Optimization and Engineering
https://doi.org/10.1007/s11081-020-09579-8

RESEARCH ARTICLE

On inexact projected gradient methods for solving variable vector optimization problems

J. Y. Bello-Cruz[1] · G. Bouza Allende[2]

Received: 7 April 2020 / Revised: 31 August 2020 / Accepted: 7 November 2020

Abstract
Variable order structures model situations in which the comparison between two points depends on a point-to-cone map. In this paper, inexact projected gradient methods for solving smooth constrained vector optimization problems on variable ordered spaces are presented. It is shown that every accumulation point of the generated sequences satisfies the first-order necessary optimality condition. Moreover, under suitable convexity assumptions for the objective function, it is proved that all accumulation points of any generated sequences are weakly efficient points. The convergence results are also derived in the particular case in which the problem is unconstrained and even if inexact directions are taken as descent directions. Furthermore, we investigate the application of the proposed method to optimization models where the domain of the variable order map coincides with the image of the objective function. In this case, similar concepts and convergence results are presented. Finally, some computational experiments designed to illustrate the behavior of the proposed inexact methods versus the exact ones (in terms of CPU time) are performed.

Keywords Gradient method · K-convexity · Variable order · Vector optimization · Weakly efficient points

Mathematics Subject Classification 90C29 · 90C52 · 65K05 · 35E10

✉ J. Y. Bello-Cruz
yunierbello@niu.edu

G. Bouza Allende
gema@matcom.uh.cu

[1] Department of Mathematical Sciences, Northern Illinois University, Watson Hall 366, DeKalb, IL 60115, USA

[2] Facultad de Matemática y Computación, Universidad de La Habana, 10400 La Habana, Cuba

Published online: 25 November 2020

1 Introduction

Variable order structures are a natural extension of the well-known fixed (partial) order given by a closed, pointed and convex cone; see Eichfelder (2014). This kind of orderings model situations in which the comparison between two points depends on a set-valued map. These problems have recently received much attention from the optimization community due to their broad application to several different areas. Variable order structures (VOS), given by a point-to-cone valued map, were well studied in Eichfelder (2014), Eichfelder (2011), Engau (2008), motivated by important applications. VOS appear in medical diagnosis (Eichfelder 2014), portfolio optimization (Wiecek 2007), capability theory of well-being (Bao et al. 2015b), psychological modeling (Bao et al. 2015a), consumer preferences (John 2001, 2006) and location theory, etc; see, for instance, Baatar and Wiecek (2006), Engau (2008). The main goal is to model elements of a certain set such that their objective function evaluation cannot be improved by the image of any other feasible point with respect to the variable order. So, their mathematical description corresponds to the so-called Optimization Problem(s) on Variable Ordered Spaces (OPVOS(s)). For the reasons mentioned above, although the variable order setting is a relatively new avenue of research, several papers and even books have been published with many real-life problems modeled via this approach.

An interesting application of vector optimization with a variable structure is given in the theory of consumer demand in economics by John (2001, 2006). These papers present a local and global theory to explain consumer behaviors. In the local approach, it is assumed that the consumer faces a nonempty set of feasible alternatives, $A \subset \mathbb{R}^n$. By contrast with the global approach, a local preference only requires that the consumer is able to rank alternatives in a small neighborhood of a given commodity bundle relative to that bundle. This idea can be represented by a comparative economic function $g\colon \mathbb{R}^n \to \mathbb{R}^n$ such that $\overline{y}$ in a neighborhood of y is interpreted to be better than y if and only if $g(\overline{y})^T(y-\overline{y}) < 0$. The choice set assigned to A in the local theory is then given by

$$C(A) := \left\{\overline{y} \in A : \forall y \in A,\ \ g(\overline{y})^T(y-\overline{y}) \geq 0\right\}.$$

This leads to a set-valued map $K\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ defined by

$$K(\overline{y}) := \left\{d \in \mathbb{R}^n : g(\overline{y})^T d \geq 0\right\}.$$

Note that the above point-to-cone map K defines the variable order and it satisfies the main assumptions of this paper under a suitable condition on the image of the feasible alternatives A by g, $g(A)$. If the consumer is interested in an alternative $\overline{y} \in A$ such that $\forall y \in A, g(\overline{y})^T(y-\overline{y}) \geq 0$, then $y-\overline{y} \in K(\overline{y})$ for all $y \in A$, i.e., $A \subseteq \{\overline{y}\}+K(\overline{y})$. Furthermore, the consumer is looking for alternatives $\bar{y} \in A$ such that

$$\forall y \in A \backslash \{\overline{y}\},\ \ g(\overline{y})^T(y-\overline{y}) < 0,$$

for all $y \in A \backslash \{\bar{y}\}$, $y - \bar{y} \notin K(\bar{y})$. This means that the consumer prefers alternatives $\bar{y} \in A$ such that $A \backslash (\bar{y} - K(\bar{y})) = \{\bar{y}\}$, i.e., minimal points of the vector optimization problem with variable domination structure $\min_K y$ $s.t.$ $y \in A$ are desired.

OPVOS(s) have been studied in Eichfelder and Duc Ha (2013), in the sense of finding a minimizer of the image of a vector function, with respect to a variable ordered structure depending on points in the image. It is a particular case of the problem described in Eichfelder (2011), where the goal of the model is to find a minimum of a set. Here we will consider a partial (variable) order defined by the cone-valued map, which is used to define our problem. We want to point out that OPVOSs generalize the classical vector optimization problems. Indeed, they correspond to the case in which the order is defined by a constant cone valued map. Many approaches have been proposed to solve the classical constrained vector optimization, such as projected gradient methods, proximal points iterations, weighting technique schemes, Newton-like and subgradient methods; see, for instance, Bello Cruz (2013), Bello Cruz and Lucambio Pérez (2014), Bonnel et al. (2005), Graña Drummond et al. (2008), Fliege et al. (2009), Jahn (1984), Graña Drummond and Iusem (2004), Bello Cruz et al. (2011), Fukuda and Graña Drummond (2011, 2013), Luc (1987), Fliege and Svaiter (2000), Graña Drummond and Svaiter (2005). It is worth noting that, as far as we know, only a few of these schemes mentioned above have been proposed and studied in the variable ordering setting; as, e.g., the steepest descent algorithm and sub-gradient-like algorithm for unconstrained problems, and a Newton-like method; see, for instance, Bento et al. (2018), Bello Cruz and Bouza Allende (2014), Bello Cruz et al. (2014). The use of extensions of these iterative algorithms to the variable ordering setting is currently a promising idea. So, it is important to find efficient solution algorithms for solving this kind of model.

In this paper, we present the projected gradient method with an inexact strategy for solving constrained variable order vector problems because of its simplicity and the adaptability to the vector structure of the problem. Moreover, we derive the convergence of the inexact projected gradient method for the unconstrained problem under variable order generalizing (Fukuda and Graña Drummond 2013). Finally, analogous results are obtained if the variable order is given by a point-to-cone map whose domain coincides with the image of the objective function.

This work is organized as follows. The next section provides some notations and preliminary results that will be used in the remainder of this paper. We also recall the concept of K-convexity of a function on a variable ordered space and present some properties of this class. Section 3 is devoted to the presentation of the inexact projected gradient algorithm. Section 4 devotes to present that every accumulation point of the generated sequence satisfies the first-order necessary optimality condition. Moreover, under K-convexity of the objective function, all accumulation points are shown to be weakly efficient in Sect. 4. Section 5 discusses the properties of this algorithm when the variable order is taken as a cone-value set from the image of the objective function. Section 6 introduces some examples illustrating the behavior of both proposed methods. Finally, some final remarks are given.

2 Preliminaries

In this section we present some preliminary results and definitions. First we introduce some useful notations. Throughout this paper, $p := q$ indicates that p is defined to be equal to q and we write $\mathbb{N}$ for the nonnegative integers $\{0, 1, 2, \ldots\}$. The canonical inner product in $\mathbb{R}^n$ will be denoted by $\langle \cdot, \cdot \rangle$ and the induced norm by $\| \cdot \|$. The closed ball centered at x with radius $r > 0$ is represented by $\mathbb{B}(x, r) := \{y \in \mathbb{R}^n : \operatorname{dist}(x, y) := \|y - x\| \leq r\}$ and also the sphere by $\mathbb{S}(x, r) := \{y \in \mathbb{B}(x, r) : \operatorname{dist}(x, y) = r\}$. Given two bounded sets A and B, we will consider $\mathrm{d}_H(A, B)$ as the *Hausdorff* distance, *i.e.*

$$\mathrm{d}_H(A, B) := \max \left\{ \sup_{a \in A} \inf_{b \in B} \operatorname{dist}(a, b),\ \sup_{b \in B} \inf_{a \in A} \operatorname{dist}(a, b) \right\},$$

or equivalently $\mathrm{d}_H(A, B) = \inf\{\epsilon \geq 0 : A \subseteq B_\epsilon \quad \text{and} \quad B \subseteq A_\epsilon\}$, where

$$D_\epsilon := \cup_{d \in D} \{x \in \mathbb{R}^n : \operatorname{dist}(d, x) \leq \epsilon\}$$

is the ϵ-enlargement of any set D. The set D^c and $\operatorname{int}(D)$ denote the complement and the interior of of D, respectively. The set $\operatorname{conv}(D)$ is used for the convex hull of D, i.e., the intersection of all convex sets containing D. If D is closed and convex, we define the orthogonal projection of x onto D, denoted by $P_D(x)$, as the unique point in D such that $\|P_D(x) - y\| \leq \|x - y\|$ for all $y \in D$. Given the partial order structure induced by a closed, convex and pointed (if $x, -x \in \mathcal{K}$ then $x = 0$) cone $\mathcal{K}$, the concept of infimum of a sequence can be defined. Indeed, for a sequence $(x^k)_{k \in \mathbb{N}}$ and a cone $\mathcal{K}$, the point x^* is $\inf_k\{x^k\}$ if and only if $(x^k - x^*)_{k \in \mathbb{N}} \subset \mathcal{K}$, and there is not x such that $x - x^* \in \mathcal{K}$, $x \neq x^*$ and $(x^k - x)_{k \in \mathbb{N}} \subset \mathcal{K}$. We say that $\mathcal{K}$ has the *Daniell* property if for all sequence $(x^k)_{k \in \mathbb{N}}$ such that $(x^k - x^{k+1})_{k \in \mathbb{N}} \subset \mathcal{K}$ and for some $\hat{x}$, $(x^k - \hat{x})_{k \in \mathbb{N}} \subset \mathcal{K}$, then $\lim_{k \to \infty} x^k = \inf_k\{x^k\}$. Here we assume that $K(x)$, $x \in \mathbb{R}^n$, is a convex, pointed, and closed cone, which guarantees that $K(x)$ has the Daniell property as was shown in Luc (2008). For each $x \in \mathbb{R}^n$, the dual cone of $K(x)$ is defined as $K^*(x) := \{w \in \mathbb{R}^m : \langle w, y \rangle \geq 0,\ \text{for all } y \in K(x)\}$. As usual, the graph of a set-valued map $K\colon \operatorname{dom}(K) \subset \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is the set $\operatorname{Gph}(K) := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m :\ y \in K(x)\}$ and its domain is $\operatorname{dom}(K) := \left\{x \in \mathbb{R}^n \,\middle|\, K(x) \neq \varnothing\right\}$. Finally, we remind that the mapping K is closed if $\operatorname{Gph}(K)$ is a closed subset of $\mathbb{R}^n \times \mathbb{R}^m$.

Next, we will define the constrained vector optimization problem on variable ordered spaces, which finds a K-minimizer of the vector function $F\colon \mathbb{R}^n \to \mathbb{R}^m$ in the set C as

$$K - \min F(x), \quad x \in C. \tag{1}$$

Here C is a nonempty convex and closed subset of $\mathbb{R}^n$ and $K : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is a point-to-cone map, where for each $x \in \mathbb{R}^n$, $K(x)$ is a pointed, convex and closed cone with nonempty interior. We say that the point $x^* \in C$ is a minimizer of problem (1) if for all $x \in C$,

$$F(x) - F(x^*) \notin -K(x^*) \backslash \{0\}.$$

The set of all minimizers (or efficient solutions) of problem (1) is denoted by S^*.

As in the case of classical vector optimization, related solution concepts such as weak efficiency and stationarity can be extended to the constrained setting. The point $x^* \in C$ is a weak solution of problem (1) iff for all $x \in C$, $F(x) - F(x^*) \notin -\mathrm{int}(K(x^*))$, S^w is the set of all weak solution points. We want to point out that this definition corresponds with the concept of weak minimizer given in Eichfelder (2011). On the other hand, if F is a continuously differentiable function, the point $x^* \in C$ is stationary, if and only if, for all $d \in C - x^* := \{v \in \mathbb{R}^n :\ v = c - x^*,\ \text{for some } c \in C\}$, we have

$$J_F(x^*)d \notin -\mathrm{int}(K(x^*)), \tag{2}$$

where J_F denotes the Jacobian matrix of F. The set of all stationary points will be denoted by S^s.

Now we present a version of Proposition 2.1 of Bello Cruz and Bouza Allende (2014), which is an extension of Lemma 5.2 of Fukuda and Graña Drummond (2011) for constrained OPVOS.

Proposition 2.1 *Let $x^* \in C$ be a weak solution of problem* (1). *If F is a continuously differentiable function (i.e., $F \in \mathcal{C}^1$), then x^* is a stationary point.*

Proof Suppose that $x^* \in C$ is a weak solution of problem (1). Fix $d \in C - x^*$. By definition there exists $c \in C$, such that $d = c - x^*$. Since C is a convex set, for all $\alpha \in [0, 1]$, $x^* + \alpha d \in C$. Since $x^* \in C$ is a weak solution of problem (1), $F(x^* + \alpha d) - F(x^*) \notin -\mathrm{int}(K(x^*))$. Hence,

$$F(x^* + \alpha d) - F(x^*) \in (-\mathrm{int}(K(x^*)))^c. \tag{3}$$

The Taylor expansion of F at x^* leads us to $F(x^* + \alpha d) = F(x^*) + \alpha J_F(x^*)d + o(\alpha)$. The last equation together with (3) implies $\alpha J_F(x^*)d + o(\alpha) \in (-\mathrm{int}(K(x^*)))^c$. Using that $(-\mathrm{int}(K(x^*)))^c$ is a closed cone, and since $\alpha > 0$, it follows that

$$J_F(x^*)d + \frac{o(\alpha)}{\alpha} \in (-\mathrm{int}(K(x^*)))^c.$$

Taking limit in the above inclusion, when α goes to 0, and using the closedness of $(-\mathrm{int}(K(x^*)))^c$, we obtain that $J_F(x^*)d \in (-\mathrm{int}(K(x^*)))^c$, establishing that $x^* \in S^s$. $\square$

In classical optimization, stationarity is also a sufficient condition for weak minimality under convexity assumptions. For vector optimization problems on variable ordered spaces, the convexity concept was introduced in Definition 3.1 of Bello Cruz and Bouza Allende (2014) as follows:

Definition 2.2 We say that F is a K-convex function on C if for all $\lambda \in [0, 1]$, $x, \bar{x} \in C$,

$$F(\lambda x + (1 - \lambda)\bar{x}) \in \lambda F(x) + (1 - \lambda)F(\bar{x}) - K(\lambda x + (1 - \lambda)\bar{x}).$$

It is worth noting that in the variable order setting the convexity of $\mathrm{epi}(F) := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m \mid F(x) \in y - K(x)\}$ is equivalent to the K-convexity of F if and only if $K(x) \equiv K$ for all $x \in \mathbb{R}^n$; see Proposition 3.1 of Bello Cruz and Bouza Allende (2014). As already shown in Bello Cruz and Bouza Allende (2014), K-convex functions have directional derivatives under natural assumptions; see Proposition 3.5 of Bello Cruz and Bouza Allende (2014). In particular, if $\mathrm{Gph}(K)$ is closed and $F \in \mathcal{C}^1$ is K-convex, then we have the gradient inclusion inequality as follows:

$$F(x) - F(\bar{x}) \in J_F(\bar{x})(x - \bar{x}) + K(\bar{x}), \quad x, \bar{x} \in C.$$

In the next proposition, we study the relation between stationarity, descent directions and the weak solution concept in the constrained sense for problem (1) extending to the variable order setting the results presented in Proposition 1 of Graña Drummond and Iusem (2004) and Lemma 5.2 of Fukuda and Graña Drummond (2011).

Proposition 2.3 *Let K be a point-to-cone and closed mapping, and $F \in \mathcal{C}^1$ be a K-convex function. Then:*

(i) The point $x^* \in C$ is a weak solution of problem (1) if and only if it is a stationary point.
(ii) If for all $d \in C - x^*$, $J_F(x^*)d \notin -K(x^*)\backslash\{0\}$, then x^* is a minimizer of problem (1).

Proof (i) Let $x^* \in S^s$, where S^s is the set of the stationary points. If $x^* \in C$ is not a weak minimizer then there exists $x \in C$ such that $-k_1 := F(x) - F(x^*) \in -\mathrm{int}(K(x^*))$. By the convexity of F, for some $k_2 \in K(x^*)$, we have

$$-k_1 = F(x) - F(x^*) = J_F(x^*)(x - x^*) + k_2.$$

It follows from the above equality that

$$J_F(x^*)(x - x^*) = -(k_1 + k_2). \tag{4}$$

Moreover, since $K(x^*)$ is a convex cone, $k_1 \in \mathrm{int}(K(x^*))$ and $k_2 \in K(x^*)$, it holds that $k_1 + k_2 \in \mathrm{int}(K(x^*))$. Thus, the last two equalities imply that $J_F(x^*)(x - x^*) \in -\mathrm{int}(K(x^*))$, which contradicts the fact that x^* is a stationary point because x belongs to C and hence $x - x^* \in C - x^*$. The converse implication was already shown in Proposition 2.1.

(ii) By contradiction suppose that there exists $x \in C$ such that $F(x) - F(x^*) = -k_1$, where $k_1 \in K(x^*)\backslash\{0\}$. Combining the previous condition with (4), it follows that

$$J_F(x^*)(x - x^*) = -(k_1 + k_2) \in -K(x^*).$$

Using that $J_F(x^*)(x - x^*) \notin -K(x^*)\backslash\{0\}$, we get that $(k_1 + k_2) = 0$, and as $k_1, k_2 \in K(x^*)$, $k_1 = -k_2$. It follows from the pointedness of the cone $K(x^*)$ that $k_1 = k_2 = 0$, contradicting the fact that $k_1 \neq 0$. $\square$

It is worth mentioning that the concept of K-convexity for F depends on the point-to-cone mapping K. Thus, this general approach covers several convexity concepts, from the scalar setting to the vector one, and it can be used to model a large number of applications; see, for instance, Bao et al. (2015a), Bao et al. (2015b), Eichfelder (2014). In Section 5 we discuss another variable order when the point-to-cone map depends on the image set of F, such kind of variable orders were introduced and studied in Bello Cruz et al. (2014), Bello Cruz and Bouza Allende (2014).

The Inexact Projected Gradient Method to solve problem (1) is presented in the next section.

3 The inexact projected gradient method

This section is devoted to present an inexact projected gradient method for solving constrained smooth problems equipped with a variable order. This method uses an Armijo-type line-search, which is done on inexact descent feasible directions. The proposed scheme here has two main differences with respect to the approach introduced in Bello Cruz and Bouza Allende (2014). (i) it solves constrained problems. (ii) it accepts approximate directions with some tolerance. It can be as an extension to the variable order setting of the one defined in Fukuda and Graña Drummond (2013) for vector optimization.

In the following, several constrained concepts and results will be presented and proved, which will be used in the convergence analysis of the proposed method below.

We start this section by presenting some definitions and basic properties of some auxiliary functions and sets, which will be useful in the convergence analysis of the proposed algorithms. Firstly, we define the set-valued mapping $G\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, which for each x, defines the set of the normalized generators of $K^*(x)$, *i.e.* $G(x) \subseteq K^*(x) \cap \mathbb{S}(0,1)$ is a compact set such that $\operatorname{conv}(G(x))$ is $K^*(x)$. Note that there smaller sets than $K^*(x) \cap \mathbb{S}(0,1)$ that fulfill those properties; see, for instance, Jahn (1986, 2004), Luc (1989). On the other hand, assuming that $F \in \mathcal{C}^1$, we consider the support function $\rho\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ as

$$\rho(x,w) := \max_{y \in G(x)} y^T w. \tag{5}$$

The function $\rho(x,w)$ was extensively studied for vector optimization in Proposition 3.1 of Fukuda and Graña Drummond (2011) and it is useful to define the useful auxiliary function $\phi\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, as

$$\phi(x,v) := \max_{y \in G(x)} y^T J_F(x) v. \tag{6}$$

Then, we are ready to introduce the following auxiliary subproblem, for each $x \in \mathbb{R}^n$ and $\beta > 0$, as

$$\min_{v \in C - x} \left\{ \frac{\|v\|^2}{2} + \beta \phi(x,v) \right\}. \tag{P_x}$$

Remark 3.1 Since $G(x)$ is compact, the function $\phi(x,\cdot)\colon \mathbb{R}^n \to \mathbb{R}$ is well defined for each $x \in \mathbb{R}^n$. Moreover, it is a continuous function.

The next proposition provides a characterization of stationarity using the auxiliary function ϕ, defined in (6). It can be as an extension for the variable vector optimization setting of Lemma 2.4 of Fukuda and Graña Drummond (2013). Moreover, the unconstrained version of the following proposition can be found in Proposition 4.1 of Bello Cruz and Bouza Allende (2014).

Proposition 3.2 *The following statements hold:*

(i) For each $x \in \mathbb{R}^n$, $\max_{y\in G(x)} y^T\hat{w} < 0$ *if and only if* $\hat{w} \in -\mathrm{int}(K(x))$.
(ii) The point x *is not stationary if and only if there exists* $v \in C - x$ *such that* $\phi(x,v) < 0$.
(iii) If $\phi(x,v) < 0$ *and* $\beta > 0$, *then there exists* $\bar{\lambda} > 0$ *such that* $\frac{\|\lambda v\|^2}{2} + \beta\phi(x,\lambda v) < 0$ *for all* $\lambda \in (0,\bar{\lambda}]$.
(iv) For each $x \in \mathbb{R}^n$, *subproblem* (P_x) *has a unique solution, denoted by* $v(x)$.

Proof (i) The result of this item follows as in Proposition 4.1(i) of Bello Cruz and Bouza Allende (2014).

(ii) Note that, fixing x, it follows from (6) that $\phi(x,v) = \rho(x, J_F(x)v)$. Then, by the definition of stationarity and item (i), the statement holds true.

(iii) It follows from the definition of $\phi(x,v)$ that $\phi(x,\cdot)$ is a positive homogeneous function. Thus, for all $\lambda > 0$,

$$\frac{\|\lambda v\|^2}{2} + \beta\phi(x,\lambda v) = \lambda\left(\lambda\frac{\|v\|^2}{2} + \beta\phi(x,v)\right). \tag{7}$$

Since $\phi(x,v) < 0$, there exists $\bar{\lambda} > 0$ small enough such that $\bar{\lambda}\frac{\|v\|^2}{2} + \beta\phi(x,v) < 0$. Hence, (7) together with the above inequality implies that $\frac{\|\lambda v\|^2}{2} + \beta\phi(x,\lambda v) < 0$, for all $\lambda \in (0,\bar{\lambda}]$, as desired.

(iv) Using the definition of the function $\phi(x,v)$, given in (7), it is easy to prove that $\phi(x,\cdot)$ is a sublinear function as well. Hence, $\phi(x,\cdot)$ is a convex function, and then, $\frac{\|v\|^2}{2} + \beta\phi(x,v)$ is a strongly convex function. Since C is a convex set, $C - x$ is also convex and therefore, subproblem (P_x) has a unique minimizer. □

Based on Proposition 3.2(iii), we can define $v(x)$ as the unique solution of subproblem (P_x) and $y(x,v)$ is an element of the compact set $G(x)$ such that $y(x,v)^T J_F(x)v = \phi(x,v)$. Next we will discuss about the continuity of the function

$$\theta_\beta(x) := \frac{\|v(x)\|^2}{2} + \beta\phi(x,v(x)), \tag{8}$$

which is related with the one defined in (35) of Graña Drummond and Iusem (2004).

The following proposition is the constrained version of Proposition 4.2 in Bello Cruz and Bouza Allende (2014). Items (i)-(ii), (iii) and (iv) are extensions for the variable vector optimization setting of Proposition 3 of Graña Drummond and Iusem (2004), Proposition 2.5 of Fukuda and Graña Drummond (2013) and Proposition 3.4 of Fukuda and Graña Drummond (2011), respectively.

Proposition 3.3 *Let $F \in \mathcal{C}^1$ and fix $\beta > 0$. Then, the following hold*

(i) $\theta_\beta(x) \leq 0$ *for all* $x \in C$.
(ii) $x \in C$ *is a stationary point if and only if* $\theta_\beta(x) = 0$.
(iii) $\|v(x)\| \leq 2\beta \|J_F(x)\|$ *for all* $x \in C$.
(iv) *If G is a closed map, then θ_β is an upper semi-continuous function on C.*

Proof (i) Note that as $0 \in C - x$ for all $x \in C$ and $\theta_\beta(x) \leq \dfrac{\|0\|^2}{2} + \beta\phi(x, 0) = 0$.

(ii) As shown in Proposition 3.2(ii), x is a non stationary point if and only if for some $v \in C - x$, $\phi(x, v) < 0$. Then, by Proposition 3.2(iii), there exists $\hat{v} \in C - x$ such that $\dfrac{\lambda^2}{2}\|\hat{v}\|^2 + \lambda\beta\phi(x, \hat{v}) < 0$ and hence $\theta_\beta(x) < 0$.

(iii) By (i), $0 \geq \theta_\beta(x) = \dfrac{\|v(x)\|^2}{2} + \beta y(x, v(x))^T J_F(x)v(x)$. Then, after some algebra, we get

$$\frac{\|v(x)\|^2}{2} \leq -\beta y(x, v(x))^T J_F(x)v(x) \leq \beta \|y(x, v(x))^T J_F(x)v(x)\|.$$

Using that $\|y(x, v(x))\| = 1$, it follows from the above inequality that

$$\frac{\|v(x)\|^2}{2} \leq \beta \|J_F(x)\| \|v(x)\|,$$

and the result follows after dividing the above inequality by the positive term $\|v(x)\|/2 \neq 0$.

(iv) Now we prove the upper semi-continuity of the function θ_β. Let $(x^k)_{k\in\mathbb{N}}$ be a sequence converging to x. Take $\hat{x} \in C$ such that $v(x) = \hat{x} - x$ and also denote $\hat{x}^k := v^k + x^k$. It is clear that, for all $k \in \mathbb{N}$, $\hat{x} - x^k \in C - x^k$, and so,

$$\begin{aligned}
\theta_\beta(x^k) &= \frac{\|\hat{x}^k - x^k\|^2}{2} + \beta\phi(x^k, \hat{x}^k - x^k) \\
&\leq \frac{\|\hat{x} - x^k\|^2}{2} + \beta\phi(x^k, \hat{x} - x^k) \\
&= \frac{\|\hat{x} - x^k\|^2}{2} + \beta y_k^T J_F(x^k)(\hat{x} - x^k).
\end{aligned} \tag{9}$$

Since each $y_k := y(x^k, \hat{x} - x^k)$ belongs to the compact set $G(x^k) \subseteq K^*(x^k) \cap \mathbb{S}(0, 1) \subseteq \mathbb{B}(0, 1)$ for all $k \in \mathbb{N}$, then the sequence $(y_k)_{k\in\mathbb{N}}$ is bounded. Therefore, there exists a convergent subsequence of $(y_k)_{k\in\mathbb{N}}$. We can assume without lost of generality that $\lim_{k\to\infty} y_k = y$, and also since G is closed, $y \in G(x)$. Taking limit in (9), we get

$$\begin{aligned}\limsup_{k\to\infty} \theta_\beta(x^k) &\le \limsup_{k\to\infty} \frac{\|\hat{x}-x^k\|^2}{2} + \beta y_k^T J_F(x^k)(\hat{x}-x^k)\\ &= \frac{\|\hat{x}-x\|^2}{2} + \beta y^T J_F(x)(\hat{x}-x)\\ &\le \frac{\|\hat{x}-x\|^2}{2} + \beta\phi(x,\hat{x}-x) = \theta_\beta(x).\end{aligned}$$

Then, the function θ_β, defined in (8), is upper semi-continuous. □

Lemma 3.4 *Consider any $x, \hat{x} \in C$ and $z \in \mathbb{R}^n$. If J_F is locally Lipschitz around x for all $x \in C$, $d_H(G(x), G(\hat{x})) \le L_G\|x-\hat{x}\|$ for some $L_G > 0$ and C is bounded, then*

$$\left|\phi(x,z) - \phi(\hat{x},z)\right| \le L\|x-\hat{x}\|,$$

for some $L > 0$. Hence, for all $v \in \mathbb{R}^n$, $\phi(\cdot, v)$ is a continuous function on C.

Proof By Proposition 4.1(iv) of Bello Cruz and Bouza Allende (2014), and using the Lipschitz assumption for G in C, $\rho(x,w)$, defined in (5), is also a Lipschitz function for all $(x,w) \in C \times W$ for any bounded subset $W \subset \mathbb{R}^n$. That is

$$|\rho(x_1,w_1) - \rho(x_2,w_2)| \le \hat{L}\|x_1-x_2\| + \|w_1-w_2\|, \tag{10}$$

for all $x_1, x_2 \in C$ and $w_1, w_2 \in W$, where $\|w_i\| \le M$, $i = 1, 2$ with $M > 0$ and $\hat{L} := L_G M$. Let $(x^k)_{k\in\mathbb{N}}$ be a sequence converging to x and $\hat{x}^k := v^k + x^k$. So, taking (10) for $x_1 = x$, $x_2 = x^k$, $w_1 = J_F(x)(\hat{x}^k - x^k)$ and $w_2 = J_F(x^k)(\hat{x}^k - x^k)$, we get

$$\begin{aligned}&\left|\rho\left(x, J_F(x)(\hat{x}^k-x^k)\right) - \rho\big(x^k, J_F(x^k)(\hat{x}^k-x^k)\big)\right|\\ &\quad\le \hat{L}\|x-x^k\| + \|(J_F(x)-J_F(x^k))(\hat{x}^k-x^k)\|\\ &\quad\le \hat{L}\|x-x^k\| + \|J_F(x)-J_F(x^k)\|\|\hat{x}^k-x^k\|,\end{aligned}$$

because of the compactness of C and the continuity of J_F, $\|J_F(x)(\hat{x}^k - x^k)\| \le M$ for all $k \in \mathbb{N}$ and $x \in C$. Noting that

$$\phi(x,\hat{x}^k-x^k) - \phi(x^k,\hat{x}^k-x^k) = \rho\left(x, J_F(x)(\hat{x}^k-x)\right) - \rho\left(x^k, J_F(x^k)(\hat{x}^k-x^k)\right),$$

and due to J_F is locally Lipschitz and (10), it follows that

$$\left|\phi(x,\hat{x}^k-x^k) - \phi(x^k,\hat{x}^k-x^k)\right| \le (\hat{L} + L_F\hat{M})\|x-x^k\|, \tag{11}$$

for all $x \in C$ with $\hat{L} := L_G M$ and L_F the Lipschitz constant of J_F and $\hat{M} > 0$ such $\|\hat{x}^k - x^k\| \le \hat{M}$ for all $k \in \mathbb{N}$. This proves the continuity of ϕ in the first argument. □

Now we can prove the lower semicontinuity of θ_β by following similar ideas of the result presented in Proposition 3.4 of Fukuda and Graña Drummond (2011) for vector optimization.

Proposition 3.5 *Let $F \in \mathcal{C}^1$ and consider any $x, \hat{x} \in C$ with C bounded. Then, if $d_H(G(x), G(\hat{x})) \le L_G\|x - \hat{x}\|$ for some $L_G > 0$ and J_F is locally Lipschitz around x for all $x \in C$, θ_β is a lower semicontinuous function on C.*

Proof We consider the function $\theta_\beta(x)$. Note further that

$$\begin{aligned}\theta_\beta(x) &\le \beta\phi(x, \hat{x}^k - x) + \frac{\|\hat{x}^k - x\|^2}{2}\\ &= \theta_\beta(x^k) + \beta\left[\phi(x, \hat{x}^k - x) - \phi(x^k, \hat{x}^k - x^k)\right] + \frac{\|\hat{x}^k - x\|^2 - \|\hat{x}^k - x^k\|^2}{2}\\ &= \theta_\beta(x^k) + \beta\left[\phi(x, \hat{x}^k - x) - \phi(x^k, \hat{x}^k - x^k)\right]\\ &\quad + \frac{1}{2}\left[-2\langle \hat{x}^k, x^k - x\rangle + \|x\|^2 - \|x^k\|^2\right].\end{aligned}$$

Thus, taking limit in the previous inequality and using Lemma 3.4, we get

$$\lim_{k\to\infty} \phi(x, \hat{x}^k - x) - \phi(x^k, \hat{x}^k - x^k) = 0.$$

Also, it is follows that $\lim_{k\to\infty} \frac{1}{2}\left[\|x\|^2 - \|x^k\|^2\right] - \langle \hat{x}^k, x^k - x\rangle = 0$. Hence,

$$\begin{aligned}\theta_\beta(x) &\le \liminf_{k\to\infty}\left\{\theta_\beta(x^k) + \beta\left[\phi(x, \hat{x}^k - x) - \phi(x^k, \hat{x}^k - x^k)\right]\right.\\ &\qquad \left. - \langle \hat{x}^k, x^k - x\rangle + \frac{\|x\|^2 - \|x^k\|^2}{2}\right\}\\ &= \liminf_{k\to\infty} \theta_\beta(x^k),\end{aligned}$$

establishing the desired result. □

Now we recall the concept of δ-approximate direction introduced in Definition 3.1 of Fukuda and Graña Drummond (2013) which in turn extends definition (4) of Fliege and Svaiter (2000).

Definition 3.6 Let $x \in C$ and $\beta > 0$. Given $\delta \in [0, 1)$, we say that v is a δ-approximate solution of subproblem (P_x) if $v \in C - x$ and $\beta\phi(x, v) + \frac{\|v\|^2}{2} \le (1 - \delta)\theta_\beta(x)$. If $v \ne 0$ we say that v is a δ-approximate direction at x.

Hence, from a numerical point of view, it would be interesting to consider algorithms in which the line-search is given over a δ-approximate direction at x for subproblem (P_x) instead of on an exact solution of it.

Remark 3.7 Note that if the solution of subproblem (P_x) is 0, then the only possible δ-approximate solution is $v = 0$. In other case, since $\theta_\beta(x) < 0$, there exist feasible directions v such that

$$\beta\phi(x, v) + \frac{\|v\|^2}{2} \in \left[\theta_\beta(x), (1-\delta)\theta_\beta(x)\right].$$

In particular $v(x)$, the solution of subproblem (P_x), is always a δ-approximate direction at x.

Next we present an inexact algorithm for solving problem (1). The algorithm requires the following exogenous parameters: $\delta \in [0, 1)$ and $\sigma, \gamma \in (0, 1)$ and $0 < \bar{\beta} \leq \hat{\beta} < +\infty$.

Inexact **P**rojected **G**radient Method (**IPG Method**). Assume that $\beta_k \in [\bar{\beta}, \hat{\beta}]$ for all $k \in \mathbb{N}$.

Initialization Take $x^0 \in \mathbb{R}^n$ and β_0.

Iterative step Given x^k and β_k, compute v^k a δ-approximate direction at x^k for

$$\min_{v \in C - x^k} \left\{ \frac{\|v\|^2}{2} + \beta_k \phi(x^k, v) \right\}. \qquad (P_{x^k})$$

If $v^k = 0$, then stop. Otherwise compute

$$j(k) := \min\left\{ j \in \mathbb{N} \colon F(x^k) - F(x^k + \gamma^j v^k) + \sigma\gamma^j J_F(x^k)v^k \in K(x^k) \right\}. \qquad (12)$$

Set $x^{k+1} = x^k + \gamma_k v^k \in C$, with $\gamma_k = \gamma^{j(k)}$.

It is worth noting that **IPG Method** extends Algorithm 3.3 of Fukuda and Graña Drummond (2013) to the variable order setting. Note that $v^k = 0$ if and only if x^k is stationary. Indeed, the fact $v^k = 0$ implies x^k stationary follows from Definition 3.6 combined with Proposition 3.3 (i)-(ii); while the converse implication is a consequence of Definition 3.6 together with the fact that $\theta_{\beta_k}(x^k) \leq \beta_k\phi(x^k, v^k) + 1/2\|v^k\|^2$. Next proposition proves that the stepsize γ_k is well defined for all $k \in \mathbb{N}$, *i.e.,* there exists a finite $j(k)$ that fulfills the Armijo-type rule given in (12) at each step of **IPG Method**. The proof of the next result uses a similar idea to the presented in Proposition 2.2 of Fukuda and Graña Drummond (2013).

Proposition 3.8 *Subproblem* (12) *has solution, i.e., there exists an index* $j(k)$ *is a nonnegative integer which is solution of* (12).

Proof If $v^k = 0$ then **IPG Method** stops. Otherwise, if $v^k \neq 0$ then by Proposition 3.3(ii), x^k is not a stationary point and $\theta_{\beta_k}(x^k) < 0$. Moreover,

$$\beta_k\phi(x^k, v^k) \leq \beta_k\phi(x^k, v^k) + \frac{\|v^k\|^2}{2} \leq (1-\delta)\theta_{\beta_k}(x^k) < 0.$$

Note further that $\phi(x^k, v^k) = \max_{y \in G(x^k)} y^T J_F(x^k)v^k < 0$. Thus, it follows from Proposition 3.2(i) that

$$J_F(x^k)v^k \in -\mathrm{int}(K(x^k)). \qquad (13)$$

Using the Taylor expansion of F at x^k, we obtain that

$$F(x^k) + \sigma\gamma^j J_F(x^k)v^k - F(x^k + \gamma^j v^k) = (\sigma - 1)\gamma^j J_F(x^k)v^k + o(\gamma^j). \quad (14)$$

Since $\sigma < 1$ and $K(x^k)$ is a cone, it follows from (13) that $(\sigma - 1)\gamma^j J_F(x^k)v^k \in \mathrm{int}(K(x^k))$. Then, there exists $\ell \in \mathbb{N}$ such that, for all $j \geq \ell$, we get $(\sigma - 1)\gamma^j J_F(x^k)v^k + o(\gamma^j) \in K(x^k)$. Combining the last inclusion with (14), we obtain $F(x^k) + \sigma\gamma^j J_F(x^k)v^k - F(x^k + \gamma^j v^k) \in K(x^k)$ for all $j \geq \ell$. Hence (12) holds for $j(k) = \ell$. □

Remark 3.9 After this proposition, it is clear that given (x^k, v^k), $j(k)$ is well-defined and it can be computed by implementing a backtracking procedure in the iterative step of **IPG Method**. Furthermore, the sequence generated by **IPG Method** is always feasible. Indeed, as x^k, $x^k + v^k \in C$, $\gamma_k \in (0, 1]$ and C is convex, $x^{k+1} = x^k + \gamma_k v^k \in C$.

4 Convergence analysis of IPG method

In this section we prove the convergence of **IPG Method** presented in the previous section. First we consider the general case and then the result is refined for K-convex functions. From now on, $(x^k)_{k\in\mathbb{N}}$ denote the sequence generated by **IPG Method**. We begin the section with the following lemma.

Lemma 4.1 *Let $F \in \mathcal{C}^1$. Assume that $\cup_{x\in C} K(x) \subseteq \mathcal{K}$, where $\mathcal{K}$ is a closed, pointed and convex cone. If x^* is an accumulation point of $(x^k)_{k\in\mathbb{N}}$, then* $\lim_{k\to\infty} F(x^k) = F(x^*)$.

Proof Let x^* be any accumulation point of the sequence $(x^k)_{k\in\mathbb{N}}$ and denote $(x^{i_k})_{k\in\mathbb{N}}$ a subsequence of $(x^k)_{k\in\mathbb{N}}$ such that $\lim_{k\to\infty} x^{i_k} = x^*$. It follows from the definition of Armijo-type line-search in (12) that

$$F(x^{k+1}) - F(x^k) - \sigma\gamma_k J_F(x^k)v^k \in -K(x^k). \quad (15)$$

Since **IPG Method** does not stop after finitely many steps, $v_k \neq 0$, which means that $\phi(x^k, v^k) < 0$. By Proposition 3.2(i), this means that $J_F(x^k)v^k \in -\mathrm{int}(K(x^k))$. Multiplying the last inclusion by $\sigma\gamma_k > 0$ and summing with (15), we get from the convexity of $K(x^k)$ that

$$F(x^{k+1}) - F(x^k) - \sigma\gamma_k J_F(x^k)v^k + \sigma\gamma_k J_F(x^k)v^k \in -\mathrm{int}(K(x^k)).$$

Thus, $F(x^{k+1}) - F(x^k) \in -\mathrm{int}(K(x^k))$. Since $\cup_{x\in C} K(x) \subseteq \mathcal{K}$ from assumption, it holds that $\mathrm{int}(K(x)) \subseteq \mathrm{int}(\mathcal{K})$ for all x, and $F(x^{k+1}) - F(x^k) \in -\mathrm{int}(\mathcal{K})$. Hence, $(F(x^k))_{k\in\mathbb{N}}$ is decreasing with respect to cone $\mathcal{K}$. The continuity of F imply that $\lim_{k\to\infty} F(x^{i_k}) = F(x^*)$. Then, to prove that the whole sequence $(F(x^k))_{k\in\mathbb{N}}$ converges to $F(x^*)$, we use the fact that the whole sequence $(F(x^k))_{k\in\mathbb{N}}$ is decreasing with respect to cone $\mathcal{K}$, which is a closed, pointed and convex cone; see, for instance, Proposition 3.1, pages 90, 91 of Peressini (1967) and Example 23.4 of Isac and Tammer (2010). Thus, we get that $\lim_{k\to\infty} F(x^k) = F(x^*)$, as desired. □

We present an analogous result as was proved in Proposition 3.3(iii) where v^k is a δ-solution of subproblem $\left(P_{x^k}\right)$, which gives us a upper bound for the norm of v^k. Next lemma is a version of Proposition 2.5 of Fukuda and Graña Drummond (2013) to the variable order setting.

Lemma 4.2 *Let $(x^k)_{k\in\mathbb{N}}$ be the sequence generated by* **IPG Method**. *Then,* $\|v^k\| \le 2\beta_k\|J_F(x^k)\|$.

Proof By the definition of δ-approximate direction $\beta_k\phi(x^k, v^k) + \dfrac{\|v^k\|^2}{2} \le (1-\delta)\theta_{\beta_k}(x^k)$. As was shown in Proposition 3.3(ii), $(1-\delta)\theta_{\beta_k}(x^k) \le 0$, since $x^k \in C$. Thus, $\dfrac{\|v^k\|^2}{2} \le -\beta_k\phi(x^k, v^k)$ and the result follows as in Proposition 3.3(iii). □

Next, we prove the stationarity of the accumulation points of the generated sequence. Some arguments used in the proof of the next theorem are similar to those of Theorem 3.5 of Fukuda and Graña Drummond (2013) and Theorem 5.1 of Bello Cruz and Bouza Allende (2014) for fixed and variable vector optimization, respectively.

Theorem 4.3 *Suppose that*

(a) $\cup_{x\in C} K(x) \subseteq \mathcal{K}$, *where $\mathcal{K}$ is a closed, pointed and convex cone.*
(b) The map G is closed.
(c) $d_H(G(x), G(\hat{x})) \le L_G\|x-\hat{x}\|$, *for all $x, \hat{x} \in C$.*
(d) J_F is a locally Lipschitz function around x, for all $x \in C$.
Then, all the accumulation points of $(x^k)_{k\in\mathbb{N}}$ the sequence generated by **IPG Method** *are stationary points of problem (1).*

Proof Let x^* be an accumulation point of the sequence $(x^k)_{k\in\mathbb{N}}$. Denote $(x^{i_k})_{k\in\mathbb{N}}$ any convergent subsequence to x^*. Since $F \in \mathcal{C}^1$, Lemma 4.2 implies that the subsequence $(v^{i_k})_{k\in\mathbb{N}}$ is also bounded and hence has a convergent subsequence. Without loss of generality, we assume that $(v^{i_k})_{k\in\mathbb{N}}$ converges to v^*, β_{i_k} and γ_{i_k} converge to $\beta_* \ge \bar{\beta}$ and γ^*, respectively. Recall that $\rho(x, w) = \max_{y\in G(x)} y^T w$.

By definition we have $F(x^{k+1}) - F(x^k) - \sigma\gamma_k J_F(x^k)v^k \in -\mathcal{K}$. Using Proposition 3.2(i), implies that $\rho\left(x^{i_k}, F(x^{k+1}) - F(x^k) - \sigma\gamma_k J_F(x^k)v^k\right) \le 0$. Since the function ρ is sublinear, as shown in Proposition 3.2 (iv), we get

$$\rho\left(x^k, F(x^{k+1}) - F(x^k)\right) \le \sigma\gamma_k\rho\left(x^k, J_F(x^k)v^k\right). \tag{16}$$

Using the semilinear property for ρ in the second argument, we can rewrite (16) as

$$\rho\left(x^k, F(x^k)\right) - \rho\left(x^k, F(x^{k+1})\right) \ge -\sigma\gamma_k\rho\left(x^k, J_F(x^k)v^k\right) \ge 0.$$

Now considering the subsequences $(x^{i_k})_{k\in\mathbb{N}}$ and $(v^{i_k})_{k\in\mathbb{N}}$, where $v^{i_k} = v(x^{i_k})$ on this inequality, we have

$$\lim_{k\to\infty}\left[\rho\left(x^{i_k}, F(x^{i_k})\right) - \rho\left(x^{i_k}, F(x^{i_k+1})\right)\right] \ge -\sigma\lim_{k\to\infty}\gamma_{i_k}\rho\left(x^{i_k}, J_F(x^{i_k})v^{i_k}\right) \ge 0.$$

As already was observed in the proof of Lemma 3.4, from **(c)** and **(d)** we have that ρ is continuous and moreover from Lemma 4.1, we have $\lim_{k\to\infty} F(x^k) = F(x^*)$. Thus,

$$\lim_{k\to\infty} \left[\rho\left(x^{i_k}, F(x^{i_k})\right) - \rho\left(x^{i_k}, F(x^{i_k+1})\right)\right] = \rho\left(x^*, F(x^*)\right) - \rho\left(x^*, F(x^*)\right) = 0.$$

These facts imply that $\lim_{k\to\infty} \gamma_{i_k} \rho\left(x^{i_k}, J_F(x^{i_k})v^{i_k}\right) = 0$. Hence we can split our analysis in two cases $\gamma^* > 0$ and $\gamma^* = 0$.

Case 1 $\gamma^* > 0$. Here

$$\lim_{k\to\infty} \phi(x^{i_k}, v^{i_k}) = \lim_{k\to\infty} \rho\left(x^{i_k}, J_F(x^{i_k})v^{i_k}\right) = 0. \tag{17}$$

Suppose that

$$\theta_{\beta_*}(x^*) = \|v(x^*)\|^2/2 + \beta_* \phi(x^*, v(x^*)) < -\epsilon < 0, \tag{18}$$

where $v(x^*) = \hat{x} - x^*$ with $\hat{x} \in C$. Due to the continuity of $\phi(\cdot, \cdot)$ in both arguments, Lemma 3.4 and (17) imply that

$$\phi(x^{i_k}, v^{i_k}) > -\frac{(1-\delta)\epsilon}{\max_{k\in\mathbb{N}} \beta_k} = -\frac{(1-\delta)\epsilon}{\hat{\beta}}$$

for k large enough. After note that $(\beta_k)_{k\in\mathbb{N}}$ is a positive and bounded sequence, then

$$\|v^{i_k}\|^2/2 + \beta_{i_k}\phi(x^{i_k}, v^{i_k}) \geq \beta_{i_k}\phi(x^{i_k}, v^{i_k}) > -\beta_{i_k}\frac{(1-\delta)\epsilon}{\hat{\beta}} \geq -(1-\delta)\epsilon. \tag{19}$$

By definition of the subsequence $(v^{i_k})_{k\in\mathbb{N}}$, we have, for all $v^{i_k} \in C - x^{i_k}$ and $v \in C - x^{i_k}$,

$$(1-\delta)\left(\frac{\|v\|^2}{2} + \beta_{i_k}\phi(x^{i_k}, v)\right) \geq (1-\delta)\theta_{\beta_{i_k}}(x^{i_k}) \geq \frac{\|v^{i_k}\|^2}{2} + \beta_{i_k}\phi(x^{i_k}, v^{i_k}). \tag{20}$$

Combining (19) and (20), we obtain that $(1-\delta)\left(\frac{\|v\|^2}{2} + \beta_{i_k}\phi(x^{i_k}, v)\right) > -(1-\delta)\epsilon$. In particular consider $\hat{v}^k = \hat{x} - x^{i_k}$. Dividing by $(1-\delta) > 0$, we obtain

$$\frac{\|\hat{v}^k\|^2}{2} + \beta_{i_k}\phi(x^{i_k}, \hat{v}^k) > -\epsilon.$$

By the continuity of function ϕ with respect to the first argument and taking limit in the previous inequality, lead us to the following inequality $\|v(x^*)\|^2/2 + \beta^*\phi(x^*, v(x^*)) \geq -\epsilon$. This fact and (18) imply

$$-\epsilon > \frac{\|v(x^*)\|^2}{2} + \beta_*\phi(x^*, v(x^*)) \geq -\epsilon,$$

which is a contradiction. Thus, we can conclude that $\theta_{\beta_*}(x^*) \geq 0$ and, hence, using Proposition 3.3, x^* is a stationary point if $\limsup_{k\to\infty} \gamma_{i_k} > 0$.

Case 2 $\gamma^* = 0$. We consider the previously defined convergent subsequences $(x^{i_k})_{k\in\mathbb{N}}$, $(\beta_{i_k})_{k\in\mathbb{N}}$, $(v^{i_k})_{k\in\mathbb{N}}$, $(\gamma_{i_k})_{k\in\mathbb{N}}$ convergent to x^*, β_*, v^* and $\gamma^* = 0$, respectively. Since $\beta_* > 0$, we get that

$$\rho\left(x^{i_k}, J_F(x^{i_k})v^{i_k}\right) \leq \rho\left(x^{i_k}, J_F(x^{i_k})v^{i_k}\right) + \frac{\|v^{i_k}\|^2}{2\beta_{i_k}}.$$

Since v^{i_k} is a δ-approximate direction at x^{i_k} for $(P_{x^{i_k}})$, see Definition 3.6, then

$$\rho\left(x^{i_k}, J_F(x^{i_k})v^{i_k}\right) + \frac{\|v^{i_k}\|^2}{2\beta_{i_k}} \leq \frac{(1-\delta)}{\beta_{i_k}}\theta_{\beta_{i_k}}(x^{i_k}) < 0.$$

Recalling that C is closed and **(b)** and **(c)** hold by Propositions 3.3 and 3.5, we have that θ is a continuous function. Taking limits above, we get $\rho(x^*, J_F(x^*)v^*) \leq -\frac{\|v^*\|^2}{2\beta_*} \leq 0$. Fix $q \in \mathbb{N}$. Then, for k large enough $F(x^{i_k} + \gamma^q v^{i_k}) \notin F(x^{i_k}) + \sigma\gamma^q J_F(x^{i_k})v^{i_k} - K(x^{i_k})$, as there exists $\hat{y}_{i_k} \in G(x^{i_k})$ such that $\left\langle F(x^{i_k} + \gamma^q v^{i_k}) - F(x^{i_k}) - \sigma\gamma^q J_F(x^{i_k})v^{i_k}, \hat{y}_{i_k}\right\rangle > 0$, it holds that

$$\rho\left(x^{i_k}, F(x^{i_k} + \gamma^q v^{i_k}) - F(x^{i_k}) - \sigma\gamma^q J_F(x^{i_k})v^{i_k}\right) \geq 0.$$

Taking limit as k tends to $+\infty$, and using that ρ is a continuous function, then

$$\rho\left(x^*, F(x^* + \gamma^q v^*) - F(x^*) - \sigma\gamma^q J_F(x^*)v^*\right) \geq 0.$$

But $\rho(x, \cdot)$ is a positive homogeneous function, so,

$$\rho\left(x^*, \frac{F(x^* + \gamma^q v^*) - F(x^*)}{\gamma^q} - \sigma J_F(x^*)v^*\right) \geq 0.$$

Taking limit as q tends to $+\infty$, we obtain $\rho\,(x^*, (1-\sigma)J_F(x^*)v^*) \geq 0$. Finally, since $\rho\,(x^*, J_F(x^*)v^*) \leq 0$, it holds $\rho\,(x^*, J_F(x^*)v^*) = 0$ and by Proposition 3.2(ii), this is equivalent to say that $x^* \in S^s$. □

The above result generalizes Theorem 5.1 of Bello Cruz and Bouza Allende (2014), where the exact steepest descent method for unconstrained problems was studied. Recall that at the exact variant of the algorithm, the direction v^k is computed as an exact solution of problem (P_{x^k}). In order to fill the gap between these two cases, we present two direct consequences of the above result, the inexact method for unconstrained problems and the exact method for the constrained problem.

Corollary 4.4 *Suppose that conditions (a)-(d) of Theorem 4.3 are fulfilled. Then all accumulation points of the sequence $(x^k)_{k\in\mathbb{N}}$ generated by the exact variant of* **IPG Method** *are stationary points of problem* (1).

Proof Apply Theorem 4.3 to the case $\delta = 0$. □

Corollary 4.5 *Suppose that conditions (a)-(d) of Theorem 4.3 are fulfilled for $C = \mathbb{R}^n$. If $(\beta_k)_{k\in\mathbb{N}}$ is a bounded sequence, then all accumulation points of $(x^k)_{k\in\mathbb{N}}$ computed by* **IPG Method** *are stationary points of problem (1).*

Proof Directly by applying Theorem 4.3 for $C = \mathbb{R}^n$. □

The result presented in Theorem 4.3 assumes the existence of accumulation points. We want to emphasize that this is a fact that takes place even when the projected gradient method is applied to the solution of classical scalar problems, i.e., $m = 1$ and $K(x) = \mathbb{R}_+$. The convergence of the whole sequence generated by the algorithm is only possible under stronger assumptions as convexity. Now, based on quasi-Fejér theory, we will prove the existence of accumulation points for the sequence generated by **IPG Method** when we assume that F is K-convex. We start by presenting its definitions and its properties.

Definition 4.6 Let S be a nonempty subset of $\mathbb{R}^n$. A sequence $(z^k)_{k\in\mathbb{N}}$ is said to be quasi-Fejér convergent to S if and only if for all $x \in S$, there exists $\bar{k}$ and a summable sequence $(\varepsilon_k)_{k\in\mathbb{N}} \subset \mathbb{R}_+$ such that $\|z^{k+1} - x\|^2 \leq \|z^k - x\|^2 + \varepsilon_k$ for all $k \geq \bar{k}$.

This definition originates in Browder (1967) and has been further elaborated in Iusem et al. (1994). A useful result on quasi-Fejér sequences is the following.

Fact 4.7 *If $(z^k)_{k\in\mathbb{N}}$ is quasi-Fejér convergent to S then,*

(i) The sequence $(z^k)_{k\in\mathbb{N}}$ is bounded.
(ii) If an accumulation point of $(z^k)_{k\in\mathbb{N}}$ belongs to S, then the whole sequence $(z^k)_{k\in\mathbb{N}}$ converges.

Proof See Theorem 1 of Burachik et al. (1995). □

Next, we introduce the following definition, which is an adaptation of Definition 4.2 of Fukuda and Graña Drummond (2013) and together with Proposition 4.9 gives us a condition for finding approximate directions for **IPG Method** without computing exact directions.

Definition 4.8 Let $x \in C$. A direction $v \in C - x$ is scalarization compatible (or simply s-compatible) at x if there exists $w \in \operatorname{conv}(G(x))$ such that $v = P_{C-x}(-\beta J_F(x)w)$.

In the following proposition we present the relation between inexact and s-compatible directions.

Proposition 4.9 *Let $x \in C$, $w \in \operatorname{conv}(G(x))$, $v = P_{C-x}(-\beta J_F(x)w)$ and $\delta \in [0, 1)$. If*

$$\beta\phi(J_F(x)v) \leq (1 - \delta)\beta\langle w, J_F(x)v\rangle - \frac{\delta}{2}\|v\|^2,$$

then v is a δ-approximate projected gradient direction.

Proof See Proposition 4.3 of Fukuda and Graña Drummond (2013). □

We start the analysis with a technical result which is an extension to the variable order case of Lemma 5.3 of Fukuda and Graña Drummond (2013).

Lemma 4.10 *Suppose that F is K-convex. Let $(x^k)_{k\in\mathbb{N}}$ be a sequence generated by* **IPG Method** *where v^k is an s-compatible direction at x^k, given by $v^k = P_{C-x^k}(-\beta_k J_F(x^k)w^k)$, with $w^k \in \operatorname{conv}(G(x^k))$ for all $k \in \mathbb{N}$. If for a given $\hat{x} \in C$ we have $F(\hat{x}) - F(x^k) \in -K(x^k)$, then*

$$\|x^{k+1} - \hat{x}\|^2 \leq \|x^k - \hat{x}\|^2 + 2\beta_k\gamma_k|\langle w^k, J_F(x^k)v^k\rangle|.$$

Proof Since $x^{k+1} = x^k + \gamma_k v^k$, we have $\|x^{k+1} - \hat{x}\|^2 = \|x^k - \hat{x}\|^2 + \gamma_k^2\|v^k\|^2 - 2\gamma_k\langle v^k, \hat{x} - x^k\rangle$. Let us analyze the rightmost term of the above expression. It follows from the definition of v^k and the obtuse angle property of projections that $\langle -\beta_k J_F(x^k)w^k - v^k, v - v^k\rangle \leq 0$, for all $v \in C - x^k$. Taking $v = \hat{x} - x^k \in C - x^k$ on the above inequality, we obtain

$$-\langle v^k, \hat{x} - x^k\rangle \leq \beta_k\langle w^k, J_F(x^k)(\hat{x} - x^k)\rangle - \beta_k\langle w^k, J_F(x^k)v^k\rangle - \|v^k\|^2.$$

Now, it follows from the convexity of F that $\langle w^k, J_F(x^k)(\hat{x} - x^k)\rangle \leq \langle w^k, F(\hat{x}) - F(x^k)\rangle$. Also the fact $F(\hat{x}) \preceq_{K(x^k)} F(x^k)$, i.e., $F(\hat{x}) - F(x^k) \in -K(x^k)$, together with $w^k \in K^*(x^k)$ imply that $\langle w^k, F(\hat{x}) - F(x^k)\rangle \leq 0$. Moreover, by using $J_F(x^k)v^k \in \operatorname{int}(-K(x^k))$ and w^k in $\operatorname{conv}(G(x^k)) = K^*(x^k)$, we have $\langle w^k, J_F(x^k)v^k\rangle < 0$. Thus, we get

$$-\langle v^k, \hat{x} - x^k\rangle \leq \beta_k|\langle w^k, J_F(x^k)v^k\rangle| - \|v^k\|^2.$$

The result follows because $\gamma_k \in (0, 1]$. □

We still need to make a couple of supplementary assumptions, which are standard in convergence analysis of classical (scalar-valued) methods and its extensions to the vector optimization setting.

Assumption 4.4 Let $(z^k)_{k\in\mathbb{N}} \in F(C)$ be a sequence such that $z^k - z^{k+1} \in K(x^k)$ for all $k \in \mathbb{N}$ and $z \in F(C)$, $z^k - z \in \mathcal{K}$ for some closed, convex and pointed cone $\mathcal{K}$, $\cup_{k\in\mathbb{N}}K(x^k) \subset \mathcal{K}$. Then there exists $\hat{x} \in C$ such that $F(\hat{x}) \preceq_{\mathcal{K}} z^k$ for all $k \in \mathbb{N}$, i.e., $F(\hat{x}) - z^k \in -\mathcal{K}$. In this case, we say that the sequence $(z^k)_{k\in\mathbb{N}}$ is bounded from below with respect to $\mathcal{K}$.

Recently, it was observed in Kim et al. (2018) that this assumption could be replaced by assuming that the restriction of F on C has compact sections. This assumption is related to the completeness of the image of F. It is important to mention that completeness is a standard assumption for ensuring the existence of efficient points in vector problems in Luc (1989).

Assumption 4.5 The search direction v^k is s-compatible at x^k, that is to say, $v^k = P_{C-x^k}(-\beta J_F(x^k)^T w^k)$, where $w^k \in \operatorname{conv}(G(x^k))$ for all $k \in \mathbb{N}$ where P_{C-x^k} is the orthogonal projection onto $C - x^k$.

This assumption holds automatically in the exact case. Moreover, it has been widely used in the literature in the vector case; see, for instance, Fukuda and Graña Drummond (2013). A version of these assumptions is also used in Fukuda and Graña Drummond (2013) when the partial order is given by a constant cone.

The next result is an extension to the variable order setting of Theorem 5.6 of Fukuda and Graña Drummond (2013).

Theorem 4.11 *Assume that F is K-convex and that Assumptions 4.4 and 4.5 hold. If $\mathrm{int}(\cap_{k\in\mathbb{N}} K(x^k)) \neq \varnothing$ and there exists $\mathcal{K}$, a pointed, closed and convex cone such that $K(x^k) \subset \mathcal{K}$ for all $k \in \mathbb{N}$, then every sequence generated by* **IPG Method** *is bounded and its accumulation points are weakly efficient solutions.*

Proof Let us consider the set $T := \{x \in C : F(x^k) - F(x) \in K(x^k),\ \text{for all } k\}$, and take $\hat{x} \in T$, which exists as consequence of Assumption 4.4 together with (12) and the fact that $\cup_{k\in\mathbb{N}} K(x^k) \subset \mathcal{K}$. Since F is a K-convex function and Assumption 4.5 holds, it follows from Lemma 4.10 that

$$\|x^{k+1} - \hat{x}\|^2 \leq \|x^k - \hat{x}\|^2 + 2\beta_k\gamma_k |\langle w^k, J_F(x^k)v^k\rangle|, \tag{21}$$

for all $k \in \mathbb{N}$. By the definition of v^k, it is a descent condition. This means that $-J_F(x^k)v^k \in K(x^k)$. Hence $\langle w^k, J_F(x^k)v^k\rangle \leq 0$. Then,

$$\|x^{k+1} - \hat{x}\|^2 - \|x^k - \hat{x}\|^2 \leq 2\beta_k\gamma_k |\langle w^k, J_F(x^k)v^k\rangle| \leq -2\beta_k\gamma_k \langle w^k, J_F(x^k)v^k\rangle. \tag{22}$$

On the other hand as $\mathcal{K}$ is a closed, convex and pointed cone with nonempty interior, $\mathcal{K}^*$ is also a closed, convex and pointed cone with nonempty interior. Since $K(x^k) \subset \mathcal{K}$, it holds that $\mathcal{K}^* \subset K^*(x^k)$. Hence $\mathcal{K}^* \subset \cap_{k\in\mathbb{N}} K^*(x^k)$. Let $\omega_1, \ldots, \omega_m \in \mathcal{K}^*$ be a basis of $\mathbb{R}^m$ which exists because $\mathrm{int}(K^*) \neq \emptyset$.

Then, there exist $\alpha_1^k, \ldots, \alpha_m^k \in \mathbb{R}$ such that $w^k = \sum_{i=1}^m \alpha_i^k \omega_i$. Substituting in (22),

$$\|x^{k+1} - \hat{x}\|^2 - \|x^k - \hat{x}\|^2 \leq -2\beta_k\gamma_k \sum_{i=1}^m \alpha_i^k \langle \omega_i, J_F(x^k)v^k\rangle. \tag{23}$$

On the other hand, since $-J_F(x^k)v^k \in K(x^k)$, $\omega_1, \ldots, \omega_m \in \mathcal{K}^* \subset K^*(x^k)$ and $\beta_k, \gamma_k > 0$ for all $k \in \mathbb{N}$, it holds $\langle \omega_i, -2\beta_k\gamma_k J_F(x^k)v^k\rangle \geq 0$. Without loss of generality, we can assume that $\|\omega_i\| = 1$ because the normalized non-null vectors are still a basis of $\mathbb{R}^n$. Then, α_i^k is uniformly bounded, i.e. there exists $M > 0$ such that for all k, i $|\alpha_i^k| \leq M$. Hence,

$$\|x^{k+1} - \hat{x}\|^2 - \|x^k - \hat{x}\|^2 \leq -2M\beta_k\gamma_k \sum_{i=1}^m \langle \omega_i, J_F(x^k)v^k\rangle. \tag{24}$$

By the Armijo-type line-search in (12), $F(x^{k+1}) - F(x^k) - \gamma_k \sigma J_F(x^k) v^k \in -K(x^k)$. Recall that $\omega_i \in \cap_{k \in \mathbb{N}} K^*(x^k)$, we obtain

$$\frac{\langle \omega_i, F(x^k) - F(x^{k+1}) \rangle}{\sigma} \geq \langle \omega_i, -\gamma_k J_F(x^k) v^k \rangle.$$

It follows from (24) that

$$\|x^{k+1} - \hat{x}\|^2 - \|x^k - \hat{x}\|^2 \leq 2 \frac{M}{\sigma} \beta_k \sum_{i=1}^{m} \langle \omega_i, F(x^k) - F(x^{k+1}) \rangle. \tag{25}$$

For the Fejér convergence of $(x^k)_{k \in \mathbb{N}}$ to T, it is enough to prove that the term $\beta_k \sum_{i=1}^{m} \langle \omega_i, F(x^k) - F(x^{k+1}) \rangle \geq 0$ is summable at all $k \in \mathbb{N}$. Since $\beta_k \leq \hat{\beta}$ for all $k \in \mathbb{N}$,

$$\sum_{k=0}^{n} \beta_k \sum_{i=1}^{m} \langle \omega_i, F(x^k) - F(x^{k+1}) \rangle \leq \hat{\beta} \sum_{i=1}^{m} \langle \omega_i, F(x^0) - F(x^{n+1}) \rangle. \tag{26}$$

As consequence of the Armijo-type line-search, we have $F(x^k) - F(x^{k+1}) \in K(x^k) \subset \mathcal{K}$. So, $(F(x^k))_{k \in \mathbb{N}}$ is a decreasing sequence with respect to $\mathcal{K}$. Furthermore by **Assumption 4.4**, it is bounded from below, also with respect to the order given by $\mathcal{K}$, by $F(\hat{x})$, where $\hat{x} \in T$. Hence, Proposition 3.1, pages 90, 91 of Peressini (1967) implies that the sequence $(F(x^k))_{k \in \mathbb{N}}$ converges and using (26) in the inequality below, we get

$$\begin{aligned}
\sum_{k=0}^{\infty} \beta_k \sum_{i=1}^{m} \langle \omega_i, F(x^k) - F(x^{k+1}) \rangle &= \lim_{n \to \infty} \sum_{k=0}^{n} \beta_k \sum_{i=1}^{m} \langle \omega_i, F(x^k) - F(x^{k+1}) \rangle \\
&\leq \hat{\beta} \lim_{n \to \infty} \sum_{i=1}^{m} \langle \omega_i, F(x^0) - F(x^{n+1}) \rangle \\
&= \hat{\beta} \sum_{i=1}^{m} \langle \omega_i, F(x^0) - \lim_{n \to \infty} F(x^{n+1}) \rangle \\
&= \hat{\beta} \sum_{i=1}^{m} \langle \omega_i, F(x^0) - F(\hat{x}) \rangle < +\infty.
\end{aligned}$$

So, the quasi-Fejér convergence is fulfilled.

Since $\hat{x}$ is an arbitrary element of T, it is clear that $(x^k)_{k \in \mathbb{N}}$ converges quasi-Fejér to T. Hence, by Fact 4.7, it follows that $(x^k)_{k \in \mathbb{N}}$ is bounded. Therefore, $(x^k)_{k \in \mathbb{N}}$ has at least one accumulation point, which, by Theorem 4.3 is stationary. By Proposition 2.3, this point is also weakly efficient solution, because F is K-convex. Moreover, since C is closed and the whole sequence is feasible, then this accumulation point belongs to C. $\square$

5 Another variable order

As was noticed in Section 6 of Bello Cruz et al. (2014) the variable order structure can be formulated in two different ways. Moreover, Examples 3.1 and 3.2 in Bello Cruz et al. (2014) illustrate the differences in considering one order or the other. Thus, the variable order for the optimization problem may also depend on a new order by using the cone value mapping $\hat{K}\colon\ \mathrm{dom}(\hat{K}) \subset \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ where $\hat{K}(y)$ is a convex, closed and pointed cone for all $y \in F(C) \subset \mathbb{R}^m$ where C is the feasible set. It is worth noting that the domain of the new mapping $\hat{K}$ is in $\mathbb{R}^m$ and to guarantee the well definition of the variable order below $F(C) \subset \mathrm{dom}(\hat{K})$. So, by asking convexity, closedness and pointedness of $\hat{K}(y)$ for all $y \in F(C)$, those conditions hold for $\hat{K}(F(x))$ for all $x \in C$. Note further that the ordering considered in the previous sections is defined by applications whose domain is $\mathbb{R}^n$. As already discussed in Bello Cruz et al. (2014), convexity can be defined and convex functions satisfy nice properties such as the existence of subgradients in the nonsmooth case.

Given a closed and convex set C, we say that $x^* \in C$ solves the optimization problem

$$\hat{K} - \min F(x) \text{ s.t. } x \in C, \tag{27}$$

if, for all $x \in C$,

$$F(x) - F(x^*) \notin -\hat{K}(F(x^*))\backslash\{0\}.$$

Here we can assume that $\hat{K}\colon\ F(C) \subseteq \mathbb{R}^m \rightrightarrows \mathbb{R}^m$. We shall mention that the main difference between the above problem and problem (1) yields in the definition of the variable order given now by $\hat{K}$. For a more detailed study of the properties of the minimal points and their characterizations and convexity concept on this case; see (Eichfelder 2014; Bello Cruz et al. 2014).

In this framework, the definitions of weak solutions and stationary points are analogous. The main difference is that instead of $K(x^*)$, the cone $\hat{K}(F(x^*))$ is considered to define the variable partial order. That is, the point x^* is stationary, if and only if for all $d \in C - x^*$, we have $J_F(x^*)d \notin -\mathrm{int}(\hat{K}(F(x^*)))$. Then, similarly as in the case of problem (1), the following holds.

Proposition 5.1 *If F is a continuously differentiable function and C is a convex set, weak solutions of problem (27) are stationary points. Moreover if F is also convex with respect to $\hat{K}$, the converse is true.*

Proof It follows the same lines of the proofs of Propositions 2.1 and 2.3. The Taylor expansion of F combined with the closedness of $\hat{K}(F(x^*))$ imply the result. □

The inexact algorithm is adapted in the following way

F-**I**nexact **P**rojected **G**radient Method (**FIPG Method**). Given $0 < \bar{\beta} \le \beta_k \le \hat{\beta} < +\infty$, $\delta \in (0, 1]$ and $\sigma, \gamma \in (0, 1)$.

Initialization Take $x^0 \in \mathbb{R}^n$ and β_0.

Iterative step Given x^k and β_k, compute v^k a δ-approximate direction at x^k for (Q_{x^k}). If $v^k = 0$, then stop. Otherwise compute

$$\ell(k) := \min\left\{\ell \in \mathbb{N}\colon\ F(x^k) + \sigma\gamma^\ell J_F(x^k)v^k - F(x^k + \gamma^\ell v^k) \in \hat{K}(F(x^k))\right\}. \tag{28}$$

Set $x^{k+1} = x^k + \gamma_k v^k \in C$, with $\gamma_k = \gamma^{\ell(k)}$.

Here the auxiliary problem (Q_{x^k}) is defined as

$$\min_{v \in C - x^k} \left\{\frac{\|v\|^2}{2} + \beta_k \phi(x^k, v)\right\}, \tag{Q_{x^k}}$$

where $\phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$,

$$\phi(x, v) := \max_{y \in G(F(x))} y^T J_F(x)v, \tag{29}$$

for $G : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ generator of $\hat{K}^*(F(x)) := \left[\hat{K}(F(x))\right]^*$.

With this ordering, the function ϕ characterizes the stationarity. Furthermore, subproblem (Q_{x^k}) has a unique solution, which is $v^k = 0$ if and only if x^k is a stationary point. Results analogous to those proven in Propositions 3.3 and 3.5 are also true. These facts imply that **FIPG Method** is well defined, i.e., if it stops, then the computed point is stationary, and in the other case, there exists $\ell(k)$ which satisfies the Armijo-type line-search (28). So, only the convergence of a sequence generated by it must be studied.

As in the last section, we analyze the convergence of the functional values sequence $(F(x^k))_{k\in\mathbb{N}}$.

Lemma 5.2 *Suppose that x^* is an accumulation point of $(x^k)_{k\in\mathbb{N}}$ of the sequence generated by* **FIPG Method**. *If $\cup_{x\in C}\hat{K}(F(x)) \subseteq \mathcal{K}$, where $\mathcal{K}$ is a closed, pointed and convex cone, then $\lim_{k\to\infty} F(x^k) = F(x^*)$.*

Proof The result is again proven by the existence of a non-increasing sequence with an accumulation point. □

Next, with the help of the last Lemma, we summarize the convergence of the generated sequence with the following result.

Theorem 5.3 *Suppose that*

(a) $\cup_{x\in C}\hat{K}(F(x)) \subset \mathcal{K}$, where $\mathcal{K}$ is a a closed, pointed and convex cone.
(b) The map $G \circ F$ is closed.
(c) $d_H(G(F(x)), G(F(\hat{x}))) \le L_{GF}\|x - \hat{x}\|$, for all $x, \hat{x} \in C$.
(d) J_F is a locally Lipschitz function around x for all $x \in C$.

Then, all accumulation points of $(x^k)_{k\in\mathbb{N}}$ generated by **FIPG Method** *are stationary points of problem (27).*

Proof It follows from the same lines of the proof of Theorem 4.3. □

We want to point out that in the last theorem the condition

$$\mathrm{d}_H(G(F(x)), G(F(\hat{x}))) \leq L_{GF}\|x - \hat{x}\|, \quad \forall x, \hat{x} \in C$$

replaces the similar one given in Theorem 4.3, i.e., $\mathrm{d}_H(G(y), G(\hat{y})) \leq L_G\|y - \hat{y}\|$, for all $y, \hat{y} \in C$. Moreover, if F is Lipschitz on C, then this last condition implies the condition (c) in Theorem 5.3 by taking $y = F(x)$ and $\hat{y} = F(\hat{x})$.

Next result is an extension to the variable order setting of Theorem 5.2 of Fukuda and Graña Drummond (2013).

Theorem 5.4 *Assume that F is $\hat{K}$-convex and additionally:*

(a) If $(z^k)_{k\in\mathbb{N}} \subset F(C)$ is a sequence such that $z^k - z^{k+1} \in \hat{K}(F(x^k))$ for all $k \in \mathbb{N}$ and $z \in C$, $z^k - z \in \mathcal{K}$ for some closed, convex and pointed cone $\mathcal{K}$, $\cup_{k\in\mathbb{N}}\hat{K}(F(x^k)) \subseteq \mathcal{K}$, then there exists $\hat{x} \in C$ such that $F(\hat{x}) \preceq z^k$ for all $k \in \mathbb{N}$.
(b) The search direction v^k is s-compatible at x^k, i.e., $v^k = P_{C-x^k}(-\beta J_F(x^k)^T w^k)$, where $w^k \in \mathrm{conv}(G(F(x^k)))$, for all $k \in \mathbb{N}$.
(c) $\mathrm{int}(\cap_{k\in\mathbb{N}}\hat{K}(F(x^k))) \neq \varnothing$.
(d) There exists $\mathcal{K}$, a pointed, closed and convex cone such that $\hat{K}(F(x^k)) \subseteq \mathcal{K}$ for all $k \in \mathbb{N}$.
Then every sequence generated by **FIPG Method** *is bounded and its accumulation points are weakly efficient solutions.*

Proof It follows from the same lines of the proof of Theorem 4.11 using now the new variable order structure. □

6 Illustrative examples

In this section, we present three examples, two for problem (1) and one for problem (27), illustrating how both proposed methods work starting at ten different random initial points. We verify our assumptions in each problem and make some comparisons between the proposed methods by changing the inexactness of the approximate gradient direction.

The algorithms were implemented in MATLAB R2012 and ran on an Intel(R) Atom(TM) CPU N270 at 1.6GHz. All starting points are not solutions and are randomly generated. The stopping criteria were $\|v^k\| < 10^{-4}$ and also when a maximum of 30-iterations is reached. The solutions were displayed with four digits, CPU time was recorded in seconds, and the number of iterations was also displayed in each case. Our numerical tests were performed with β_k constant equal 1. Despite the fact that it may not be an easy task to compute the positive dual cone of a given cone, the computation of (approximate) directions is, in general, complicated. Indeed, after the definition, the exact optimal value of problem (P_x) must be known. The use of s-compatible directions at iteration k of the proposed methods, see Definition 4.8, is recommended in the case in which the exact projection onto the set $C - x^k$ is not too complicated.

This is the case for the set of feasible solutions in the next example. Clearly, in all examples below, the defining order cone-valued mappings are closed maps, and their generators are Lipschitz with respect to the Hausdorff distance.

Example 6.1 We consider the vector problem as (1) with

$$K - \min\ F(x) = \Big(x+1, x^2+1\Big),\ \text{s.t.}\ \ x \in [0, 1],$$

where $F : \mathbb{R} \to \mathbb{R}^2$ and the variable order is given by $K : \mathbb{R} \rightrightarrows \mathbb{R}^2$,

$$K(x) := \Big\{(z_1, z_2) \in \mathbb{R}^2 : z_1 \geq 0,\ \ (x^2+1)z_1 - (x+1)z_2 \leq 0\Big\}.$$

In this model the closed interval $[0, \sqrt{2}-1] \approx [0,\ 0.4142]$ is the set of minimizers.

The **IPG Method** was run ten times by using ten random initial points, outside of the set of minimizers, and each time it ended at solution points, which have been obtained after the verification of the stopping criterion.

The method gives the following data:

Note that in all cases above optimal solutions were obtained. Moreover, to illustrate how the inexactness of the **IPG Method** affects the performance (average of CPU time and the average of the number of iterations for 10 instances) of the convergence is considering different values of the approximate direction parameter δ.

The above table shows that when the inexactness increases, the number of iterations increases. However, the CPU time decreases from $\delta = 0$ to $\delta = 0.5$, hence in the last value of δ the CPU time slightly increases.

The next example is a non-convex problem corresponding to the model studied in the previous section.

Example 6.2 [cf. Example 4.13 of Eichfelder and Duc Ha (2013)] Consider the following vector problem as problem (27)

$$\hat{K} - \min\ F(x_1, x_2) = \Big(x_1^2, x_2^2\Big),\ \text{s.t.}\ \pi \leq x_1^2 + x_2^2 \leq 2\pi,$$

where $F : \mathbb{R}^2 \to \mathbb{R}^2$ and the variable order is given by the cone $\hat{K} : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$,

$$\hat{K}(y) := \left\{ z = (z_1, z_2) \in \mathbb{R}^2 : \|z\|_2 \leq \left[\begin{pmatrix} 2 & 1 \\ -1 & -1 \end{pmatrix} y\right]^T z/\pi \right\}.$$

The set of solutions (stationary points) of this problem is directly computed by using the definition of S^s given in (2) as

$$\Big\{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 = \pi \quad \text{or} \quad x_1^2 + x_2^2 = 2\pi\Big\}.$$

The **FIPG Method** computes the following points:

Note that the solutions computed at all iterations of the proposed method belong to the set of optimal solutions.

Instances	1	2	3	4	5	6	7	8	9	10
Initial Points	0.6557	0.6948	0.8491	0.9340	0.6787	0.7577	0.7431	0.4387	0.6555	0.9502
Solutions	0.4115	0.4128	0.4140	0.4135	0.4116	0.4131	0.4127	0.4136	0.4114	0.4130
CPU Time	0.0001	0.0250	0.0001	0.0156	0.0001	0.0001	0.0001	0.1094	0.0156	0.0781
N$^{\underline{o}}$ Iterations	16	19	23	26	17	20	20	4	16	28

δ-Approximation Parameters	$\delta = 0$	$\delta = 0.25$	$\delta = 0.5$	$\delta = 0.75$
Average of CPU Time	0.4819	0.2232	0.0244	0.0251
Average of N$^{\underline{o}}$ Iterations	9	15	19	21

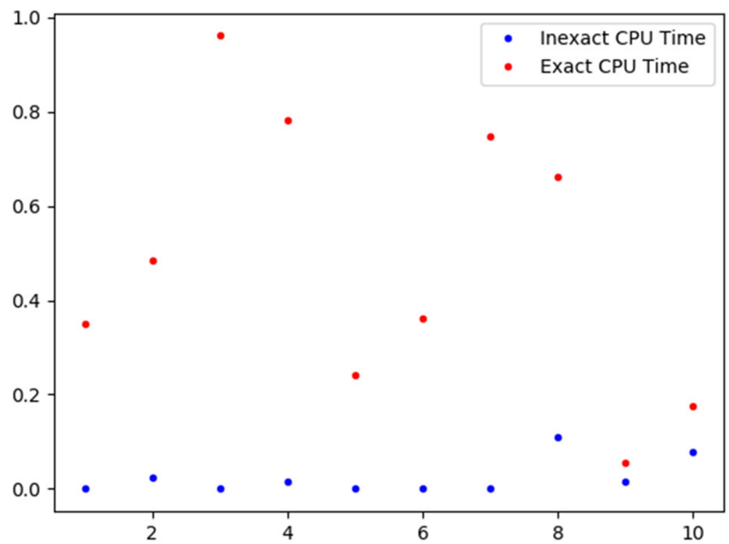

Fig. 1 Example 6.1

The last example from Section 9.2 of the book (Eichfelder 2014) is widely studied.

Example 6.3 [cf. Example 9.5 of Eichfelder (2014)] Consider the following problem

$$K - \min \ F(x) = (x_1, x_2), \ \text{ s.t. } \ x \in C$$

where

$$C := \left\{ x = (x_1, x_2) \in [0, \pi] \times [0, \pi] \;\middle|\; \begin{array}{l} x_1^2 + x_2^2 - 1 - \dfrac{1}{10} \cos \left(16 \arctan \left(\dfrac{x_1}{x_2} \right) \right) \geq 0, \\ (x_1 - 0.5)^2 + (x_2 - 0.5)^2 \leq 0.5. \end{array} \right\}.$$

Consider the map $K : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$, given by

$$K(x) := \left\{ z \in \mathbb{R}^2 : \|z\|_2 \leq \frac{2}{\min_{i=1,2} x_i} x^T z \right\}.$$

The **IPG Method** stating at ten random initial points performs as follows:

In this case, the maximum number of iterations is achieved. Nevertheless, good approximations to minimal elements have been computed at each of the ten instances presented in the above table.

For all the above examples, we compare the exact and inexact versions by taking the same initial points (10 in total on the "x" axes) in term of the CPU time (in seconds on the "y" axes); see figures below (Figs. 1, 2, 3).

Instances	1	2	3	4	5	6	7	8	9	10
Initial Points	1.8650	1.7525	2.4190	1.9573	0.7931	1.2683	1.8135	−2.0485	−0.6446	−0.8561
	1.6400	1.6350	0.0835	0.2813	−2.0321	−1.6814	0.3050	0.3229	1.9606	2.1011
Solutions	1.1632	0.9850	1.7705	1.7492	0.8634	1.4208	1.7456	−1.7438	−0.8016	−0.9535
	2.2204	2.3050	0.0841	0.2859	−2.3532	−2.0650	0.3074	0.3172	2.3750	2.3182
CPU Time	0.2969	0.2344	0.2188	0.1719	0.1563	0.5625	0.2656	0.1875	0.2031	0.3125
$N^{\underline{o}}$ Iterations	2	4	2	2	4	17	3	2	2	5

Instances	1	2	3	4	5	6	7	8	9	10
Initial Point	0.9735	0.7932	0.8403	0.9847	0.7508	0.9786	0.9790	0.9679	0.8082	0.8965
	0.6608	0.9050	0.8664	0.6228	0.9326	0.6448	0.6433	0.6762	0.8937	0.8046
Solution	0.9011	0.7407	0.7854	0.9096	0.7004	0.9050	0.9054	0.8967	0.7551	0.7754
	0.5589	0.7916	0.7541	0.5228	0.8182	0.5437	0.5423	0.5735	0.7806	0.5859
CPU Time	0.0938	0.0313	0.0625	0.0313	0.0469	0.0156	0.0313	0.0313	0.0313	0.0156

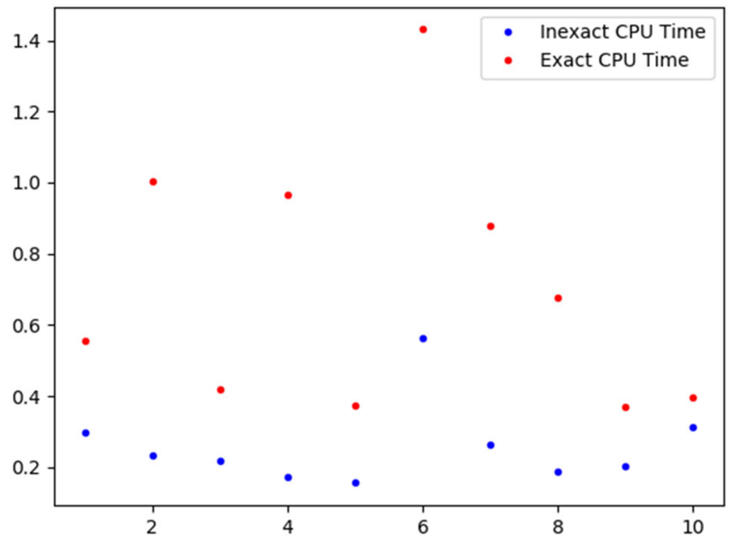

Fig. 2 Example 6.2

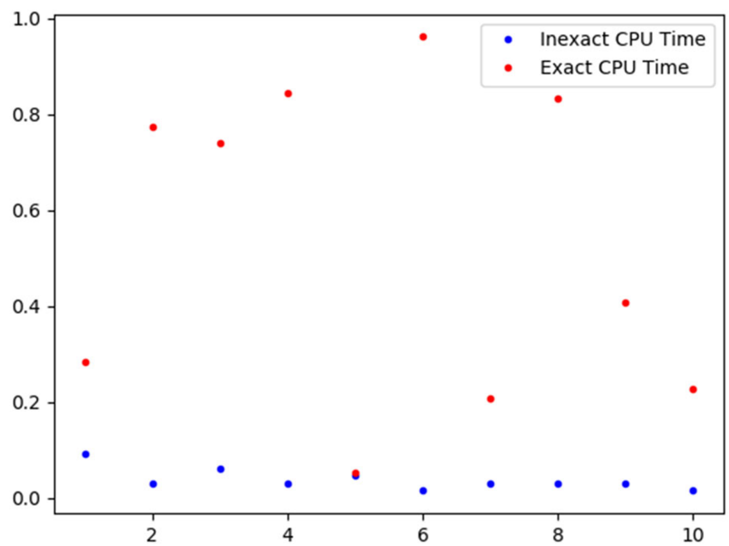

Fig. 3 Example 6.3

This shows that the inexact versions are significantly faster (CPU Time), in almost all instances than the exact ones.

However, for all initial points of the above examples, the exact versions of the proposed methods take fewer iterations than the inexact ones to converge to a solution. It is worth emphasizing that the exact versions have to solve exactly the harder subproblems P_{x^k} and Q_{x^k} to find the descent direction at each iteration k. This is a serious drawback for the computational implementation point of view (avoidable for the above examples), making the exact implementation inefficient in general.

7 Final remarks

The projected gradient method is one of the classical and basic schemes for solving smooth constrained optimization problems. In this paper, we have extended the inexact scheme proposed in Fukuda and Graña Drummond (2013) for constrained vector optimization problems. The proposed methods solve smooth and constrained vector optimization problems under a variable ordering by taking inexact descent directions. Both methods are able to find the solution of problem (1) using less effort from a computational viewpoint, and it considers all the possible combinations: constrained-unconstrained, and exact-inexact in the variable order setting. This inexact projected approach promotes future research on other efficient variants for these kinds of problems.

As it is shown in the examples above, it is more effective and implementable than the exact one. Moreover, constrained variable optimization problems can now be solved by using inexact directions, which improves the result presented in Bello Cruz and Bouza Allende (2014). However, the full convergence of the generated sequence to a weakly efficient solution is still an open problem in variable order settings.

Another important solution concept is non-domination, which is used for optimizing set-valued maps under the variable structure. A variant of the proposed approach can also be considered using the condition $J_F(x^k)v \in -\cup_{x\in C} K(x)$ as the descent direction at each step k. However, this approach will converge to a stationary point, as in the case we have proposed. Non-dominated points are also minimizers under well-known conditions. A future research direction would be to find, in this case, how we can guarantee that the algorithm converges to non-dominated points. Set-valued optimization is also an interesting topic and finding algorithms that compute points that at least fulfill conditions, such as those presented in Theorem 3.3 in Durea et al. (2015), is important.

Future work will also be addressed to investigate for some particular instances of this problem the case in which the objective function is a non-smooth function, extending the projected subgradient method proposed in Bello Cruz (2013) to the variable order setting.

The numerical behavior of these approaches under K-convexity of the non-smooth objective function remains open, and it is a promising direction to be investigated. Despite its computational shortcomings, it hopefully sets the foundations of future research into more efficient and general algorithms for this setting.

Acknowledgements JYBC was partially supported by the National Science Foundation (NSF) Grant DMS-1816449 and by Research & Artistry grant from NIU. The results of this paper were developed during the stay of the second author at the University of Halle-Wittenberg supported by the Humboldt Foundation.

References

Baatar D, Wiecek MM (2006) Advancing equitability in multiobjective programming. Comput Math Appl 2:225–234

Bao TQ, Mordukhovich BS, Soubeyran A (2015) Variational analysis in psychological modeling. J Optim Theory Appl 164:290–315

Bao TQ, Mordukhovich BS, Soubeyran A (2015) Fixed points and variational principles with applications to capability theory of wellbeing via variational rationality. Set-Valued Var Anal 23:375–398
Bello Cruz JY (2013) A subgradient method for vector optimization problems. SIAM J Optim 23:2169–2182
Bello Cruz JY, Bouza Allende G (2014) A steepest descent-like method for variable order vector optimization problems. J Optim Theory Appl 162:371–391
Bello Cruz JY, Bouza Allende G, Lucambio Pérez LR (2014) Subgradient algorithms for solving variable inequalities. Appl Math Comput 247:1052–1063
Bello Cruz JY, Lucambio Pérez LR (2014) A subgradient-like algorithm for solving vector convex inequalities. J Optim Theory Appl 162:392–404
Bello Cruz JY, Lucambio Pérez LR, Melo JG (2011) Convergence of the projected gradient method for quasiconvex multiobjective optimization. Nonlinear Anal 74:5268–5273
Bento GC, Bouza Allende G, Pereira YR (2018) A Newton-like method for variable order vector optimization problems. J Optim Theory Appl 177:201–221
Bonnel H, Iusem AN, Svaiter BF (2005) Proximal methods in vector optimization. SIAM J Optim 15:953–970
Browder FE (1967) Convergence theorems for sequences of nonlinear operators in Banach spaces. Math Z 100:201–225
Burachik R, Graña Drummond LM, Iusem AN, Svaiter BF (1995) Full convergence of the steepest descent method with inexact line searches. Optimization 32:137–146
Durea M, Strugariu R, Tammer C (2015) On set-valued optimization problems with variable ordering structure. J Global Optim 61:745–767
Eichfelder G (2014) Vector optimization in medical engineering. In: Pardalos PM, Rassias TM (eds) Mathematics without boundaries. Springer, Berlin, pp 181–215
Eichfelder G, Duc Ha TX (2013) Optimality conditions for vector optimization problems with variable ordering structures. Optimization 62:597–627
Eichfelder G (2014) Variable ordering structures in vector optimization. Springer, Berlin
Eichfelder G (2011) Optimal elements in vector optimization with variable ordering structure. J Optim Theory Appl 151:217–240
Engau A (2008) Variable preference modeling with ideal-symmetric convex cones. J Global Optim 42:295–311
Fliege J, Graña Drummond LM, Svaiter BF (2009) Newton's method for multiobjective optimization. SIAM J Optim 20:602–626
Fliege J, Svaiter BF (2000) Steepest descent methods for multicriteria optimization. Math Methods Oper Res 51:479–494
Fukuda EH, Graña Drummond LM (2013) Inexact projected gradient method for vector optimization. Comput Optim Appl 54:473–493
Fukuda EH, Graña Drummond LM (2011) On the convergence of the projected gradient method for vector optimization. Optimization 60:1009–1021
Graña Drummond LM, Iusem AN (2004) A projected gradient method for vector optimization problems. Comput Optim Appl 28:5–30
Graña Drummond LM, Maculan N, Svaiter BF (2008) On the choice of parameters for the weighting method in vector optimization. Math Prog 111:201–216
Graña Drummond LM, Svaiter BF (2005) A steepest descent method for vector optimization. J Comput Appl Math 175:395–414
Isac G, Tammer C (2010) Application of a vector-valued Ekeland-type variational principle for deriving optimality conditions. In: Nonlinear analysis and variational problems: in Honor of George Isac, Springer, Berlin, vol 35, pp 343–365
Iusem AN, Svaiter BF, Teboulle M (1994) Entropy-like proximal methods in convex programming. Math Oper Res 19:790–814
Jahn J (2004) Vector optimization: theory, applications and extensions. Springer, Berlin
Jahn J (1986) Mathematical vector optimization in partially ordered linear spaces. Verlag Peter D. Lang, Frankfurt
Jahn J (1984) Scalarization in vector optimization. Math Prog 29:203–218
John R (2001) The concave nontransitive consumer. J Glob Optim 20:297–308
John R (2006) Local and global consumer preferences. In: Konnov I, Luc DT, Rubinov A (eds) Generalized convexity and related topics. Springer, Heidelberg, pp 315–326326

Kim DS, Pham TS, Tuyen NV (2018) On the existence of Pareto solutions for polynomial vector optimization problems. Math. Program. 1–21

Luc DT (2008) Pareto optimality, game theory and equilibria. Pareto Optim Spring Optim Appl 17:481–515

Luc DT (1989) Theory of vector optimization. Lecture Notes in Economics and Mathematical Systems 319. Springer, Berlin

Luc DT (1987) Scalarization of vector optimization problems. J Optim Theory Appl 55:85–102

Peressini AL (1967) Ordered topological vector space. Harper and Row

Wiecek MM (2007) Advances in cone-based preference modeling for decision making with multiple criteria. Decis Mak Manuf Serv 1:153–173