On the convergence of the projected gradient method for vector optimization

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

September 30, 2009

Abstract

In 2004, Graña Drummond and Iusem proposed an extension of the projected gradient method for constrained vector optimization problems. In that method, an Armijo-like rule, implemented with a backtracking procedure, was used in order to determine the steplengths. The authors just showed stationarity of all cluster points and, for another version of the algorithm (with exogenous steplengths), under some additional assumptions, they proved convergence to weakly efficient solutions. In this work, first we correct a slight mistake in the proof of a certain continuity result on that 2004 article, and then we extend its convergence analysis. Indeed, under some reasonable hypotheses, for convex objective functions with respect to the ordering cone, we establish full convergence to optimal points of any sequence produced by the projected gradient method with an Armijo-like rule, no matter how poor the initial guesses may be.

Keywords: vector optimization; weak efficiency; projected gradient method; convexity with respect to cones; quasi-Féjer convergence

AMS Subject Classification: 90C29; 90C30

*Instituto de Matemática e Estatística, Universidade de São Paulo (IME-USP). Rua do Matão 1010, Cidade Universitária, São Paulo, SP, 05508-090, Brazil (ellen@ime.usp.br).

†Faculdade de Administração e Ciências Contábeis, Universidade Federal do Rio de Janeiro (FACC-UFRJ). Av. Pasteur 250, Rio de Janeiro, RJ, 22290-240, Brazil (lmgd@facc.ufrj.br).

1 Introduction

Finding efficient points for the preference order induced by the Paretian cone $\mathbb{R}^m_+$ is, of course, a very relevant problem on many areas, such as engineering [8], statistics [7], environmental analysis [10, 20], space exploration [24, 27], management science [17, 25, 28] and design [12]. For orders induced by other cones, the problem of finding efficient points (weakly or not) is certainly not as frequent as the one concerning the point-wise partial order, but, nevertheless, it is not just an extension of the Paretian case and has its own importance. Let us just mention, for instance, that finding weakly efficient solutions for portfolio selection in security markets vector-valued problems, with underlaying cones different from the nonnegative orthant, have been studied in [1] and [2]. Thus, the matter of solving vector optimization problems is also relevant and deserves our attention.

Among the most popular strategies for solving vector optimization problems, we mention the scalarization approach, where one or more real-valued optimization problems are solved, in such a way that all optima of the scalar problems are solutions of the vector-valued one [19, 21, 23]. The so-called weighting method is one of the most widely used scalarization techniques. Basically, one minimizes a linear combination of the objectives, where the vector of "weights" is not known a priori and, so, this procedure may lead to unbounded numerical problems, which, therefore, may lack minimizers [14].

In this work we follow a research line which consists of extending classical (i.e., scalar-valued) optimization methods to the vector-valued setting. Up to now, successful versions of the steepest descent method, the projected gradient method (with exogenous chosen steplengths) and the Newton method have been proposed in [9, 16], [13] and [11, 15], respectively. These methods do not scalarize the vector-valued problem and work on the image space, providing adequate search directions with respect to the preference order. Other type of algorithms which do not scalarize have been recently developed for multicriteria optimization [4, 5].

Here we improve the results on the convergence of the projected gradient method, implemented with an Armijo-like rule, proposed by Graña Drummond and Iusem in [13]. At each iteration of the method, while a certain first-order optimality condition is not satisfied, a search direction is computed in the image space by means of an auxiliary nonsmooth problem (the minimization of the max ordering scalarization). In [13], the authors just prove stationarity of accumulation points of the sequences produced by the method, while here we establish full convergence of all sequences generated by the method. Under standard hypotheses, for convex objective functions with respect to the ordering cone, we show convergence to weakly efficient solutions for arbitrary initial points.

We point out that many of the results that we will need were already proved on [13], so we will just state them and give proper references for their proofs. The outline of this article is as follows. In Section 2 we present the problem and a necessary condition for optimality. In Section 3 we expose the main ideas of the projected gradient method for vector optimization and we present some results which will be used in our work. We also correct the proof of [13, Proposition 4]. In Section 4 we introduce the algorithm and show a couple of results of [13] which will be used on the sequel. Finally, in Section 5, we state and prove our main result: assuming convexity of the objective function with respect to the ordering cone, and under some reasonable assumptions, we show that, for any starting point, all sequences produced by the projected gradient method converge to weakly efficient solutions.

2 The problem

Let $K \subset \mathbb{R}^m$ be a closed and convex cone, with nonempty interior and such that K is pointed, that is to say $K \cap (-K) = \emptyset$ or, equivalently, K does not contain straight lines. Consider the partial order induced by K in $\mathbb{R}^m$: $u \preceq v$ (alternatively, $v \succeq u$) if $v - u \in K$, as well as the following stronger relation: $u \prec v$ (alternatively, $v \succ u$) if $v - u \in \operatorname{int}(K)$, where, as usual, $\operatorname{int}(K)$ stands for the topological interior of 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. We are interested in the following problem:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in C, \end{array}$$

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

A necessary, but in general non-sufficient, condition for optimality of a point $\bar{x} \in C$ is *stationarity* (or *criticality*):

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

where $Jf(\bar{x})$ stands for the Jacobian matrix of f at $\bar{x}$. So $\bar{x} \in C$ is *stationary* for f if, and only if, for all $v \in C - \bar{x}$ we have $Jf(\bar{x})v \not\prec 0$. Note that for $m = 1$ we retrieve the classical stationarity condition for constrained scalar-valued optimization:

$$\langle \nabla f(\bar{x}), x - \bar{x} \rangle \geq 0 \quad \text{for all } x \in C,$$

where $\langle \cdot, \cdot \rangle$ is the canonical inner product.

Clearly, a point $x \in C$ is nonstationary if there exists a direction $v \in C - x$ such that $Jf(x)v \prec 0$. The next result shows not only that such v is a K-descent direction, but it also estimates the K-decrease on a nontrivial segment beginning at x with direction v. As we will see, this result allows us to implement an Armijo-like rule on the algorithm.

Proposition 2.1. *([13, Proposition 1]) Let $\sigma \in (0, 1)$, $x \in C$ and $v \in C - x$ such that $Jf(x)v \prec 0$. Then, there exists $\bar{t} > 0$ such that $f(x + tv) \prec f(x) + \sigma t Jf(x)v$ for all $t \in (0, \bar{t}]$.*

3 Preliminaries

Let us introduce some necessary notations and recall a couple of results. We will denote K^* as the *positive polar* (or *dual*) cone of K, that is,

$$K^* := \big\{y \in \mathbb{R}^m \colon \langle z, y\rangle \geq 0 \text{ for all } z \in K\big\}.$$

Note that, since K is a nonempty closed convex cone, we have $K^{**} = K$ [26, Theorem 14.1].

Let $G \subset K^*$ be a compact set with $0 \notin G$ and such that the cone generated by its convex hull is K^*, i.e., $\operatorname{cone}(\operatorname{conv}(G)) = K^*$. Observe that G always exists; for instance, we can take $G := \{w \in K^* \colon \|w\| = 1\}$, where $\|\cdot\|$ denotes Euclidean norm. In general, we can consider much smaller sets: in the multiobjective case, as $(\mathbb{R}^m)^* = \mathbb{R}^m$, G can be taken as the canonical basis of $\mathbb{R}^m$; in the case K is a polyhedral cone, since K^* is also polyhedral, G can be chosen as the (finite) set of its extreme rays.

For a given such set G, we will denote by $\varphi\colon \mathbb{R}^m \to \mathbb{R}$ its *support function*, which is defined as:

$$\varphi(y) := \max_{w \in G} \langle w, y\rangle. \tag{1}$$

Since G spans all the nontrivial directions of the cone K^* and $K^{**} = K$, we have

$$-K = \{y \in \mathbb{R}^m \colon \varphi(y) \leq 0\}, \tag{2}$$

$$-\operatorname{int}(K) = \{y \in \mathbb{R}^m \colon \varphi(y) < 0\}, \tag{3}$$

that is, φ gives us a scalar representation of $-K$ and its interior $-\operatorname{int}(K)$. Note now that if $\bar{x} \in C$ is stationary, then, from (3), it follows that

$$\varphi(Jf(\bar{x})v) \geq 0 \quad \text{for all } v \in C - \bar{x}. \tag{4}$$

The following properties of the support function φ will be used on the sequel.

Proposition 3.1. *([13, Proposition 2]) Let $\varphi\colon \mathbb{R}^m \to \mathbb{R}$ be the support function of G. Then,*

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

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

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

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

Now we begin to describe the projected gradient method for vector optimization, proposed in [13]. For a given $x \in C$, the *projected gradient direction* for f at x is given by $v\colon C \to \mathbb{R}$,

$$v(x) := \operatorname*{argmin}_{v \in C - x} h_x(v), \tag{5}$$

where $h_x(v) := \hat{\beta}\varphi(Jf(x)v) + \|v\|^2/2$, with $\hat{\beta} > 0$. Observe that the objective function h_x depends not just upon x, but also on $\hat{\beta}$. Also note that, in view of the strong convexity of h_x, the projected gradient direction is well defined and is the optimal solution of a nonsmooth scalar-valued optimization problem. Nevertheless, in some important cases, as the multicriteria problems ($K = \mathbb{R}^m_+$), we can replace this problem by the minimization of a single variable subject to finitely many inequalities plus the original constraint (see [11, problem (3.8)]). This can also be done in the general polyhedral case, since K^* inherits polyhedrality from K.

It is interesting to observe that we are in fact extending the real-valued projected gradient direction. Indeed, for $m = 1$, since $K = [0, +\infty)$ and $G = \{1\}$, we retrieve the following identities:

$$v(x) = P_{C-x}\big(-\hat{\beta}\nabla f(x)\big) = P_C\big(x - \hat{\beta}\nabla f(x)\big) - x,$$

As we will see, the minimizer of (5) and its (optimal) functional value somehow measure proximity to stationarity; in fact, they are used in the stopping criterion of the algorithm. Define $\theta \colon C \to \mathbb{R}$ by

$$\theta(x) := \min_{v \in C-x} h_x(v) = h_x(v(x)) = \hat{\beta}\varphi\big(Jf(x)v(x)\big) + \frac{\|v(x)\|^2}{2}.$$

Observe that, since $0 \in C - x$, we have

$$\theta(x) \leq 0 \quad \text{for all } x \in C. \tag{6}$$

Let us now give a characterization of stationarity in terms of $\theta(\cdot)$ and $v(\cdot)$.

Proposition 3.2. *([13, Proposition 3]) For $x \in C$, the following conditions are equivalent:*

i) *The point x is not stationary.*

ii) $\theta(x) < 0$.

iii) $v(x) \neq 0$.

In particular, $x \in C$ is stationary if and only if $\theta(x) = 0$.

Note that for a nonstationary point x, from the above characterization, we see that $\theta(x) < 0$ and so, from the definition of θ, we get $\varphi(Jf(x)v(x)) < 0$, which, from (3), in turn implies that $Jf(x)v(x) \prec 0$. Whence, Proposition 2.1 gives us

$$f(x + tv(x)) \prec f(x) + \sigma\, tJf(x)v(x), \tag{7}$$

for all $t \in (0, \bar{t}]$ and, in particular, $v(x)$ is a descent direction, i.e., $f(x + tv(x)) \prec f(x)$ in that nontrivial interval.

The following result gives a bound for the projected gradient direction at a point x in terms of $Jf(x)$ and G.

Lemma 3.3. *For all $x \in C$, it holds that $\|v(x)\| \leq 2\hat{\beta}M\|Jf(x)\|$, where $M = \max_{w\in G}\|w\|$.*

Proof. As $\theta(x) \leq 0$ for all $x \in C$, we have, for any $w \in G$,

$$\frac{\|v(x)\|^2}{2} \leq -\hat{\beta}\varphi(Jf(x)v(x)) \leq -\hat{\beta}\langle w, Jf(x)v(x)\rangle,$$

where the last inequality follows from (1). Using the Cauchy-Schwartz inequality, we conclude that $\|v(x)\| \leq 2\hat{\beta}M\|Jf(x)\|$ for all $x \in C$. □

Now we show that θ is a continuous function. We point out that this fact was already stated in [13, Proposition 4], but the proof given there has a slight mistake, which we correct here. Actually, that proof works just for unconstrained vector-valued problems, while ours is valid for the general constrained case.

Proposition 3.4. *The function $\theta\colon C \to \mathbb{R}$ is continuous.*

Proof. Take $x \in C$ and let $\{x^{(k)}\} \subset C$ be a sequence such that $\lim_{k\to\infty} x^{(k)} = x$. Since $v(x) \in C-x$, we have $v(x)+x-x^{(k)} \in C-x^{(k)}$, and so, from (5) and the definition of θ, we obtain $\theta(x^{(k)}) \leq h_{x^{(k)}}(v(x) + x - x^{(k)})$ for all k. Hence, from Proposition 3.1 (ii), we get

$$\begin{aligned}\theta(x^{(k)}) \quad \leq \quad & \hat{\beta}\varphi\big(Jf(x^{(k)})v(x)\big) + \hat{\beta}\varphi\big(Jf(x^{(k)})(x - x^{(k)})\big) \\ & +\frac{||v(x)||^2}{2} + \frac{||x - x^{(k)}||^2}{2} + \langle v(x), x - x^{(k)}\rangle.\end{aligned}$$

Since f is continuously differentiable and, as established in Proposition 3.1 (iv), φ is continuous, taking $\limsup_{k\to\infty}$ on both sides of the above inequality yields

$$\limsup_{k\to\infty} \theta(x^{(k)}) \leq \theta(x)\,. \tag{8}$$

On the other hand, using now the optimality of $v(x)$, the fact that $v(x^{(k)})+x^{(k)}-x \in C - x$ and the definition of θ, we have $\theta(x) \leq h_x(v(x^{(k)}) + x^{(k)} - x)$, and thus, again from Proposition 3.1 (ii),

$$\begin{aligned}\theta(x) \quad \leq \quad & \hat{\beta}\varphi\big(Jf(x)v(x^{(k)})\big) + \hat{\beta}\varphi\big(Jf(x)(x^{(k)} - x)\big) \\ & +\frac{||v(x^{(k)})||^2}{2} + \frac{||x^{(k)} - x||^2}{2} + \langle v(x^{(k)}), x^{(k)} - x\rangle.\end{aligned}$$

So, taking $\liminf_{k\to\infty}$ on both sides of the above inequality we have:

$$\begin{aligned}\theta(x) \quad & \leq \quad \liminf_{k\to\infty} \hat{\beta}\varphi\big(Jf(x)v(x^{(k)})\big) + \|v(x^{(k)})\|/2 \\ & = \quad \liminf_{k\to\infty} \theta(x^{(k)}) + \hat{\beta}\Big[\varphi\big(Jf(x)v(x^{(k)})\big) - \varphi\big(Jf(x^{(k)})v(x^k)\big)\Big] \\ & \leq \quad \liminf_{k\to\infty} \theta(x^{(k)}) + \hat{\beta}M\|Jf(x) - Jf(x^{(k)})\|\|v(x^{(k)})\|,\end{aligned}$$

where the second inequality holds in view of Proposition 3.1 (iv) and $M = \max_{w\in G} \|w\|$. Now, since f is continuously differentiable and $\{x^{(k)}\}$ is convergent, we conclude from Lemma 3.3 that $\{v(x^k)\}$ is bounded. Then, since $x^{(k)} \to x$, using once again that f is continuously differentiable, we get

$$\theta(x) \leq \liminf_{k\to\infty} \theta(x^{(k)}),$$

which, together with (8), completes the proof. □

4 The algorithm

Let us now summarize the projected gradient method for vector optimization, proposed in [13], where $G := \{y \in K^*: \|y\| = 1\}$.

Algorithm 4.1. *The exact projected gradient method for vector optimization.*

1. Choose $x^0 \in C$ and the parameters $\hat{\beta} > 0$, $\nu > 1$, $\sigma \in (0, 1)$. Set $k = 0$.
2. Compute $v(x^k)$, the projected gradient direction for f at $x = x^k$, as in (5).
3. If $\theta(x^k) = 0$, then stop.
4. Choose t_k as the largest $t \in \{\nu^{-j}: j = 0, 1, 2, \dots\}$ such that
$$f\big(x^k + tv(x^k)\big) \preceq f(x^k) + \sigma\, t Jf(x^k)v(x^k).$$
5. Set $x^{k+1} = x^k + t_k v(x^k)$, $k = k + 1$ and go to step 2.

Some comments are in order. First of all, let us mention that, in view of Proposition 3.2 and (7), step 4 is well defined and so is the whole method. Second, note that, for each choice of the parameters in step 1, the method generates an infinite sequence of nonstationary points or it ends up with a stationary point. Third, observe that, in view of (7), $\{f(x^k)\}$ is a K-decreasing sequence, i.e., $f(x^{k+1}) \prec f(x^k)$ for all k. Finally, as we will state in the next proposition, the generated sequence is feasible.

Proposition 4.2. *([13, Proposition 5]) If $\{x^k\}$ is a sequence generated by Algorithm 4.1, then $x^k \in C$ for all k.*

If $x^* \in C$ is an accumulation point of $\{x^k\}$, by ([13, Theorem 1]), we have that x^* is stationary, therefore, from Proposition 3.2, we see that $\theta(x^*) = 0$. So, since θ is continuous, if $x^k \to x^*$, necessarily $\theta(x^k) \to 0$. In the next lemma we will see how fast does $t_k|\theta(x^k)|$ converge to zero. We will need the following estimate. Recall that from the Armijo condition and the properties of the support function φ (Proposition 3.1), we have

$$\begin{aligned} \varphi\big(f(x^{k+1})\big) &\leq \varphi\big(f(x^k)\big) + \sigma t_k \varphi\big(Jf(x^k)v(x^k)\big) \\ &\leq \varphi\big(f(x^k)\big) + \frac{\sigma t_k}{\hat{\beta}}\left[\theta(x^k) - \frac{\|v(x^k)\|^2}{2}\right]. \end{aligned} \tag{9}$$

Before the announced lemma, we need to introduce the notion of K-boundedness, which will be needed on our convergence analysis. Given any sequence $\{x^{(k)}\} \subset C$, we say that $\{f(x^{(k)})\} \subset \mathbb{R}^m$ is *K-bounded* (from below) if there exists $\bar{f} \in \mathbb{R}^m$ such that $\bar{f} \preceq f(x^{(k)})$ for all k. The following result will be essential for showing convergence of any sequence $\{x^k\}$ generated by Algorithm 4.1.

Lemma 4.3. *If the sequence $\{f(x^k)\}$ is K-bounded from below, then*

$$\sum_{k=0}^{\infty} t_k |\theta(x^k)| < \infty.$$

Proof. Adding up, from $k = 0$ to $k = N$ at (9), we get

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

where the equality holds by virtue of (6). Since $\{f(x^k)\}$ is K-bounded (say by $\bar{f}$), $\sigma \in (0, 1)$, from Proposition 3.1 (iii), we obtain

$$\sum_{k=0}^{N} t_k |\theta(x^k)| \leq \frac{\hat{\beta}}{\sigma} \left[\varphi\big(f(x^0)\big) - \varphi(\bar{f})\right] \quad \text{for all } N = 0, 1, \ldots,$$

and the result follows immediately. □

5 Convergence of the method

The main result on convergence of the projected gradient method presented in [13] is the following.

Theorem 5.1. *([13, Theorem 1]) Every accumulation point, if any, of the sequence $\{x^k\}$ generated by the projected gradient method is a (feasible) stationary point.*

We now begin to show that every sequence produced by the method converges to a weakly efficient point, no matter how poor is the initial point. From now on, we will assume that the method does not stop, i.e., that it generates an infinite sequence of iterates. We will also assume that our objective function f is *K-convex* on C, i.e.,

$$f(\lambda x + (1-\lambda)u) \preceq \lambda f(x) + (1-\lambda) f(u), \quad \text{for all } x, u \in C \text{ and } \lambda \in [0, 1]. \tag{10}$$

Observe that, for $w \in K^*$, we have that $x \mapsto \langle w, f(x)\rangle$ is a smooth convex scalar function and so it overestimates its linear approximation. Therefore, since $\nabla_x \langle w, f(x)\rangle = Jf(x)^\top w$, we have

$$\langle w, f(z)\rangle + \langle w, Jf(z)(u - z)\rangle \leq \langle w, f(u)\rangle \quad \text{for all } z, u \in C. \tag{11}$$

Next we will see that, as in the scalar case, under convexity, stationarity is also a sufficient condition for optimality.

Lemma 5.2. *Assume that $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is K-convex. Then $\bar{x} \in \mathbb{R}^n$ is stationary if and only if $\bar{x}$ is weakly efficient.*

Proof. Assume that $\bar{x}$ is stationary. Using (4) with $v = x - \bar{x} \in C - \bar{x}$, we have $\varphi(Jf(\bar{x})(x - \bar{x})) \geq 0$. Therefore, taking $\bar{w} \in G$ such that $\langle \bar{w}, Jf(\bar{x})(x - \bar{x}) \rangle = \varphi(Jf(\bar{x})(x - \bar{x}))$ and applying (11) with $w = \bar{w}, z = \bar{x}$ and $u = x$, we get

$$\langle \bar{w}, f(\bar{x}) \rangle \leq \langle \bar{w}, f(x) \rangle \quad \text{for all } x \in C.$$

If $\bar{x}$ is not weakly efficient, there exists $x \in C$ such that $f(x) \prec f(\bar{x})$. But $\bar{w} \in G \subset K^*$, so $\langle \bar{w}, f(x) \rangle < \langle \bar{w}, f(\bar{x}) \rangle$, in contradiction with the above inequality.

The converse is immediate, since, as we have already seen, for any nonstationary point, there exists a descent direction (Proposition 2.1). □

Supposing that the K-decreasing sequence $\{f(x^k)\}$ is K-bounded from below on $f(C)$, we can prove the following technical result, which is an immediate consequence of (11).

Lemma 5.3. *Assume that f is K-convex and there exists $\hat{x} \in C$ such that $f(\hat{x}) \preceq f(x^k)$ for all k. Then*

$$\left\langle Jf(x^k)^\top w, \hat{x} - x^k \right\rangle \leq 0 \quad \textit{for all } w \in K^*. \tag{12}$$

Proof. Let $w \in K^*$. From (11) with $u = \hat{x}$ and $z = x^k$, we have

$$\langle w, f(x^k) \rangle + \langle w, Jf(x^k)(\hat{x} - x^k) \rangle \leq \langle w, f(\hat{x}) \rangle. \tag{13}$$

On the other hand, from $f(\hat{x}) \preceq f(x^k)$, we get

$$\langle w, f(\hat{x}) \rangle \leq \langle w, f(x^k) \rangle. \tag{14}$$

So, combining (13) with (14), we have $\langle Jf(x^k)^\top w, \hat{x} - x^k \rangle \leq 0$, and the proof is complete. □

In order to prove convergence of the sequences produced by Algorithm 4.1, 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 and its consequences. We 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 $\varepsilon_k := \varepsilon_k(x) \geq 0$ for all k and $\sum_{k=0}^{\infty} \varepsilon_k < \infty$, such that

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

The following result is the main tool on our convergence proof.

Theorem 5.4. *([6, 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 $\lim_{k\to\infty} x^{(k)} = x^*$.*

Now observe that $v(x)$ is the optimum of $\min_{v\in C-x} h_x(v)$, where $h_x\colon \mathbb{R}^m \to \mathbb{R}$, given by $h_x(v) = \hat{\beta}\max_{w\in G}\langle w, Jf(x)v\rangle + \|v\|^2/2$, is a convex function and $C-x$ is a convex subset of $\mathbb{R}^m$. Hence, as a consequence of [3, Proposition 4.7.2], there exists ξ in the subdifferential of h_x at $v(x)$, such that

$$\langle \xi, v - v(x)\rangle \geq 0 \quad \text{for all } v \in C - x. \tag{15}$$

Let us also recall that, for the max-type function h_x, the subdifferential at its minimizer $v(x)$ is given by the convex hull of the active functions gradients at $v(x)$ (see [18, Theorem D.4.4.2]). More precisely, the subdifferential at $v(x)$ is given by

$$\operatorname{conv}\big(\big\{\nabla_v\langle w, Jf(x)v\rangle\colon \langle w, Jf(x)v(x)\rangle = \varphi(Jf(x)v(x))\big\}\big) + v(x). \tag{16}$$

In order to prove convergence of the sequences produced by Algorithm 4.1, we will also assume the following condition, which is standard in the analysis of classical methods extensions to vector optimization [9, 16, 15].

Assumption 5.5. *For all $\{y^k\} \subset f(C)$ K-decreasing, there exists $x \in C$ such that $f(x) \preceq y^k$, $k = 1, 2, \dots$*

Finally, we state and proof our main result.

Theorem 5.6. *Assume that f is K-convex and that Assumption 5.5 holds. Then, every sequence produced by the projected gradient method converges to a weakly efficient point.*

Proof. 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 5.5. We have the following equality:

$$\|x^{k+1} - \hat{x}\|^2 = \|x^k - \hat{x}\|^2 + \|x^{k+1} - x^k\|^2 + 2\langle x^k - x^{k+1}, \hat{x} - x^k\rangle. \tag{17}$$

On the other hand, from step 5 of Algorithm 4.1, $x^{k+1} = x^k + t_k v(x^k)$, so

$$\langle x^k - x^{k+1}, \hat{x} - x^k\rangle = -t_k\langle v(x^k), \hat{x} - x^k\rangle. \tag{18}$$

Also note that, from (5) with $x = x^k$, we get

$$v(x^k) = \operatorname*{argmin}_{v\in C-x^k}\left\{\hat{\beta}\varphi(Jf(x^k)v) + \frac{\|v\|^2}{2}\right\},$$

so, by virtue of (15) and (16) for $x = x^k$, there exist scalars $\alpha_1^k, \dots, \alpha_r^k \in \mathbb{R}$ and vectors $w^1, \dots, w^r \in I(v(x^k))$, with $\sum_{i=1}^r \alpha_i^k = 1$, $\alpha_i^k \geq 0$ for all $i = 1, \dots, r$, where r is a positive integer and $I(v(x^k)) = \{\bar{w} \in G\colon \langle \bar{w}, Jf(x^k)v(x^k)\rangle = \max_{w\in G}\langle w, Jf(x^k)v(x^k)\rangle\}$, such that

$$\left\langle v(x^k) + \hat{\beta}\sum_{i=1}^r \alpha_i^k Jf(x^k)^\top w^i, v - v(x^k)\right\rangle \geq 0 \quad \text{for all } v \in C - x^k.$$

In particular, for $v = \hat{x} - x^k$, we obtain

$$\left\langle v(x^k) + \hat{\beta}\sum_{i=1}^{r} \alpha_i^k Jf(x^k)^\top w^i, \hat{x} - x^k - v(x^k) \right\rangle \geq 0.$$

So, from the above inequality we get

$$\begin{aligned}
-t_k\langle v(x^k), \hat{x} - x^k\rangle &\leq t_k\hat{\beta}\sum_{i=1}^{r} \alpha_i^k \langle Jf(x^k)^\top w^i, \hat{x} - x^k\rangle \\
&\quad - t_k\left\langle v(x^k) + \hat{\beta}\sum_{i=1}^{r} \alpha_i^k Jf(x^k)^\top w^i, v(x^k)\right\rangle,
\end{aligned}$$

which, together with (18), gives

$$\begin{aligned}
\langle x^k - x^{k+1}, \hat{x} - x^k\rangle &\leq t_k\hat{\beta}\sum_{i=1}^{r} \alpha_i^k \langle Jf(x^k)^\top w^i, \hat{x} - x^k\rangle \\
&\quad - t_k\left\langle v(x^k) + \hat{\beta}\sum_{i=1}^{r} \alpha_i^k Jf(x^k)^\top w^i, v(x^k)\right\rangle. \qquad (19)
\end{aligned}$$

From K-convexity of f, by virtue of Lemma 5.3, we have

$$\langle Jf(x^k)^\top w^i, \hat{x} - x^k\rangle \leq 0 \quad \text{for all } i = 1, \ldots, r. \qquad (20)$$

Hence, from (17), (19), (20) and the fact that $t_k\hat{\beta} \geq 0$, it follows that

$$\begin{aligned}
\|x^{k+1} - \hat{x}\|^2 &\leq \|x^k - \hat{x}\|^2 + \|x^{k+1} - x^k\|^2 \\
&\quad - 2t_k\left(\|v(x^k)\|^2 + \hat{\beta}\sum_{i=1}^{r} \alpha_i^k \langle w^i, Jf(x^k)v(x^k)\rangle\right). \qquad (21)
\end{aligned}$$

Now let us take a look to the last part of the right hand side expression in (21):

$$\begin{aligned}
&\|x^{k+1} - x^k\|^2 - 2t_k\left(\|v(x^k)\|^2 + \hat{\beta}\sum_{i=1}^{r} \alpha_i^k \langle w^i, Jf(x^k)v(x^k)\rangle\right) \\
\leq\ & t_k\|v(x^k)\|^2 - 2t_k\left(\|v(x^k)\|^2 + \hat{\beta}\sum_{i=1}^{r} \alpha_i^k \langle w^i, Jf(x^k)v(x^k)\rangle\right) \\
=\ & -2t_k\left(\frac{\|v(x^k)\|^2}{2} + \hat{\beta}\varphi(Jf(x^k)v(x^k))\right) \\
=\ & -2t_k\theta(x^k) \\
=\ & 2t_k|\theta(x^k)|, \qquad (22)
\end{aligned}$$

where the inequality is a consequence of the definition of x^{k+1} and the fact that $t_k \in [0,1]$, the first equality holds since $w^i \in I(v(x^k))$ for $i = 1, \dots, r$ and $\sum_{i=1}^r \alpha_i^k = 1$ and, finally, the last one follows from nonstationarity of x^k together with Proposition 3.2.

From Lemma 4.3 and inequalities (21) and (22), we conclude that $\{x^k\}$ converges quasi-Fejér to T. Using now Theorem 5.4, we see that $\{x^k\}$ is bounded. Therefore, $\{x^k\}$ has accumulation points, which, by Theorem 5.1 and Lemma 5.2, are weakly efficient. Let x^* be one of these accumulation points and say that $x^{k_j} \to x^*$. Let $\bar{k}$ be a fixed but arbitrary nonnegative integer; for j large enough,

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

As f is continuous and K is closed, letting $j \to \infty$ in the above inequality yields

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

Since $\bar{k}$ was an arbitrary nonnegative integer, we conclude that the accumulation point x^* belongs to T. So, using again Theorem 5.4, we see see that $\{x^k\}$ converges to x^*, a weakly efficient solution. □

6 Final remarks

In this work we proved global convergence of the projected gradient method for vector optimization with an Armijo-like rule for choosing step sizes, implemented with a backtracking procedure. The method was originally proposed in [13], but in that paper the convergence analysis was incomplete, since the only result established on that issue was stationarity of cluster points. So we can see our work as an improvement upon the results of [13].

According to Theorem 5.6, the method converges globally to weakly efficient points in the partial order induced by the cone $K \subset \mathbb{R}^m$. Note that if $\hat{K}$ is a closed convex pointed cone with $K \setminus \{0\} \subset \text{int}(\hat{K})$ 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 $f(x) \neq f(x^*)$. For instance, consider Q a set of representants of all nontrivial directions of K, that is to say, Q is compact, $0 \notin Q$ and $K = \text{cone}(\text{conv}(Q))$. One could obtain such an augmented $\hat{K}$ by simply considering the cone generated by the set of points such that their distance to $\text{conv}(Q)$ is greater than or equal to a certain small enough tolerance tolerance so that pointedness is not lost. Clearly, such a construction may have its drawbacks, like, for example, the fact that if K has finitely many extremal rays, the enlarged cone $\hat{K}$ could lose polyhedrality. Of course, other constructions of such a $\hat{K}$ may be considered, and whenever the enlargement is tractable, we can use it instead of K, perform the method and obtain efficient points for the original problem. Note that, unless these optima are already $\hat{K}$-efficient, such points are not properly K-efficient (see [22, Definition 2.1]). Without considering new cones, we can also guarantee convergence of the method to efficient points if the objective function f is *strictly K-convex*, i.e., if it satisfies (10) with strict inequality (see [15, Proposition 2.2]).

Finally, it is worth to point out that our goal is not to find all the points in the optimal set; 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 points. We leave for a future work the numerical experiments which might help to confirm the last assertion. The study of the convergence properties of an inexact version of this method, as done in [16] for the unconstrained vector optimization problem, is also left as a matter for a future research.

Acknowledgements

We thank Alfredo Iusem and Andreas Hamel for their constructive comments. This research was supported by FAPESP under grant 2007/53471-0, FAPERJ under grant Cientistas do Nosso Estado 2008, and CNPq Universal under grant 480101/2008-6.

References

[1] C.D. Aliprantis, M. Florenzano, V.F. Martins da Rocha and R. Tourky, *General equilibrium in infinite security markets*, J. Math. Econ. 40 (2004), pp. 683–699.

[2] C.D. Aliprantis, M. Florenzano and R. Tourky, *General equilibrium analysis in ordered topological vector spaces*, J. Math. Econ. 40 (2004), pp. 247–269.

[3] D. Bertsekas, A. Nedić and E. Ozdaglar, *Convex Analysis and Optimization*, Athena Scientific, Belmont, 2003.

[4] J. Branke, K. Deb, K. Miettinen and R. Slowinski, *Seminar 04461, Practical approaches to multi-objective optimization*, Dagstuhl Seminar, Germany, 2005.

[5] J. Branke, K. Deb, K. Miettinen and R. Slowinski, *Seminar 06501, Practical approaches to multi-objective optimization*, Dagstuhl Seminar, Germany, 2007.

[6] R. Burachik, L.M. Graña Drummond, A.N. Iusem and B.F. Svaiter, *Full convergence of the steepest descent method with inexact line searches*, Optimization. 32 (1995), pp. 137–146.

[7] E. Carrizosa and J.B.G. Frenk. *Dominating sets for convex functions with some applications*, J. Optim. Theory Appl. 96 (1998), pp. 281–295.

[8] H. Eschenauer, J. Koski and A. Osyczka, *Multicriteria Design Optimization*, Springer, Berlin, 1990.

[9] J. Fliege and B.F. Svaiter, *Steepest descent methods for multicriteria optimization*, Math. Methods Oper. Res. 51 (2000), pp. 479–494.

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

[11] J. Fliege, L.M. Graña Drummond and B.F. Svaiter, *Newton's method for multi-objective optimization*, SIAM J. Optim. 20 (2009), pp. 602–626.

[12] Y. Fu and U.M. Diwekar, *An efficient sampling approach to multiobjective optimization*, Ann. Oper. Res. 132 (2004), pp. 109–134.

[13] L.M. Graña Drummond and A.N. Iusem, *A projected gradient method for vector optimization problems*, Comput. Optim. Appl. 28 (2004), pp. 5–29.

[14] L.M. Graña Drummond, N. Maculan and B.F. Svaiter, *On the choice of parameters for the weighting method in vector optimization*, Math. Program. 111 (2008), pp. 201–216.

[15] L.M. Graña Drummond, F.M.P. Raupp and B.F. Svaiter, *A Newton method for vector optimization*. Submitted for publication.

[16] L.M. Graña Drummond and B.F. Svaiter, *A steepest descent method for vector optimization*, J. Comput. Appl. Math. 175 (2005), pp. 395–414.

[17] M. Gravel, J.M. Martel, R. Nadeau, W. Price and R. Tremblay, *A multicriterion view of optimal ressource allocation in job-shop production*, Eur. J. Oper. Res. 61 (1992), pp. 230–244.

[18] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis*, Springer, Berlin, 2001.

[19] J. Jahn, *Scalarization in vector optimization*, Math. Program. 29 (1984), pp. 203–218.

[20] T.M. Leschine, H. Wallenius and W.A. Verdini, *Interactive multiobjective analysis and assimilative capacity-based ocean disposal decisions*, Eur. J. Oper. Res. 56 (1992), pp. 278–289.

[21] D.T. Luc, *Scalarization of vector optimization problems*, J. Optim. Theory Appl. 55 (1987), pp. 85–102.

[22] D.T. Luc, *Theory of Vector Optimization*, Springer, Berlin, 1989.

[23] K.M. Miettinen, *Nonlinear Multiobjective Optimization*, Kluwer, Norwell, 1999.

[24] G. Palermo, C. Silvano, S. Valsecchi and V. Zaccaria, *A system-level methodology for fast multi-objective design space exploration*, 13th ACM Great Lakes Symposium on VLSI, New York, 2003.

[25] D. Prabuddha, J.B. Ghosh and C.E. Wells, *On the minimization of completion time variance with a bicriteria extension*, Oper. Res. 40 (1992), pp. 1148–1155.

[26] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, New Jersey, 1970.

[27] M. Tavana, *A subjective assessment of alternative mission architectures for the human exploration of Mars at NASA using multicriteria decision making*, Comp. Oper. Res. 31 (2004), pp. 1147–1164.

[28] D.J. White, *Epsilon-dominating solutions in mean-variance portfolio analysis*, Eur. J. Oper. Res. 105 (1998), pp. 457–466.