Inexact projected gradient method for vector optimization

Ellen H. Fukuda* L. M. Graña Drummond†

Abstract

In this work, we propose an inexact projected gradient-like method for solving smooth constrained vector optimization problems. In the unconstrained case, we retrieve the steepest descent method introduced by Graña Drummond and Svaiter. In the constrained setting, the method we present extends the exact one proposed by Graña Drummond and Iusem, since it admits relative errors on the search directions. At each iteration, a decrease of the objective value is obtained by means of an Armijo-like rule. The convergence results of this new method extend those obtained by Fukuda and Graña Drummond for the exact version. Basically, for antisymmetric and non-antisymmetric partial orders, under some reasonable hypotheses, global convergence to weakly efficient points of all sequences produced by the inexact projected gradient method is established for convex (respect to the ordering cone) objective functions. In the convergence analysis we also establish a connection between the so-called weighting method and the one we propose.

Keywords: Weak efficiency; Multiobjective optimization; Projected gradient method; Vector optimization

1 Introduction

Let $K \subset \mathbb{R}^m$ be a closed, convex and pointed (i.e., $K \cap (-K) = \{0\}$) cone with nonempty interior $\mathrm{int}(K)$. The partial order induced by K in $\mathbb{R}^m$ is defined by $u \preceq v$ (alternatively, $v \succeq u$) if $v - u \in K$. We also consider the following stronger relation: $u \prec v$ (alternatively, $v \succ u$) if $v - u \in \mathrm{int}(K)$. Let $C \subseteq \mathbb{R}^n$ be a closed and convex set and $F \colon \mathbb{R}^n \to \mathbb{R}^m$ a continuously differentiable function. Here we are interested in solving the following problem:

$$\begin{array}{ll} \min & F(x) \\ \text{s.t.} & x \in C, \end{array} \tag{1}$$

*Institute of Mathematics, Statistics and Computer Science, State University of Campinas. Rua Sérgio Buarque de Holanda 651, 13083-859, Campinas, SP, Brazil (`ellen@ime.usp.br`).

†Faculty of Business and Administration, Federal University of Rio de Janeiro. Av. Pasteur 250, 22290-240, Rio de Janeiro, RJ, Brazil (`lmgd@facc.ufrj.br`.)

understood in the weak Pareto sense. A point $x^* \in C$ is a *weakly efficient* (or *weak Pareto*) solution of the above problem if there exists no $x \in C$ such that $F(x) \prec F(x^*)$.

We will refer to (1) as a smooth constrained *vector optimization* problem. In the particular case of the Paretian cone $K = \mathbb{R}^m_+$, the nonnegative orthant of $\mathbb{R}^m$, "$\preceq$" is the usual component-wise partial order on $\mathbb{R}^m$. In this case, (1) is called *multiobjective* (or *multicriteria*) *optimization* problem. Moreover, if $m = 1$, the partial order "$\preceq$" is the usual (complete) order on $\mathbb{R}$, induced by the cone of nonnegative real numbers, and (1) is a typical constrained scalar-valued optimization problem. For a general treatment of vector optimization problems see [18, 19, 23, 28].

For the preference order induced by the Paretian cone, the problem of finding weakly (or strongly) efficient solutions emerges in different areas. As some examples, we cite design [11, 20], engineering [7], environmental analysis [8, 21], management science [3, 26, 31], space exploration [25, 30] and statistics [6]. For orders induced by different cones, the vector optimization problem does not arise as frequently as the one concerning the point-wise partial order. However, it is also important, as it can be seen in [1, 2], with issues concerning portfolio selection in security markets. Therefore, vector optimization problems, such as (1), are also relevant and deserve our attention.

One of the most popular strategies for solving vector optimization problems is the *scalarization* approach. Roughly speaking, it consists on replacing the vector-valued problem by a family of scalar-valued problems, in such a way that all optima of the real-valued problems are solutions of the original one [17, 22, 24]. Among the scalarization techniques, we have the *weighting method*, where basically one minimizes a linear combination of the objective functions. Its main drawback is that the vector of "weights" is not known a priori and so, the method may lead to unbounded, and thus unsolvable, scalar problems, even if the vector optimization problem does have weakly efficient solutions [14]. Of course, other procedures for solving multicriteria and vector optimization problems are also available, like the goal programming and interactive methods [24].

Here we follow another path, which consists on extending classical scalar optimization methods to the vector-valued setting. Versions of steepest descent, projected gradient and Newton methods have been already studied in [10, 16], [12, 13] and [9, 15], respectively. These methods do not scalarize the vector-valued problem and so they do not require prior knowledge of weighting factors or some ordering information of the objectives, and their convergence analysis is quite satisfactory. Different type of algorithms which do not scalarize have been also developed for multicriteria optimization [29].

In this work we extend the (exact) projected gradient method, studied by Graña Drummond and Iusem [13] and Fukuda and Graña Drummond [12], admitting relative errors in the search directions. We recall that at each iteration of that exact method, a search direction is computed by means of an auxiliary nonsmooth problem (the minimization of the max ordering scalarization). In the inexact version proposed here, while a certain first-order optimality condition is not satisfied, an approximation of the exact search direction is computed at each iteration. As in [12], we also show global convergence of the generated sequences, under the same assumptions. Indeed, under standard hypotheses, for convex objective functions with respect to the ordering cone, we show convergence to weakly efficient solutions for arbitrary initial points. Finally, we mention that the proposed method can also be seen as a direct extension of the steepest descent method for unconstrained vector-valued problems with relative errors proposed in [16].

The outline of this article is as follows. In Section 2 we expose the main ideas behind the exact version of the projected gradient method for vector optimization [13] and we also present some results which will be used in this work. We define, in Section 3, the inexact version of such projected gradient method and study some properties of the generated sequences. In Section 4 we analyze the connection between the weighting method and the one we propose. In Section 5, assuming convexity of the objective function with respect to the ordering cone and under some reasonable assumptions, we show that all sequences produced by the inexact projected gradient method are globally convergent to weakly efficient solutions. Dropping the pointedness hypothesis on the cone and adding some new conditions, we establish convergence of the method in Section 6. We finish the article in Section 7 with some brief comments on implementation issues; in particular, we give some conditions on the cone in order to perform the method and obtain efficient points (instead of just weak optima).

2 Preliminaries

In this section we describe the exact projected gradient method proposed in [13] and studied in [12]. The vast majority of the results of this section can be found in these works, so we just state them without proofs. We point out that some of this results will also be needed for the inexact version of the method. Furthermore, we state and proof a technical result, which cannot be found on the aforementioned papers, useful for interpreting the method as an extension of the classical scalar one.

First, let us introduce some definitions and notations. The *dual cone* (or *positive polar cone*) of K is given by

$$K^* \doteq \left\{ y \in \mathbb{R}^m \colon \langle z, y \rangle \geq 0 \text{ for all } z \in K \right\},$$

where $\langle \cdot, \cdot \rangle$ is the Euclidean inner product. Since K is closed, convex and nonempty, we have $K^{**} = K$ [27, Theorem 14.1].

Let us now give a scalar representation of K^*. Consider $G \subset K^*$ a fixed (but arbitrary) compact set with $0 \notin G$ and such that the cone generated by its convex hull $\operatorname{conv}(G)$ is K^*. We can take, for example, $G \doteq \{w \in K^* \colon \|w\| = 1\}$, where $\|\cdot\|$ denotes the Euclidean norm. However, we can consider much smaller sets in general. In the multiobjective case, since $(\mathbb{R}^m_+)^* = \mathbb{R}^m_+$, G can be taken as the canonical basis of $\mathbb{R}^m$. If K is a polyhedral cone, since K^* is also polyhedral, G can be chosen as the finite set of its extreme rays.

In the general case, for a given such set G, we denote by $\varphi\colon \mathbb{R}^m \to \mathbb{R}$ its *support function*, that is

$$\varphi(y) \doteq \max_{w \in G} \langle w, y \rangle. \tag{2}$$

Since $K^{**} = K$ and G spans all the nontrivial directions of the cone K^*, we have the following characterizations of $-K$ and $-\mathrm{int}(K)$:

$$-K = \{y \in \mathbb{R}^m \colon \varphi(y) \leq 0\} \quad \text{and} \quad -\mathrm{int}(K) = \{y \in \mathbb{R}^m \colon \varphi(y) < 0\}. \tag{3}$$

For the sake of completeness, we list here some properties of the support function φ that will be used on the sequel.

Proposition 2.1. *[16, Lemma 3.1] Consider $G \subset K^*$ and $\varphi\colon \mathbb{R}^m \to \mathbb{R}$ its the support function. Then:*

(i) The function φ is positively homogeneous of degree 1.

(ii) For all $y, z \in \mathbb{R}^m$, we have $\varphi(y+z) \leq \varphi(y) + \varphi(z)$ and $\varphi(y) - \varphi(z) \leq \varphi(y-z)$.

(iii) For all $y, z \in \mathbb{R}^m$, if $y \preceq z$ ($y \prec z$), then $\varphi(y) \leq \varphi(z)$ $\big(\varphi(y) < \varphi(z)\big)$.

(iv) The function φ is Lipschitz continuous with constant $\max_{w \in G} \|w\|$.

Let us now go back to the vector optimization problem (1) and introduce a necessary (but in general not sufficient) condition for optimality. We say that a point $\bar{x} \in C$ is *stationary* for F if

$$-\mathrm{int}(K) \cap \big\{JF(\bar{x})(x - \bar{x})\colon x \in C\big\} = \emptyset,$$

where $JF(\bar{x})$ is the Jacobian of F at $\bar{x}$. Thus, $\bar{x} \in C$ is stationary for F if, and only if, $JF(\bar{x})v \not\prec 0$ for all $v \in C - \bar{x}$. Using (3), this is also equivalent to $\varphi\big(JF(\bar{x})v\big) \geq 0$, for all $v \in C - \bar{x}$. Note that for $m = 1$ and $K = \mathbb{R}_+$, we retrieve the well-known stationarity condition for constrained scalar-valued optimization: $\langle \nabla F(\bar{x}), x - \bar{x} \rangle \geq 0$ for all $x \in C$.

Observe also that if $x \in C$ is nonstationary, there exists $v \in C - x$ with $JF(x)v \prec 0$. So, since

$$F(x + tv) = F(x) + tJF(x)v + o(t),$$

with $\lim_{t \to 0} o(t)/t = 0$, we can take t positive and sufficiently small, in order to still have $JF(x)v + o(t)/t \in -\mathrm{int}(K)$. Therefore, for any nonstationary $x \in C$, there exists $v \in C - x$, a *descent direction* for F at x, i.e., $F(x + tv) \prec F(x)$ for all $t \in (0, t_0]$, for some $t_0 > 0$. Actually, for any nonstationary point $x \in C$, we have a stronger result which will allow us to implement an Armijo-like rule at each step of the algorithm.

Proposition 2.2. *[13, Proposition 1] Let $\delta \in (0,1)$, $x \in C$ and $v \in C - x$ such that $JF(x)v \prec 0$. Then, there exists $\bar{t} > 0$ such that $F(x + tv) \prec F(x) + \delta t JF(x)v$ for all $t \in (0, \bar{t}]$.*

Now we can briefly describe the exact projected gradient method for vector optimization [13]. The *exact projected gradient direction* for F at $x \in C$ is given by

$$v(x) \doteq \operatorname*{argmin}_{v \in C-x} h_x(v), \quad \text{where} \quad h_x(v) \doteq \hat{\beta}\varphi\big(JF(x)v\big) + \frac{\|v\|^2}{2} \tag{4}$$

and $\hat{\beta} > 0$. Note that, in view of the strong convexity of h_x, the direction $v(x)$ is well defined and its defining scalar-valued optimization problem is in general nonsmooth, because of the maximum function. However, in some important cases, as multiobjective optimization, this problem can be replaced by the minimization of a single variable subject to the original constraint plus finitely many smooth inequalities. This is also possible for the general polyhedral case.

For a given $x \in C$, let us define the optimal value of the minimization problem from (4) as

$$\theta(x) \doteq \min_{v \in C-x} h_x(v) = h_x\big(v(x)\big).$$

Since $0 \in C - x$, we have

$$\theta(x) \leq 0 \quad \text{for all } x \in C. \tag{5}$$

Also we have the following characterization of stationarity.

Proposition 2.3. *[13, Proposition 3] For any $x \in C$, the following three conditions are equivalent: (i) The point x is nonstationary for F; (ii) $\theta(x) < 0$ and (iii) $v(x) \neq 0$. In particular, $x \in C$ is stationary for F if and only if $\theta(x) = 0$.*

At iteration k of the exact projected gradient method for vector optimization, the search direction is obtained by computing $v(x^k)$. By (3), this is a descent direction, since $\theta(x^k) < 0$ implies $\varphi(JF(x^k)v(x^k)) < 0$. So, by means of a backtracking procedure, a steplength t_k is computed in order to define a new iterate $x^{k+1} \doteq x^k + t_k v(x^k)$, in such a way that an Armijo-like rule (as in Proposition 2.2) is satisfied. As noted in [13], unless the (feasible) sequence produced by the method terminates with a stationary point, it consists of infinitely many nonstationary points. In particular, $F(x^{k+1}) \prec F(x^k)$ for all k.

The following auxiliary result will be widely used afterwards in this work, and right now it will allow us to see that the method we are describing effectively extends to the vector-valued case the classical projected gradient method for constrained scalar optimization.

Lemma 2.4. *Let $w \in G \subset K^*$ and $\hat{\beta} > 0$. Then,*

$$\operatorname*{argmin}_{v \in C-x} \left\{ \hat{\beta}\langle w, JF(x)v \rangle + \frac{\|v\|^2}{2} \right\} = P_{C-x}\big(-\hat{\beta}JF(x)^\top w\big) = P_C\big(x - \hat{\beta}JF(x)^\top w\big) - x,$$

where P_C and P_{C-x} denote the orthogonal projection onto C and $C - x$, respectively.

Proof. Consider $v_w \doteq \operatorname{argmin}_{v\in C-x}\left\{\hat{\beta}\langle w, JF(x)v\rangle + \|v\|^2/2\right\}$. Then,

$$\begin{aligned} v_w &= \operatorname*{argmin}_{v\in C-x}\left\{\frac{\hat{\beta}^2\|JF(x)^\top w\|^2}{2} + \hat{\beta}\langle w, JF(x)v\rangle + \frac{\|v\|^2}{2}\right\} \\ &= \operatorname*{argmin}_{v\in C-x}\frac{\|v+\hat{\beta}JF(x)^\top w\|^2}{2} \;=\; P_{C-x}\big(-\hat{\beta}JF(x)^\top w\big). \end{aligned}$$

Using the obtuse angle property of projections, we have

$$\big\langle -\hat{\beta}JF(x)^\top w - v_w, v - v_w\big\rangle \le 0 \quad \text{for all } v\in C-x,$$

which can be written as

$$\big\langle x - \hat{\beta}JF(x)^\top w - (x+v_w), u - (x+v_w)\big\rangle \le 0 \quad \text{for all } u\in C,$$

which in turn means

$$x + v_w = P_C\big(x - \hat{\beta}JF(x)^\top w\big),$$

and the proof is complete. □

Observe that in the constrained scalar-valued case, we have $JF(x) = \nabla F(x)^\top$, where $\nabla F(x)$ stands for the gradient of F in x. Also, we can take $G = \{1\}$ and, thus, from the above lemma,

$$v(x) = \operatorname*{argmin}_{v\in C-x}\left\{\hat{\beta}\langle\nabla F(x), v\rangle + \frac{\|v\|^2}{2}\right\} = P_C\big(x - \hat{\beta}\nabla F(x)\big) - x,$$

which is precisely the direction considered in the projected gradient method for real-valued optimization.

We end this section stating a couple of results once again concerning $v(\cdot)$ and $\theta(\cdot)$. The first one gives an upper bound for $\|v(x)\|$ in terms of $JF(x)$. The other one establishes the continuity of $\theta(\cdot)$.

Proposition 2.5. *[12, Lemma 3.3] For all $x\in C$, it holds that*

$$\|v(x)\| \le 2\hat{\beta}\left(\max_{w\in G}\|w\|\right)\|JF(x)\|.$$

Proposition 2.6. *[12, Proposition 3.4] The function $\theta : C \to \mathbb{R}$ is continuous.*

3 The inexact method

In the previous section, we have seen that the direction $v(x)$ computed in each iteration of the (exact) projected gradient method is the minimizer of the problem $\min_{v\in C-x} h_x(v)$, where $h_x(v) = \hat{\beta}\varphi(JF(x)v) + \|v\|^2/2$. Therefore, from a numerical point of view, it would be interesting to see what happens if we take an approximate solution of this problem. As in the steepest descent method for vector optimization [16], we will consider now approximate search directions.

Definition 3.1. *Let $x \in C$ and $\sigma \in [0,1)$. Then, $v \in C - x$ is a σ-approximate (projected gradient) direction at x if*

$$h_x(v) \leq (1-\sigma)\theta(x).$$

Some comments are in order. Observe first that by virtue of (5), $v \in C - x$ is a σ-approximate direction at x if, and only if, the relative error of h_x at v is at most σ, that is to say,

$$\left|\frac{h_x(v) - h_x(v(x))}{h_x(v(x))}\right| = \left|\frac{h_x(v) - \theta(x)}{\theta(x)}\right| \leq \sigma.$$

Clearly, the exact direction $v = v(x)$ is always σ-approximate at x for any $\sigma \in [0,1)$. Moreover, $v(x)$ is the unique 0-approximate direction at x. Also note that given a nonstationary point $x \in C$ and $\sigma \in [0,1)$, a σ-approximate direction v is always a descent direction. Indeed, from Proposition 2.3 and the definition of σ-approximation, it follows that $\varphi(JF(x)v) < 0$, since $\|v\|^2 \geq 0$ and $\hat{\beta} > 0$. Whence, from (3), it follows that $JF(x)v \prec 0$ and so, by Proposition 2.2, v is a descent direction.

We present now a result which, for a given point $x \in C$, establishes the degree of proximity between an approximate direction v and the exact one $v(x)$, in terms of $\theta(x)$. Before that, we need to recall a general fact: for a strongly convex numerical function g, with modulus of convexity κ, if x^* is a minimizer of g, then $g(x) - g(x^*) \geq (\kappa/2)\|x - x^*\|^2$ for all x. So, since h_x is strongly convex with modulus 1 and $\theta(x) = h_x(v(x)) = \min_{v \in C-x} h_x(v)$, we have

$$h_x(v) - \theta(x) \geq \frac{\|v - v(x)\|^2}{2}. \tag{6}$$

We now state the announced result.

Lemma 3.2. *Let $x \in C$, $\sigma \in [0,1)$ and $v \in C - x$ a σ-approximate direction at x. Then*

$$\|v - v(x)\|^2 \leq 2\sigma|\theta(x)|.$$

Proof. The result follows directly from (5), (6) and the definition of σ-approximation. □

We are in a position to describe the inexact projected gradient method, which, as we will see, is a direct extension of the exact version proposed in [13]. For the sake of simplicity, let us allow to use the same notations as in the exact case: if $\{x^k\}$ is the generated sequence, we call $h_{x^k}(v^k) = \hat{\beta}\varphi(JF(x^k)v^k) + \|v^k\|^2/2$, where v^k is a σ-approximate direction produced by the method at iteration k. Also, $\mathbb{N}$ denotes the set of nonnegative integer numbers. Now we formally state the inexact method.

Algorithm 3.3. *Inexact projected gradient method for vector optimization.*

1. Choose $x^0 \in C$, $\hat{\beta} > 0$, $\tau > 1$, $\delta \in (0,1)$ and $\sigma \in [0,1)$. Set $k \leftarrow 0$.

2. Compute a σ-approximate direction v^k at x^k.

3. If $h_{x_k}(v^k) = 0$, then stop.

4. Choose t_k as the largest $t \in \{\tau^{-j} : j = 0, 1, 2, \dots\}$ such that

$$F\left(x^k + tv^k\right) \preceq F(x^k) + \delta t JF(x^k)v^k.$$

5. Set $x^{k+1} = x^k + t_k v^k$, $k \leftarrow k+1$ and go to step 2.

We must observe that we do not compute at each iteration k the exact direction $v(x^k)$ and, thus, neither the exact solution $\theta(x^k)$. So, instead of $\theta(x^k) = 0$ as stopping criterion, we use $h_{x_k}(v^k) = 0$, which is still equivalent to the stationarity of x^k. In fact, if $h_{x_k}(v^k) = 0$, then by the definition of σ-approximation, $\theta(x^k) \geq 0$. But from (5), it follows that $\theta(x^k) = 0$, which, by Proposition 2.3, means that x^k is a stationary point. On the other hand, if x^k is stationary, then $\theta(x^k) = 0$, and therefore $h_{x_k}(v^k) = 0$ follows directly because $\theta(x^k) \leq h_{x_k}(v^k) \leq (1-\sigma)\theta(x^k)$.

As we have already seen, if x^k is nonstationary, $JF(x^k)v^k \prec 0$. So, thanks to Proposition 2.2, we can use the Armijo-like rule with a backtracking procedure, as stated in step 4. Moreover, if the method does not stop, the sequence $\{F(x^k)\}$ is K-decreasing, i.e., it satisfies $F(x^{k+1}) \prec F(x^k)$ for all k; otherwise, this condition holds until it stops. An important issue, from a practical point of view, is the matter of how can we generate σ-approximate directions without knowing the exact projected gradient direction. In Section 4 we give a sufficient condition for σ-approximation.

From now on, we assume that Algorithm 3.3 produces an infinite sequence $\{x^k\}$ of nonstationary points. From Proposition 2.3, we have $\theta(x^k) < 0$ for all k, and in order to see that every accumulation point x^* of $\{x^k\}$ is stationary it is enough to prove that $\theta(x^*) = 0$. This will be done in our next theorem, which is an application of the standard convergence argument for the classical steepest descent method. Before that, we establish feasibility of the sequence $\{x^k\}$.

Proposition 3.4. *If $\{x^k\} \subset \mathbb{R}^n$ is a sequence generated by Algorithm 3.3, then $x^k \in C$ for all k.*

Proof. We prove that $\{x^k\} \subset C$ by induction in k. The initial iterate x^0 belongs to C by virtue of step 1 of Algorithm 3.3. Assuming that $x^k \in C$, since C is convex and, by the definition of the inexact algorithm, $t_k \in (0,1]$ and $v^k \in C - x^k$, we conclude that $x^{k+1} \in C$. □

Now we can show that every accumulation point of any sequence produced by the inexact method is stationary. We point out that this theorem is different from its counterpart in [13], because there the authors obtain the same conclusion for the exact method, while here we deal with inexact search directions.

Theorem 3.5. *Every accumulation point, if any, of a sequence $\{x^k\}$ generated by Algorithm 3.3 is a feasible stationary point.*

Proof. Assume that $x^* \in \mathbb{R}^n$ is an accumulation point of the sequence $\{x^k\}$. Feasibility of x^* follows from Proposition 3.4, since C is a closed set. According to Proposition 2.3, it is enough to see that $\theta(x^*) = 0$. As $F(x^{k+1}) \preceq F(x^k)$ for all k, from Proposition 2.1(iii), it follows that

$$\varphi(F(x^{k+1})) \leq \varphi(F(x^k)) \quad \text{for all } k.$$

Since F is continuously differentiable and φ is continuous from Proposition 2.1(iv), there exists a subsequence of $\{\varphi(F(x^k))\}$ converging to $\varphi(F(x^*))$. Whence, from the above inequality it follows that

$$\lim_{k\to\infty} \varphi(F(x^k)) = \varphi(F(x^*)). \tag{7}$$

On the other hand, from the Armijo condition, the fact that $\delta > 0$ and Proposition 2.1(i)–(iii), we have

$$-\delta t_k \varphi\big(JF(x^k)v^k\big) \leq \varphi(F(x^k)) - \varphi(F(x^{k+1})).$$

Now, using the fact that $JF(x^k)v^k \prec 0$, together with (3), (7) and the above inequality, we obtain

$$0 \leq -\lim_{k\to\infty} \delta t_k \varphi\big(JF(x^k)v^k\big) \leq \lim_{k\to\infty} \Big[\varphi(F(x^k)) - \varphi(F(x^{k+1}))\Big] = 0.$$

Therefore

$$\lim_{k\to\infty} t_k \varphi\big(JF(x^k)v^k\big) = 0. \tag{8}$$

Consider now a subsequence $\{x^{k_j}\}_j$ converging to x^*. Taking into consideration that $t_k \in (0,1]$ for all k, we just have the following two alternatives:

$$\text{(a) } \limsup_{j\to\infty} t_{k_j} > 0 \qquad \text{or} \qquad \text{(b) } \limsup_{j\to\infty} t_{k_j} = 0. \tag{9}$$

First, assume that (9)(a) holds. Then, there exists a subsequence $\{t_{k_{j_i}}\}_i$ converging to some $t^* > 0$, and such that $\lim_{i\to\infty} x^{k_{j_i}} = x^*$. So, from (8) it follows that $\lim_{i\to\infty} \varphi\big(JF(x^{k_{j_i}})v^{k_{j_i}}\big) = 0$. Hence, we have

$$0 = \lim_{i\to\infty} \varphi\big(JF(x^{k_{j_i}})v^{k_{j_i}}\big) \leq \frac{1}{\hat{\beta}} \lim_{i\to\infty} \theta(x^{k_{j_i}}) = \frac{1}{\hat{\beta}} \theta(x^*),$$

where the inequality is due to the fact that

$$\hat{\beta}\varphi\big(JF(x^{k_{j_i}})v(x^{k_{j_i}})\big) \leq \hat{\beta}\varphi\big(JF(x^{k_{j_i}})v(x^{k_{j_i}})\big) + \frac{\|v(x^{k_{j_i}})\|^2}{2} = \theta(x^{k_{j_i}}),$$

with $\hat{\beta} > 0$, and the rightmost equality follows from Proposition 2.6. Then, since $\theta(x^*) \leq 0$ from (5), we get $\theta(x^*) = 0$.

Now assume that (9)(b) holds. We know that $\{x^{k_j}\}_j$ is a bounded sequence of nonstationary points, so, since $JF(x^{k_j})$ is bounded, due to the fact that F is continuously differentiable, using Proposition 2.5, we conclude that there exists $K_1 > 0$ such that

$$\|v(x^{k_j})\| \leq 2\hat{\beta}\left(\max_{w\in G}\|w\|\right)\|JF(x^{k_j})\| \leq K_1 \quad \text{for all } j.$$

We claim that the sequence of directions $\{v^{k_j}\}_j$ is bounded too. Indeed, since $\{x^{k_j}\}$ is convergent, from the continuity of θ, established in Proposition 2.6, we see that there exists $K_2 > 0$, such that $|\theta(x^{k_j})| \leq K_2$. Hence, using the triangular inequality, Lemma 3.2 and the above inequality, we have:

$$\|v^{k_j}\| \leq \|v(x^{k_j})\| + \sqrt{2\sigma|\theta(x^{k_j})|} \leq K_1 + \sqrt{2\sigma K_2} \doteq K, \quad \text{for all } j.$$

Therefore, we can take subsequences $\{x^{k_{j_i}}\}_i$, $\{v^{k_{j_i}}\}_i$ and $\{t_{k_{j_i}}\}_i$ converging to x^*, v^* and 0, respectively. Observe that

$$\varphi\big(JF(x^{k_{j_i}})v^{k_{j_i}}\big) \leq \frac{1}{\hat{\beta}}\theta(x^{k_{j_i}}) < 0 \quad \text{for all } i,$$

where the strict inequality follows from Proposition 2.3 and the fact that $\{x^k\}$ is a sequence of nonstationary points. Then, letting $i \to \infty$ in the above inequality, from the continuity of φ and θ (Propositions 2.1(iv) and 2.6), we get

$$\varphi\big(JF(x^*)v^*\big) \leq \frac{1}{\hat{\beta}}\theta(x^*) \leq 0. \tag{10}$$

Take now some fixed but arbitrary positive integer q. Since $\lim_{i\to\infty} t_{k_{j_i}} = 0$, for i large enough we have $t_{k_{j_i}} < \tau^{-q}$, which means that the Armijo condition does not hold at $x^{k_{j_i}}$ for $t = \tau^{-q}$, i.e.,

$$F\left(x^{k_{j_i}} + \tau^{-q}v^{k_{j_i}}\right) \not\preceq F\left(x^{k_{j_i}}\right) + \delta\tau^{-q}JF\left(x^{k_{j_i}}\right)v^{k_{j_i}}. \tag{11}$$

Whence, for all large enough positive integer i, there exists $w^i \in G$ such that

$$\left\langle w^i, F\left(x^{k_{j_i}} + \tau^{-q}v^{k_{j_i}}\right) - F(x^{k_{j_i}}) - \delta\tau^{-q}JF(x^{k_{j_i}})v^{k_{j_i}}\right\rangle > 0,$$

otherwise, from (3), we would have $F(x^{k_{j_i}} + \tau^{-q}v^{k_{j_i}}) \preceq F(x^{k_{j_i}}) + \delta\tau^{-q}JF(x^{k_{j_i}})v^{k_{j_i}}$, in contradiction with (11). Clearly, we do not lose generality if we assume that $\{w^i\}_i$ converges to some $\bar{w} \in G$, since G is compact. So letting $i \to \infty$ in the above numerical inequality, we get

$$\left\langle \bar{w}, F(x^* + \tau^{-q}v^*) - F(x^*) - \delta\tau^{-q}JF(x^*)v^*\right\rangle \geq 0.$$

Hence, using the definition of the support function (2),

$$\varphi\left(F(x^* + \tau^{-q}v^*) - F(x^*) - \delta\tau^{-q}JF(x^*)v^*\right) \geq 0.$$

Since this inequality holds for any positive integer q, using Proposition 2.2 and (3), we conclude that $JF(x^*)v^* \not\prec 0$, i.e. $\varphi(JF(x^*)v^*) \geq 0$. Combining this with (10), we get that $\theta(x^*) = 0$, which completes the proof. $\square$

We know that if $x^* \in C$ is an accumulation point of $\{x^k\}$, then $\theta(x^*) = 0$. So, since θ is continuous, if $x^k \to x^*$, necessarily $\theta(x^k) \to 0$. In the next lemma we see, among other things, how fast does $t_k|\theta(x^k)|$ converge to zero, where t_k is the steplength taken from x^k. We will need the following estimate. From the Armijo condition and the properties of the support function φ (Proposition 2.1), we have

$$\begin{aligned}\varphi\big(F(x^{k+1})\big) &\leq \varphi\big(F(x^k)\big) + t_k\delta\varphi\big(JF(x^k)v^k\big) &(12)\\ &\leq \varphi\big(F(x^k)\big) + t_k\frac{\delta}{\hat{\beta}}\left[(1-\sigma)\theta(x^k) - \frac{\|v^k\|^2}{2}\right], &(13)\end{aligned}$$

where the last inequality follows from the fact that v^k is a σ-approximate direction at $x^k \in C$. Before the announced lemma, we still need to introduce a natural concept: we say that a sequence $\{z^k\} \subset \mathbb{R}^m$ is *K-bounded* (from below) if there exists $\bar{z} \in \mathbb{R}^m$ such that $\bar{z} \preceq z^k$ for all k.

Lemma 3.6. *If the sequence $\{F(x^k)\}$ is K-bounded from below, then:*

(i) $\displaystyle\sum_{k=0}^{\infty} t_k\big|\theta(x^k)\big| < \infty$;

(ii) $\displaystyle\sum_{k=0}^{\infty} t_k\|v^k\|^2 < \infty$;

(iii) $\displaystyle\sum_{k=0}^{\infty} t_k\big|\varphi(JF(x^k)v^k)\big| < \infty$;

(iv) $\displaystyle\sum_{k=0}^{\infty} t_k\big|\langle w, JF(x^k)v^k\rangle\big| < \infty$, *for all* $w \in \operatorname{conv}(G)$.

Proof. Adding up, from $k = 0$ to $k = N$ at (13), we get

$$\begin{aligned}\varphi\big(F(x^{N+1})\big) &\leq \varphi\big(F(x^0)\big) + \sum_{k=0}^{N} t_k\frac{\delta}{\hat{\beta}}\left[(1-\sigma)\theta(x^k) - \frac{\|v^k\|^2}{2}\right]\\ &= \varphi\big(F(x^0)\big) - \sum_{k=0}^{N} t_k\frac{\delta}{\hat{\beta}}\left[(1-\sigma)|\theta(x^k)| + \frac{\|v^k\|^2}{2}\right],\end{aligned}$$

where the equality holds by virtue of the nonstationarity of x^k together with Proposition 2.3. Since $\{F(x^k)\}$ is K-bounded (say by $\bar{F}$), $\delta \in (0,1)$, $\hat{\beta} > 0$, from Proposition 2.1(iii), we obtain

$$\sum_{k=0}^{N} t_k\left[(1-\sigma)|\theta(x^k)| + \frac{\|v^k\|^2}{2}\right] \leq \frac{\hat{\beta}}{\delta}\left[\varphi\big(F(x^0)\big) - \varphi(\bar{F})\right] \quad \text{for all } N.$$

Using now the fact that $\sigma \in [0,1)$, we immediately get (i) and (ii). The proof of (iii) is analogous, but instead of (13), we use (12).

As for (iv), we observe from the Armijo condition, that for all $w \in \mathrm{conv}(G) \subset K^*$,

$$-t_k\delta\langle w, JF(x^k)v^k\rangle \leq \langle w, F(x^k)\rangle - \langle w, F(x^{k+1})\rangle.$$

Adding up k from 0 to N at the above inequality, and observing that $\langle w, JF(x^k)v^k\rangle < 0$, because $JF(x^k)v^k \prec 0$, we get

$$\sum_{k=0}^{N} t_k|\langle w, F(x^k)v^k\rangle| \leq \frac{1}{\delta}\left[\langle w, F(x^0)\rangle - \langle w, F(x^{N+1})\rangle\right].$$

Since $\delta > 0$, using once again the fact that $\bar{F} \preceq F(x^k)$ for all k, we conclude that (iv) is also true. □

4 Relation between the projected gradient method and the weighting method

A well-known technique for solving problem (1) is the so-called *scalarization* procedure, which roughly speaking consists on replacing the vector optimization problem by one or more scalar-valued problems. A very simple version of this strategy is the so-called weighting method, which consists on taking $w \in K^*$ and solving the following scalar-valued problem

$$\begin{array}{ll} \min & \langle w, F(x)\rangle \\ \text{s.t.} & x \in C. \end{array} \tag{14}$$

The optima of the above problem are also weakly efficient solutions of the original vector optimization problem. Although this procedure has the advantage of requiring to solve smooth scalar minimization problems, for which we have fast and reliable methods, it is well-known that many choices of w can lead to unbounded problems, even if the vector-valued problem has weakly efficient solutions. It is not an easy task to find $w \in K^*$ such that the corresponding scalarized problem is guaranteed to have an optimal solution [14].

In this section, we will combine the ideas of previous sections with the weighting method. First observe that $\nabla_x\langle w, F(x)\rangle = JF(x)^\top w$, so if the (scalar-valued) projected gradient method is applied to solve the problem (14), at iteration k, the search direction will be $P_C(x^k - \hat{\beta}JF(x^k)^\top w) - x^k$. The next proposition shows that for a suitable $w \in K^*$, the projected gradient direction for the scalarized problem coincides with the (exact) direction computed by the projected gradient method for vector optimization. So the exact method can be seen as if, at each iteration, it implicitly performs a scalarization (for an unknown weight $w \in \mathbb{R}^m$) and computes the projected gradient direction for the corresponding scalarized problem. Also, observe that $\varphi(y) = \max_{w\in G}\langle w, y\rangle = \max_{w\in \mathrm{conv}(G)}\langle w, y\rangle$.

Proposition 4.1. *Take $x \in C$ and let $v(x)$ be the (exact) projected gradient direction for F at x, as defined in* (4)*. Then, there exists $\tilde{w} \in \operatorname{conv}(G) \subset K^*$ such that $v(x)$ is the projected gradient direction for the scalarized problem* (14)*, with $w = \tilde{w}$.*

Proof. Let us recall that the (primal) problem which defines $v(x)$ is

$$\min_{v \in C-x} \max_{w \in \operatorname{conv}(G)} \left\{ \hat{\beta} \langle w, JF(x)v \rangle + \frac{\|v\|^2}{2} \right\}.$$

The dual of this problem is

$$\max_{w \in \operatorname{conv}(G)} \min_{v \in C-x} \left\{ \hat{\beta} \langle w, JF(x)v \rangle + \frac{\|v\|^2}{2} \right\}.$$

Using Lemma 2.4, we can simplify the dual problem as

$$\max_{w \in \operatorname{conv}(G)} \left\{ \hat{\beta} \langle JF(x)^\top w, P_{C-x}(-\hat{\beta} JF(x)^\top w) \rangle + \frac{\|P_{C-x}(-\hat{\beta} JF(x)^\top w)\|^2}{2} \right\}. \tag{15}$$

Since $\operatorname{conv}(G)$ is convex and compact, the dual problem has always a solution, which is not necessarily unique. So, let $\tilde{w} \in \operatorname{conv}(G)$ be a dual solution. Let us now show that there is no duality gap. First note that, for all $v \in C - x$,

$$\theta(x) = \max_{w \in \operatorname{conv}(G)} \hat{\beta} \langle w, JF(x)v(x) \rangle + \frac{\|v(x)\|^2}{2} \leq \max_{w \in \operatorname{conv}(G)} \hat{\beta} \langle w, JF(x)v \rangle + \frac{\|v\|^2}{2},$$

because $v(x)$ is the primal solution. In particular, taking $v = P_{C-x}(-\hat{\beta} JF(x)^\top w)$,

$$\theta(x) \leq \max_{w \in \operatorname{conv}(G)} \left\{ \hat{\beta} \langle JF(x)^\top w, P_{C-x}(-\hat{\beta} JF(x)^\top w) \rangle + \frac{\|P_{C-x}(-\hat{\beta} JF(x)^\top w)\|^2}{2} \right\}.$$

The strong duality follows from the above inequality, (15) and the weak duality relation. Thus, we conclude that $(v(x), \tilde{w})$ is a saddle point of $\hat{\beta} \langle w, JF(x)v \rangle + (1/2)\|v\|^2$ in $(C - x) \times \operatorname{conv}(G)$ [4, Proposition 2.6.1]. In particular, we have:

$$\hat{\beta} \langle \tilde{w}, JF(x)v(x) \rangle + \frac{\|v(x)\|^2}{2} \leq \hat{\beta} \langle \tilde{w}, JF(x)v \rangle + \frac{\|v\|^2}{2},$$

for all $v \in C - x$. So, using again Lemma 2.4, we have

$$v(x) = P_{C-x}\big(- \hat{\beta} JF(x)^\top \tilde{w} \big) = P_C\big(x - \hat{\beta} JF(x)^\top \tilde{w} \big) - x, \tag{16}$$

which is the (scalar-valued) projected gradient direction for the scalarized problem when w is replaced by $\tilde{w}$ in (14), because $JF(x)^\top \tilde{w} = \nabla_x \langle \tilde{w}, F(x) \rangle$. □

We now extend to the constrained vector optimization setting a concept considered in [16] for the unconstrained case.

Definition 4.2. *Let $x \in C$. A direction $v \in C - x$ is* scalarization compatible *(or simply s-*compatible*) at x if there exists $w \in \text{conv}(G)$ such that*

$$v = P_{C-x}(-\hat{\beta} JF(x)^\top w).$$

Note that *s*-compatible directions are well defined, since C is closed and convex. Moreover, if v is *s*-compatible at x, by Lemma 2.4, $v = P_C(x - \hat{\beta} JF(x)^\top w) - x$ for some $w \in \text{conv}(G)$. Observe that the exact search direction $v(x)$ is *s*-compatible as it was shown in (16).

Let us now go back to the inexact method. As step 2 of Algorithm 3.3 requires the computation of a σ-approximate projected gradient direction at x^k, it would be interesting to see for which "weights" $w \in \text{conv}(G)$ we have that $v = P_{C-x}(-\hat{\beta} JF(x)^\top w)$ is a σ-approximate direction at x^k. The next proposition, which is an extension of [16, Proposition 5.2], gives us a sufficient condition for an *s*-compatible direction to be a σ-approximation.

Proposition 4.3. *Let $x \in C$, $w \in \text{conv}(G)$, $v = P_{C-x}(-\hat{\beta} JF(x)^\top w)$ and $\sigma \in [0,1)$. If*

$$\hat{\beta}\varphi(JF(x)v) \leq (1-\sigma)\hat{\beta}\langle w, JF(x)v\rangle - \frac{\sigma}{2}\|v\|^2, \tag{17}$$

then v is a σ-approximate projected gradient direction.

Proof. Adding $\|v\|^2/2$ on both sides of (17), we obtain

$$h_x(v) = \hat{\beta}\varphi(JF(x)v) + \frac{\|v\|^2}{2} \leq (1-\sigma)\left\{\hat{\beta}\langle w, JF(x)v\rangle + \frac{\|v\|^2}{2}\right\}.$$

Since $v = P_{C-x}(-\hat{\beta} JF(x)^\top w)$, we see that the right hand side of the above inequality is the objective function of problem (15) (multiplied by $1 - \sigma > 0$). As $\theta(x)$ is the primal optimal value, the result follows from weak duality. □

We end this section by noting that, for $\sigma > 0$, the exact projected gradient direction $v(x)$ satisfies condition (17) with strict inequality.

5 Convergence analysis

In this section we can finally show that, under reasonable hypotheses, every sequence produced by the inexact projected gradient method converges globally to a weakly efficient solution of problem (1). Recall that we are assuming that the method does not stop, i.e., that it generates an infinite sequence $\{x^k\}$ of nonstationary points.

From now on, we will also assume that F is *K-convex* on C, i.e.,

$$F(\lambda x + (1-\lambda)u) \preceq \lambda F(x) + (1-\lambda)F(u), \tag{18}$$

for all $x, u \in C$ and all $\lambda \in [0,1]$. Observe that, for $w \in K^*$, we have that $x \mapsto \langle w, F(x) \rangle$ is a smooth convex real-valued function, so, since $\nabla_x \langle w, F(x) \rangle = JF(x)^\top w$, we have

$$\langle w, F(x) \rangle + \langle w, JF(x)(u-x) \rangle \leq \langle w, F(u) \rangle \quad \text{for all } x, u \in C. \tag{19}$$

Note that, since w is an arbitrary vector in K^*, by linearity of $\langle w, \cdot \rangle$, the above condition is equivalent to the following:

$$F(x) + JF(x)(u-x) \preceq F(u) \quad \text{for all } x, u \in C.$$

The above K-inequality is a general property for K-convex continuously differentiable functions, which clearly extends the well-known fact that every smooth convex scalar function overestimates its linear approximations [23, Lemma 5.2].

For the sake of completeness, let us now state another well-known fact, namely that, as in the scalar case, under convexity of the objective function, stationarity is equivalent to optimality.

Lemma 5.1. *[12, Lemma 5.2] Assume that $F \colon \mathbb{R}^n \to \mathbb{R}^m$ is K-convex. Then $x \in \mathbb{R}^n$ is stationary if and only if x is weakly efficient.*

In order to prove convergence of Algorithm 3.3, we will use a standard technique on extensions of classical scalar optimization methods to the vector-valued case: the notion of quasi–Féjer convergence to a set. Let us recall that a sequence $\{x^{(k)}\} \subset \mathbb{R}^n$ is *quasi-Féjer convergent* to a nonempty set $T \subset \mathbb{R}^n$ if, for every $x \in T$ there exists a sequence $\{\varepsilon_k\} \subset \mathbb{R}$, with $\varepsilon_k \doteq \varepsilon_k(x) \geq 0$ for all k and $\sum_{k=0}^{\infty} \varepsilon_k < \infty$, such that

$$\|x^{(k+1)} - x\|^2 \leq \|x^{(k)} - x\|^2 + \varepsilon_k \quad \text{for all } k.$$

The following result is the main tool for the convergence proof.

Theorem 5.2. *[5, Theorem 1] Let $\{x^{(k)}\}$ be a quasi-Féjer convergent sequence to a nonempty set $T \subset \mathbb{R}^n$. Then, $\{x^{(k)}\}$ is bounded. If, in addition, an accumulation point x^* of $\{x^{(k)}\}$ belongs to T, then $\lim_{k\to\infty} x^{(k)} = x^*$.*

We begin the convergence analysis with a technical result, necessary for establishing quasi-Féjer convergence of any sequence $\{x^k\}$ produced by the inexact method to the set $T \doteq \{x \in C \colon F(x) \preceq F(x^k) \text{ for all } k\}$.

Lemma 5.3. *Suppose that F is K-convex and let v^k be an s-compatible direction at x^k, given by $v^k = P_{C-x^k}(-\hat{\beta} JF(x^k)^\top w^k)$, with $w^k \in \operatorname{conv}(G)$. If for a given $\hat{x} \in C$ we have $F(\hat{x}) \preceq F(x^k)$, then*

$$\|x^{k+1} - \hat{x}\|^2 \leq \|x^k - \hat{x}\|^2 + 2\hat{\beta} t_k \big| \langle w^k, JF(x^k) v^k \rangle \big|.$$

Proof. As $x^{k+1} = x^k + t_k v^k$, we have

$$\|x^{k+1} - \hat{x}\|^2 = \|x^k - \hat{x}\|^2 + t_k^2 \|v^k\|^2 - 2t_k \big\langle v^k, \hat{x} - x^k \big\rangle. \tag{20}$$

Let us analyze the rightmost term of the above expression. From the definition of v^k and the obtuse angle property of projections, we get

$$\left\langle -\hat{\beta} JF(x^k)^\top w^k - v^k, v - v^k \right\rangle \leq 0 \quad \text{for all } v \in C - x^k.$$

Taking $v = \hat{x} - x^k \in C - x^k$ on the above inequality, we obtain

$$-\left\langle v^k, \hat{x} - x^k \right\rangle \leq \hat{\beta} \left\langle w^k, JF(x^k)(\hat{x} - x^k) \right\rangle - \hat{\beta} \left\langle w^k, JF(x^k) v^k \right\rangle - \|v^k\|^2. \qquad (21)$$

Now, from (19),

$$\left\langle w^k, JF(x^k)(\hat{x} - x^k) \right\rangle \leq \left\langle w^k, F(\hat{x}) - F(x^k) \right\rangle \leq 0,$$

where the last inequality follows because $F(\hat{x}) \preceq F(x^k)$ and $w^k \in K^*$. Also, since $JF(x^k)v^k \prec 0$, we have $\left\langle w^k, JF(x^k)v^k \right\rangle < 0$. Thus we can rewrite (21) as

$$-\left\langle v^k, \hat{x} - x^k \right\rangle \leq \hat{\beta} \left| \langle w^k, JF(x^k) v^k \rangle \right| - \|v^k\|^2,$$

which, together with (20) and the fact that $t_k \in (0,1]$, gives us the desired result. □

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

Assumption 5.4. *Every K-decreasing sequence $\{z^k\} \subset F(C)$ is K-bounded from below by an element of $F(C)$, i.e., there exists $\hat{x} \in C$ such that $F(\hat{x}) \preceq z^k$ for all k, and any $\{z^k\}$ with $z^{k+1} \prec z^k$ for all k.*

Assumption 5.5. *The search direction v^k is s-compatible at x^k, that is to say, $v^k = P_{C-x^k}(-\hat{\beta} JF(x^k)^\top w^k)$, where $w^k \in \operatorname{conv}(G)$ for all k.*

As mentioned in [16], Assumption 5.4, known as *K-completeness* [23, Section 19], is standard for guaranteeing existence of efficient solutions for vector-valued optimization problems. In the scalar unconstrained case, it is clearly equivalent to the existence of an optimum. As for Assumption 5.5, it clearly deals with the sequence itself rather than with the objective function F. The search direction prescribed by Algorithm 3.3 is σ-approximate at x^k, and we are now asking it to be also s-compatible at the current iterate.

Let us now finally state and prove the main convergence result.

Theorem 5.6. *Assume that F is K-convex and that Assumptions 5.4 and 5.5 hold. Then, every sequence produced by the* inexact *projected gradient method converges to a weakly efficient solution.*

Proof. Let us consider the set

$$T = \{x \in C\colon F(x) \preceq F(x^k) \text{ for all } k\},$$

and take $\hat{x} \in T$, which exists by Assumption 5.4. Since K-convexity of F and Assumption 5.5 hold, it follows from Lemma 5.3 that

$$\|x^{k+1} - \hat{x}\|^2 \leq \|x^k - \hat{x}\|^2 + 2\hat{\beta} t_k |\langle w^k, JF(x^k)v^k\rangle| \quad \text{for all } k. \tag{22}$$

Let us prove first that the sequence $\{x^k\}$ is quasi-Féjer convergent to the set T. From Assumption 5.5 , we have that $v^k = P_{C-x^k}(-\hat{\beta} JF(x^k)^\top w^k)$, where $w^k \in \mathrm{conv}(G)$, for all k. Besides, as K is pointed, $\mathrm{int}(K^*) \neq \emptyset$ [28, Propositions 2.1.4, 2.1.7(i)] and therefore, the dual cone K^* contains a (Hammel) basis of $\mathbb{R}^m$, say $\{\tilde{w}^1, \ldots, \tilde{w}^m\}$. Without loss of generality, assume that $\{\tilde{w}^1, \ldots, \tilde{w}^m\} \subset \mathrm{conv}(G)$. Thus, for each k, there exist $\eta_i^k \in \mathbb{R}$, $i = 1, \ldots, m$, such that

$$w^k = \sum_{i=1}^m \eta_i^k \tilde{w}^i.$$

In view of the compactness of $\mathrm{conv}(G)$, all scalars η_i^k are uniformly bounded, which means that there exists $L > 0$, such that $|\eta_i^k| \leq L$ for all i and k. So, from (22) we get

$$\|x^{k+1} - \hat{x}\|^2 \leq \|x^k - \hat{x}\|^2 + 2\hat{\beta} t_k L \sum_{i=1}^m |\langle \tilde{w}^i, JF(x^k)v^k\rangle| \quad \text{for all } k.$$

Defining $\varepsilon_k \doteq 2\hat{\beta} t_k L \sum_{i=1}^m |\langle \tilde{w}^i, JF(x^k)v^k\rangle|$, we have that $\varepsilon_k \geq 0$ and

$$\|x^{k+1} - \hat{x}\|^2 \leq \|x^k - \hat{x}\|^2 + \varepsilon_k.$$

In view of Assumption 5.4, $\{F(x^k)\}$ is K-bounded, so from Lemma 3.6(iv), we have $\sum_{k=0}^\infty \varepsilon_k < \infty$. Since $\hat{x}$ is an arbitrary element of T, we see that $\{x^k\}$ converges quasi-Féjer to T.

Hence, by virtue of Theorem 5.2, it follows that $\{x^k\}$ is bounded. Therefore, $\{x^k\}$ has at least one accumulation point, which, by Theorem 3.5 is stationary. By virtue of Lemma 5.1, this point is also weakly efficient, because F is K-convex. Moreover, since C is closed and the sequence is feasible, this accumulation point belongs to C. Let $x^* \in C$ be one of these accumulation points and $\{x^{k_j}\}_j$ be a subsequence converging to x^*. Let $\bar{k}$ be a fixed but arbitrary nonnegative integer. For large enough j we have

$$F(x^{k_j}) \preceq F(x^{\bar{k}}).$$

As K is closed, letting $j \to \infty$, we obtain

$$F(x^*) \preceq F(x^{\bar{k}}).$$

Since $\bar{k}$ is arbitrary, $x^* \in T$. So, once again from Theorem 5.2, we conclude that $\{x^k\}$ converges to x^*, which is a weakly efficient solution. □

Clearly, whenever Assumptions 5.4 and 5.5 hold, the previous theorem encompasses the most important case of vector-valued problems: multicriteria, i.e., when $K = \mathbb{R}^m_+$. Actually, this theorem also shows the convergence of the inexact method in the (pointed) polyhedral case, i.e., whenever the ordering cone is pointed and has finitely many extreme rays.

6 Convergence analysis without pointedness

Observe that, up to now, we just explicitly used the pointedness of K at the beginning of the proof of Theorem 5.6, while showing that $\mathrm{int}(K^*) \neq \emptyset$. In this section we analyze the convergence of the inexact method for a non-antisymmetric partial order, i.e. for a nonpointed cone K. Recall that, a *partial order* on $\mathbb{R}^m$ is a reflexive and transitive binary relation on $\mathbb{R}^m$, which is consistent with the algebraic sum of vectors and the multiplication by nonnegative scalars. We know that every partial order is induced by a cone and, if this cone is pointed, the relation is also antisymmetric (see [19]). As we said, now we drop this assumption and study the convergence of the method for this weaker partial order, i.e., for a cone which contains nontrivial subspaces. As it is natural, in order to get similar results as those of last section with a weaker ordering structure, we will need supplementary assumptions. In our next theorem, we present a couple of alternative hypotheses.

Theorem 6.1. *Let K be nonpointed and F K-convex. If Assumptions 5.4 and 5.5 hold, as well as one of the following conditions:*

(i) there exist $\hat{w}^1, \dots, \hat{w}^r \in \mathrm{conv}(G)$ such that

$$\{w^k\} \subset \mathrm{conv}\{\hat{w}^1, \dots, \hat{w}^r\} \quad \textit{with } r \in \mathbb{N},$$

(ii) there exist $\rho > 0$ and $k_0 \in \mathbb{N}$ such that

$$\varphi\big(JF(x^k)w^k\big) \leq \rho \big\langle w^k, JF(x^k)v^k \big\rangle \quad \textit{for all } k \geq k_0,$$

then, every sequence produced by the inexact projected gradient method converges to a weakly efficient solution.

Proof. Consider the set $T = \{x \in C \colon F(x) \preceq F(x^k) \text{ for } k = 0, 1, \dots\}$ and take $\hat{x} \in T$, which exists by Assumption 5.4. So, using Assumption 5.5 and Lemma 5.3, it follows that

$$\|x^{k+1} - \hat{x}\|^2 \leq \|x^k - \hat{x}\|^2 + 2\hat{\beta} t_k \big|\langle w^k, JF(x^k)v^k \rangle\big| \quad \text{for all } k. \tag{23}$$

Now we will see that, under hypotheses (i) or (ii), the sequence $\{x^k\}$ is quasi-Féjer convergent to the set T.

Let us first suppose that hypothesis (i) holds. By Assumption 5.5, $w^k \in \operatorname{conv}(G)$ and then there exist $\alpha_j^{(k)} \geq 0$, for $j = 1, \ldots, r$, with $\sum_{j=1}^r \alpha_j^{(k)} = 1$ such that $w^k = \sum_{j=1}^r \alpha_j^{(k)} \hat{w}^j$ for all k. Whence, from (23) and the triangle inequality, it follows that

$$\|x^{k+1} - \hat{x}\|^2 \leq \|x^k - \hat{x}\|^2 + \sum_{j=1}^r 2\hat{\beta} t_k \big|\langle \hat{w}^j, JF(x^k)v^k \rangle\big| \quad \text{for all } k.$$

Now observe that, applying Lemma 3.6(iv) to $w = \hat{w}^j$, with $j = 1, \ldots, r$, we see that $\sum_{k=0}^\infty \varepsilon_k < \infty$, where $\varepsilon_k \doteq \sum_{j=1}^r 2\hat{\beta} t_k |\langle \hat{w}^j, JF(x^k)v^k \rangle|$. Since $\hat{x}$ was an arbitrary element of the set T, the sequence $\{x^k\}$ converges quasi-Féjer to T.

Now let us assume that hypothesis (ii) holds. Since we are assuming that $\{x^k\}$ is a sequence of nonstationary points, we have that $JF(x^k)v^k \prec 0$, which combined with the fact that $w^k \in \operatorname{conv}(G) \subset K^*$, gives us $\langle w^k, JF(x^k)v^k \rangle < 0$. So condition (ii) can be rewritten as

$$|\langle w^k, JF(x^k)v^k \rangle| \leq \rho^{-1} |\varphi(JF(x^k)v^k)|.$$

Whence, $\sum_{k=0}^\infty 2\hat{\beta} t_k \big|\langle w^k, JF(x^k)v^k \rangle| < \infty$, by virtue of Lemma 3.6(iii). Therefore, from (23), we conclude that $\{x^k\}$ converges quasi-Féjer to T.

Up to now, we have seen that, under conditions (i) or (ii), the sequence $\{x^k\}$ is quasi-Féjer convergent to the set T. Hence, in both cases, (i) or (ii), proceeding as in the end of the proof of Theorem 5.6, we see that $\{x^k\}$ converges to a weakly efficient solution. $\square$

Note that for the above theorem other hypotheses, similar to (ii), can be derived from Lemma 3.6. For instance, one could ask, instead of (ii), that for some $\rho > 0$ and $k_0 \in \mathbb{N}$, $-\|v^k\|^2 \leq \rho \langle w^k, JF(x^k)v^k \rangle$ for all $k \geq k_o$.

As we will now see, Theorem 6.1 encompasses the polyhedral nonpointed case, i.e., it shows us that, among other cases, we also have convergence for any nonpointed ordering cone with finitely many extreme rays.

Corollary 6.2. *Let K be nonpointed. Under polyhedrality of K, K-convexity of F, Assumptions 5.4 and 5.5, we have that any sequence generated by the inexact method converges to a weakly efficient solution.*

Proof. If K is polyhedral, then K^* is also polyhedral. Then, there exists $r \in \mathbb{N}$ such that $\hat{w}^1, \ldots, \hat{w}^r \in K^* \setminus \{0\}$ and $G = \{\hat{w}^1, \ldots, \hat{w}^r\}$. Since, by Assumption 5.5, $w^k \in \operatorname{conv}(G)$ for all k, condition (i) of Theorem 6.1 holds and so the result follows. $\square$

Now we show that for the exact gradient projected method we do not need to ask all assumptions of Theorem 6.1 in order to have convergence, since, as we saw in (16), the exact direction $v(x^k)$ is always s-compatible, so it automatically satisfies Assumption 5.5. Under pointedness of K, a similar convergence result was established in [12, Theorem 5.6].

Corollary 6.3. *Let K be nonpointed. Under K-convexity of F and just Assumption 5.4, all sequences produced by the* exact *method converge to a weakly efficient solution.*

Proof. We know that $v(x^k)$, the exact projected gradient direction, is s-compatible at x^k for all k, that is to say, Assumption 5.5 is satisfied. Furthermore, $v(x^k)$ verifies condition (17) for all k. Moreover, since $\langle w^k, JF(x^k)v(x^k)\rangle < 0$ for all k, (17) implies condition (ii) of Theorem 6.1. Hence, the conclusion follows from that theorem. □

7 Final remarks

Let us make some final comments on issues concerning the implementation of the method. From a practical point of view, the user should try to reach condition (17). Indeed, this condition, as shown in Proposition 4.3, guarantees not only that the s-compatible direction $v^k = P_C(-\hat{\beta} JF(x^k)w^k)$ is σ-approximate but, whenever K is not pointed, according to Theorem 6.1, it also ensures the convergence of the corresponding sequence, because, as we have already seen, (17) implies hypotheses (ii) of that theorem. Now, since the exact projected gradient direction $v(x^k)$ satisfies condition (17) (with strict inequality, for $0 < \sigma$), it seems reasonable to try to approximately solve (4). It is worth to point out that, as usual with this kind of extensions, our goal is not to obtain the whole set of optimal points; we are just concerned with finding a single optimum. Nevertheless, from a numerical point of view, we can expect to somehow approximate the solution set by just performing our method for different initial guesses.

Observe that, as was noted in [12], if $\hat{K}$ is a closed convex pointed cone with $K \setminus \{0\} \subset \mathrm{int}(\hat{K})$, such that Assumption 5.4 holds and x^* is $\hat{K}$-weakly efficient, then x^* is *efficient* for the original partial order, that is, there does not exist $x \in C$ such that $F(x) \preceq F(x^*)$ and $F(x) \neq F(x^*)$. Some practical considerations on how to obtain such $\hat{K}$ are given in that work. Without considering new cones, as was also mentioned in [12] for the exact method, we can guarantee convergence of all sequences produced by the inexact method to efficient points whenever the objective function F is *strictly K-convex*, i.e., when it satisfies (18) with strict K-inequality "$\prec$" (see [15, Proposition 2.2]).

Acknowledgements. This research was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (grant no. 2010/20572-0), Fundação de Amparo à Pesquisa do Estado de Rio de Janeiro (grant Cientistas do Nosso Estado 2008) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (Universal grant no. 480101/2008-6).

References

[1] Aliprantis, C.D., Florenzano, M., Martins da Rocha, V.F., Tourky, R.: Equilibrium analysis in financial markets with countably many securities. Journal of Mathematical Economics **40**(6), 683–699 (2004)

[2] Aliprantis, C.D., Florenzano, M., Tourky, R.: General equilibrium analysis in ordered topological vector spaces. Journal of Mathematical Economics **40**(3-4), 247–269 (2004)

[3] Bagchi, U.: Simultaneous minimization of mean and variation of flow time and waiting time in single machine systems. Operations Research **37**, 118–125 (1989)

[4] Bertsekas, D.P.: Convex Analysis and Optimization. Athena Scientific, Belmont, MA (2003)

[5] 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)

[6] Carrizosa, E., Frenk, J.B.G.: Dominating sets for convex functions with some applications. Journal of Optimization Theory and Applications **96**, 281–295 (1998)

[7] Eschenauer, H., Koski, J., Osyczka, A.: Multicriteria Design Optimization. Springer, Berlin (1990)

[8] Fliege, J.: OLAF – A general modeling system to evaluate and optimize the location of an air polluting facility. OR Spektrum **23**, 117–136 (2001)

[9] Fliege, J., Graña Drummond, L.M., Svaiter, B.F.: Newton's method for multiobjective optimization. SIAM Journal on Optimization **20**(2), 602–626 (2009)

[10] Fliege, J., Svaiter, B.F.: Steepest descent methods for multicriteria optimization. Mathematical Methods of Operations Research **51**(3), 479–494 (2000)

[11] Fu, Y., Diwekar, U.: An efficient sampling approach to multiobjective optimization. Mathematical Methods of Operations Research **132**, 109–134 (2004)

[12] Fukuda, E.H., Graña Drummond, L.M.: On the convergence of the projected gradient method for vector optimization. Optimization (2011). Accepted for publication

[13] Graña Drummond, L.M., Iusem, A.N.: A projected gradient method for vector optimization problems. Computational Optimization and Applications **28**(1), 5–29 (2004)

[14] Graña Drummond, L.M., Maculan, N., Svaiter, B.F.: On the choice of parameters for the weighting method in vector optimization. Mathematical Programming **111**(1-2) (2008)

[15] Graña Drummond, L.M., Raupp, F.M.P., Svaiter, B.F.: A quadratically convergent Newton method for vector optimization (2011). Submitted for publication

[16] Graña Drummond, L.M., Svaiter, B.F.: A steepest descent method for vector optimization. Journal of Computational and Applied Mathematics **175**(2), 395–414 (2005)

[17] Jahn, J.: Scalarization in vector optimization. Mathematical Programming **29**, 203–218 (1984)

[18] Jahn, J.: Mathematical vector optimization in partially ordered linear spaces. Verlag Peter D. Lang, Frankfurt (1986)

[19] Jahn, J.: Vector Optimization – Theory, Applications and Extensions. Springer, Erlangen (2003)

[20] Jüschke1, A., Jahn, J., Kirsch, A.: A bicriterial optimization problem of antenna design. Computational Optimization and Applications **7**(3), 261–Ű276 (1997)

[21] Leschine, T., Wallenius, H., Verdini, W.: Interactive multiobjective analysis and assimilative capacity-based ocean disposal decisions. European Journal of Operational Research **56**, 278–289 (1992)

[22] Luc, D.T.: Scalarization of vector optimization problems. Journal of Optimization Theory and Applications **55**(1), 85–102 (1987)

[23] Luc, D.T.: Theory of vector optimization. In: Lecture Notes in Economics and Mathematical Systems, 319. Springer, Berlin (1989)

[24] Miettinen, K.M.: Nonlinear Multiobjective Optimization. Kluwer Academic Publishers, Norwell (1999)

[25] Palermo, G., Silvano, C., Valsecchi, S., Zaccaria, V.: A system-level methodology for fast multi-objective design space exploration. ACM, New York (2003)

[26] Prabuddha, D., Ghosh, J., Wells, C.: On the minimization of completion time variance with a bicriteria extension. Operations Research **40**, 1148–1155 (1992)

[27] Rockafellar, R.T.: Convex Analysis. Princeton University Press, New Jersey (1970)

[28] Sawaragi, Y., Nakayama, H., Tanino, T.: Theory of Multiobjective Optimization. Academic Press, Orlando (1985)

[29] Tamiz, M., Jones, D.F., El Darzi, E.: A review of goal programming and its applications. Annals of Operations Research **58**, 39–53 (1993)

[30] Tavana, M.: A subjective assessment of alternative mission architectures for the human exploration of mars at NASA using multicriteria decision making. Computational and Operations Research **31**, 1147–1164 (2004)

[31] White, D.J.: Epsilon-dominating solutions in mean-variance portfolio analysis. European Journal on Operations Research **105**, 457–466 (1998)