A proximal point method for difference of convex functions in multi-objective optimization with application to group dynamic problems

Glaydston de Carvalho Bento, Sandro Dimy Barbosa Bitar, João Xavier da Cruz Neto, Antoine Soubeyran, João Carlos de Oliveira Souza

A proximal point method for difference of convex functions in multi-objective optimization with application to group dynamic problems

G.C. Bento · S.D.B. Bitar · J.X. Cruz Neto · A. Soubeyran · J.C.O. Souza

Abstract We consider the constrained multi-objective optimization problem of finding Pareto critical points of difference of convex functions. The new approach proposed by Bento et al. [SIAM J. Optim., 28 (2018), pp. 1104-970] to study the convergence of the proximal point method is applied. Our method minimizes at each iteration a convex approximation instead of the (non-convex) objective function constrained to a possibly non-convex set which assures the vector improving process. The motivation comes from the famous Group Dynamic Problem in Behavioral Sciences where, at each step, a group of (possible badly informed) agents tries to increase his joint payoff, in order to be able to increase the payoff of each of them. In this way, at each step, this ascent process guarantees the stability of the group. Some encouraging preliminary numerical results are reported.

Keywords Multi-objective programming · proximal point method · DC function · variational rationality · behavioral sciences

Glaydston de Carvalho Bento
IME, Federal University of Goiás, Goiânia, Brazil
E-mail: glaydstonc@gmail.com

Sandro Dimy Barbosa Bitar
ICE, Federal University of Amazonas, Manaus, MA , Brazil
E-mail: sandrobitar@ufam.edu.br

João Xavier da Cruz Neto
CCN, DM, Federal University of Piauí, Teresina, PI , Brazil
E-mail: jxavier@ufpi.edu.br

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

João Carlos de Oliveira Souza
CCN, DM, Federal University of Piauí, Teresina, PI , Brazil
E-mail: joaocos.mat@ufpi.edu.br

1 Introduction

In this paper, we consider the multi-objective optimization problem of finding a Pareto critical point for a special class of non-convex vector functions which can be written as difference of two convex vector functions (called DC functions). Multi-objective optimization is the process of simultaneously optimizing two or more real-valued objective functions. Usually, no single point will minimize all given objective functions at once (i.e., there is no ideal minimizer), and so some concepts have to be adapted. There is a wide research program consisting of extension to the vector setting of several iterative methods for scalar functions, for instance the steepest descent method (see Graña Drummond and Svaiter [27] and Bento et al. [6]), projected methods (see Fukuda and Graña Drummond [24], Graña Drummond and Iusem [26] and Brito et al. [11]), subgradient method (see Bello Cruz [3,19]), Newton's method (see Fliege et al. [22] and Qu et al. [51]) and proximal methods (see Bonnel et al. [10], Ceng and Yao [13], Bento et al. [5], Apolinário et al. [1] and Qu et al. [52]). Multi-objective optimization has application in economics, industry, agriculture, location of public facilities, among others. For more details see, for example, Jahn [33], Luc [38], Miettinen [41] and Ross and Soland [54].

There are a lot of works devoted to the theory of DC functions in different contexts; for instance, duality and optimality conditions (e.g. Hartman [29], Hiriart-Urruty [30]), proximal algorithms (e.g. Moudafi and Maingé [45], Souza and Oliveira [60], Sun et al. [61], Wen et al. [67]), subgradient methods (e.g. Tao and Souad [62]), cutting angle and branch-and-bound methods (e.g. Ferrer et al. [21], Tuy and Horst [65]), vector optimization (e.g. Flores-Bazán and Oettli [23], Guo and Li [28]), variational inequality problems (e.g. Muu and Quoc [49]), equilibrium problems (e.g. Dinh et al. [20]), trust-region problem (e.g. Tao and An [63]), transportation problem (e.g. Holmberg and Tuy [31]).

The aim of this paper is to study convergence properties of a proximal point method, for DC vector functions, adding a vectorial improving constraint which is not convex. In a finite dimensional multi-objective setting our non-convex multi-objective proximal descent method generalizes the (exact) convex multi-objective proximal point descent method introduced by Bonnel et al. [10]. Our non-convex descent property also implies the one obtained by Ji et al. [34]. It is worth to mention that even if DC function is a locally Lipschitz function our method is different to the one considered by Bento et al. [4] because our algorithm minimizes at each iteration a linear approximation of the objective function instead of minimizing the objetive function as [4] does. As we will see in Section 6 our approach have important applications in Behavioral Sciences.

A motivation to study the proximal point method involving DC vector functions and a step by step vectorial improving constraint comes from the very important "group dynamic" problem in Behavioral Sciences, where at each step, a group of agents tries to increase his joint payoff, in order to be able to increase the payoff of each of them. In this way, our ascent method guarantees, at each step, the stability of the group. There is a huge literature in Behavioral Sciences on “group dynamics” (see, for instance, the force field theory in Lewin [35,36] in Psychology, organization change theories in Management Sciences, see Poole and Van de Ven [50], and dynamic cooperative games in Economics). As emphasized by Lewin [35,36], a major aspect of the group dynamic problem is the presence of resistance to change, given that becoming able to improve and then improving is costly. In order to modelize this resistance to change aspect, we consider an original vectorial proximal algorithm where the perturbation term includes the square of a quasi distance. This model of resistance to change comes from the recent variational rationality approach of human dynamics (see Soubeyran [55–59]).

The paper is organized as follows. Section 2 presents the definition and basic properties of quasi distances. Section 3 states formally the problem and introduces some required definitions and results in multi-objective optimization. The algorithm is presented in Section 4. Section 5 contains the convergence analysis of the algorithm. An application to “group dynamic” problems in Behavioral Sciences is given in Section 6. Finally, some numerical experiments are presented in Section 7 to show the efficiency of the method and its performance compared to another proximal method for DC functions.

2 Quasi distance

The motivation to handle with quasi distances as the regularization term in proximal methods comes from applications. In Management Sciences costs of moving (being able to move and move) from a position x to a position y is not symmetric, i.e., $C(y,x) \neq C(x,y)$. Thus, cost functions involving quasi distances are better adapted to applications in Behavioral Sciences.

A quasi metric space is a pair (X,q) such that X is a non-empty set and $q : X \times X \to \mathbb{R}_+$, called a quasi metric or quasi distance, is a mapping satisfying:

1. For all $x, y \in X$, $q(x,y) = q(y,x) = 0 \Leftrightarrow x = y$;
2. For all $x, y, z \in X$, $q(x,z) \leq q(x,y) + q(y,z)$.

Therefore, metric spaces are quasi metric spaces satisfying the symmetric property $q(x,y) = q(y,x)$. Quasi distances are not necessarily convex, continuously differentiable or coercive functions. Examples of quasi distances can be found in [44] and references therein. In this paper, we consider quasi distances satisfying the following condition:

Condition 1: There are positive real numbers $\alpha > 0$ and $\beta > 0$ such that

$$\alpha||x-y|| \leq q(x,y) \leq \beta||x-y||, \quad \forall x,y \in \mathbb{R}^n. \tag{1}$$

Remark 1 This condition has been used to prove convergence of proximal point algorithms for non-convex and non-smooth functions, see for instance [7,8,44]. Moreno et al. [44] present several examples of quasi distances highlighting some that satisfy Condition 1.

Proposition 1 *Let $q : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ be a quasi distance that verifies* (1)*. Then, for each $z \in \mathbb{R}^n$ the functions $q(\cdot, z)$, $q(z, \cdot)$ are Lipschitz and $q^2(\cdot, z)$, $q^2(z, \cdot)$ are locally Lipschitz on $\mathbb{R}^n$.*

Proof See [44, Propositions 3.6 and 3.7]. □

3 Multi-objective optimization

In this section we recall some basic definitions and properties of multi-objective optimization which can be found, for instance, in Jahn [33] and Luc [38]. We state and prove some results that will allow to define the algorithm and study its convergence properties.

Let $\mathbb{R}^m$ be the m-dimensional Euclidean space with the partial order " $\preceq$ " induced by the Paretian cone $\mathbb{R}^m_+$, given by $y \preceq z$ (or $z \succeq y$) iff $z - y \in \mathbb{R}^m_+$ with its associate relation " $\prec$ ", given by $y \prec z$ (or $z \succ y$) iff $z - y \in \mathbb{R}^m_{++}$, where

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

and $\mathcal{I} = \{1, \ldots, m\}$. Given a vector function $F : \mathbb{R}^n \to \mathbb{R}^m$, the problem of finding a weak Pareto point of F consists of finding a point $x^* \in \mathbb{R}^n$ such that there exists no other $x \in \mathbb{R}^n$ with $F(x) \prec F(x^*)$. More generally, given a non-empty set $\Omega \subset \mathbb{R}^n$, a point $x^* \in \Omega$ is a weak Pareto point of the vector function F constrained to the set Ω iff there is no $x \in \Omega$ with $F(x) \prec F(x^*)$. We denote this problem by

$$\min{}_w\{F(x) \ : \ x \in \Omega\}, \tag{2}$$

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

Let F be a vector function given by $F(x) := (f_1(x), \ldots, f_m(x))$. The Jacobian $m \times n$ matrix of F at $x \in \mathbb{R}^n$ is defined by

$$JF(x) := (\nabla f_1(x), \ldots, \nabla f_m(x))^\top.$$

For a vector function $F : \mathbb{R}^n \to \mathbb{R}^m$ we say that F is $\mathbb{R}^m_+$-convex if and only if, for every $x, y \in \mathbb{R}^n$ and $\lambda \in (0, 1)$

$$F(\lambda x + (1 - \lambda)y) \preceq \lambda F(x) + (1 - \lambda)F(y).$$

Note that this concept is equivalent to component-wise convexity.

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

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

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

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

see Clarke [16].

Proposition 2 *Let $f_1, f_2 : \mathbb{R}^n \to \mathbb{R}$ be Lipschitz continuous at a point $x \in \mathbb{R}^n$. One has*

$$\partial(f_1 f_2)(x) \subset f_2(x)\partial f_1(x) + f_1(x)\partial f_2(x).$$

Proof See [17, Proposition 2.3.13]. □

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

$$\partial F(x) := \{U \in \mathbb{R}^{m\times n} \ : \ U^\top d \preceq F^\circ(x; d), \quad \forall d \in \mathbb{R}^n\}, \tag{3}$$

where $F^\circ(x; d) := (f_1^\circ(x; d), \ldots, f_m^\circ(x; d))$. It is worth to point out that an equivalent definition has appeared, in a more general context, in Thibault [64]. If F is $\mathcal{C}$-convex, for some ordering cone $\mathcal{C}$, a similar definition can be found in Luc [37].

Let Ω be a closed and convex set. A point $x \in \Omega$ is said to be a Pareto-Clarke critical point (or Pareto critical point) of F in Ω iff, for any $y \in \Omega$, there exists $i \in \mathcal{I}$ such that

$$f_i^\circ(x, y - x) \geq 0. \tag{4}$$

Remark 2 Note that if $m = 1$ in the previous definition, we retrieve the (classical) definition of critical points for non-smooth functions: $0 \in \partial f(x)$. It is worth to notice that, combining (4) with Clarke [16, Proposition 1.4], we have the following alternative definition: a point $x \in \mathbb{R}^n$ is a Pareto critical point of F in Ω if, for any $y \in \Omega$, there exist $i \in \mathcal{I}$ and $\xi \in \partial f_i(x)$ such that $\langle \xi, y - x\rangle \geq 0$. Thus, if x is not a Pareto critical point of F in Ω, then there exists $y \in \Omega$ such that

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

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

Lemma 1 *Let $a, b \in \mathbb{R}^m_+$ be such that at least one of the vectors is not equal to zero. Assume that Ω is a non-empty, closed and convex set. If $-(U^\top a+V^\top b) \in N_\Omega(x)$, for some $U, V \in \partial F(x)$, then x is a Pareto critical point of F.*

Proof We will prove just the case where $a, b \in \mathbb{R}^m_+\backslash\{0\}$. The proof of the case where one of the vectors $a, b \in \mathbb{R}^m_+$ is equal to zero follows the same argument. Take $x \in \Omega$ and $a, b \in \mathbb{R}^m_+\backslash\{0\}$ such that $-(U^\top a + V^\top b) \in N_\Omega(x)$. Let us suppose, by contradiction, that x is not a Pareto critical point of F. From Remark 2, there exists $y \in \Omega$ such that

$$U(y - x) \prec 0 \quad \text{and} \quad V(y - x) \prec 0.$$

Thus, we have $\langle a, U(y - x)\rangle < 0$ and $\langle b, V(y - x)\rangle < 0$ because $a, b \in \mathbb{R}^m_+\backslash\{0\}$. Therefore, we have that $\langle U^\top a, y - x\rangle < 0$ and $\langle V^\top b, y - x\rangle < 0$. Adding last two inequalities, we obtain

$$\langle U^\top a + V^\top b, y - x\rangle < 0,$$

which contradicts the fact that $-(U^\top a + V^\top b) \in N_\Omega(x)$, and hence the desired result is proved. □

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

$$d_C(x) := \inf\{||x - c|| \ : \ c \in C\}. \tag{5}$$

The next result presents a necessary condition for a point x^* to be a solution of (2) in the particular case where $\mathcal{D} \subset \mathbb{R}^n$ is a non-empty, closed and convex set, and

$$\Omega = \{x \in \mathcal{D} : s_j(x) \leq 0, \quad j \in J\} \quad \text{with} \quad s_j : \mathbb{R}^n \to \mathbb{R}. \tag{6}$$

This kind of results have been extended to more general contexts such as vector variational inequalities (see Mai and Luu [39]) and others minimal concept in vector optimization such as relative minimal points; see Mordukhovich [48, Definition 9.3].

Theorem 1 *Let $\mathcal{D} \subset \mathbb{R}^n$ be a non-empty, closed and convex set. Assume that the set Ω in (2) is given as in (6), and the functions $f_j, s_j : \mathbb{R}^n \to \mathbb{R}$, $j \in \mathcal{I}$, are locally Lipschitz. If $x^* \in \Omega$ is a weak Pareto solution of (2), then there exist real numbers $u_j \geq 0$, $v_j \geq 0$, with $j \in \mathcal{I}$, and $\tau > 0$ such that*

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

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

Proof The proof follows from Minami [42, Theorem 3.1]. □

Remark 3 As remarked by Minami [42, Remark 3.1], since $\mathcal{D}$ is a closed and convex set, the cone $\{w \ : \ w \in \tau\partial d_{\mathcal{D}}(x^*), \ \tau > 0\}$ is the normal cone $N_{\mathcal{D}}(x^*)$.

Remark 4 It is worth to mention that a stronger version of the above theorem can be found in Mordukhovich [47, Theorem 5.61] and Mordukhovich [48, Theorem 9.22] by using the Mordukhovich's subdifferential; see also Mordukhovich [46]. In [47,48], necessary optimality conditions for local minimizers are given for constrained multi-objective problems under different perspectives such as Pareto and weak Pareto points but also to other concepts of minimal points in multi-objective optimization, for instance, relative, intrinsic relative and quasi-relative minimal points. A visible disadvantage of weak Pareto points is the non-empty interior requirement on the order cone in both finite and infinite dimensional frameworks. In this case, the concept of relative interior points seems to be reasonable. The relative-type minimal points handle with a possible empty interior requirement of the order cone and they intermediate the position between Pareto and weak Pareto minimal points (equivalence between them holds for instance in finte dimensional spaces with a non-empty interior order cone).

4 The algorithm

The (scalar) proximal point method was introduced in the literature by Moreau [43], Martinet [40] and later popularized by Rockafellar [53] who performs the method for the problem of finding zeros of operators. We recall that Bonnel et al. [10] have proposed an extension of the proximal point method to vector optimization, i.e., when other underlying ordering cones are used instead of the non-negative orthant $\mathbb{R}^m_+$. If we restrict the analysis to the finite dimensional multi-objective setting, the method proposed in [10] generates a sequence satisfying

$$x^{k+1} \in \operatorname{argmin}_w \left\{ F(x) + \frac{\lambda_k}{2}||x - x^k||^2 \varepsilon^k \ : \ x \in \Omega_k \right\}, \tag{7}$$

where $\{\lambda_k\} \subset \mathbb{R}_{++}$, $\{\varepsilon^k\} \subset \mathbb{R}^m_+$, $\Omega_k = \{x \in \mathbb{R}^n \ : \ F(x) \preceq F(x^k)\}$ and "argmin_w" denotes the set of weak Pareto solutions. The constrained set Ω_k assures the vector improving process in the sense that $F(x^{k+1}) \preceq F(x^k)$.

Recently, Ji et al. [34] have studied a proximal point method for vector functions $F = (f_1, \ldots, f_m) : \mathbb{R}^n \to \mathbb{R}^m$ such that, for each $i \in \mathcal{I}$, $f_i(\cdot)$ can be written as difference of two convex function, namely, $f_i(x) = g_i(x) - h_i(x)$. At the kth iterate, the algorithm proposed in [34] computes a point $x^{k+1} \in S$ solution of the following (scalar) subproblem

$$\min_{x\in S} \max_{i\in\mathcal{I}} \left[g_i(x) - \langle v_i^k, x - x^k \rangle \right] + \frac{\theta_k}{2}||x - x^k||^2, \tag{8}$$

where $v_i^k \in \partial h_i(x^k)$, with ∂h_i denoting a subdifferential of h_i at x^k, $\theta_k > 0$ for all k and, S is a compact and convex set. Note that (8) can be written as

$$\min_{x\in S} \xi \left(F_k(x) + \frac{\theta_k}{2}||x - x^k||^2 e \right), \tag{9}$$

where $F_k(x) = G(x) - V_k(x - x^k)$, $V_k \in \partial H(x^k)$, $e = (1, \ldots, 1) \in \mathbb{R}^m$ and $\xi : \mathbb{R}^m \to \mathbb{R}$ is a scalarization function given by $\xi(x) = \max_{1\le j\le m}\langle x, \beta_j \rangle$ where $\{\beta_j\}$ is the canonical base of the space $\mathbb{R}^m$. The finite termination of an algorithm given by (9) was analyzed by Bento et al. [5]. Clearly, the sequence generated by (9) does not satisfies $F(x^{k+1}) \preceq F(x^k)$, for all $k \in \mathbb{N}$ as in (7). This descent property plays an important role for a large class of application problems; see for instance Bento and Soubeyran [7]. Other authors have considered variants of the proximal method proposed by Bonnel et al. [10] using the improving constraint $F(x^{k+1}) \preceq_{\mathcal{C}} F(x^k)$, for some ordering cone $\mathcal{C}$; for instance Apolinário et al. [1], Bento et al. [5], Ceng and Yao [13], Ceng et al. [14], Choung et al. [15], Villacorta and Oliveira [66]. However, the constrained set Ω_k is convex in all these works.

From now on, we assume the vector functions $G, H : \mathbb{R}^n \to \mathbb{R}^m$ are $\mathbb{R}^m_+$-convex with H continuously differentiable. Let F be a DC vector function given by $F(x) := G(x) - H(x)$. Take a vector $z \in \mathbb{R}^m_+\backslash\{0\}$ fixed, and auxiliar sequences $\{\lambda_k\} \subset \mathbb{R}_{++}$ and $\{\varepsilon^k\} \subset \mathbb{R}^m_{++}$ such that $||\varepsilon^k|| = 1$ for all k. In this case, the proximal point method generates a sequence $\{x^k\}$ 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

$$x^{k+1} \in \arg\min \left\{ \langle G(x) - JH(x^k)(x - x^k) + \frac{\lambda_k}{2} q^2(x^k, x)\varepsilon^k, z\rangle \ : \ x \in \Omega_k \right\}, \tag{10}$$

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

Next, we present some interesting comments and existing results about proximal point methods in vector optimization literature.

Remark 5 As mentioned before, the first extension of the proximal point method to vector optimization was proposed by Bonnel, Iusem and Svaiter [10] (see (7)). They assume that each iterate of the sequence is a weak Pareto solution of a (constrained) vector optimization problem. Note that Algorithm 1 computes at each subproblem a solution of a scalar problem. This is the well-known scalarization approach for solving a multi-objective optimization problem. As mentioned in [10, Remark 5] the difference between the presentation of the iterative steps in (10) and (7) is not substantial because every solution of the scalar subproblems (10) is a weak Pareto solution of the subproblem (7) with $F(x) = G(x) - JH(x^k)(x - x^k)$; see [38, Proposition 2.2].

Remark 6 Note that in (7), for each $k \in \mathbb{N}$, the mapping $F_k(x) = F(x) + \frac{\lambda_k}{2}||x - x^k||^2\varepsilon^k$ is $\mathbb{R}^m_+$- convex, and hence, its Pareto critical points are weak Pareto point; see [25, Lemma 3.5]. This is why the authors in [10] take as next iterate a point x^{k+1} which is a weak Pareto point of F_k (constrained to Ω_k) besides the fact that the set of weak Pareto contains the Pareto ones. In this context, it is proved that the algorithm (7) converges to a weak Pareto solution of F. In some applications it is often the case that only Pareto solutions (instead of weak Pareto ones) are of interest; see [33]. In order to obtain convergence of the proximal point method to a Pareto solution (instead of weak Pareto point), the authors in [10] consider a more restrictive choice of x^{k+1}, namely, they take x^{k+1} a properly Pareto point of F_k (see this definition in [10, page 957]). One has that a properly Pareto point is a Pareto point, and hence, it is a weak Pareto point.

Remark 7 Recall that Algorithm 1 satisfies $x^{k+1} \in \Omega_k$, for all $k \geq 0$, i.e.,

$$F(x^{k+1}) \preceq F(x^k), \tag{11}$$

which is an important property for some application problems, as we will see in Section 6, while the algorithm in [34] satisfies

$$\sum_{i \in \mathcal{I}} \alpha_i (f_i(x^{k+1}) - f_i(x^k)) \leq 0, \tag{12}$$

for some $\alpha_i \geq 0$ with $\sum_{i\in\mathcal{I}} \alpha_i = 1$. Clearly, the descent property (11) implies (12).

Remark 8 It is known that every DC function is a locally Lipschitz function; see [30]. The proximal point method for locally Lipschitz functions in multi-objective optimization was analyzed by Bento et al. [4] who firstly considered a possibly non-convex vector improving constraints Ω_k. However, we mention that Algorithm 1 (applied for DC functions) offers an additional flexibility for considering, at each step, a linear approximation $G(x) - JH(x^k)(x - x^k)$ instead of minimize the non-convex function $F(x) = G(x) - H(x)$ directly. Minimizing a linear approximation is very important for instance for applications in Behavioral Sciences when agents are badly informed, i.e., agents have limited or partial knowledge. An interesting application by using a linear (local) approximation of each (scalar) cost function was considered by Cruz Neto et al. [18].

Remark 9 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 $F \succ 0$. On the other hand, any non-negative DC function admits a non-negative DC decomposition; see [30]. Therefore, we will assume throughout of this paper that $G \succeq 0$.

Proposition 3 *Algorithm 1 is well-defined.*

Proof The proof will be made by induction on k. Let $\{x^k\}$ be a sequence defined in (10) and $\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ given by

$$\phi(x) = \langle G(x), z\rangle - [JH(x^k)(x - x^k)]^\top z + \frac{\lambda_k}{2}\langle \varepsilon^k, z\rangle q^2(x^k, x) + I_{\Omega_k}(x). \quad (13)$$

Since $G \succeq 0$ and (1) holds, it follows that ϕ is coercive taking into account that $(\lambda_k/2)\langle \varepsilon^k, z\rangle > 0$. Then, as Ω_k is closed, there exists $\tilde{x} \in \Omega_k$ such that

$$\tilde{x} \in \operatorname{argmin}_{x\in\Omega_k} \phi(x).$$

Therefore, we can take $x^{k+1} := \tilde{x}$ and the induction is done, which ends the proof. □

From now on, we consider $\{x^k\}$, $\{\lambda_k\}$ and $\{\varepsilon^k\}$ the sequences appearing in Algorithm 1. Next, we explore deeply the structure of the vector problem by using the necessary condition for a weak Pareto point of a multi-objective problem given by Theorem 1. The following result will be used in our main convergence result.

Proposition 4 *For all $k \in \mathbb{N}$, there exist $A_k, B_k \in \mathbb{R}^{m\times n}$, $u^k, v^k \in \mathbb{R}^m_+$ with $||u^k + v^k||_1 = 1$, $w^k \in \mathbb{R}^m$ and $\tau_k \in \mathbb{R}_{++}$ such that*

$$[A_k - JH(x^{k-1})]^\top u^k + \lambda_{k-1}\langle \varepsilon^{k-1}, u^k\rangle q(x^{k-1}, x^k)\eta^k + B_k^\top v^k + \tau_k w^k = 0, \quad (14)$$

where

$$A_k \in \partial G(x^k),\ B_k \in \partial F(x^k),\ w^k \in B[0,1] \cap N_{\mathcal{D}}(x^k), \eta^k \in \partial q(x^{k-1}, \cdot)(x^k). \quad (15)$$

Proof It follows from definition of the algorithm that, for each $k \in \mathbb{N}$, x^k is a solution of the scalar problem

$$\min \left\{ \langle G(x) - JH(x^{k-1})(x - x^{k-1}) + \frac{\lambda_{k-1}}{2} q^2(x^{k-1}, x)\varepsilon^{k-1}, z\rangle \ : \ x \in \Omega_{k-1} \right\}.$$

Thus, from [38, Proposition 2.2], we have that x^k is a weak Pareto solution of the vector problem

$$\min_w \{G(x) - JH(x^{k-1})(x - x^{k-1}) + \frac{\lambda_{k-1}}{2} q^2(x^{k-1}, x)\varepsilon^{k-1} \ : \ x \in \Omega_{k-1}\}.$$

Note that

$$(f_k)_j(x) = g_j(x) - \langle \nabla h(x^k), x - x^k\rangle + \frac{\lambda_k}{2} q^2(x^k, x)\varepsilon^k_j, \text{ with } j \in \mathcal{I}, \quad (16)$$

and

$$(s_k)_j(x) = f_j(x) - f_j(x^k), \text{ with } j \in \mathcal{I}, \quad (17)$$

are locally Lipschitz functions, for each $j \in \mathcal{I}$, because each function in (16) is a sum of the convex function $g_j(\cdot) - \langle \nabla h(x^k), \cdot - x^k\rangle$ with the locally Lipschitz function $\frac{\lambda_k}{2} q^2(x^k, \cdot)\varepsilon^k_j$ and the functions in (17) are DC functions, for each $j \in \mathcal{I}$, and hence locally Lipschitz. Besides, taking into account that $x^k \in \mathcal{D}$, we have that $\partial d_{\mathcal{D}}(x^k) = B[0,1] \cap N_{\mathcal{D}}(x^k)$; see Burke et al. [12]. For each $k \in \mathbb{N}$ fixed applying Theorem 1 for $f_j(x) = (f_k)_j(x)$ and $s_j(x) = (s_k)_j(x)$ given by (16) and (17), respectively, we obtain that there exist $A_k, B_k \in \mathbb{R}^{m\times n}$, $u^k, v^k \in \mathbb{R}^m_+$ with $||u^k + v^k||_1 = 1$, $w^k \in \mathbb{R}^m$ and $\tau_k \in \mathbb{R}_{++}$ such that

$$0 \in [A_k - JH(x^{k-1})]^\top u^k + B_k^\top v^k + \tau_k w^k + \frac{\lambda_{k-1}}{2}\langle \varepsilon^{k-1}, u^k\rangle \partial q^2(x^{k-1}, \cdot)(x^k). \quad (18)$$

Applying Proposition 2 for $f_1(x) = f_2(x) = q(x^k, x)$, we have

$$\partial q^2(x^{k-1}, \cdot)(x^k) \subset 2q(x^{k-1}, x^k)\partial q(x^{k-1}, \cdot)(x^k).$$

Combining this fact with (18) we have that there exists $\eta^k \in \partial q(x^{k-1}, \cdot)(x^k)$ such that (15) holds and the proof is finished. □

Remark 10 Note that from (15), $\{u^k\}$, $\{v^k\}$ and $\{w^k\}$ are bounded sequences. As mentioned by Bolte et al. [9, Remark 1], ∂f_j is bounded on compact sets. So, we have that $\{A_k\}$, $\{B_k\}$ and $\{\eta^k\}$ are bounded sequences as long as $\{x^k\}$ is bounded. Therefore, if $\{\lambda_k\}$ and $\{x^k\}$ are bounded sequences, it follows from (14) 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 (14) implies

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

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

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

Hence, the desired result follows by using Lemma 1 and taking into account that $v^{k_0} \in \mathbb{R}^m_+ \backslash \{0\}$. □

As in [10], the stopping rule in Algorithm 1 can be changed by the following, 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 $p \in \mathbb{N}$. Similar to Corollary 1 last proposition combined with Lemma 1 allows to see that this condition is sufficient to get the stopping rule given in Algorithm 1.

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

5 Convergence analysis

In this section, we analyze the convergence of Algorithm 1. Note that, if the algorithm terminates after a finite number of iterations, it terminates at a Pareto critical point. Throughout this paper, we suppose that $\{x^k\}$ is an infinite sequence. Thus, we assume that $x^{k+1} \neq x^k$ and $u^k \neq 0$ for all $k \in \mathbb{N}$, in view of Corollaries 2 and 1, respectively. Additionally, we assume that $\{\lambda_k\}$ is a bounded sequence such that

$$\liminf_{k} \lambda_k \langle \varepsilon^k, z \rangle > 0, \tag{19}$$

which is easily verified if $z \in \mathbb{R}^m_{++}$ and $\inf_{k \in \mathbb{N}} \lambda_k > 0$.

Proposition 5 *The following properties hold:*

i) *$\{\langle F(x^k), z \rangle\}$ is decreasing;*
ii) $\lim\limits_{k \to +\infty} q(x^k, x^{k+1}) = 0$.

Proof Since H is a differentiable and convex function, we have

$$\langle JH(x^k)(x^{k+1} - x^k), z \rangle \leq \langle H(x^{k+1}) - H(x^k), z \rangle, \quad \forall k \in \mathbb{N}. \tag{20}$$

On the other hand, from (10), we have, for each $k \in \mathbb{N}$,

$$\langle G(x^{k+1}) - JH(x^k)(x^{k+1} - x^k) + \frac{\lambda_k}{2} q^2(x^k, x^{k+1}) \varepsilon^k, z \rangle \leq \langle G(x^k), z \rangle, \tag{21}$$

which, combined with (20) and $F(x) = G(x) - H(x)$, implies that

$$\langle F(x^{k+1}), z\rangle + \frac{\lambda_k}{2}\langle \varepsilon^k, z\rangle q^2(x^k, x^{k+1}) \leq \langle F(x^k), z\rangle, \quad \forall k \in \mathbb{N}. \tag{22}$$

Thus, using that $(\lambda_k/2)\langle \varepsilon^k, z\rangle > 0$ and $x^{k+1} \neq x^k$ for all $k \in \mathbb{N}$, item i) is proved. Now, taking into account that $F \succeq 0$ and $z \in \mathbb{R}^m_+\backslash\{0\}$, from item i), we have that $\{\langle F(x^k), z\rangle\}$ is a convergent sequence. From (22), we have

$$0 \leq \frac{\lambda_k}{2}\langle \varepsilon^k, z\rangle q^2(x^k, x^{k+1}) \leq \langle F(x^k), z\rangle - \langle F(x^{k+1}), z\rangle, \quad \forall k \in \mathbb{N}. \tag{23}$$

Thus, taking $k \to +\infty$ in last inequality, we obtain that the right hand-side of (23) vanishes which combined with (19) leads to the desired result. □

Theorem 2 *Every cluster point of $\{x^k\}$, if any, is a Pareto critical point of F.*

Proof Let $\hat{x}$ be a cluster point of $\{x^k\}$ and consider a subsequence $\{x^{k_l}\}$ of $\{x^k\}$ converging to $\hat{x}$. From Proposition 4, there exist sequences $\{A_k\}, \{B_k\} \subset \mathbb{R}^{m\times n}$, $\{u^k\}, \{v^k\} \subset \mathbb{R}^m_+$, $\{w^k\} \subset \mathbb{R}^m$, $\{\tau_k\} \subset \mathbb{R}_{++}$ and $\{\eta^k\}$, with $\eta^k \in \partial q(x^{k-1}, \cdot)(x^k)$, satisfying (14). Since $\{x^{k_l}\}$ is convergent and $\{\lambda_k\}$ is bounded, it follows from Remark 10 that $\{A_k\}$, $\{B_k\}$, $\{u^k\}$, $\{v^k\}$, $\{w^k\}$, $\{\tau_k\}$ and $\{\eta^k\}$ are bounded. Thus, we can assume without loss of generality that $A_{k_l} \to \hat{A}$, $B_{k_l} \to \hat{B}$, $u^{k_l} \to \hat{u}$, $v^{k_l} \to \hat{v}$, $w_{k_l} \to \hat{w}$ and $\tau_{k_l} \to \hat{\tau}$ (we may extract other subsequences if necessary). From (14) we have

$$[A_{k_l} - JH(x^{k_l-1})]^\top u^{k_l} + \gamma_{k_l} q(x^{k_l-1}, x^{k_l})\eta^{k_l} + B_{k_l}^\top v^{k_l} + \tau_{k_l} w^{k_l} = 0, \tag{24}$$

where $\gamma_{k_l} = \lambda_{k_l-1}\langle \varepsilon^{k_l-1}, u^{k_l}\rangle$. Note that $\{\gamma_k\}$ is bounded. This fact combined with Proposition 5 implies that $\{\gamma_{k_l} q(x^{k_l-1}, x^{k_l})\eta^{k_l}\}$ converges to 0 as $l \to +\infty$. Note that $\partial f_i(x) = \partial g_i(x) - \nabla h_i(x)$, for each $i \in \mathcal{I}$; see Bačák and Borwein [2]. Thus, taking the limit as $l \to +\infty$ in (24) and having in mind the closedness property of $\partial f(\cdot)$ and $N_{\mathcal{D}}(\cdot)$ (in the sense that if $w^k \to w$, $x^k \to x$ and $w^k \in \mathcal{X}(x^k)$, then $w \in \mathcal{X}(x)$), we have

$$\hat{C}^\top \hat{u} + \hat{B}^\top \hat{v} + \hat{\tau}\hat{w} = 0, \tag{25}$$

where $\hat{C} := (\hat{A} - JH(\hat{x})) \in \partial F(\hat{x})$, $\hat{B} \in \partial F(\hat{x})$ and $\hat{\tau}\hat{w} \in N_{\mathcal{D}}(\hat{x})$. Hence, from (25) we obtain

$$-(\hat{C}^\top \hat{u} + \hat{B}^\top \hat{v}) \in N_{\mathcal{D}}(\hat{x}). \tag{26}$$

Note that $\hat{u}, \hat{v} \in \mathbb{R}^m_+$ and from (15), we have that $\hat{u} \neq 0$ or $\hat{v} \neq 0$. Therefore, the desired result follows from (26) combined with Lemma 1. □

6 The "group dynamic" problem

In this section, we have in mind a group of producers $\mathcal{I} = \{1, 2, \dots, m\}$. The objective of each of them is a "to be increased" payoff (profit, utility), $\widetilde{f}_i(x) = \widetilde{g}_i(x) - \widetilde{h}_i(x)$, where $\widetilde{g}_i(x) \in \mathbb{R}_+$ and $\widetilde{h}_i(x) \in \mathbb{R}_+$ refer to their scalar revenue and cost function. The decision variables of the group is the vector $x \in \mathbb{R}^n$ which must satisfy some constraints. Each agent wants a payoff as high as possible. The goal of the group $\mathcal{I}$ is to approach and reach a Pareto point, or a Pareto critical point. The vectorial objective of the group is $\widetilde{F}(x) = \widetilde{G}(x) - \widetilde{H}(x) \in \mathbb{R}^m$, where $\widetilde{F}(x) = (\widetilde{f}_1(x), \dots, \widetilde{f}_m(x))$, $\widetilde{G}(x) = (\widetilde{g}_1(x), \dots, \widetilde{g}_m(x))$ and $\widetilde{H}(x) = (\widetilde{h}_1(x), \dots, \widetilde{h}_m(x))$. In a dynamic cooperative setting, all agents of the group will accept to change from the last position $x = x^k$ to the next, $y = x^{k+1}$ only if their payoffs do not decrease, i.e., if $\widetilde{f}_i(x^{k+1}) \geq \widetilde{f}_i(x^k)$, for all $i \in \mathcal{I}$. In the opposite case, some agents will quit the group or resist to change. This defines a cooperative improving dynamic $x^{k+1} \in \Omega(x^k)$, where $\Omega(x^k) = \left\{ y \in \mathcal{D} \;:\; \widetilde{F}(y) \succeq \widetilde{F}(x^k) \right\}$. The cooperative group dynamic problem is to find a cooperative improving dynamic $x^{k+1} \in \Omega(x^k)$ which converges to (approaches and reaches) a Pareto critical point.

In real life, in a dynamic setting, most of the time, increasing returns prevail, coming from fixed costs, learning by doing, ... This means that each marginal revenue and cost function is decreasing. Then, each revenue $\widetilde{g}_i$ and cost $\widetilde{h}_i$ function is concave. In this case the payoff of each agent is a difference between two concave functions. We consider "to be decreased" payoffs $f_i(y) = -\widetilde{f}_i(y)$, where $g_i(y) = -\widetilde{g}_i(y)$ and $h_i(y) = -\widetilde{h}_i(y)$, which are the difference of two convex functions $f_i(y) = g_i(y) - h_i(y)$. Then, we will provide a multi-objective DC algorithm which converges to a Pareto critical point, with a very important added requirement: the algorithm must follow a cooperative improving dynamic $x^{k+1} \in \Omega(x^k)$. In this behavioral setting, the papers of Bento et al. [5], Bonnel et al. [10] and Choung et al. [15] make "as if" they consider this case of a cooperative improving dynamic. However, the aforementioned works use convexity of the improving sets $\Omega(x^k)$, for all $k \in \mathbb{N}$. A motivation, in a dynamic context, to consider the constrained set $\Omega(x^k)$ is given in Bento et al. [5]. In the present paper, we drop this important convexity hypothesis, allowing for non-convex improving sets with DC multi-objective payoff functions. The algorithm proposed in Ji et al. [34] cannot be viewed as a cooperative improving dynamic and this is why algorithm (10) is better adapted to applications, for instance in behavioral sciences, than algorithm (8).

In this paper, we suppose that agents have limited (partial) knowledge. We also suppose that they do not know their cost function, while they are aware of their revenue function. This is why we consider approximate objective functions, using a linear local approximation of each cost function (see Cruz Neto et al. [18] for a more precise justification in the scalar case of one agent, with one objective).

6.1 Worthwhile to change processes

For applications, consider the behavioral context of the dynamics of human behavior, at the level of an agent with several objectives, or at the level of an organization where each agent has his own objective. We will use the recent variational rationality approach (see Soubeyran [55–59]) where, at each period, agents accept to change if these changes are worthwhile. This is the case when their (vectorial) motivation to change $M = U[A] \in \mathbb{R}^m$ are high enough with respect to their (vectorial) resistance to change $R = D[I] \in \mathbb{R}^m_+$, i.e., $M \geq \xi R$. High enough means that the threshold level $\xi > 0$ is high enough (it represents a satisfying ratio).

Let us show very briefly how, in this variational rationality context, our algorithm (10) represents a worthwhile stay and change process.

This paper considers the linear-quadratic case where the utilities $U[A] = A$ of their advantages to change $A \in \mathbb{R}^m$ are higher enough with respect to the disutilities $D[I] = I^2$ of their inconveniences to change $I \in \mathbb{R}^m_+$, i.e., $A \geq \xi I^2$, where I^2 is the vector of squares of each component of vector I. At each period, advantages to change moving from "repeating the last action x" to "doing a new action y" are defined as the difference $A = A(x, y) = P(y) - P(x)$ between the current (to be increased) vectorial payoff $P(y) \in \mathbb{R}^m$ to do the current action y and the current vectorial payoff $P(x) \in \mathbb{R}^m$ to repeat the last action x. In this paper $F(\cdot) = -P(\cdot)$ represents a vector of unsatisfied needs (to be decreased). Then, $A(x, y) = F(x) - F(y)$. Inconveniences to change $I(x, y) = C(x, y) - C(x, x)$ refer to the difference between costs $C(x, y) \in \mathbb{R}^m_+$ of being able to change from x to y and costs $C(x, x) \in \mathbb{R}^m_+$ of being able to repeat action x. In our paper $C(x, y) = q^2(x, y)\varepsilon$, where $\varepsilon \in \mathbb{R}^m_{++}$ is a vector of shares. Then, a change from x to y is worthwhile if $A(x, y) \geq \xi C(x, y)^2$, i.e.,

$$F(x) - F(y) \succeq \xi q^2(x, y)\varepsilon.$$

Let $z \in \mathbb{R}^n_{++}$ be a given vector of weights which helps to add different advantages to change and different inconveniences to change, and allow to compare their scalarized formulations. In this case a change is worthwhile if

$$\langle F(x) - F(y), z\rangle \geq \xi q^2(x, y)\langle \varepsilon, z\rangle.$$

Let k and $k + 1$ be the previous and current periods, where $x = x^k$ and $y = x^{k+1}$. Then, in the current period, a change from repeating the previous action x^k to perform the new action x^{k+1} is worthwhile if

$$\langle F(x^k) - F(x^{k+1}), z\rangle \geq \xi_k q^2(x^k, x^{k+1})\langle \varepsilon^k, z\rangle,$$

where $\{\varepsilon^k\}$ and $\xi_k = \lambda_k/2$ are as in (10).

This proves that our algorithm is a specific instance of a worthwhile stay and change process (see (22)). Our paper shows when, in the context of cooperative group dynamics where all agents require, at each period, that their payoffs do not decrease, a worthwhile stay and change transition approaches and reaches a Pareto critical point. Future research will consider the case of weak Pareto points, to better fits with applications to group dynamics ending in some trap. Of course the convex case ends in weak Pareto points which are traps; see [10, Theorem 3.1].

7 Numerical experiments

In this section we present some numerical results to verify the practical efficiency of the proposed method. We consider some academic test problems with smooth and non-smooth multi-objective DC functions. These functions are introduced in the sequel and after that we report our numerical experiments analyzing our results under different point of view. We coded our simulations in MATLAB on a machine AMD Athlon(tm) X2 with a Dual Core Processor and 1.20 GHz CPU. In all tests we consider Algorithm 1 with $\mathcal{D} = \mathbb{R}^2$, $\lambda_k = 1$, for all $k \in \mathbb{N}$, $q(x, y) = ||x - y||$ and the stopping rule $||x^{k+1} - x^k|| < 10^{-4}$.

The proximal point method has proved to be an efficient tool in several instances, be it through direct application or through methods which can be derived from it. The performance of the method strongly depends on the algorithm used to solve the subproblems. In this situation, it makes little sense to compare the proximal method with other methods in terms of computational efficiency, unless a specific procedure is chosen for solving the subproblems. In this paper, we refrain from discussing algorithms to solve the subproblems, and hence we skip a discussion of comparing it with other methods. However, we compare the performance of our method with the proximal-type method (8) for multi-objective DC function proposed in [34] (called Algorithm 2). In our simulations the subproblems are solved using the MATLAB routine "fmincon".

Problem 1 Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be a non-smooth multi-objective DC function given by $F(x, y) = G(x, y) - H(x, y)$, where

$$G(x,y) = \left((x-3)^2 + (y-4)^2 + 4, \sqrt{(x+6)^2 + (y+5)^2} + 4\right)$$

and

$$H(x,y) = \left(0.3(x-3)^2 + 0.2(y-4)^2, 0.01\left[(x+6)^2 + (y+5)^2\right]\right).$$

Problem 2 Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be a multi-objective DC function given by $F(x, y) = G(x, y) - H(x, y)$, where

$$G(x,y) = \left(\frac{1}{2}(x^2+y^2)^2, x^2+y^2\right)$$

and

$$H(x,y) = \left(\frac{1}{2}(x^4+y^4), 0.5x^2 + xy + 0.5y\right).$$

Problem 3 Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be a multi-objective DC function given by $F(x, y) = G(x, y) - H(x, y)$, where

$$G(x, y) = \left(x^2 + y^2, x^2 + y^2\right)$$

and

$$H(x, y) = \left(x^2 + y^2 - xy, 0.5x^2 + xy + 0.5y^2\right).$$

7.1 Computing Pareto critical points

Next, we perform Algorithm 1 for 30 randomly generated starting points. In this subsection, we consider $z = (1, 2)$ and $\varepsilon_k = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, for all $k \in \mathbb{N}$. The results are presented in Table 1. The column Iteration (k) refers to the number of iteration until the stopping rule is satisfied. The column Time(s) shows the time in second from the starting point is generated to the algorithm stoped. The other columns present the starting point used, the limit point obtained and its objective value, respectively.

Table 1 Running 30 times Algorithm 1 for Problem 1 with random starting points

#	Iteration (k)	Time (s)	x^0	x^*	$F(x^*) = (F_1(x^*), F_2(x^*))$
1	31	8.564	(-23.1101,20.7594)	(1.801711408,2.933901786)	(5.914379229,13.88901642)
2	8	4.025	(-19.657,-27.781)	(-5.590530497,-4.535074441)	(113.9360465,4.615694718)
3	31	8.065	(-11.2467,19.0368)	(2.174280382,3.269234602)	(4.904483488,14.2755238)
4	26	6.942	(-15.9218,5.2465)	(0.1859623857,1.437630995)	(14.79575333,12.13090998)
5	30	8.595	(25.4511,-10.1971)	(2.374066594,3.447416184)	(4.518533936,14.47985931)
6	14	5.538	(-17.6675,-8.1308)	(-3.795903734,-2.566031225)	(70.81922814,7.175810354)
7	24	7.644	(28.5429,13.0136)	(2.655948458,3.697500387)	(4.156064836,14.76505697)
8	15	5.881	(8.5986,16.954)	(2.826545431,3.8479396)	(4.039558432,14.935818)
9	18	5.71	(-2.0111,-16.0634)	(-2.01363433,-0.7154761526)	(39.38414278,9.509718382)
10	30	8.518	(-10.9273,17.3275)	(2.143692966,3.241899263)	(4.973056628,14.24408509)
11	11	4.54	(8.0666,9.587)	(2.933917755,3.942037931)	(4.005744483,15.04246158)
12	25	7.925	(2.255,25.153)	(2.517563127,3.575056862)	(4.307383071,14.62557106)
13	30	8.907	(16.8304,-10.3974)	(2.298136341,3.379692696)	(4.652653733,14.40236386)
14	28	8.237	(14.3249,-3.736)	(2.351373695,3.427171633)	(4.557007124,14.45671761)
15	27	8.19	(5.5078,-23.1294)	(0.4645940327,1.700468139)	(12.73007583,12.44372627)
16	27	7.909	(-10.8501,7.3112)	(1.029760389,2.227648398)	(9.230275067,13.06591714)
17	16	6.396	(23.6732,27.8152)	(2.8987099,3.911178252)	(4.013493221,15.00753443)
18	12	5.101	(-28.7962,-23.3609)	(-4.457223424,-3.278822358)	(85.31213082,6.257982982)
19	17	5.85	(-5.8188,-24.1683)	(-2.539302582,-1.251101629)	(47.53796585,8.841718983)
20	29	8.455	(-12.4171,10.7225)	(1.543743805,2.699572005)	(6.837367862,13.61730037)
21	29	8.752	(28.7875,-6.034)	(2.434734324,3.501410105)	(4.422541203,14.54156822)
22	31	9.236	(19.2531,-22.1028)	(2.253850821,3.340097721)	(4.738093825,14.35701921)
23	31	9.454	(20.6935,-16.0364)	(2.291705828,3.373978638)	(4.664698637,14.39580307)
24	17	6.225	(-25.9707,-10.1129)	(-3.08244084,-1.81301414)	(56.93016736,8.134077027)
25	26	7.924	(4.5141,-21.7908)	(0.2061560526,1.456622946)	(14.63890829,12.15364477)
26	30	8.97	(22.0679,-10.244)	(2.349298103,3.425342324)	(4.560574224,14.45461217)
27	15	4.477	(-0.7199,-4.1078)	(-2.362582838,-1.070028803)	(44.69426,9.06819408)
28	32	9.173	(-14.3675,28.9802)	(2.219850717,3.310015134)	(4.806906338,14.32233114)
29	20	6.974	(27.9419,26.911)	(2.794198542,3.8193532)	(4.055754574,14.90348543)
30	15	5.413	(-2.8967,-10.0984)	(-3.013352958,-1.741393723)	(55.68317121,8.224860327)

Algorithm 1 solves in each subproblem a multi-objective problem subject to the constraint set Ω_k which imposes a descent process. Such a condition does not make the algorithm unpractical or even costly. Actually, our analysis shows that Algorithm 1 has better performance (in all the 30 experiments) in number of iterations and running time than Algorithm 2. The performance of Algorithm 2 is presented in Table 2. In Algorithm 2, we use the same MATLAB routine to solve the subproblems and the same initial points considered in Table 1. The subproblems are solved subject to the set $S = \{(x, y) \in \mathbb{R}^2 : a \leq x \leq b,\ c \leq x \leq d\}$ large enough to include the solution set.

Table 2 Running 30 times Algorithm 2 for Problem 1 with the same starting points used in Algorithm 1

#	Iteration (k)	Time (s)	x^0	x^*	$F(x^*) = (F_1(x^*), F_2(x^*))$
1	46	56.019	(-23.1101,20.7594)	(0.6978730274,1.919117585)	(11.17410955,12.70254544)
2	24	23.463	(-19.657,-27.781)	(0.6978733322,1.919118982)	(11.17412519,12.70254271)
3	47	43.383	(-11.2467,19.0368)	(0.6978732346,1.919116249)	(11.17409894,12.70254729)
4	43	38.47	(-15.9218,5.2465)	(0.6978733502,1.919118773)	(11.17412379,12.70254296)
5	51	49.998	(25.4511,-10.1971)	(0.697870314,1.919088557)	(11.17381005,12.70259762)
6	40	33.649	(-17.6675,-8.1308)	(0.6978731952,1.919116781)	(11.17410326,12.70254654)
7	52	48.438	(28.5429,13.0136)	(0.6978706872,1.919089272)	(11.17381954,12.70259597)
8	45	41.122	(8.5986,16.954)	(0.6978734559,1.919117146)	(11.17410915,12.70254551)
9	41	37.534	(-2.0111,-16.0634)	(0.6978705264,1.919088559)	(11.17381152,12.70259737)
10	46	41.496	(-10.9273,17.3275)	(0.6978733094,1.919118079)	(11.17411692,12.70254415)
11	40	36.379	(8.0666,9.587)	(0.6978702805,1.919088136)	(11.17380551,12.70259842)
12	49	45.194	(2.255,25.153)	(0.6978730399,1.91911907)	(11.17412397,12.70254293)
13	49	44.491	(16.8304,-10.3974)	(0.6978704721,1.91908895)	(11.17381488,12.70259678)
14	50	43.555	(14.3249,-3.736)	(0.6978709245,1.919089005)	(11.17381898,12.70259607)
15	45	40.607	(5.5078,-23.1294)	(0.6978704927,1.919088007)	(11.17380594,12.70259834)
16	43	37.237	(-10.8501,7.3112)	(0.6978733224,1.919118831)	(11.17412434,12.70254286)
17	47	43.119	(23.6732,27.8152)	(0.6978704995,1.919088169)	(11.17380758,12.70259805)
18	39	34.772	(-28.7962,-23.3609)	(0.6978731226,1.919117273)	(11.17410761,12.70254578)
19	41	35.631	(-5.8188,-24.1683)	(0.6978707302,1.919089952)	(11.17382637,12.70259478)
20	45	38.516	(-12.4171,10.7225)	(0.6978732224,1.919116208)	(11.17409815,12.70254743)
21	51	47.112	(28.7875,-6.034)	(0.6978705631,1.919088)	(11.17380621,12.70259829)
22	49	45.442	(19.2531,-22.1028)	(0.6978706227,1.91908848)	(11.17381226,12.70259724)
23	50	47.097	(20.6935,-16.0364)	(0.6978707782,1.919088881)	(11.17381676,12.70259646)
24	41	36.114	(-25.9707,-10.1129)	(0.6978733163,1.919118396)	(11.17411966,12.70254368)
25	45	39.718	(4.5141,-21.7908)	(0.6978706845,1.919089198)	(11.17381916,12.70259604)
26	51	46.411	(22.0679,-10.244)	(0.6978709668,1.919089126)	(11.1738205,12.7025958)
27	38	29.952	(-0.7199,-4.1078)	(0.6978706621,1.919089802)	(11.17382461,12.70259509)
28	48	44.382	(-14.3675,28.9802)	(0.6978734495,1.919117127)	(11.17410888,12.70254556)
29	48	44.71	(27.9419,26.911)	(0.6978699387,1.919088207)	(11.17380254,12.70259893)
30	39	32.292	(-2.8967,-10.0984)	(0.697870396,1.919090211)	(11.17382542,12.70259495)

It is worth to mention that Algorithm 2 finds the same Pareto critical point of Problem 1 in all the 30 experiments. On the other hand, our method has been shown to be efficient to obtain an approximation of the solution set when it is performed many times for different initial points. This is illustrated in Figures 1, 2 and 3 where the blue squares, red stars and blue stars denote the generated initial points, the points of the sequence and the limit points of the sequence generated by Algorithm 1, respectively.

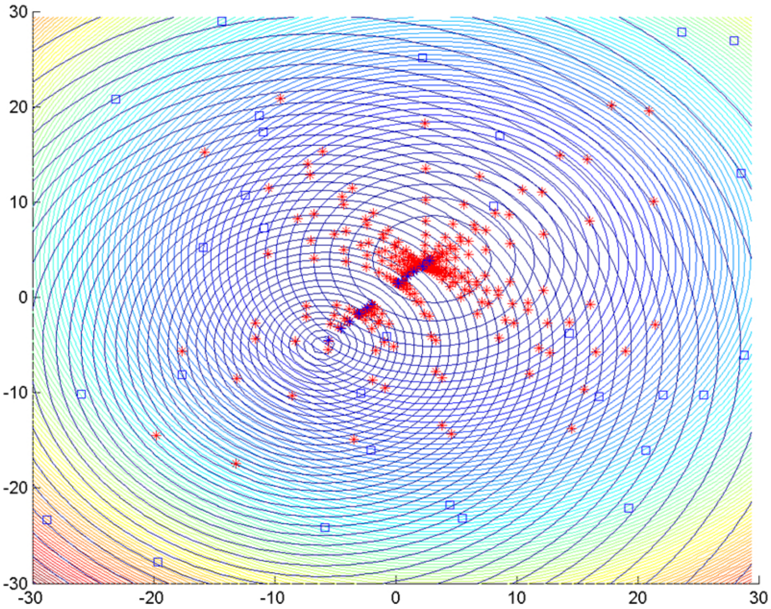

Fig. 1 Approximation of the solution set of Problem 1

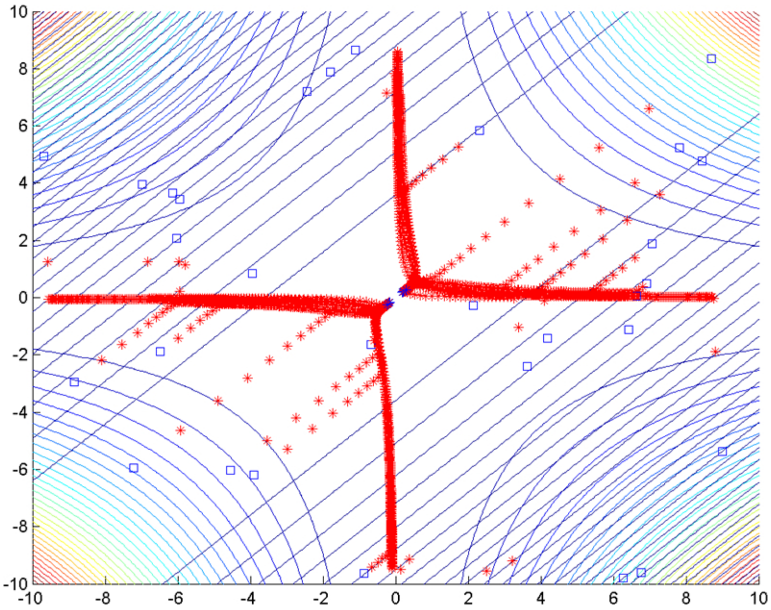

Fig. 2 Approximation of the solution set of Problem 2

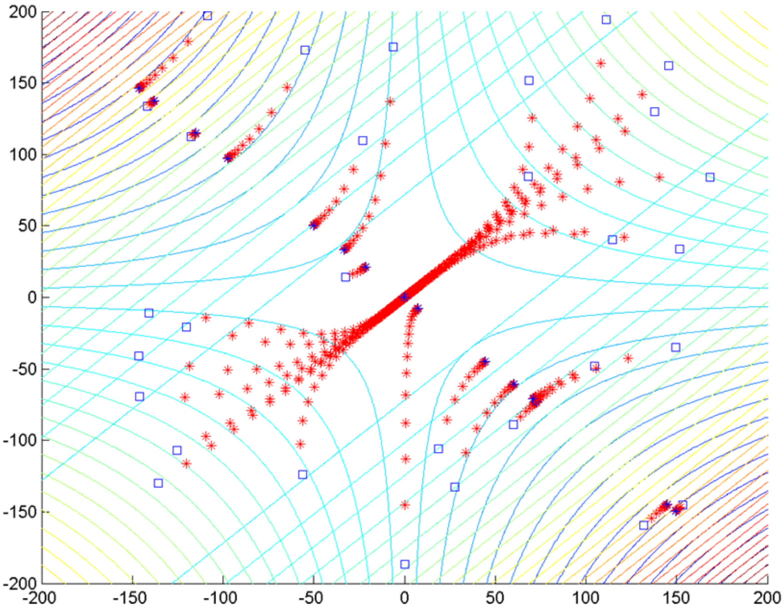

Fig. 3 Approximation of the solution set of Problem 3

We also mention that Algorithm 2 does not necessarily decrease all the components of the objective function as our method does. This is shown in Table 3 and 4 where the algorithms are applied to solve Problem 1 with initial point $x^0 = (20, -20)$. In last column of Table 4 we report the values $F(x^{k+1}) - F(x^k)$ for Algorithm 2 proving that the method is not a descent process.

7.2 Analysis of the parameters ε^k and z

In this subsection we analyze the performance of Algorithm 1 taking into account the variation of the parameters ε^k and z for Problem 1. Firstly, we analyze the influence of the parameter ε^k in the time of convergence of Algorithm 1 by considering it fixed and variable in all iterations. For this case, we run the algorithm 30 times with randomly generated initial points. Each time, it is considered $\varepsilon^k = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ fixed, for all $k \in \mathbb{N}$, and ε^k randomly variable for each $k \in \mathbb{N}$. The results are presented in Figure 4. One can note that considering the parameter ε^k variable provided a better performance of the method.

Secondly, we study the influence of parameter z in the time of convergence of the algorithm. To this end, we run the algorithm 30 times for each one of the following four cases: case 1 the initial point is $(-20, -20)$; case 2 the initial point is $(20, 20)$; case 3 the initial point is $(-20, 20)$; case 4 the initial point is $(20, -20)$. In all these cases, we consider the parameter ε^k randomly variable for each $k \in \mathbb{N}$ and for each one of the 30 times that the algorithm was performed we take $z = (10\cos(\frac{n\pi}{60}), 10\sin(\frac{n\pi}{60}))$, for $n = 1, 2\ldots, 30$. One can note that the choice of the parameter z does not significantly change the performance of the method (worst performance) except if it is taken close to the y-axis.

Table 3 Algorithm 1 applied to Problem 1 from the starting point $x^0 = (20, -20)$

Iteration (k)	time (s)	$\|\|x^{k+1} - x^k\|\|$	x^k	$F(x^k) = (F_1(x^k), F_2(x^k))$
1	0.344	91.355	(15.07009249,-12.30890954)	(318.7654171,21.32808068)
2	0.339	63.359	(11.46189075,-7.100753285)	(152.7038953,18.49449446)
3	0.296	43.751	(8.83401866,-3.602828076)	(74.06743741,16.67966911)
4	0.312	30.085	(6.934829845,-1.269588972)	(37.05287448,15.64975494)
5	0.305	20.636	(5.572649163,0.2804881576)	(19.70078128,15.10234905)
6	0.329	14.144	(4.601146173,1.308425417)	(11.59022734,14.81434204)
7	0.33	0.96991	(3.910751674,1.989659653)	(7.813802678,14.65680253)
8	0.309	0.66589	(3.421126776,2.440962486)	(6.068621808,14.56397469)
9	0.344	0.45786	(3.074246629,2.739804181)	(5.274333595,14.50424937)
10	0.314	0.31535	(2.828596791,2.937552554)	(4.923601002,14.46269997)
11	0.359	0.2176	(2.654637849,3.068279015)	(4.777975705,14.43215578)
12	0.359	0.15044	(2.531419453,3.154591954)	(4.725469221,14.40897532)
13	0.331	0.10422	(2.444108011,3.21149476)	(4.713703543,14.39112213)
14	0.25	0.072117	(2.384114703,3.251515485)	(4.713703545,14.37973172)
15	0.278	0.049534	(2.344644073,3.28144279)	(4.713703545,14.37434991)
16	0.287	0.033816	(2.318576009,3.302982706)	(4.713703545,14.37183977)
17	0.187	0.023001	(2.301273056,3.318136737)	(4.713703548,14.37067808)
18	0.218	0.01561	(2.28973383,3.32864975)	(4.713703549,14.37014301)
19	0.223	0.010577	(2.282010948,3.335876748)	(4.71370355,14.36989735)
20	0.22	0.0071596	(2.276827812,3.340815901)	(4.71370355,14.36978478)
21	0.205	0.0048432	(2.273342172,3.344178539)	(4.713703551,14.36973328)
22	0.218	0.0032749	(2.270994746,3.346462043)	(4.713703551,14.36970973)
23	0.219	0.0022137	(2.269412284,3.348010101)	(4.713703551,14.36969897)
24	0.179	0.0014914	(2.2683481,3.349054953)	(4.713703715,14.36969398)
25	0.172	0.0010094	(2.267628752,3.349763082)	(4.713703791,14.36969169)
26	0.172	0.00068307	(2.267142392,3.350242705)	(4.713703825,14.36969065)
27	0.187	0.00046217	(2.266813509,3.350567421)	(4.713703841,14.36969017)
28	0.187	0.00031274	(2.266591055,3.350787235)	(4.713703848,14.36968995)
29	0.14	0.0002783	(2.266393133,3.350982876)	(4.713703906,14.3696898)
30	0.11	0.00015983	(2.266279497,3.351095267)	(4.713703926,14.36968975)
31	0.11	9.1935E-005	(2.266214143,3.351159926)	(4.713703932,14.36968973)

8 Final remarks

Convergence properties of a proximal point method for vector-valued DC functions adding a (non-convex) vectorial improving constraint have been proved. Bento et al. [4] handled with a proximal algorithm for multi-objective optimization involving locally Lipchitz vector functions and (non-convex) vectorial improving constraints. Although we use the same approach of convergence analysis our method is different to the one considered in [4] because it minimizes at each iteration a convex approximation instead of the (non-convex) objective function. Our method generalizes (in a finite dimensional multi-objective

Table 4 Algorithm 2 applied to Problem 1 from the starting point $x^0 = (20, -20)$

Iteration (k)	time (s)	$\|\|x^{k+1} - x^k\|\|$	x^k	$F(x^{k+1}) - F(x^k)$
1	2.008	540.2782	(12.562497,-8.000002916)	(-483.89099980,-5.73896609)
2	1.406	396.3238	(8.378802864,-1.99999502)	(-130.15698383,-2.73676509)
3	0.861	108.2335	(6.025582502,1.000009166)	(-35.44415573,-0.89777979)
4	0.548	28.0215	(5.144387567,1.498727569)	(-5.38389099,-0.39665191)
5	0.733	1.1728	(4.610249688,1.187252036)	(-0.07969592,-0.46260705)
6	0.641	0.88719	(4.067198442,0.8705688819))	(0.48763633,-0.47816945)
7	0.502	0.83687	(3.527535025,0.7120754666)	(0.21125337,-0.42110684
8	0.504	0.68461	(3.043118796,0.7434488424)	(-0.35776162,-0.30941565)
9	0.521	0.57117	(2.614787298,0.803043836)	(-0.20510666,-0.25709370)
10	0.516	0.47313	(2.252940969,0.9022951263)	(-0.21300673,-0.19125228)
11	0.578	0.38491	(1.952390871,1.016024702))	(-0.17576213,-0.13957153
12	0.77	0.31081	(1.707198679,1.133397375)	(-0.14766312,-0.09763097)
13	0.543	0.25051	(1.508999834,1.245808462)	(-0.11924930,-0.06655698
14	0.439	0.2022	(1.349762745,1.348786032)	(-0.09516737,-0.04419338)
15	0.484	0.16364	(1.2221836 ,1.440223332)	(-0.07503942,-0.02867450)
16	0.464	0.13283	(1.120031828,1.519695262)	(-0.05888158,-0.01815865)
17	0.42	0.10802	(1.038228736,1.587677475)	(-0.04610299,-0.01117334)
18	0.394	0.087993	(0.9726313697,1.645176892)	(-0.03611252,-0.00662644)
19	0.405	0.071745	(0.9199493651,1.693397949)	(-0.02835212,-0.00371806)
20	0.393	0.058524	(0.877569491,1.733578141)	(-0.02232555,-0.00189737)
21	0.394	0.047747	(0.8434215179,1.766894321)	(-0.01764201,-0.00078698
22	0.417	0.038954	(0.8158644645,1.794414643)	(-0.01399127,-0.00013417)
23	0.408	0.031775	(0.7935956034,1.817080566)	(-0.01113498,0.00022864)
24	0.385	0.025915	(0.7755783156,1.835705368)	(-0.00889078,0.00041115)
25	0.396	0.021131	(0.7609856084,1.850981736)	(-0.00711987,0.00048440)
26	0.345	0.017228	(0.749155217,1.863494331)	(-0.00571666,0.00049383)
27	0.36	0.014042	(0.7395576356,1.873730913)	(-0.00460079,0.00046820)
28	0.359	0.011444	(0.7317663605,1.882097829)	(-0.00370959,0.00042505)
29	0.343	0.0093247	(0.7254380348,1.888931632)	(-0.00299615,0.00037520)
30	0.323	0.0075974	(0.7202956245,1.894510011)	(-0.00242332,0.00032481)
31	0.345	0.0061895	(0.7161152716,1.899061535)	(-0.00196234,0.00027728)
32	0.345	0.0050422	(0.7127158586,1.902773916)	(-0.00159064,0.00023429)
33	0.374	0.0041075	(0.7099506166,1.905801125)	(-0.00129045,0.00019641)
34	0.326	0.0033462	(0.7077005881,1.908269211)	(-0.00104770,0.00016369)
35	0.361	0.0027263	(0.7058691266,1.910281391)	(-0.00085115,0.00013580)
36	0.329	0.0022222	(0.7043774777,1.911922317)	(-0.00069195,0.00011227)
37	0.375	0.0018085	(0.7031644616,1.913258462)	(-0.00056303,0.00009260)
38	0.551	0.0014687	(0.7021800244,1.914343991)	(-0.00045720,0.00007604)
39	0.55	0.0011961	(0.7013784133,1.915228136)	(-0.00037062,0.00006216)
40	0.527	0.00097422	(0.7007257972,1.915948489)	(-0.00030196,0.00005102)
41	0.532	0.00079357	(0.7001943229,1.916535355)	(-0.00024561,0.00004174)
42	0.527	0.00064653	(0.699761428,1.917013555)	(-0.00019999,0.00003414)
43	0.583	0.00052688	(0.6994087086,1.917403296)	(-0.00016283,0.00002791)
44	0.567	0.00042954	(0.6991211918,1.91772106)	(-0.00013266,0.00002280)
45	0.547	0.0003504	(0.6988866665,1.917980293)	(-0.00010812,0.00001864)
46	0.296	0.00028612	(0.6986951772,1.918191971)	(-0.00008819,0.00001523)
47	0.141	0.0013963	(0.6977027884,1.919174066)	(-0.00008819,0.00001523)
48	0.205	0.00022594	(0.697905313,1.919118593)	(-0.00046806,0.00008155)
49	0.153	5.0025e-005	(0.6978705211,1.919087548)	(0.000215490,-0.00003755)

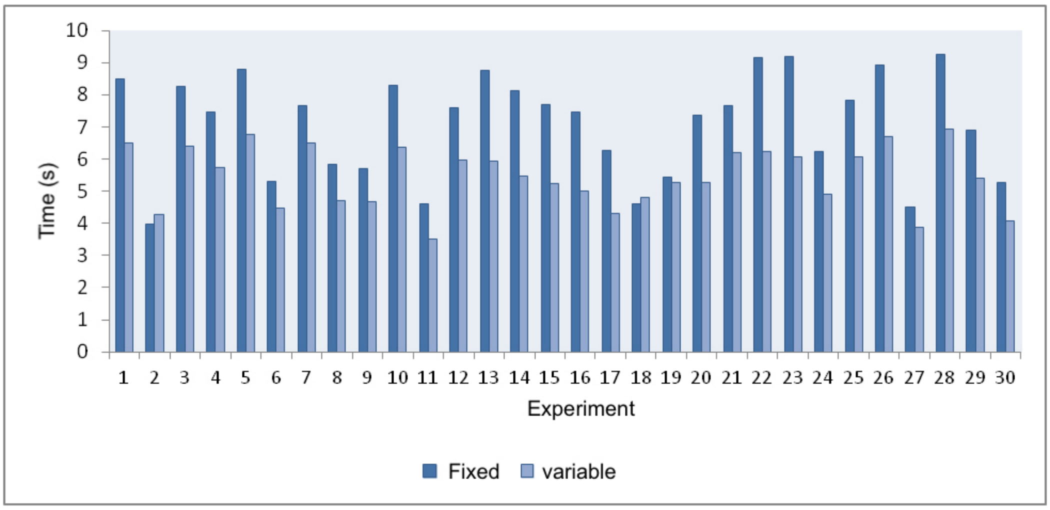

Fig. 4 Parameter ε^k in Algorithm 1: fixed × variable

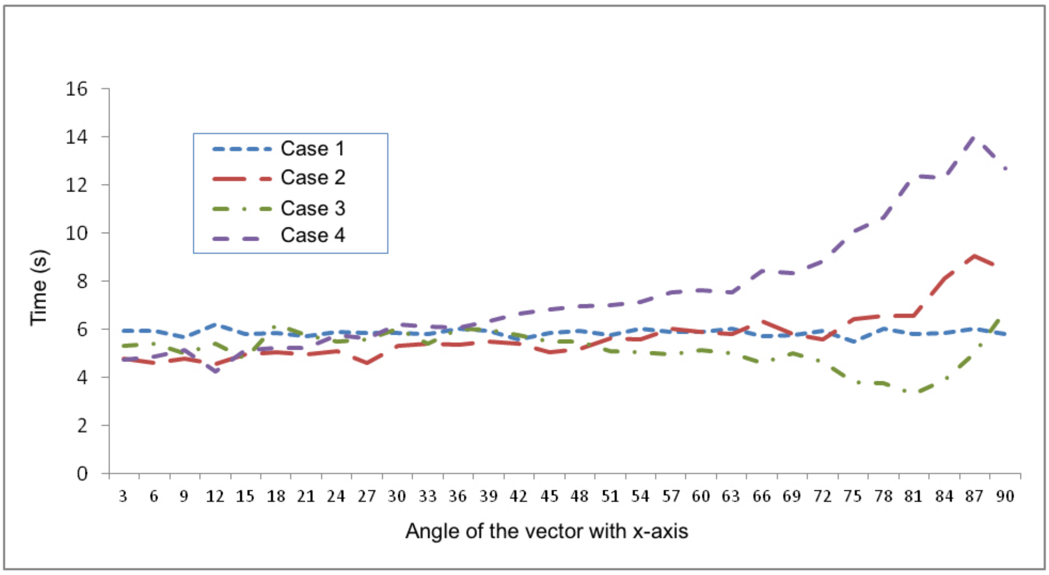

Fig. 5 Variety of parameters z in Algorithm 1

setting) the exact proximal point method introduced by Bonnel et al. [10] and our descent property implies the one obtained by Ji et al. [34] who do not consider vectorial improving constraints. Preliminary computational experiments have been conducted to check that the non-convex constraints does not make the method costly. The performance of the method has been compared to the proximal method considered by [34]. It also has been shown how our proximal method can greatly help to solve a "group dynamic" problem by using the recent variational rationality approach. In view of Remarks 4 and 6, it seems to be interesting to extend the approach proposed in [4] to infinite dimensional space and consider the concepts of relative minimizers in multi-objective optimization given by Mordukhovich [47,48] which generalize the concepts of Pareto and weak Pareto points. We will consider this topic in a future research.

References

1. Apolinário, H.C.F., Papa Quiroz, E.A., Oliveira, P.R.: A scalarization proximal point method for quasiconvex multiobjective minimization. J. Glob. Optim. 64, 79–96 (2016)
2. Bačák, M., Borwein, J.M.: On difference convexity of locally Lipschitz functions. Optimization 60, 961–978 (2011)
3. Bello Cruz, J.Y.: A subgradient method for vector optimization problems. SIAM J. Optim. 23, 2169–2182 (2013)
4. Bento, G.C., Cruz Neto, J.X., López, G., Soubeyran, A., Souza, J.C.O.: The proximal point method for locally Lipschitz functions in multiobjective optimization with application to the compromise problem, SIAM J. Optim. 28(2), 1104–1120 (2018)
5. Bento, G.C., Cruz Neto, J.X., Soubeyran, A.: A proximal point-type method for multicriteria optimization. Set-Valued Var. Anal. 22, 557–573 (2014)
6. Bento, G. C., Cruz Neto, J.X., Oliveira, P. R., Soubeyran, A.: The self regulation problem as an inexact steepest descent method for multicriteria optimization. Eur. J. Oper. Res. 235, 494–502 (2014)
7. Bento, G.C., Soubeyran, A.: Generalized inexact proximal algorithms: routine's formation with resistance to change, following worthwhile changes. J. Optim. Theory Appl. 166, 172–187 (2015)
8. Bento, G.C., Soubeyran, A.: A generalized inexact proximal point method for nonsmooth functions that satisfies Kurdyka-Lojasiewicz inequality. Set-Valued Var. Anal. 23, 501–517 (2015)
9. Bolte, J., Danilidis, A., Lewis, A., Shiota, M.: Clarke critical values of subanalytic Lipschitz continuous functions. Ann. Polon. Math. 87, 13–25 (2005)
10. Bonnel, H., Iusem, A.N., Svaiter, B.F.: Proximal methods in vector optimization. SIAM J. Optim. 15, 953–970 (2005)
11. Brito, A. S., Cruz Neto, J.X., Santos, P. S. M., Souza, S. S.: A relaxed projection method for solving multiobjective optimization problems. Eur. J. Oper. Res. 256, 17–23 (2017)
12. Burke, J.V., Ferris, M.C., Qian, M.: On the Clarke subdifferential of the distance function of a closed set. J. Math. Anal. Appl. 166, 199–213 (1992)
13. Ceng, L.C., Yao, J.C.: Approximate proximal methods in vector optimization. Eur. J. Oper. Res. 183, 1–19 (2007)
14. Ceng, L.C., Mordukhovich, B.S., Yao, J.C.: Hybrid approximate proximal method with auxiliary variational inequality for vector optimization. J. Optim. Theory Appl. 146, 267–303 (2010)
15. Choung, T.D., Mordukhovich, B.S., Yao, J.C.: Hybrid approximate proximal algorithms for efficient solutions in vector optimization. J. Nonlinear Convex Anal. 12, 257–286 (2011)
16. Clarke, F.H.: Generalized gradients and applications. Trans. Amer. Math. Soc. 205, 247–262 (1975)

17. Clarke, F.H.: Optimization and Nonsmooth Analysis. Classics in Applied Mathematics, v. 5, New York, Wiley (1990)
18. Cruz Neto, J.X., Oliveira, P.R., Soubeyran, A, Souza, J.C.O.: A generalized proximal linearized algorithm for DC functions with application to the optimal size of the firm problem. Preprint (2018)
19. Cruz Neto, J.X., Silva, G.J.P., Ferreira, O. P., Lopes, J.O.: A subgradient method for multiobjective optimization. Comput Optim Appl. 54, 461–472 (2013)
20. Dinh, N., Strodiot J.J., Nguyen, V.H.: Duality and optimality conditions for generalized equilibrium problems involving DC functions. . Glob. Optim. 48, 183–208 (2010)
21. Ferrer, A., Bagirov, A., Beliakov, G.: Solving DC programs using the cutting angle method. J. Glob. Optim. 61, 71–89 (2015)
22. Fliege, J., Graña Drummond, L. M., Svaiter, B. F.: Newton's method for multiobjective optimization. SIAM J. Optim. 20(2), 602–626 (2009)
23. Flores-Bazán, F., Oettli, W.: Simplified optimality conditions for minimizing the difference of vector-valued functions. J. Optim. Theory Appl. 108, 571–586 (2001)
24. Fukuda, E. H., Graña Drummond, L. M.: On the convergence of the projected gradient method for vector optimization. Optimization 60, 1009–1021 (2011)
25. Fukuda, E. H., Graña Drummond, L. M.: A survey on multiobjective descent methods. Pesqui. Oper., 34, 585–620 (2014)
26. Graña Drummond, L. M., Iusem, A. N.: A projected gradient method for vector optimization problems. Comput Optim Appl. 28, 5–29 (2004)
27. Graña Drummond, L.M., Svaiter, B.F.: A steepest descent method for vector optimization. J. Comput. Appl. Math. 175, 395–414 (2005)
28. Guo, X.L., Li, S.J.: Optimality conditions for vector optimization problems with difference of convex maps. J. Optim. Theory Appl. 162, 821–844 (2014)
29. Hartman, P.: On functions representable as a difference of convex functions. Pac. J. Math. 9, 707–713 (1959)
30. Hiriart-Urruty, J.B.: Generalized differentiabity, duality and optimization for problems dealing with difference of convex functions, Convexity and Duality in Optimization. Lecture Notes in Economics and Mathematical Systems 256, 37–70 (1986)
31. Holmberg, K., Tuy, H.: A production-transportation problem with stochastic demand and concave production costs. Math. Program. 85, 157–179 (1999)
32. Huang, X.X., Yang, X.Q.: Duality for multiobjective optimization via nonlinear Lagrangian functions. J. Optim. Theory Appl. 120, 111–127 (2004)
33. Jahn, J.: Vector Optimization: Theory, Applications and Extensions. Springer, Berlin (2004)
34. Ji, Y., Goh, M., De Souza, R.: Proximal point algorithms for multi-criteria optimization with the difference of convex objective functions. J. Optim. Theory Appl. 169, 280–289 (2016)
35. Lewin, K.: Frontiers in group dynamics: concept, method and reality in social science; social equilibria and social change. Human Relations 1, 5–41 (1947)
36. Lewin, K.: Field Theory in Social Science. Harper torchbooks. New York, Harper and Row (1964)
37. Luc, D.T., Tan, N.X., Tinh, P.N.: Convex Vector Functions and their Subdifferential. Acta Math Vietnam 23, 107–127 (1998)
38. Luc, D.T.: Theory of Vector Optimization. Lecture Notes in Econom. and Math. Syst., New York, Springer-Verlag (1989)
39. Mai, T. T., Luu, D.V.: Optimality conditions for weakly efficient solutions of vector variational inequalities via convexificators, J. Nonlinear Var. Anal. 2, 379–389 (2018)
40. Martinet, B.: Regularisation d'inéquations variationelles par approximations succesives. Rev. Francaise d'Inform. Recherche Oper. 4, 154–159 (1970)
41. Miettinen, K. M.: Nonlinear Multiobjective Optimization. Norwell, Kluwer Academic (1999)
42. Minami, M.: Weak Pareto-optimal necessary conditions in a nondifferentiable multiobjective program on a Banach space. J. Optim. Theory Appl. 41, 451–461 (1983)
43. Moreau, J.J.: Proximité et dualité dans un espace Hilbertien. Bull. Soc. Math. France 93, 273–299 (1965)
44. Moreno, F.G., Oliveira, P.R., Soubeyran, A.: A proximal point algorithm with quasi distance. Application to habit's formation. Optimization 61, 1383–1403 (2012)

45. Maingé, P-E., Moudafi, A.: Convergence of new inertial proximal methods for DC programming. SIAM J. Optim. 19, 397–413 (2008)
46. Mordukhovich B.S.: Variational Analysis and Generalized Differentiation. I. Basic Theory, Grundlehren der Mathematischen Wissenschaften, vol. 330. Springer, Berlin (2006)
47. Mordukhovich B.S.: Variational Analysis and Generalized Differentiation. II. Applications, Grundlehren der Mathematischen Wissenschaften, vol. 331. Springer, Berlin (2006)
48. Mordukhovich B.S.: Variational Analysis and Applications, Springer, New York (2018)
49. Muu, L.D., Quoc, T.D.: One step from DC optimization to DC mixed variational inequalities. Optimization 59, 63–76 (2010)
50. Poole, M.S., Van de Ven, A.H.: Handbook of Organizational Change and Innovation. New York, Oxford University Press (2004)
51. Qu, S., Liu, C., Goh, M., Li, Y., Ji, Y.: Nonsmooth multiobjective programming with quasi-Newton methods. Eur. J. Oper. Res. 235, 503–510 (2014)
52. Qu, S., Goh, M., Ji, Y., De Souza, R.: A new algorithm for linearly constrained c-convex vector optimization with a supply chain network risk application. Eur. J. Oper. Res. 247, 359–365 (2015)
53. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control. Optim. 14, 877–898 (1976)
54. Ross, G. T., Soland, R. M.: A multicriteria approach to the location of public facilities. Eur. J. Oper. Res. 4, 307–321 (1980)
55. Soubeyran, A.: Variational rationality, a theory of individual stability and change: worthwhile and ambidextry behaviors. GREQAM, Aix Marseillle University, France. Preprint (2009)
56. Soubeyran, A.: Variational rationality and the "unsatisfied man": routines and the course pursuit between aspirations, capabilities and beliefs. GREQAM, Aix Marseillle University, France. Preprint (2010)
57. Soubeyran, A.: Variational rationality. A theory of worthwhile stay and change approach-avoidance transitions ending in traps. GREQAM-AMSE, Aix Marseille University, France. Preprint (2016)
58. Soubeyran, A.: Variational rationality.1. An adaptive theory of the unsatisfied man. GREQAM-AMSE, Aix Marseille University, France. Preprint (2019)
59. Soubeyran, A.: Variational rationality. 2. A general theory of goals and intentions as satisficing worthwhile moves. GREQAM-AMSE, Aix Marseille University, France. Preprint (2019)
60. Souza, J.C.O., Oliveira, P.R.: A proximal point algorithm for DC functions on Hadamard manifolds. J. Glob. Optim. 63, 797–810 (2015)
61. Sun, W., Sampaio, R.J.B., Candido, M.A.B.: Proximal point algorithm for minimization of DC Functions. Journal of Computational Mathematics 21, 451–462 (2003)
62. Tao, P.D., Souad, E.B.: Algorithms for solving a class of nonconvex optimization problems: methods of subgradient. Fermat Days 85: Mathematics for Optimization 249–270 (1986)
63. Tao, P.D., An, L.T.H.: A DC Optimization algorithm for solving the trust region subproblem. SIAM J. Optim. 8, 476–505 (1998)
64. Thibault, L.: Subdifferentials of nonconvex vector-valued functions. J. Math. Anal. Appl. 86, 319–344 (1982)
65. Tuy, H., Horst, R.: Convergence and restart in branch-and-bound algorithms for global optimization. Application to concave minimization and dc optimization problems. Math. Program. 41, 161–183 (1988)
66. Villacorta, K.D.V., Oliveira, P.R.: An interior proximal method in vector optimization. Eur. J. Oper. Res. 214, 485–492 (2011)
67. Wen, B., Chen, X., Pong, T.K.: A proximal difference-of-convex algorithm with extrapolation. Comput Optim Appl. 69, 297–324 (2018)