of F in \mathbb{R}_+^2 , we will apply our inexact multi-objective proximal point method with an appropriate proximal distance d with respect to the convex set $C = \mathbb{R}_{++}^2$. In fact, there are several examples of such distances, some of them satisfy (d5), see [11]. Now, since $(1/2, 0) \preceq F(x, y)$ for all $(x, y) \in \mathbb{R}^2$, and the gradients of $(x, y) \mapsto x^2 + y^2$ and $(x, y) \mapsto -(x^2 + y^2) + 1$ are Lipschitz continuous on \mathbb{R}^2 , the assumptions **A1** and **A2** are satisfied. Note that for all $x^p \in \mathbb{R}^2$ with $(1/2, 0) \preceq F(x^p) \prec (1, 0)$, the level set $\Omega_p := \{(x, y) \in \mathbb{R}^2 : F(x, y) \preceq F(x^p)\}$ is not convex, once $(-1/\sqrt{2}, 0), (1/\sqrt{2}, 0) \in \Omega_p$ and $1/2(-1/\sqrt{2}, 0) + 1/2(1/\sqrt{2}, 0) \notin \Omega_p$. Choosing x^0 with $(1/2, 0) \preceq F(x^0) \prec (1, 0)$, the results of Bento [3, Theorem 1] cannot be applied for this example. From Theorem 4.1, $x^{k+1} \in C \cap \Omega_k$, for $k = 0, 1, \dots$. Hence, $0 \preceq F(x^{k+1}) \preceq F(x^k) \preceq F(x^0)$, and so, $\{F(x^k)\}$ is bounded. If $x^k := (x_1^k, x_2^k)$, from definition of F , it follows that both $\{x_1^k\}$ and $\{x_2^k\}$ are bounded. Then, $\{x^k\}$ is bounded. Now, we can apply Theorem 4.1 to guarantee the existence of a Pareto–Clarke critical of F in \bar{C} . Note that, in this case, the inexact multi-objective proximal point method (9) becomes

$$0 \in \partial \left(\left\langle F(\cdot), z^k \right\rangle + \gamma_k d(\cdot, x^k) \right) (x^{k+1}) + \varepsilon_k \mathbb{B} + \gamma_k \nabla_1 d(x^{k+1}, x^k) + N_{\Omega_k}(x^{k+1}),$$

$k = 0, 1, \dots$. Hence, there exist $u^{k+1} \in \mathbb{B}$ and $v^{k+1} \in N_{\Omega_k}(x^{k+1})$ satisfying

$$0 \in \partial \left(\left\langle F(\cdot), z^k \right\rangle + \gamma_k d(\cdot, x^k) \right) (x^{k+1}) + \varepsilon_k u^{k+1} + \gamma_k \nabla_1 d(x^{k+1}, x^k) + v^{k+1},$$

$k = 0, 1, \dots$. Putting $z^k := (z_1^k, z_2^k)$, there exists a sequence $\{\alpha_k\} \subset [0, 1]$ such that, for all $k = 0, 1, \dots$,

$$0 = \left(2\alpha_k(x_1^k, x_2^k) - 2(1 - \alpha_k)(x_1^k, x_2^k) \right) z_1^k + 2(x_1^k, x_2^k)z_2^k + \varepsilon_k u^{k+1} \\ + 2\gamma_k \nabla_1 d(x^{k+1}, x^k) + v^{k+1}.$$

Example 4.2 Define $F : \mathbb{R}^2 \rightarrow \mathbb{R}$ by $F(x, y) = \max \left\{ e^{-x^2-y^2}, e^{x^2+y^2-2} \right\}$. As $0 \leq F(x, y)$ for all points $(x, y) \in \mathbb{R}^2$, and the gradients of $(x, y) \mapsto e^{-x^2-y^2}$ and $(x, y) \mapsto e^{x^2+y^2-2}$ are Lipschitz continuous on \mathbb{R}^2 , the assumptions **A1** and **A2** are satisfied. Moreover, note that for all $x^p \in \mathbb{R}^2$ with $e^{-1} \leq F(x^p) < 1$, the level set $\Omega_p := \{(x, y) \in \mathbb{R}^2 : F(x, y) \leq F(x^p)\}$ is not convex, once $(-1, 0), (1, 0) \in \Omega_p$ and the combination $1/2(-1, 0) + 1/2(1, 0) \notin \Omega_p$. Choosing x^0 with $e^{-1} \leq F(x^0) \leq 1$, the results of Bento [3, Theorem 1] cannot be applied for this example. Similarly, Theorem 4.1 guarantees the existence of a Pareto–Clarke critical of F in \bar{C} .

4.1 Convex Case

In this subsection, we establish convergence of the inexact multi-objective proximal point method (9) for the convex case. For this, let us assume that F is convex, i.e.,

$$F(\alpha x + (1 - \alpha)y) \leq \alpha F(x) + (1 - \alpha)F(y), \quad x, y \in \mathbb{R}^n, \quad \alpha \in [0, 1].$$

To set the convergence of any sequence generated by (9), we need to make further assumptions on the induced proximal distance H , which were also considered in [11]. Let $(d, H) \in \Phi_+(\bar{C}) \in \Phi(\bar{C})$ be such that the function H satisfies the following two additional properties: For $y \in \bar{C}$ and $\{y^k\} \subset C$,

- (Ha) $\lim_{k \rightarrow +\infty} y^k = y$, whenever $\{y^k\}$ is bounded and $\lim_{k \rightarrow +\infty} H(y, y^k) = 0$;
- (Hb) $\lim_{k \rightarrow +\infty} H(y, y^k) = 0$, whenever $\lim_{k \rightarrow +\infty} y^k = y$.

We also make the following assumption:

$$\mathbf{A3.} \quad \Omega := \bigcap_{k=0}^{+\infty} \Omega_k \neq \emptyset.$$

Under these assumptions, we prove that the inexact multi-objective proximal point algorithm (9) converges to a Pareto–Clarke critical of F .

Theorem 4.2 *Let $(d, H) \in \Phi_+(\bar{C})$ and let $\{x^k\}$ be a sequence generated by (9). If $\{x^k\}$ is bounded, $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is convex and $\sum_k \varepsilon_k < \infty$, then $\{x^k\}$ converges to a Pareto–Clarke critical point of F in \bar{C} .*

Proof Since **A3** holds, Ω_{k+1} is nonempty. Then, consider $y \in \Omega_{k+1}$. Since F is convex, we have

$$f_j^\circ(x^{k+1}; y - x^{k+1}) \leq f_j(y) - f_j(x^{k+1}), \quad j \in I, \quad k \geq 0.$$

Hence, owing to $y \in \Omega_{k+1}$,

$$f_j^\circ(x^{k+1}; y - x^{k+1}) \leq 0, \quad j \in I, \quad k \geq 0. \quad (15)$$

On the other hand, by using similar arguments as in the proof of Theorem 4.1, we conclude that there exist $M > 0$ and $v^{k+1} \in N_{\Omega_k}(x^{k+1})$ satisfying

$$\begin{aligned} & H(y, x^k) - H(y, x^{k+1}) + M\varepsilon_k \\ & \geq \frac{1}{2\gamma_k} \left(- \sum_{j=1}^m z_j^k f_j^\circ(x^{k+1}; y - x^{k+1}) - \langle v^{k+1}, y - x^{k+1} \rangle \right), \quad k \geq 0. \end{aligned} \quad (16)$$

Considering (15), it follows from Proposition 3.1 that $y - x^{k+1} \in T_{\Omega_k}(x^{k+1})$, for all $k \geq 0$. Moreover, since $v^{k+1} \in N_{\Omega_k}(x^{k+1})$ and $y - x^{k+1} \in T_{\Omega_k}(x^{k+1})$, $k \geq 0$, we have $\langle v^{k+1}, y - x^{k+1} \rangle \leq 0$, $k \geq 0$. Hence, we can combine the last inequality with (15) and (16) to obtain

$$M\varepsilon_k + H(y, x^k) \geq H(y, x^{k+1}), \quad k \geq 0.$$

Applying Lemma 4.1, we conclude that $\{H(y, x^k)\}$ converges to some point $\beta(y)$. Let x^* be an accumulation point of $\{x^{k+1}\}$. Since $x^{k+1} \in \Omega_k \cap C$, we have $F(x^{k+1}) \leq F(x^k)$ for all k . From the continuity of F , we can conclude that $F(x^*) \leq F(x^{k+1})$ for all k , which means that $x^* \in \Omega_{k+1} \cap \bar{C}$. Based on **(Ha)**, we obtain $\lim_{\ell \rightarrow +\infty} H(x^*, x^{k\ell+1}) = 0$. Considering that $\{H(y, x^k)\}$ converges, we conclude that $\lim_{k \rightarrow +\infty} H(x^*, x^{k+1}) = 0$. Now, by **(Hb)** it follows that $\{x^k\}$ converges to x^* . Therefore, from Theorem 4.1, x^* is a Pareto–Clarke critical point of F in \bar{C} , which proves the theorem. \square

Remark 4.2 It is worth to point out that **A1** was used to guarantee the well-definedness of proximal-type methods in the nonconvex setting. However, under the assumption of convexity, we do not need **A1**.

5 A Static Rawls Solution to the Problem of Distributive Justice

In economics and philosophy, the literature on social welfare is intended to express a statement of objectives of the society. Bentham’s utilitarian approach [18] measures social welfare as the total sum of individual incomes, $\sum_{i \in I} g_i$, or the average of individual incomes, $(1/n) \sum_{i \in I} g_i$, where n refers to the number of individuals. In this case, society tries to maximize the total incomes of all people in the society, without considering how incomes are distributed. It does not distinguish between an income transfer from the rich to the poor people and vice versa. If an income transfer from the poor to the rich people ends in a bigger increase in the utility of the rich than to a decrease in the utility of the poor people, the utilitarian approach advocates that the society must accept such a transfer because the total utility of society has increased as a whole.

In contrast, in philosophy, Rawls [19] proposed measuring the social welfare of society by the well-being of the least well-off individual member of society, $\min \{g_i; i \in I\}$. Hence, Rawls advocates that the society must maximize the income of the poorest persons in society without regard to the income of other individuals. In this way, he developed a theory of distributive justice, that is, a theory of the “good” as justice, and justice conceived as fairness. Let us propose a simple model for [19] this max-min principle.

Let X be the universal space of actions (doings), havings, or beings, depending on the applications. X includes all past elements and all the new items that can be discovered as time evolves. Consider a society made of a list of agents $i \in I$. This society is divided into nonoverlapping groups $J = I_1 \cup \dots \cup I_m$. Each agent i belongs to only one group $I_j \subset J$, that is, $J = \cup_{j=1}^m I_j$ and $I_h \cap I_j = \emptyset, h \neq j$. An institution spends public resources $x \in X$ to improve the redistribution of wealth within this society.

Each agent $i \in I_j$ of a group I_j earns a gain $g_{i,j} = g_{i,j} [a_i, a_{j-\{i\}}, x] \in \mathbb{R}_+, a_i \in A_{i,j}(x)$, which depends of his own action a_i ; the profile of actions $a_{j-\{i\}}$ done by other agents of his group j ; and rules, norms, taxes, subventions, and public investments $x \in X$ set and spent by an institution. His feasible set of actions $A_{i,j}(x)$ depends itself on the group j it belongs to and on the chosen collective action x of the institution.

Each period, the institution first sets an action $x \in \Omega \subset X$. Then, within each group j , agents play a noncooperative game or a cooperative game that determines a unique profile of actions $a_i^*(x) \in A_{i,j}(x)$ within each group. The equilibrium gain of each agent $i \in I_j$ is $g_{i,j}(x) = g_{i,j} [a_i^*(x), a_{j-\{i\}}^*(x)]$. Within each group $j \in J$, the minimum gain earned by an agent of this group is $g_j(x) = \min \{g_{i,j}(x) : i \in I_j\}$. Then, the vector of minimum gains earned within each group of the society is $G(x) = (g_1(x), \dots, g_m(x))$. The institution follows the famous max-min rule of the Rawls theory of justice if it maximizes (in a vectorial sense) this vector of minimum gains, that is, if it solves the multi-objective optimization problem $\max_{x \in \Omega} G(x)$. A solution $x^* \in \Omega$ is a weak Pareto point of the multi-objective maximization problem if there is no $x \in \Omega$ that strictly dominates x^* , that is, such that $G(x) \geq G(x^*)$, i.e., such that $g_j(x) > g_j(x^*), j \in I$.

Therefore, with reference to a multi-objective distributive justice program that maximizes minimum objectives, our multi-objective program refers to the minimization of maximum objectives.

6 A Lewin–Rawls Group Dynamic of Distributive Justice

In this section, we show how our semi-local search inexact multi-objective proximal algorithm models, in a crude but interesting way, a dynamic group process that converges to a solution of the static Rawlsian max-min problem of distributive justice, using our model given in Sect. 5. For this purpose, we will take advantage of a recent VR approach to human behavior (Soubeyran, [6–8]).

6.1 Lewin Theory of Group Dynamics

The famous psychologist Kurt Lewin [20,21] is the pioneer of research on group dynamics, action research, and change processes. His “force field analysis theory”

focused attention on the way groups and individuals act and react to changing. Lewin [20] is the father of approach-avoidance dynamics (see Townsend and Busemeyer [22] and Elliot [23]). Approach motivation is the energization of behavior by, or the direction of behavior toward, positive stimuli (objects, events, and possibilities), whereas avoidance motivation is the energization of behavior by, or the direction of action away from, negative stimuli (objects, events, and possibilities).

In this context, Lewin [20,21] emphasized that two kinds of forces influence a temporary situation: forces that drive the movement toward a goal (motivational and helping forces) and forces that block movement toward a goal (resistance to change and hindering forces). A “force” can be people, resources, attitudes, traditions, values, needs, desires, etc. Lewin’s famous balancing principle states that human behavior can be modeled as a succession of temporary equilibria. That is, when driving and restraining forces are of almost equal size, the quasi-equilibrium is not broken, and to be able to change from one temporary equilibrium to the next one, an agent or a group must create an imbalance between forces, i.e., to increase the drivers and to reduce the restraints. In this dynamical context, Lewin [20,21] emphasizes the significant role of resistance in changing, using, for a typical unit change, a three-stage model of change (unfreezing, change, and freezing).

6.2 The Variational Rationality Approach

Recently, the VR approach [6–8] offers, in a dynamical context, a general model of individual human behaviors and group dynamics. It models many human behaviors as worthwhile “stay and changes” approach-avoidance dynamics. These dynamics include a starting point, a transition, and some end (if any). They start from some undesirable initial states. They follow acceptable transitions made of a succession of stays and changes because agents do not always stay or always change. Stays refer to habits, routines, norms, rules, and exploitation phases, while changes refer to exploration, search, learning, training, and innovations. These transitions can end (if so) in some desired ends or much before, in some traps where agents, being stuck in the middle, can fail to reach their desires. Such transitions are worthwhile, and hence are acceptable, when each successive stay or change is itself worthwhile. This means that, in each period, an agent prefers to change rather than to stay if his motivation to change is high enough relative to his resistance to change. Otherwise, he will prefer to stay. An end is a (variational) trap that is worth approaching and reaching, but not worthwhile to leave. Quite surprisingly, the VR approach, whose initial goal was to propose a general model of human behavior in a dynamical context, not only unified a lot of stay and change dynamics in behavioral sciences, but also unified many variational principles and methods in optimization theory in terms of few behavioral principles.

6.2.1 A Formal Presentation

In the VR approach (see [6–8]), one considers, successively, **(1)** individuals $i \in I$ (agents) and groups of agents I_j ; **(2)** different positions $x, y \in X$ for individuals and groups of agents, where positions can be actions (doings), havings (possessions), or

beings; **(3)** some change $x \rightsquigarrow y$ from x to $y \neq x$ or stay $x \rightsquigarrow y = x$ for agents and groups; **(4)** human stay and change dynamics including (i) a given initial position x^0 , (ii) a transition $\{x^0, x^1, \dots, x^k, x^{k+1}, \dots\}$ defined as a succession of changes and stays $x^k \rightsquigarrow x^{k+1}$, where $x^{k+1} \neq x^k$ or $x^k = x^k$, and (iii) a possible end x^* (if any); **(5)** acceptable change and stay as worthwhile change and stay (where the term worthwhile will be defined below); **(6)** worthwhile human transitions, where each change or stay $x^k \rightsquigarrow x^{k+1}$, with $x^{k+1} \neq x^k$ or $x^k = x^k$, is worthwhile; **(7)** ends $x^* \in X$, defined as variational traps, which are both (a) aspiration points, worthwhile to reach from some or any position of the transition, and (b) stationary trap, not worthwhile to leave; and **(8)** desired ends or desires $x^{**} \in X$, which are positions preferred to any other position belonging to some subset.

Then, the VR approach emphasizes that, when final solutions cannot be reached in one step, transitions are necessary and so must be acceptable.

6.2.2 Worthwhile Change

Let us define, successively, in the simplest VR structure (see [8]) the following: **(1)** vectorial payoffs $G(x) = (g_1(x), \dots, g_m(x)) \in \mathbb{R}^m$, where $j \in I$ represent different groups of agents and $g_j(x) \in \mathbb{R}^m$ is the payoff of group j , given the chosen action $x \in X$; **(2)** vectorial advantages to change from x to y , $A(x, y) = G(y) - G(x) \in \mathbb{R}^m$; **(3)** vectorial inconveniences to change $I(x, y) = C(x, y) - C(x, x)$ where **(4)** the costs of being able to change from x to y are $C(x, y) = (c_1(x, y), \dots, c_m(x, y)) \in \mathbb{R}_+^m$ and **(5)** the costs of being able to stay at x are $C(x, x) = (c_1(x, x), \dots, c_m(x, x)) \in \mathbb{R}_+^m$; **(6)** scalar motivation to change $M(x, y) = U[A(x, y)] \in \mathbb{R}_+$, where $U[A]$ is the utility of vectorial advantages to change $A = A(x, y)$; and **(7)** scalar resistance to change $R(x, y) = D[I(x, y)] \in \mathbb{R}_+$, where $D[I]$ is the disutility of vectorial inconveniences to change $I = I(x, y)$.

In this simple VR model, a change from x to y will be said to be worthwhile if motivation to change is high enough relative to resistance to change, that is, $M(x, y) \geq \xi R(x, y)$, where $\xi > 0$ is a chosen worthwhile to change ratio. The subset of worthwhile changes at x is $W_\xi(x) := \{y \in X : M(x, y) \geq \xi R(x, y)\} \subset X$. Then, a worthwhile stay and change transition is $x^{k+1} \in W_{\xi^{k+1}}(x^k) = \{y \in X, M(x^k, y) \geq \xi^{k+1} R(x^k, y)\}$, $k \geq 0$, where the worthwhile to change ratio $\xi^{k+1} > 0$ is adapted each current period $k + 1$. In this setting,

- (i) an aspiration point $x^* \in X$ is such that $x^* \in W_{\xi^{k+1}}(x^k)$, $k \in \mathbb{N}$, $k \geq k_0$;
- (ii) a stationary trap $x^* \in X$ is such that $W_{\xi^*}(x^*) = \{x^*\}$, where $\xi^* > 0$ is the worthwhile to change ratio chosen at the end.

6.3 The Present Paper

As an application, we consider the specific VR structure, where utilities $M = U[A]$ and disutilities $R = D[I]$ are weighted sums of the utilities and disutilities of advantages A_j and inconveniences I_j to change over all groups $j \in I$, that is, $U[A] = \sum_{j=1}^m z_j U_j[A_j]$ and $D[I] = \sum_{j=1}^m z_j D_j[I_j]$. In this paper, we assume

that utilities and disutilities are linear, that is, $U_j [A_j] = A_j$ and $D_j [I_j] = I_j$, $j \in I$. Weights are $z = (z_1, \dots, z_j, \dots, z_m) \in \mathbb{R}_+^m$, $z \neq 0$. Then, in terms of scalar products, $M = U [A] = \langle A, z \rangle$ and $R = D [I] = \langle I, z \rangle$, where, for convenience, $I(x, y) = (I_1, \dots, I_m)$. For simplification, we suppose that $C(x, x) = 0$ (for further details, see [8]). Then, a change is worthwhile if $\langle G(y) - G(x), z \rangle \geq \xi \langle C(x, y), z \rangle$. In our mathematical part, it is assumed that $F(x) = -G(x)$, $C(x, y) = d(x, y)e$, and $\xi = \lambda/2 > 0$, where $\gamma = \lambda \langle e, z \rangle$, which implies that the set of worthwhile changes at x becomes

$$W_{\lambda/2}(x) = \{y \in \mathbb{R}^n : \langle F(x) - F(y), z \rangle \geq (\lambda/2) \langle e, z \rangle d(x, y)\}. \quad (17)$$

In our distributive justice example, $G(x) = (g_1(x), \dots, g_m(x)) \in \mathbb{R}^m$ refers to a vector of payoffs (gains), where $g_j(x)$ is the gain of group $j \in J$. The mathematical part takes into account dissatisfactions or losses $F(x) = (f_1(x), \dots, f_m(x))$, where $f_j(x) = -g_j(x)$, $j \in J$.

In this paper, a very nice aspect of our worthwhile transition ending in a final distributive justice solution is that it is adaptive. That is, weights $\{z^{k+1}\}$ given for each current period $k+1$ to the different objectives are variable. They can be chosen all along a worthwhile to change process. Our model is a mix between a Rawlsian approach and a utilitarian (Bentham) approach. Each period, each group tries to maximize the minimum payoff of all its members. Then, an institution chooses weights, each period, to maximize a weighted sum of their payoffs. The VR approach shows the importance, for convergence, of the strength of resistance to change. In our example, resistance to change of all groups refers, in the current period $k+1$, to $R(x^k, x) = (\gamma_k/2)d(x, x^k)$. Then, the higher the value of $\gamma_k = \lambda_k \langle e^k, z^k \rangle$, the higher is resistance to change. The proof of the convergence result shows that convexity of the (to be decreased) proximal payoff $P_k(x^k, x) = \langle F(x), z^k \rangle + (\gamma_k/2)d(x, x^k)$ depends of the size of $\gamma_k > 0$. The higher $\gamma_k > 0$ is, the more convex is the proximal payoff. Moreover, as seen above, the higher $\gamma_k > 0$ is, the stronger is resistance to change.

In the present mathematical method, it is assumed that worthwhile stay and changes improve the dynamic $x^{k+1} \in \Omega_k = \{x \in X : F(x) \leq F(x^k)\}$. That means that, each period, the institution tries to improve the proximal payoff of every group (with no group accepting a loss, including costs to be able to change). Then, no group can raise an objection to the proposed change. In this way, the transition toward a final distributive justice solution is acceptable for each group as a worthwhile stay and change dynamic. The costs of being able to change are usually asymmetric. They are modeled as proximal distances. See Sect. 3 and also [6–8] for more details on this topic. In this paper, we also suppose that the costs of being able to change are shared in some proportion of a total cost of being able to change, where these proportions can be chosen, each period. A weak variational trap is both a limit point of a worthwhile transition and a stationary trap not worthwhile to leave. Such a trap models the approach and the end of a worthwhile stay and change process. Usually, a critical point is not a stationary trap, but, in the next result, we will show that any accumulation point of any sequence generated by our (semi) local search inexact multi-objective proximal method (9) is a stationary trap.

Theorem 6.1 *Suppose that all assumptions of Theorem 4.1 hold. If $d(x, y) = H(x, y) = 1/2\|x - y\|^2$, then each accumulation point of $\{x^{k+1}\}$ is a stationary trap x^* , relative to the worthwhileness of changing set $W_{\lambda^*}(x^*)$, for any $\lambda^* > \tilde{\lambda}$. Moreover if F is convex, $\{x^k\}$ converges to a stationary trap.*

Proof Let x^* be an accumulation point of $\{x^{k+1}\}$. Now, let us prove that x^* is a stationary trap relative to the worthwhileness of changing set $W_{\lambda^*}(x^*)$. From Theorem 4.1, $x^* \in S_{\tilde{c}}^*(F)$. Then, there exists $j_0 \in I$ such that $0 \in \partial^\circ f_{j_0}(x^*)$, and we have $\partial^\circ f_{j_0}(x^*) = \partial^\circ \left(f_{j_0}(\cdot) + \left(\lambda_k e_{j_0}^k / 2 \right) \|\cdot - x^*\|^2 \right) (x^*)$, and, consequently, $0 \in \partial^\circ \left(f_{j_0}(\cdot) + \left(\lambda_k e_{j_0}^k / 2 \right) \|\cdot - x^*\|^2 \right) (x^*)$. As $f_j(\cdot) + \left(\lambda_k e_j^k / 2 \right) \|\cdot - x^*\|^2$ is a strict convex function for all $j \in I$, $k \in \mathbb{N}$, we deduce from the last inclusion that $f_{j_0}(x) + \left(\lambda_k e_{j_0}^k / 2 \right) \|x - x^*\|^2 > f_{j_0}(x^*)$, $x \in \mathbb{R}^n$. In view of (17), we have that $W_{\lambda^*}(x^*) = \{x^*\}$ for all $\lambda^* > \tilde{\lambda}$. Therefore, x^* is a stationary trap. The second part follows immediately from Theorem 4.2. \square

7 Conclusions

In this paper, we present an inexact multi-objective proximal method in a proximal metric setting, whose regularization term is a proximal distance. Our main convergence result is restricted to suitable multi-objective nondifferentiable functions. In our approach, we do not have the convexity of the set defined in (9) and we use the definition of the tangent cone of the improving constraint omega set, which allows us to reach a structure similar to that in [2]. Our motivation has been to give, in the context of group dynamics, a new model and resolution of a distributive justice problem that balances, in each step, value creation at the group level and value appropriation at the individual level. In future research, we intend to investigate our inexact method when F satisfies the Kurdyka-Łojasiewicz property. In [24], the authors presented an abstract descent method in the quasi-metric setting. Although both methods, Method 3.1 in [24] and iterative scheme (9), are defined in the scalar and multi-objective setting, respectively, these schemes are similar in two aspects: both are proximal-type and applied for nonconvex functions. However, the full convergence of Method 3.1 in [24] is obtained for Kurdyka-Łojasiewicz functions and in the present paper for convex. More attention will need to extend the results of [24] to the multi-objective setting. We foresee further progress in this topic in the near future.

Acknowledgements The work was supported by CAPES, CNPq, MathAmSud (CAPES) 88881.117595/2016-01 and the ANR GREEN-Econ research project (ANR-16-CE03-0005).

References

1. Mordukhovich, B.: Variational Analysis and Generalized Differentiation. I. Basic Theory. Grundlehren der Mathematischen Wissenschaften, vol. 330. Springer, Berlin (2006)
2. Bonnel, H., Iusem, A.N., Svaiter, B.F.: Proximal methods in vector optimization. SIAM J. Optim. **15**(4), 953–970 (2005)

3. Bento, G.C., Ferreira, O.P., Sousa Junior, V.L.: Proximal point method for a special class of nonconvex multiobjective optimization functions. *Optim. Lett.* **12**(2), 311–320 (2018)
4. Aarts, E., Lenstra, K.: *Local Search*. Princeton University Press, Princeton (2003)
5. Attouch, H., Soubeyran, A.: Local search proximal algorithms as decision dynamics with costs to move. *Set Valued Var. Anal.* **19**(1), 157–177 (2011)
6. Soubeyran, A.: Variational rationality, a theory of individual stability and change: worthwhile and ambidexterity behaviors. GREQAM, Aix-Marseille University, France (2009, Pre-print)
7. Soubeyran, A.: Variational rationality and the “unsatisfied man”: routines and the course pursuit between aspirations, beliefs. GREQAM, Aix-Marseille University, France (2010, Pre-print)
8. Soubeyran, A.: Variational rationality. Worthwhile stay and change approach-avoidance human dynamics ending in traps. GREQAM, Aix-Marseille University, France (2016, Pre-print)
9. Hiriart-Urruty, J.B., Lemaréchal, C.: *Convex Analysis and Minimization Algorithms*, Published in Two Volumes. Springer, Berlin (1993)
10. Clarke, F.H.: *Optimization and Nonsmooth Analysis Volume 5 of Classics in Applied Mathematics*, vol. 2. Society for Industrial and Applied Mathematics (SIAM), Philadelphia (1990)
11. Auslender, A., Teboulle, M.: Interior gradient and proximal methods for convex and conic optimization. *SIAM J. Optim.* **16**(3), 697–725 (2006)
12. Burachik, R., Dutta, J.: Inexact proximal point methods for variational inequality Problems. *SIAM J. Optim.* **20**(5), 2653–2653 (2010)
13. Custódio, A.L., Madeira, J.F.A., Vaz, A.I.F., Vicente, L.N.: Direct multisearch for multiobjective optimization. *SIAM J. Optim.* **21**(3), 1109–1140 (2011)
14. Chen, Z., Huang, X.X., Yang, X.Q.: Generalized proximal point algorithms for multi-objective optimization problems. *Appl. Anal.* **90**(6), 935–949 (2011)
15. Polyak, B.T.: *Introduction to Optimization*. Optimization Software, New York (1987)
16. Rockafellar, R.T., Wets, R.: *Variational Analysis*. Grundlehren der Mathematischen Wissenschaften. Springer, New York (1998)
17. Rockafellar, R.T.: *Convex Analysis*. Princeton Mathematical Series, vol. 28. Princeton University Press, Princeton (1970)
18. Arrow, K.J.: *Social Choice and Individual Values*. Yale University Press, New Haven (1951)
19. Rawls, J.: *A Theory of Justice*. Harvard University Press, Cambridge (1971)
20. Lewin, K.: *A Dynamic Theory of Personality*. McGraw-Hill, New York (1935)
21. Lewis, K.: *Field theory in Social Science*. Harper, New York (1951)
22. Townsend, J.T., Bussemeyer, J.R.: Approach-avoidance: Return to dynamic decision behavior. The Tulane Flowerree Symposia on Cognition, Psychology Press, In: *Current Issues in Cognitive Processes* (2014)
23. Elliot, A.J.: The hierarchical model of approach-avoidance motivation. *Motiv. Emot.* **30**(2), 111–116 (2006)
24. Bento, G.C., Cruz Neto, J.X., Soubeyran, A., Sousa Junior, V.L.: Dual descent methods as tension reduction systems. *J. Optim. Theory Appl.* **171**(1), 209–277 (2016)