manuscript No.
(will be inserted by the editor)

A Steepest Descent-like Method for Variable Order Vector Optimization Problems

J. Y. Bello Cruz · **G. Bouza Allende**

Received: date / Accepted: date

Abstract In some applications, the comparison between two elements may depend on the point leading to the so called variable order structure. Optimality concepts may be extended to this more general framework. In this paper, we extend the steepest descent-like method for smooth unconstrained vector optimization problem under a variable ordering structure. Roughly speaking, we obtain that every accumulation point of the generated sequence satisfies a necessary first order condition. We discuss the consequence of this fact in the convex case.

Keywords gradient-like method · K-convexity · variable order · vector optimization · weakly efficient points

Mathematics Subject Classification (2010) 90C29 · 90C30

1 Introduction

Classical vector optimization finds minimizers of a vector function with respect to an order given by a fixed cone, which is a pointed, convex and closed. However, it is not able to model situations in which the set of points whose image is better, depends on the point.

This variable order structure is given by a set valued application whose image is a proper, pointed, convex and closed cone, for all n- dimensional real vector. So, the problem of minimizing a vector function F with respect to the set valued order is to find a point such that there does not exist another point with a better (and different) value of the objective function.

This problem has been treated in [1] in the sense of finding a minimizer of the image of the objective function with respect to an ordered structure depending on the point in the image. It is a particular case of the problem described in [2], where the goal of the model is to find a minimum of a set.

This research was supported by Project CAPES-MES-CUBA 226/2012 "Modelos de Otimização e Aplicações". First author was partially supported by PROCAD-nf-UFG/UnB/IMPA research and PRONEX-CNPq-FAPERJ. Optimization research.

J. Y. Bello Cruz
Universidade Federal de Goiás, Goiânia, Brazil.
E-mail: yunier@impa.br

G. Bouza Allende (**Corresponding author**)
Facultad de Matemática y Computación
Universidad de La Habana, La Habana, Cuba.
San Lázaro y L, Vedado.
CP: 10400, La Habana, Cuba E-mail: gema@matcom.uh.cu

Important applications of optimization problems with variable order structure appear in medical diagnosis, portfolio optimization and location theory. These applications are discussed in [3–5], where this structure is used for modelling the variability of the preference. For instance, in medical diagnosis when obtaining information from images, the collected data are transformed into another set and, from it, diagnosis is done. For determining the best transformation given the original data and the desired pattern, different measuring criteria can be used, leading to the optimization of a vector of functions.

The solution of this model using a classical weighting technique, leads to wrong results. Indeed, solutions with large values at some objectives functions are obtained. However, if the set of weights depends on the point, better results are reported; see [5].

Extensions of iterative methods for classical vector optimization models, to the variable order case is a promising idea. Although many approaches such as proximal points, weighting techniques, Newton-like and subgradient methods may be considered; see [6–13]. Due to its simplicity and the adaptability to the structure of the vectorial problem, we will focus on the steepest descent algorithm. This approach appeared in [12] for solving multicriteria models. It was extended in [13] for convex vectorial models and, recently, in [14] the convergence for the quasiconvex case was obtained.

In this work, we will present a steepest descent-like algorithm for solving vector optimization problems with variable order. We obtain the properties of the limit points of the sequence generated by the proposed algorithm. Then, under a convex like hypothesis, we guarantee that the sequence is bounded and all its accumulation points are solutions of the problem.

This paper is organized as follows: After some preliminary results, we extend the concept of convexity of a function to the variable ordered case and obtain some properties of this class of functions. Section 4 is devoted to the presentation of the algorithm and the continuity of the involved operators. Finally the convergence of the steepest descent method is shown in Section 5.

2 Preliminaries

In this section we will present some previous results and definitions. We begin with some notations.

The inner product is denoted by $\langle \cdot, \cdot \rangle$ and the norm by $\| \cdot \|$. The ball centered at x with radius r is $B(x,r) := \{y : \|y-x\| \leq r\}$. Given two sets A and B, we will consider, $\mathrm{dist}(A,B)$, as the Hausdorff distance, *i.e.*

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

$bd(A)$ denotes the boundary of the set A and A^c, its complement.

The variable order structure in $\mathbb{R}^n$ is given by the set valued application $K : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, where $K(x)$ is a proper, pointed, convex and closed cone, for all $x \in \mathbb{R}^n$. 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 application K is the set $Gr(K) := \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m : \ y \in K(x)\}$. We recall that the mapping K is closed iff $Gr(K)$ is a closed subset of $\mathbb{R}^n \times \mathbb{R}^m$. Given the variable order structure, the K- vector optimization problem is

$$K - \min F(x). \tag{1}$$

As in the case of classical vector optimization, related solution concepts such as weakly efficient and stationary points can be extended. We assume that $int(K(x^*)) \neq \varnothing$, for all $x \in \mathbb{R}^n$. The point x^* is a weak solution of Problem (1) iff for all $x \in \mathbb{R}^n$, $F(x) - F(x^*) \notin -int(K(x^*))$, S^w is the set of all weak solution points. The set of all minimizers of Problem (1) is denoted by S^*. We want to point out, that this definition corresponds with the concept of weak minimizers given in [2]. On the other hand, if F is a continuously differentiable function the point x^* is stationary, iff $Im(\nabla F(x^*)) \cap -int(K(x^*)) = \emptyset$, here S^s denotes the set of all stationary points.

A necessary optimality condition is given as follows:

Proposition 2.1 *Let x^* be a weak solution of Problem (1). If F is a continuously differentiable function, then x^* is a stationary point.*

Proof Since x^* is a weak solution of Problem (1), $F(x^* + \alpha d) - F(x^*) \notin -int(K(x^*))$, for all $\alpha > 0$, $d \in \mathbb{R}^n$. So,

$$F(x^* + \alpha d) - F(x^*) \in (-int(K(x^*)))^c. \tag{2}$$

The Taylor expansion of F at x^*, leads to

$$F(x^* + \alpha d) = F(x^*) + \alpha \nabla F(x^*)d + o(\alpha).$$

Combining (2) with the above equation, we have

$$\alpha \nabla F(x^*)d + o(\alpha) \in (-int(K(x^*)))^c.$$

Using that $(-int(K(x^*)))^c$ is a closed cone, and since $\alpha > 0$, it follows that

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

Taking limits as α goes to 0 and using the closedness of $(-int(K(x^*)))^c$, we obtain:

$$\nabla F(x^*)d \in (-int(K(x^*)))^c,$$

establishing that $x^* \in S^s$. □

Next we deal with the so called quasi-Fejér convergence and its properties.

Definition 2.1 *Let S be a nonempty subset of $\mathbb{R}^n$. A sequence $\{x^k\} \subset \mathbb{R}^n$ is said to be quasi-Fejér convergent to S, iff for all $x \in S$, there exists $\bar{k}$ and a summable sequence $\{\delta_k\} \subset \mathbb{R}_+$ such that*

$$\|x^{k+1} - x\|^2 \leq \|x^k - x\|^2 + \delta_k,$$

for all $k \geq \bar{k}$.

This definition originates in [15] and has been further elaborated in [16]. A useful result on quasi-Fejér sequences is the following.

Theorem 2.1 *If $\{x^k\}$ is quasi-Fejér convergent to S then,*

i) The sequence $\{x^k\}$ is bounded,
ii) if a cluster point of the sequence $\{x^k\}$ belongs to S, then the whole sequence $\{x^k\}$ converges.

Proof See Lemma 6 of [15] and Theorem 1 of [17]. □

3 K-convexity

Convexity is a very helpful concept in optimization. Convex functions fulfill nice properties such as existence of directional derivative and subgradient.

For classical vector optimization problems, *i.e.* $K(x) \equiv K$, for all x, we recall that F is convex if for all $\lambda \in [0,1]$, $x, \bar{x} \in \mathbb{R}^n$, it holds that

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

see [18–20]. In the variable order framework, we extend the concept of convexity as follows.

Definition 3.1 *We say that F is a K-convex function iff for all $\lambda \in [0,1]$, $x, \bar{x} \in \mathbb{R}^n$,*

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

The geometric variant of convexity is defined via the epigraph of the function F, denoted as $epi(F) := \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m : F(x) \in y - K(x)\}$; see [18, 21]. In non-variable orders, the convexity of $epi(F)$ is equivalent to the convexity of F. However, as the next proposition shows, in the variable order setting, this does not hold.

Proposition 3.1 *Suppose that F is a K-convex function. Then, $epi(F)$ is convex if and only if $K(x) = K$, for all $x \in \mathbb{R}^n$.*

Proof Suppose that there exists $z \in K(x_1) \setminus K(x_2)$. Take the points

$$(x_1, F(x_1) + 2\alpha z)$$

and

$$(2x_2 - x_1, F(2x_2 - x_1)),$$

with $\alpha > 0$. They belong to $epi(F)$.
Take the following convex combination:

$$\frac{(x_1, F(x_1) + 2\alpha z)}{2} + \frac{(2x_2 - x_1, F(2x_2 - x_1))}{2} = \left(x_2, \frac{F(x_1) + F(2x_2 - x_1)}{2} + \alpha z\right).$$

This point belongs to $epi(F)$ if and only if

$$F(x_2) = \frac{F(x_1) + F(2x_2 - x_1)}{2} + \alpha z - k(\alpha),$$

where $k(\alpha) \in K(x_2)$. By the K-convexity of F,

$$F(x_2) = \frac{F(x_1) + F(2x_2 - x_1)}{2} - k_1,$$

where $k_1 \in K(x_2)$. So,

$$\alpha z + k_1 = k(\alpha). \tag{3}$$

Since $K(x_2)$ is closed and convex, and $z \notin K(x_2)$, $\{z\}$ and $K(x_2)$ may be strictly separated in $\mathbb{R}^m$ by a hyperplane, *i.e.* there exists some $p \in \mathbb{R}^m \setminus \{0\}$ such that

$$p^T k \geq 0 > p^T z, \tag{4}$$

for all $k \in K(x_2)$. Therefore, after multiplying (3) by p^T and using (4) with $k = k(\alpha) \in K(x_2)$, we obtain that

$$\alpha p^T z + p^T k_1 = p^T k(\alpha) \geq 0.$$

Taking limits as α goes to ∞, the contradiction is established, since $0 \leq \alpha p^T z + p^T k_1 \to -\infty$. Hence, $K(x) \equiv K$ for all $x \in \mathbb{R}^n$. □

In the following we present some analytical properties of K-convex functions. For the nondifferentiable model, we generalize the classical assumptions given in the case of constant cones; see [21, 22]. Let us first present the definition of Daniell cone.

Definition 3.2 *We say that a convex cone $\mathcal{K}$ is Daniell cone, iff for all sequence $\{x^k\} \subset \mathbb{R}^n$ satisfying $\{x^k - x^{k+1}\} \subset \mathcal{K}$ and for some $\hat{x} \in \mathbb{R}^n$, $\{x^k - \hat{x}\} \subset \mathcal{K}$, then $\lim_{k\to\infty} x^k = \inf_k \{x^k\}$.*

Given the partial order structure induced by a cone $\mathcal{K}$, the concept of infimum of a sequence can be defined. Indeed, for a sequence $\{x^k\}$ and a cone $\mathcal{K}$, the point x^* is $\inf_k\{x^k\}$ if $x^k - x^* \in \mathcal{K}$, for all k and there is not x such that $x^* - x \in \mathcal{K}$ and $x^k - x \in \mathcal{K}$, for all k. It is well known that every pointed, closed and convex cone in a finite dimensional space is a Daniell cone; see [23].

Proposition 3.2 *Suppose that for each $\bar{x}$ there exists $\varepsilon > 0$ such that $\cup_{x\in B(\bar{x},\varepsilon)}K(x) \subset \mathscr{K}$, where $\mathscr{K}$ is a Daniell cone. Then*

$$F'(\bar{x};x-\bar{x}) = \lim_{t\to 0^+} \frac{F(\bar{x}+td)-F(\bar{x})}{t},$$

i.e. the directional derivative of F at $\bar{x}$ exists along $d = (x-\bar{x})$.

Proof By the convexity of F, for all $0 < t_1 < t_2 < \varepsilon$,

$$F(\bar{x}+t_1d) - \frac{t_1}{t_2}F(\bar{x}+t_2d) - \frac{t_2-t_1}{t_2}F(\bar{x}) \in -K(\bar{x}+t_1d).$$

Dividing by t_1, we have that

$$\frac{F(\bar{x}+t_1d)-F(\bar{x})}{t_1} - \frac{F(\bar{x}+t_2d)-F(\bar{x})}{t_2} \in -K(\bar{x}+t_1d) \subset -\mathscr{K}.$$

Similarly, as

$$F(\bar{x}) - \frac{t_1}{t_1+1}F(\bar{x}-d) - \frac{1}{t_1+1}F(\bar{x}+t_1d) \in -K(\bar{x}),$$

it holds that

$$\frac{F(\bar{x}+t_1d)-F(\bar{x})}{t_1} + F(\bar{x}-d) - F(\bar{x}) \in K(\bar{x}) \subset \mathscr{K}.$$

Since $\mathscr{K}$ is a Daniell cone, $\frac{F(\bar{x}+t_1d)-F(\bar{x})}{t_1}$ has a limit as t_1 goes to 0. Hence, the directional derivative exists. □

Let us present the definition of subgradient.

Definition 3.3 *We say that $\varepsilon_{\bar{x}} \in \mathbb{R}^{m\times n}$ is a subgradient of F at $\bar{x}$ iff for all $x \in \mathbb{R}^n$,*

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

Denote the set of all subgradients of F at $\bar{x}$ as $\partial F(\bar{x})$.

Proposition 3.3 *If for all $x \in \mathbb{R}^n$, $\partial F(x) \neq \emptyset$, then F is K-convex.*

Proof Since $\partial F(x) \neq \emptyset$, there exists $\varepsilon_{\lambda x_1+(1-\lambda)x_2}$ and $k_1, k_2 \in K(\lambda x_1 + (1-\lambda)x_2)$, such that

$$F(x_2) - F(\lambda x_1 + (1-\lambda)x_2) = \lambda\varepsilon_{\lambda x_1+(1-\lambda)x_2}(x_2-x_1) + k_1,$$

and

$$F(x_1) - F(\lambda x_1 + (1-\lambda)x_2) = (\lambda-1)\varepsilon_{\lambda x_1+(1-\lambda)x_2}(x_2-x_1) + k_2.$$

Multiplying the previous equalities by λ and $(1-\lambda)$ respectively, their addition leads to

$$\lambda F(x_1) + (1-\lambda)F(x_2) - F(\lambda x_1 + (1-\lambda)x_2) = \lambda k_2 + (1-\lambda)k_1.$$

Since $K(\lambda x_1 + (1-\lambda)x_2)$ is convex, the result follows. □

From now on we assume that F is a continuously differentiable function. The existence of a subgradient and the first order sufficient optimality can be obtained directly as follows.

Proposition 3.4 *Let F be a K-convex function. If $Gr(K)$ is closed, then for all $\bar{x} \in \mathbb{R}^n$, $\nabla F(\bar{x}) = \partial F(\bar{x})$.*

Proof First we show that $\nabla F(\bar{x})$ belongs to $\partial F(\bar{x})$. Since F is a continuously differentiable function, fixed x, we get

$$F(\lambda\bar{x}+(1-\lambda)x)=F(\bar{x})+(1-\lambda)\nabla F(\bar{x})(x-\bar{x})+o\left((1-\lambda)\|x-\bar{x}\|\right).$$

By K-convexity

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

So,

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

Since K is a cone, it follows that

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

By taking limits as λ goes to 1 and recalling that K is a closed map, it holds that

$$F(x)-F(\bar{x})-\nabla F(\bar{x})(x-\bar{x})\in K(\bar{x}),$$

and hence, $\nabla F(\bar{x})\in\partial F(\bar{x})$.
Suppose that $\varepsilon\in\partial F(\bar{x})$. Fixed d, we get that, for all $\lambda>0$,

$$F(\bar{x}+\lambda d)-F(\bar{x})=\lambda\nabla F(\bar{x})d+o(\lambda)\in\lambda\varepsilon d+k(\lambda),$$

where $k(\lambda)\in K(\bar{x})$. Dividing by $\lambda>0$, and taking limits as λ approaches 0 in the above inclusion, it follows that $[\nabla F(\bar{x})-\varepsilon]d\in K(\bar{x})$, since $K(\bar{x})$ is a closed set. Repeating the same analysis for $-d$, we obtain that $-[\nabla F(\bar{x})-\varepsilon]d\in K(\bar{x})$.
Taking into account that $K(\bar{x})$ is a pointed cone, $[\nabla F(\bar{x})-\varepsilon]d=0$, this implies that

$$\nabla F(\bar{x})=\varepsilon.$$

□

In classical smooth unconstrained vector optimization, the convexity of the function implies that stationary points are minimizers. Let us prove an analogous result.

Proposition 3.5 *Let F be a K-convex function and K be a closed mapping such that $int(K(x^*))\neq\emptyset$. Then,*

(i) the point x^* is a weak solution of Problem (1) if and only if $Im(\nabla F(x^*))\cap -int(K(x^*))=\emptyset$.
(ii) If $Im(\nabla F(x^*))\cap -K(x^*)=\{0\}$, then x^* is a minimizer.

Proof (i) The necessity was already shown in Proposition 2.1.
Conversely, suppose that $Im(\nabla F(x^*))\cap -int(K(x^*))=\emptyset$ and that for some x,

$$F(x)-F(x^*)=k_1\in -int(K(x^*)).$$

As already shown

$$k_2=F(x)-F(x^*)-\nabla F(x^*)(x-x^*)\in K(x^*).$$

So, $\nabla F(x^*)(x-x^*)=k_1-k_2$. Recalling that $K(x^*)$ is a convex cone, it follows that

$$\nabla F(x^*)(x-x^*)\in -int(K(x^*)),$$

contradicting the hypothesis.
(ii) Suppose that $Im(\nabla F(x^*))\cap -K(x^*)=\{0\}$ and for some x, $F(x)-F(x^*)=k_1\in -K(x^*)\setminus\{0\}$. Since

$$k_2 = F(x) - F(x^*) - \nabla F(x^*)(x - x^*) \in K(x^*),$$

it holds that $\nabla F(x^*)(x-x^*) = k_1 - k_2$. Again, due to the convexity of $K(x^*)$,

$$\nabla F(x^*)(x-x^*) \in -K(x^*).$$

Henceforth, there exist $k_1 \in -K(x^*) \setminus \{0\}$ and $k_2 \in K(x^*)$ such that

$$\nabla F(x^*)(x-x^*) = k_1 - k_2 \in -K(x^*).$$

Using that $\nabla F(x^*)(x-x^*) \cap -K(x^*) = \{0\}$, we have that $k_1 = k_2$ and, hence, $k_1 \in -K(x^*) \cap K(x^*)$. Since $K(x^*)$ is a pointed cone, it holds that $k_1 = k_2 = 0$, implying that x^* belongs to S^*. □

Given basic properties of K-convex functions, we will now present the proposed algorithm.

4 A steepest Descent-like Method

This section is devoted to presenting a steepest descent-like algorithm for solving unconstrained smooth problems with variable order. Some definitions, the algorithm and some basic properties of the involved functions will be given.

Our algorithm makes use of the set valued mapping $G : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, which for each x, defines the set of the normalized generators of $K^*(x)$, *i.e.* $G(x) \subseteq K^*(x) \cap bd\,(B(0,1))$ is a compact set, such that the cone generated by its convex hull is $K^*(x)$. Although $K^*(x) \cap bd\,(B(0,1))$ fulfills those properties, in general it is possible to take smaller sets; see [21, 24, 25].

On the other hand, we consider function $\phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$,

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

and for each $x \in \mathbb{R}^n$, the auxiliary problem

$$\min_{v \in \mathbb{R}^n} \left\{ \frac{\|v\|^2}{2} + \beta_k \phi(x,v) \right\}. \tag{P^k}$$

Fixed the constants: $\sigma \in (0,1)$, $\delta > 0$ and $\beta_k \geq \delta > 0$ for all k, the algorithm is defined as follows:

Algorithm A

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

Iterative step: Given x^k and β_k, compute v^k, solution of (P^k).
If $v^k = 0$, then stop. Otherwise compute

$$j(k) := \min \left\{ j \in \mathbb{Z}_+ : F(x^k) + \sigma 2^{-j} \nabla F(x^k) v^k - F(x^k + 2^{-j} v^k) \in K(x^k) \right\}. \tag{5}$$

Set

$$x^{k+1} = x^k + \gamma_k v^k,$$

with $\gamma_k = 2^{-j(k)}$.

Remark 4.1 *Compared with the methods proposed in [12, 13] for multicriteria and vector optimization models respectively, Algorithm A is the natural extension of the steepest descent method proposed to the variable order case.*

Due to the variability of the order, we use the auxiliary function $\rho : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, such that

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

Let us now present some properties of the functions ρ and ϕ.

Proposition 4.1 *If $int(K(x^*)) \neq \emptyset$, for all $x \in \mathbb{R}^n$, then for Algorithm A, the following statements hold:*

(i) For each $x \in \mathbb{R}^n$, $\rho(x,\hat{w}) < 0$ if and only if $\hat{w} \in -int(K(x))$.
(ii) The point x is not stationary if and only if there exists $v \in \mathbb{R}^n$ such that $\phi(x,v) < 0$.
(iii) For each $x \in \mathbb{R}^n$, Problem (P^k) has a unique solution, $v(x)$.
(iv) Suppose that $W \subset \mathbb{R}^m$ is a bounded set. If for some $L \geq 0$, $dist(G(x_1),G(x_2)) \leq L\|x_1 - x_2\|$ for all $x_1, x_2 \in \mathbb{R}^n$, then $\rho(x,w)$ is a Lipschitz function for all $(x,w) \in \mathbb{R}^n \times W$.

Proof (i) If $\rho(x,\hat{w}) < 0$, then for all $y \in G(x)$, it holds that $y^T\hat{w} < 0$ and, hence, $\hat{w} \in -K(x)$. As $G(x)$ is compact and $\rho(x,w)$ is a continuous function of w, we get that $\rho(x,w) < 0$, for all w in a neighborhood of $\hat{w}$. As a consequence, $w \in -K(x)$. So, $\hat{w} \in -int(K(x))$.

Conversely, take $\hat{w} \in -int(K(x))$. Then, $\hat{w} \in -K(x)$ and $y^T\hat{w} \leq 0$, for all $y \in G(x)$. Thus, it follows that $\rho(x,\hat{w}) \leq 0$. Consider $y \in G(x)$, such that $\rho(x,\hat{w}) = y^T\hat{w}$. If $y^T\hat{w} = 0$, as $0 \notin G(x)$, we obtain that,

$$\rho(x,\hat{w}+\alpha y) \geq y^T(\hat{w}+\alpha y) = \alpha\|y\|^2 > 0, \text{ for all } \alpha > 0.$$

For α small enough $\hat{w}+\alpha y \in -K(x)$, which means that $0 < \rho(x,\hat{w}+\alpha y) \leq 0$, establishing a contradiction.
(ii) Note that fixed x, $\phi(x,v) = \rho(x,\nabla F(x)v)$. So, the statement follows from the definition of stationarity and (i).
(iii) From the definitions, it is easy to see that $\phi(x,\cdot)$ is a positive homogeneous and sublinear function. So, $\phi(x,\cdot)$ is a convex function, and hence, $\frac{\|v\|^2}{2} + \beta_k\phi(x,v)$ is a strongly convex function for β_k positive. As a consequence, Problem (P^k) has a unique minimizer.
(iv) Analogously $\rho(x,\cdot)$ is a sublinear function and

$$\rho(x_1,w_2) - \rho(x_1,w_2-w_1) \leq \rho(x_1,w_1) \leq \rho(x_1,w_2) + \rho(x_1,w_1-w_2). \tag{6}$$

We will concentrate in the left inequality. It can be equivalently written as

$$\rho(x_1,w_2) - \rho(x_2,w_2) - \rho(x_1,w_2-w_1) \leq \rho(x_1,w_1) - \rho(x_2,w_2). \tag{7}$$

By the Cauchy-Schwartz inequality,

$$-\|w\| \leq \rho(x,w) \leq \|w\|. \tag{8}$$

Thus, combining (7) and (8), we obtain that

$$\rho(x_1,w_2) - \rho(x_2,w_2) - \|w_1 - w_2\| \leq \rho(x_1,w_1) - \rho(x_2,w_2). \tag{9}$$

Now we focus on the expression $\rho(x_1,w_2) - \rho(x_2,w_2)$. As $G(x_2)$ is compact, there exists $y_2 \in G(x_2)$, satisfying

$$\rho(x_2,w_2) = y_2^T w_2.$$

Taking into account that $dist(G(x_1),G(x_2)) \leq L\|x_1 - x_2\|$, there exist $z_1 \in G(x_1)$ and $\eta_1 \in B(0,1)$ such that $y_2 = z_1 + L\|x_1 - x_2\|\eta_1$.
Hence,

$$\rho(x_1,w_2) - \rho(x_2,w_2) = y_1^T w_2 - y_2^T w_2 = \left(y_1^T - z_1^T - \eta_1^T L\|x_1 - x_2\|\right) w_2.$$

Recalling that $z_1 \in G(x_1)$, and using the definition of ϕ, we get that

$$[y_1^T - z_1^T]w_2 \geq 0.$$

So,

$$\rho(x_1,w_2) - \rho(x_2,w_2) \geq -\eta_1^T L\|x_1 - x_2\|w_2 \geq -L\|x_1 - x_2\|\|w_2\|. \tag{10}$$

Note that as W is bounded, for some $M > 0$, $\|w\| \leq M$ for all $w \in W$. Combining the inequalities (9) and (10), and defining $\hat{L} = LM$, we obtain

$$\rho(x_1,w_1) - \rho(x_2,w_2) \geq -\hat{L}\|x_1 - x_2\| - \|w_1 - w_2\|. \tag{11}$$

On the other hand, the right inequality of (6) leads to

$$\rho\,(x_1,w_1) - \rho\,(x_2,w_2) \leq \rho\,(x_1,w_2) - \rho\,(x_2,w_2) + \rho\,(x_1,w_1 - w_2)\,.$$

Again, using (8), we obtain

$$\rho(x_1,w_1) - \rho(x_2,w_2) \leq \rho(x_1,w_2) - \rho(x_2,w_2) + \|w_1 - w_2\|. \tag{12}$$

Analogously, taking $z_2 \in G(x_2)$ and $\eta_2 \in B(0,1)$ such that $y_1 = z_2 + L\|x_1 - x_2\|\eta_2$

$$\rho(x_1,w_2) - \rho(x_2,w_2) = y_1^T w_2 - y_2^T w_2 = \left(\eta_2^T L\|x_1 - x_2\| + z_2^T - y_2^T\right) w_2$$

and

$$[z_2^T - y_2^T]w_2 \leq 0.$$

So,

$$\rho(x_1,w_2) - \rho(x_2,w_2) \leq \eta_2^T L\|x_1 - x_2\|w_2 \leq L\|x_1 - x_2\|\|w_2\|. \tag{13}$$

Taking $\hat{L} = LM$, the combination of (12) and (13), implies

$$\rho(x_1,w_1) - \rho(x_2,w_2) \leq \hat{L}\|x_1 - x_2\| + \|w_1 - w_2\|.$$

Together with (11), evidently it follows that

$$|\rho(x_1,w_1) - \rho(x_2,w_2)| \leq \hat{L}\|x_1 - x_2\| + \|w_1 - w_2\|. \tag{14}$$

□

Remark 4.2 *As a consequence if ∇F is locally Lipschitz, then $\phi(x,v)$ is also locally Lipschitz and therefore, a continuous function.*

Based on Proposition 4.1(iii), we define $v(x)$ as the unique solution of problem (P^k), and $y(x,v) \in G(x)$ is such that $y(x,v)^T \nabla F(x)v = \phi(x,v)$. We will discuss the continuity of $\theta_\beta(x) := \frac{\|v(x)\|^2}{2} + \beta\phi(x,v(x))$ with $\beta > 0$.

Proposition 4.2 *Consider $x \in \mathbb{R}^n$ and fix $\beta > 0$, then the following hold*

(i) $\theta_\beta(x) \leq 0$ *and x is a stationary point if and only if* $\theta_\beta(x) = 0$.
(ii) $\|v(x)\| \leq 2\beta\|\nabla F(x)\|$.
(iii) If G is a closed map, then θ_β is an upper semicontinuous function.
(iv) If $dist(G(x_1),G(x_2)) \leq L\|x_1 - x_2\|$ for some $L > 0$ and ∇F is locally Lipschitz, then θ_β is lower semicontinuous.

Proof (i) As shown in Proposition 4.1(ii), x is a non stationary point if and only if for some v, $\phi(x,v) < 0$. Take $\lambda \geq 0$. By the positive homogeneity of ϕ, $\phi(x,\lambda v) = \lambda\phi(x,v)$. So, $\theta_\beta(x) \leq \frac{\lambda^2}{2} + \lambda\beta\phi(x,v)$ and for λ small enough

$$\theta_\beta(x) \leq \frac{\lambda^2}{2} + \lambda\beta\phi(x,v) < 0.$$

If $\theta_\beta(x) < 0$, as $\beta > 0$, it holds that $\phi(x,v(x)) = \frac{\theta_\beta(x) - \|v\|}{\beta} < 0$.
(ii) Since $\theta_\beta(x) = \frac{\|v(x)\|^2}{2} + \beta\phi(x,v(x))$ and by (i), $\theta_\beta(x) \leq 0$. Moreover

$$\|v(x)\|^2 \leq -2\beta\phi(x,v(x)) \leq 2\beta\|\nabla F(x)v(x)\| \leq 2\beta\|\nabla F(x)\|\|v(x)\|,$$

by the Cauchy Schwartz inequality. Then,

$$\|v(x)\| \leq 2\beta\|\nabla F(x)\|,$$

which leads to the result.
(iii) We now prove the upper semi-continuity of θ_β. Let $\{x^k\}$ be a sequence converging to x. Take $\hat{x}$ such that $v(x) = \hat{x} - x$. Clearly,

$$\begin{aligned}\theta_\beta(x^k) &\leq \frac{\|\hat{x}-x^k\|^2}{2} + \beta\phi(x^k,\hat{x}-x^k)\\ &= \frac{\|\hat{x}-x^k\|^2}{2} + \beta y(x^k,\hat{x}-x^k)^T\nabla F(x^k)(\hat{x}-x^k).\end{aligned} \tag{15}$$

As $y(x^k,\hat{x}-x^k)$, is bounded there exists a convergent subsequence. Without loss of generality, we assume that $\lim_{k\to\infty} y(x^k,\hat{x}-x^k) = y$. Since G is closed, $y \in G(x)$.
Taking limits in (15),

$$\begin{aligned}\limsup_{k\to\infty}\theta_\beta(x^k) &\leq \limsup_{k\to\infty}\frac{\|\hat{x}-x^k\|^2}{2} + \beta y(x^k,\hat{x}-x^k)\nabla F(x^k)(\hat{x}-x^k)\\ &= \frac{\|\hat{x}-x\|^2}{2} + \beta y(x,v(x))\nabla F(x)(\hat{x}-x)\\ &\leq \frac{\|\hat{x}-x\|^2}{2} + \beta\phi(x,\hat{x}-x) = \theta_\beta(x).\end{aligned}$$

So, θ_β is upper semi-continuous. (iv) Consider $\phi(x,\hat{x}^k - x) - \phi(x^k,\hat{x}^k - x^k)$. As $\{x^k\}$ is a convergent sequence and F is a smooth function, both sequences $\{x^k\}$ and $\{\nabla F(x^k)\}$ are also bounded. Moreover, by (ii),

$$\|\hat{x}^k - x^k\| \leq 2\beta\|\nabla F(x^k)\|. \tag{16}$$

Hence, $\{\nabla F(x)(\hat{x}^k - x^k)\}$ and $\{\nabla F(x^k)(\hat{x}^k - x^k)\}$ are bounded. By Proposition 4.1 (iv), (14) holds for $x_1 = x$, $x_2 = x^k$, $w_1 = \nabla F(x)(\hat{x}^k - x^k)$ and $w_2 = \nabla F(x^k)(\hat{x}^k - x^k)$. That is

$$\begin{aligned}\left|\rho\left(x,\nabla F(x)(\hat{x}^k-x^k)\right) - \rho\left(x^k,\nabla F(x^k)(\hat{x}^k-x^k)\right)\right| &\leq L\|x-x^k\|\|\nabla F(x)(\hat{x}^k-x^k)\|\\ &\quad+\left\|\left(\nabla F(x)-\nabla F(x^k)\right)(\hat{x}^k-x^k)\right\|\\ &\leq \hat{L}\|x-x^k\| + \|\nabla F(x)-\nabla F(x^k)\|\|(\hat{x}^k-x^k)\|,\end{aligned}$$

where $\hat{L} = LM$ and $\|\nabla F(x)(\hat{x}^k - x^k)\| \leq M$.
Noting that

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

and since $\nabla F(x)$ is locally Lipschitz, by (16) and Remark 4.2, it follows that

$$\left|\phi(x,\hat{x}^k - x) - \phi(x^k,\hat{x}^k - x^k)\right| \leq \hat{L}\|x - x^k\|.$$

In particular $\lim_{k\to\infty} \phi(x,\hat{x}^k - x) - \phi(x^k,\hat{x}^k - x^k) = 0$. Now consider $\theta_\beta(x)$. It holds that

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

Taking limits, as k tends to ∞, it follows that

$$\theta_\beta(x) \leq \liminf_{k\to\infty}\left\{\theta_\beta(x^k) + \beta\left(\phi(x,\hat{x}^k - x) - \phi(x^k,\hat{x}^k - x^k)\right) - \langle \hat{x}^k, x^k - x\rangle + \frac{\|x\|^2 - \|x^k\|^2}{2}\right\}.$$

Since

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

and

$$\lim_{k\to\infty} -\langle \hat{x}^k, x^k - x\rangle + \frac{1}{2}\left(\|x\|^2 - \|x^k\|^2\right) = 0,$$

we obtain that

$$\theta_\beta(x) \leq \liminf_{k\to\infty} \theta_\beta(x^k).$$

□

Now we show that Algorithm A is well defined.

Proposition 4.3 *Given x^k, either there exists $j(k)$ solution of (5) or $v^k = 0$.*

Proof If $v^k \neq 0$, then by Proposition 4.2(i), x^k is not a stationary point and $\theta_\beta(x^k) < 0$. In particular, $\phi(x^k, v(x^k)) < 0$. As $\phi(x^k, v(x^k)) = \rho\left(x^k, \nabla F(x^k)v(x^k)\right)$, by Proposition 4.1 (i),

$$\nabla F(x^k)v(x^k) \in -int(K(x^k)). \tag{17}$$

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

$$F(x^k) + \sigma 2^{-j}\nabla F(x^k)v^k - F(x^k + 2^{-j}v^k) = (\sigma - 1)2^{-j}\nabla F(x^k)v^k + o(2^{-j}).$$

As $\sigma < 1$ and $K(x^k)$ is a cone, by (17), it follows that

$$(\sigma - 1)2^{-j}\nabla F(x^k)v^k \in int(K(x^k)).$$

Combining this fact and the previous equation, we get that for all j, sufficiently large,

$$(\sigma - 1)2^{-j}\nabla F(x^k)v^k + o(2^{-j}) \in K(x^k).$$

Hence, (5) has a solution. □

After discussing the continuity of the involved functions, we will now present the convergence results.

5 Convergence of the Method

In this section we obtain the convergence of the Algorithm A, as presented in the previous section. First we consider the general case and subsequently, the result is refined for K-convex functions. From now on $\{x^k\}$ denotes the sequence generated by Algorithm A. We begin with the following lemma.

Lemma 5.1 *Assume that*

(i) $\cup_{x\in\mathbb{R}^n}K(x) \subset \mathscr{K}$*, where* $\mathscr{K}$ *is a Daniell cone.*
(ii) The application $G(x)$ *is closed.*
(iii) $dist(G(x),G(\bar{x})) \leq L\|x-\bar{x}\|$*, for all* $x,\bar{x}\in\mathbb{R}^n$*.*

If x^* *is an accumulation point of* $\{x^k\}$*, then* $\lim_{k\to\infty} F(x^k) = F(x^*)$*.*

Proof Take $\lim_{k\to\infty} x^{i_k} = x^*$ a subsequence of $\{x^k\}$. By its definition it holds that

$$F(x^{k+1}) - F(x^k) - \sigma\gamma_k \nabla F(x^k)v^k \in -\mathscr{K}.$$

By Proposition 4.1 (i), implies that $\rho\left(x^{i_k}, F(x^{k+1}) - F(x^k) - \sigma\gamma_k\nabla(F(x^k)v^k)\right) \leq 0$. However ρ is a sublinear function, as shown in Proposition 4.1 (iv), so,

$$\rho\left(x^k, F(x^{k+1}) - F(x^k)\right) \leq \sigma\gamma_k\rho\left(x^k, \nabla F(x^k)v^k\right). \tag{18}$$

But $\rho\left(x^k, \nabla F(x^k)v^k\right) = \phi(x^k, v^k) < 0$. So, $\rho\left(x^k, F(x^{k+1}) - F(x^k)\right) < 0$ and, again by Proposition 4.1(ii), this inequality is equivalent to

$$F(x^k) - F(x^{k+1}) \in int(K(x^k)).$$

As $\cup_{x\in\mathbb{R}^n}K(x) \subset \mathscr{K}$, it holds that for all x, $int(K(x)) \subset int(\mathscr{K})$. So,

$$F(x^k) - F(x^{k+1}) \in int(\mathscr{K}).$$

or equivalently: $F(x^k)$ is a decreasing sequence with respect to the cone $\mathscr{K}$.
Using that F is continuous,

$$\lim_{k\to\infty} F(x^{i_k}) = F(x^*).$$

Particularly, $F(x^*)$ is an accumulation point, so, due to $\mathscr{K}$ is a Daniell cone; see [26, 27], it follows that

$$\lim_{k\to\infty} F(x^k) = F(x^*).$$

□

Theorem 5.1 *Suppose that*

(i) $\cup_{x\in\mathbb{R}^n}K(x) \subset \mathscr{K}$*, where* $\mathscr{K}$ *is a Daniell cone.*
(ii) The application $G(x)$ *is closed.*
(iii) $dist(G(x),G(\hat{x})) \leq L\|x-\hat{x}\|$*, for all* $x,\hat{x}\in\mathbb{R}^n$*.*
(iv) $\nabla F(x)$ *is a locally Lipschitz function.*

If β_k *is bounded, then all accumulation points of* $\{x^k\}$ *are stationary points of Problem (1).*

Proof Rewrite (18) as

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

and consider the subsequences $\{x^{i_k}\}$ and $\{v^{i_k}\}$, $v^{i_k} = v(x^{i_k})$. Then,

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

By Proposition 4.1(iv), ρ is continuous. As already shown in Lemma 5.1, $\lim_{k\to\infty} F(x^k) = F(x^*)$. Therefore,

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

These facts imply that

$$0 \geq -\sigma\lim_{k\to\infty}\gamma_{i_k}\rho\left(x^{i_k}, \nabla F(x^{i_k})v^{i_k}\right) \geq 0.$$

Hence,

$$\lim_{k\to\infty}\gamma_{i_k}\rho\left(x^{i_k}, \nabla F(x^{i_k})v^{i_k}\right) = 0.$$

As $\gamma_k \in (0,1)$ for all k, $\limsup_{k\to\infty}\gamma_{i_k} \geq 0$. Due to the boundedness of $\{\beta_{i_k}\}$, and the convergence of $\{x^{i_k}\}$, recall that F is smooth, by Proposition 4.2(ii), $\{v^{i_k}\}$ is also bounded. Without loss of generality, suppose that for the subsequence $\{i_k\}$ it holds that $\{x^{i_k}\}$, $\{\beta_{i_k}\}$, $\{v^{i_k}\}$ and $\{\gamma_{i_k}\}$ converge to x^*, β^*, v^* and to $\gamma^* = \limsup_{k\to\infty}\gamma_{i_k}$, respectively. We consider two cases $\gamma^* > 0$ and $\gamma^* = 0$.
Case 1: $\gamma^* > 0$. Hence, $\lim_{k\to\infty}\phi(x^{i_k}, v^{i_k}) = \lim_{k\to\infty}\rho\left(x^{i_k}, \nabla F(x^{i_k})v(x^{i_k})\right) = 0$. Suppose that

$$\theta_{\beta^*}(x^*) = \|v(x^*)\|^2/2 + \beta\phi(x^*, v) < -\varepsilon < 0.$$

Due to the continuity of $\phi(\cdot,\cdot)$, recall Remark 4.2, for k large enough $\phi(x^{i_k}, v(x^*)) < -\varepsilon/2$. On the other hand, for k sufficiently large

$$\frac{\|v^{i_k}\|^2}{2} + \beta_{i_k}\phi(x^{i_k}, v^{i_k}) > \beta_{i_k}\phi(x^{i_k}, v^{i_k}) > -\varepsilon/4. \tag{19}$$

By definition of v^{i_k}

$$\frac{\|v\|^2}{2} + \beta_{i_k}\phi(x^{i_k}, v) \geq \frac{\|v^{i_k}\|^2}{2} + \beta_{i_k}\phi(x^{i_k}, v^{i_k}). \tag{20}$$

Combining (19) and (20), we obtain:

$$\frac{\|v\|^2}{2} + \beta_{i_k}\phi(x^{i_k}, v) > -\varepsilon/4.$$

Taking limit as k goes to ∞, recalling that $\phi(x^{i_k}, v)$ is a continuous function, we obtain the following contradiction

$$-\varepsilon > \frac{\|v\|^2}{2} + \beta^*\phi(x, v) > -\varepsilon/4.$$

Therefore, we may conclude that $\theta_{\beta^*}(x^*) \geq 0$ and hence, using Proposition 4.2, if $\limsup_{k\to\infty}\gamma_{i_k} > 0$, x^* is stationary.
Case 2: $\gamma^* = 0$.
Using the previously defined convergent subsequences $\{x^{i_k}\}$, $\{\beta_{i_k}\}$, $\{v^{i_k}\}$, $\{\gamma_{i_k}\}$, we get that

$$\phi\left(x^{i_k}, \nabla F(x^{i_k})v^{i_k}\right) \leq \theta(x^{i_k}) < 0.$$

Taking limits:

$$\phi(x^*, \nabla F(x^*)v^*) \le \theta(x^*) \le 0.$$

Fix $q \in \mathbb{N}$. Then, for k sufficiently large

$$F(x^{i_k} + 2^{-q} v^{i_k}) \notin F(x^{i_k}) + \frac{\sigma \nabla F(x^{i_k})}{2^q} - K(x^{i_k}),$$

as there exists $\hat{y}_{i_k} \in G(x^{i_k})$ such that

$$\left\langle F(x^{i_k} + 2^{-q} v^{i_k}) - F(x^{i_k}) - \frac{\sigma \nabla F(x^{i_k})}{2^q}, \hat{y}_{i_k} \right\rangle > 0,$$

it holds that

$$\rho\left(x^{i_k}, F(x^{i_k} + 2^{-q} v^{i_k}) - F(x^{i_k}) - \frac{\sigma \nabla F(x^{i_k})}{2^q}\right) \ge 0.$$

Taking limits as k tends to ∞, and recalling that ρ is a continuous function, then

$$\rho\left(x^*, F(x^* + 2^{-q} v^*) - F(x^*) - \frac{\sigma \nabla F(x^*) v^*}{2^q}\right) \ge 0.$$

But ρ is a positive homogeneous function, so,

$$\rho\left(\frac{x^*, F(x^* + 2^{-q} v^*) - F(x^*)}{2^{-q}} - \sigma \nabla F(x^*) v^*\right) \ge 0.$$

Taking limits as q tends to ∞, we obtain

$$\rho\left(x^*, (1-\sigma)\nabla F(x^*) v^*\right) \ge 0.$$

Finally as $\rho\left(x^*, \nabla F(x^*) v^*\right) \le 0$, it holds

$$\rho\left(x^*, \nabla F(x^*) v^*\right) = 0.$$

and by Proposition 4.1(ii), this is equivalent to say that $x^* \in S^s$. □

This result needs the existence of an accumulation point. Based on quasi-Féjer theory, we prove this in the convex case.

Define $\beta_k = \frac{\alpha_k}{\xi_k}$, where $\alpha_k \ge 0$, $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$ and $\xi_k = \max_{y \in G(x^k)} \|y^T \nabla F(x^k)\|$. We begin with a result that provides an upper bound on $\|x^{k+1} - x^k\|^2 - \|x^k - x\|^2$.

Lemma 5.2 *Take $x \in \mathbb{R}^n$. If F is a K-convex function and K a closed map, then*

$$\|x^{k+1} - x\|^2 \le \|x^k - x\|^2 + \beta \sum_{i=1}^{\ell_k} \lambda_i^k [y_i^k]^T [F(x) - F(x^k)] + \gamma_k^2 \alpha_k^2,$$

for some $y_i^k \in G(x^k)$ and $\lambda_i^k \ge 0$, $i = 1, \ldots, \ell_k$, such that $\sum_{i=1}^{\ell_k} \lambda_i^k = 1$.

Proof First note that

$$\|x^{k+1}-x\|^2 - \|x^k - x\|^2 = 2\gamma_k \langle v^k, x^k - x\rangle + \gamma_k^2 \|v^k\|^2. \tag{21}$$

But v^k is a solution of

$$\min_{v\in\mathbb{R}^n} \frac{\|v\|^2}{2} + \beta_k \rho\left(x^k, \nabla F(x^k)v\right),$$

and, as already shown in Proposition 4.1 (iii), the objective function of this problem is convex. Hence,

$$0 = v^T + \beta_k \sum_{i=1}^{\ell_k} \lambda_i^k [y_i^k]^T \nabla F(x^k),$$

for some $\lambda_i^k \geq 0$ and some $y_i^k \in G(x^k)$, such that $\sum_{i=1}^{\ell_k} \lambda_i^k = 1$ and

$$[y_i^k]^T \nabla F(x^k) v^k = \rho\left(x^k, \nabla F(x^k)v\right).$$

Substituting in (21), it follows that

$$\|x^{k+1}-x\|^2 - \|x^k - x\|^2 = 2\beta_k\gamma_k \sum_{i=1}^{\ell_k} \lambda_i^k [y_i^k]^T \nabla F(x^k)(x - x^k) + \gamma_k^2 \|v^k\|^2. \tag{22}$$

As F is a K-convex function, by Proposition 3.4, $F(x) - F(x^k) - \nabla F(x^k)(x - x^k) \in K(x^k)$. Moreover,

$$[y_i^k]^T\left(F(x) - F(x^k)\right) \geq [y_i^k]^T \nabla F(x^k)(x - x^k),$$

since $y_i^k \in G(x^k)$, for $i = 1, \ldots, \ell_k$.
Combining the previous equation with (22) and taking into account that $0 < \gamma_k < 1$, we get

$$\|x^{k+1}-x\|^2 - \|x^k - x\|^2 \leq \beta_k \sum_{i=1}^{\ell_k} \lambda_i^k [y_i^k]^T [F(x) - F(x^k)] + \gamma_k^2 \|v(x^k)\|^2. \tag{23}$$

On the other hand, as $v(x^k)$ is a minimizer of the function $\frac{\|v\|^2}{2} + \beta_k\rho\left(x^k, \nabla F(x^k)v\right)$,

$$\frac{\|v(x^k)\|^2}{2} + \beta_k \rho\left(x^k, \nabla F(x^k) v(x^k)\right) \leq 0.$$

Thus,

$$\begin{aligned} \|v(x^k)\|^2 &\leq -2\beta_k [y_i^k]^T \nabla F(x^k) v^k \\ &\leq 2\frac{\alpha_k}{\xi_k} \max_{y\in G(x^k)} \|y^T \nabla F(x^k)\| \|v^k\| = 2\alpha_k \|v(x^k)\|. \end{aligned}$$

Combining (23) and the previous inequality, it follows that

$$\|x^{k+1}-x^k\|^2 - \|x^k - x\|^2 \leq \beta_k \sum_{i=1}^{j} \lambda_i^k [y_i^k]^T [F(x) - F(x^k)] + \gamma_k^2 \alpha_k^2.$$

□

Now we prove the existence of accumulation points of the sequence generated by Algorithm A for K-convex functions. First we define the set $T = \{x \in \mathbb{R}^n : F(x^k) - F(x) \in K(x^k),\ \forall k\}$ is nonempty. We assume that $T \neq \emptyset$. This hypothesis is closely related to the completeness of $Im(F(x))$, and, as reported in [21], the completeness of $Im(F(x))$ ensures the existence of efficient points. Indeed, this is because $T \neq \emptyset$ is assumed in order to prove the convergence of several methods for solving classical vector optimization problems; see [12–14, 19, 20, 28].

Theorem 5.2 *Let F be a K-convex function and $T \neq \emptyset$. Assume that*

(i) $\cup_{x\in\mathbb{R}^n} K(x) \subset \mathscr{K}$*, where* $\mathscr{K}$ *is a Daniell cone.*
(ii) The mapping $G(x)$ *is closed.*
(iii) $dist(G(x), G(\hat{x})) \leq L\|x-\hat{x}\|$*, for all* $x, \hat{x} \in \mathbb{R}^n$.
(iv) $\nabla F(x)$ *is a locally Lipschitz function.*

Then there exists x^* *such that* $\lim_{k\to\infty} x^{i_k} = x^*$ *and all accumulation points of* $\{x^k\}$ *are weak solutions of Problem (1).*

Proof Suppose that $x \in T$. Then by Lemma 5.2, we obtain

$$\|x^{k+1}-x\|^2 \leq \|x^k - x\|^2 + \beta \sum_{i=1}^{\ell_k} \lambda_i^k [y_i^k]^T [F(x) - F(x^k)] + \gamma_k^2 \alpha_k^2,$$

where $\lambda_i^k \geq 0$, $\sum_{i=1}^{\ell_k} \lambda_i^k = 1$, $y_i^k \in G(x^k)$, $i = 1, \ldots, \ell_k$ and

$$[y_i^k]^T \nabla F(x^k) v^k = \rho\left(x^k, \nabla F(x^k) v\right) = \max_{y\in G(x^k)} \{y^T \nabla F(x^k) v(x^k)\}.$$

Since $x \in T$, it follows that $[F(x^k) - F(x)] \in K(x^k)$ and

$$[y_i^k]^T [F(x) - F(x^k)] \leq 0.$$

Therefore,

$$\|x^{k+1}-x\|^2 \leq \|x^k - x\|^2 + \gamma_k^2 \alpha_k^2.$$

As $\gamma_k \leq 1$ and $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$, $\{x^k\}$ is a quasi- Féjer sequence with respect to T, and, by Theorem 2.1(i), x^k is bounded. So, it has an accumulation point, which will be denoted as x^*.

By Theorem 5.1, x^* is a stationary point and $\lim_{k\to\infty} F(x^k) = F(x^*)$. Hence, using Proposition 3.5, x^* is a weak solution of (1). □

Remark 5.1 *In order to prove the convergence of the whole sequence using Theorem 2.1 (ii), we need to show that there exists an accumulation point of* $\{x^k\}$ *belonging to* T*. The main difficulty is that, due to the variability of the cones, we can only guarantee* $F(x^*) - F(x) \in -bd(K(x^*))$*, for all* $x \in T$*. However, in vector optimization, the main goal is to reconstruct the set of solution. So, if* $\{x^k\}$ *has many accumulation points, a clustering technique such as K-means; see [29], can be used to identify them.*

6 Concluding Remarks

As there exist many applications of variable ordering structure optimization models, it is important to have efficient solution algorithms. As far as we know, the proposed algorithm is a first attempt to apply a continuous descent algorithm to the resolution of this kind of problems. We recall that our proposal extends the steepest descent method, which is one of the oldest, classical and most basic schemes for solving optimization problems. Despite its computational shortcomings, it sets the foundations of future more efficient algorithms, like projected gradient, Newton-like, proximal method, and so on, which have been extended only for vector optimization problems with non-variable order structure; see [6, 7, 19, 20]. The use of these approaches to variable ordering structures is a field of future research. The proposed algorithm is a starting point for the implementation of more complex solution procedures to this more general setting.

Acknowledgements The authors would like to thanks to Prof. Dr. Ole Peter Smith and to the anonymous referees whose suggestions helped us to improve the presentation of this paper.

References

1. Eichfelder, G., Duc Ha, T. X.: Optimality conditions for vector optimization problems with variable ordering structures. Optimization, DOI: 10.1080/02331934.2011.575939 (2011)
2. Eichfelder, G.: Optimal elements in vector optimization with variable ordering structure. J. Optim. Theory Appl. **151**, 217-240 (2011)
3. Baatar, D., Wiecek, M.M.: Advancing equitability in multiobjective programming. Comput. Math. Appl. **2**, 225-234 (2006)
4. Engau, A.: Variable preference modeling with ideal-symmetric convex cones, J. Global Optim. **42**, 295- 311 (2008)
5. Wiecek, M.M.: Advances in cone-based preference modeling for decision making with multiple criteria. Decis. Mak. Manuf. Serv. **1**, 153-173 (2007)
6. Bonnel, H., Iusem, A.N., Svaiter, B.F.: Proximal methods in vector optimization. SIAM J. Optim. **15**, 953-970 (2005)
7. Fliege, J., Graña Drummond, L.M., Svaiter, B.F.: Newton's method for multiobjective optimization. SIAM J. Optim. **20**, 602-626 (2009)
8. Graña Drummond, L.M., Maculan, N., Svaiter, B.F.: On the choice of parameters for the weighting method in vector optimization. Math. Program. **111**, 201-216 (2008)
9. Jahn, J.: Scalarization in vector optimization. Math. Program. **29**, 203-218 (1984)
10. Luc, D.T.: Scalarization of vector optimization problems. J. Optim. Theory Appl. **55**, 85-102 (1987)
11. Bello Cruz, J.Y., Lucambio Pérez, L.R.: Convergence of a projected gradient method variant for quasiconvex objectives. Nonlinear Anal. **9**, 2917-2922 (2010)
12. Fliege, J., Svaiter, B.F.: Steepest descent methods for multicriteria optimization. Math. Methods Oper. Res. **51**, 479-494 (2000)
13. Graña Drummond, L.M., Svaiter, B.F.: A steepest descent method for vector optimization. J. Comput. Appl. Math. **175**, 395-414 (2005)
14. Bello Cruz, J.Y., Lucambio Pérez, L.R., Melo, J.G.: Convergence of the projected gradient method for quasiconvex multiobjective optimization. Nonlinear Anal. **74**, 5268-5273 (2011)
15. Browder, F.E.: Convergence theorems for sequences of nonlinear operators in Banach spaces. Math. Z. **100**, 201-225 (1967)
16. Iusem, A.N., Svaiter, B.F., Teboulle, M.: Entropy-like proximal methods in convex programming. Math. Oper. Res. **19**, 790-814 (1994)
17. Burachik, R., Graña Drummond, L.M., Iusem, A.N., Svaiter, B.F.: Full convergence of the steepest descent method with inexact line searches. Optimization **32**, 137-146 (1995)
18. Luc, D.T., Tan, N.X., Tinh, P.N.: Convex vector functions and their subdifferential. Acta Math. Vietnam. **23**, 107-127 (1998)
19. Fukuda, E.H., Graña Drummond, L.M.: On the convergence of the projected gradient method for vector optimization. Optimization **60**, 1009-1021 (2011)
20. Graña Drummond, L.M., Iusem, A.N.: A projected gradient method for vector optimization problems. Comput. Optim. Appl. **28**, 5-30 (2004)
21. Luc, D.T.: Theory of vector optimization. Lecture Notes in Economics and Mathematical Systems **319**, Springer, Berlin (1989)
22. Chen, G.Y., Johannes Jahn.: Optimality conditions for set-valued optimization problems. Math. Meth. Oper. Res. **48**, 187-200 (1998)
23. Luc, D. T.: Pareto Optimality, Game Theory And Equilibria. In Pareto Optimality, Springer Optimization and its Applications **17**, 481-515 (2008)
24. Jahn, J.: Vector optimization - Theory, applications and extensions. Springer, Berlin (2004)
25. Jahn, J.: Mathematical vector optimization in Partially Ordered Linear Spaces. Verlag Peter D. Lang, Frankfurt (1986)
26. Isac, G., Tammer, C.: 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 Optimization and Appications **35** 343-365 (2010)
27. Peressini, A.L.: Ordered Topological Vector Space. Harper and Row, (1967)
28. Kiwiel, K.C.: The efficiency of subgradient projection methods for convex optimization I. General level methods. SIAM J. Control Optim. **34**, 660-676 (1996)
29. Kogan, J.: Introduction to clustering large and high-dimensional data. Cambridge University Press, Cambridge (2007)