A subgradient method for vector optimization problems

J.Y. Bello Cruz*

December 5, 2012

Abstract

Vector optimization problems are a significant extension of scalar optimization, and have many real life applications. We consider an extension of the projected subgradient method to convex vector optimization, which works directly with vector-valued functions, without using scalar-valued objectives. We eliminate the scalarization approach, a popular strategy for solving vector optimization problems, exploring strongly the structure of these kinds of problems. Under suitable assumptions, we show that the sequence generated by the algorithm converges to a weakly efficient optimum point.

Keywords: Nonsmooth optimization, weakly efficient points, projected subgradient method, vector optimization.

Mathematical Subject Classification (2010): 90C29, 90C30.

1 Introduction

First we briefly describe our notation. Let C be a nonempty, closed and convex subset of $\mathbb{R}^n$. The inner product in $\mathbb{R}^n$ is denoted by $\langle \cdot, \cdot \rangle$ and the norm determined by the inner product by $\| \cdot \|$. The closed ball centered at $x \in \mathbb{R}^n$ with radius ρ will be denoted by $B[x;\rho]$, i.e., $B[x;\rho] := \{y \in \mathbb{R}^n \colon \|y - x\| \le \rho\}$.

Given a proper, closed, convex and pointed cone K in $\mathbb{R}^m$, we consider the partial order " $\preceq$ " defined as $x \preceq y$ ($x \prec y$) if and only if $y - x \in K$ ($y - x \in \mathrm{int}(K)$). In similar way, we define the partial orders $\succeq$ ($\succ$). Consider a function $f : \mathbb{R}^n \to \mathbb{R}^m$. We are interested in the problem

$$\min_K \quad f(x) \tag{1}$$

$$s.t. \quad x \in C, \tag{2}$$

with the following meaning: a vector $x^* \in \mathbb{R}^n$ is solution of Problem (1)-(2), which is called a weakly efficient point, if and only if $x^* \in C$ and there does not exist $\hat{x} \in C$ such that $f(\hat{x}) \prec f(x^*)$, i.e. x^* is a solution if $f(x) - f(x^*) \notin -\mathrm{int}(K)$ for all $x \in C$. We denote the solution set of our problem by S^*, i.e.

$$S^* = \{x^* \in C \, : \, \nexists \ \hat{x} \in C \quad \text{such that} \quad f(\hat{x}) \prec f(x^*)\}.$$

When $m = 1$ and $K = \mathbb{R}_+$, $\preceq$ is the usual linear order in $\mathbb{R}$ and the problem will be called a scalar-valued optimization problem.

The positive polar cone of K, denoted by K^*, is given by $K^* = \{y \in \mathbb{R}^m \, : \, y^T x \ge 0 \ \forall x \in K\}$. Our algorithm makes use of a compact set G of normalized generators of K^*, i.e. $G \subset K^*$ is compact and such that the cone generated by its convex hull is K^*. Such a G always exists; one can take for example

*Universidade Federal de Goiás, IME, CEP 74001-970 GO, Goiânia, Brazil, e-mail: yunier@impa.br

$G = \{y \in K^* : \|y\| = 1\}$, but in general it is possible to take much smaller sets; see [18, 17, 24]. In the multiobjective case, $K = \mathbb{R}^m_+$, G can be taken as the canonical basis of $\mathbb{R}^m$ because $(\mathbb{R}^m_+)^* = \mathbb{R}^m_+$. If K is a polyhedral cone, then K^* is also polyhedral, so G can be chosen as the finite set of its extreme rays.

This paper is part of a wider research program consisting of the extension of several iterative methods for scalar-valued to multiobjective and vector-valued optimization; see [13, 15, 12, 11, 8, 6, 10]. We present a similar extension for the case of the subgradient method for scalar-valued convex optimization.

The sequence generated by the subgradient method is, in general, nondecreasing in its functional values. Due to the fact that $\mathbb{R}^m$ does not expose a total order, non-monotone methods, such as a subgradient methods, encounters major difficulties in the convergence analysis. The technique developed in this paper may be useful in other optimization problems, such as the variational inequality problems, equilibrium problems, saddle points problems, and their variations.

It is important to mention that in almost all methods that which have been extended for the context of the vector optimization, the monotony of the functional values plays an essential role in the convergence analysis; see [13, 15, 12, 11, 6].

We assume that $f : \mathbb{R}^n \to \mathbb{R}^m$ is K-convex, i.e.,

$$f(\lambda x + (1-\lambda)y) \preceq \lambda f(x) + (1-\lambda) f(y),$$

for all $x, y \in \mathbb{R}^n$ and all $\lambda \in [0,1]$. Note that for $y \in K^* \backslash \{0\}$, we have $\phi_y : \mathbb{R}^n \to \mathbb{R}$ such that $\phi_y(x) = y^T f(x)$ is a convex scalar function, then $v \in \partial\phi_y(x)$ if and only if $v \in \mathbb{R}^n$ and $\phi_y(z) \geq \phi_y(x) + \langle v, z - x\rangle$, for all $z \in \mathbb{R}^n$. It follows that exists $U \in \mathbb{R}^{m\times n}$ with $v = U^T y$ and so

$$y^T f(z) \geq y^T f(x) + \langle v, z - x\rangle = y^T f(x) + \langle U^T y, z - x\rangle = y^T f(x) + y^T U(z - x).$$

Since y is an arbitrary vector in $K^* \setminus \{0\}$, we get

$$f(z) \succcurlyeq f(x) + U(z - x), \tag{3}$$

for all z, $x \in \mathbb{R}^n$. So, we consider $\partial f : \mathbb{R}^n \rightrightarrows \mathbb{R}^{m\times n}$ defined as

$$\partial f(x) = \{U \in \mathbb{R}^{m\times n} : f(z) \succcurlyeq f(x) + U(z - x), \ \forall z \in \mathbb{R}^n\}.$$

In analogy with the scalar case, we will call this set the subdifferential and its elements the subgradients of f at x.

If we consider $K = \mathbb{R}^m_+$ and all functions f_i are convex in the usual sense, then any matrix $U \in \mathbb{R}^{m\times n}$ with the property that its ith-line is a subgradient of f_i at x, for $i = 1, \ldots, m$, belongs to $\partial f(x)$. An excellent survey of other concepts for subdifferentials of vector functions can be found in [25].

The outline of this article is as follows. In Section 2 we present the projected subgradient method for vector optimization. In Section 3 we state some basic definitions and some preliminary materials. Finally, Section 4 contains the convergence analysis of the algorithm.

2 A subgradient method for vector optimization

Our algorithm requires an exogenous sequence $\{\beta_k\}$ satisfying $\beta_k > 0$, $\sum_{k=0}^{\infty} \beta_k = \infty$ and

$$\sum_{k=0}^{\infty} \beta_k^2 < \infty. \tag{4}$$

This selection rule has been considered several times for similar methods; see for example [28, 2, 1, 4]. The algorithm is defined as:

Algorithm A

Initialization step: Take $x^0 \in C$.

Iterative step: Given $x^k \in C$, take $U^k \in \partial f(x^k)$ and compute

$$v^k = \arg\min_{w \in C_k} \left\{ \frac{1}{2}\|w\|^2 + \frac{\beta_k}{\eta_k} \max_{y \in G} \left\{ y^T U^k w \right\} \right\}, \tag{5}$$

with $C_k := C - x^k$, β_k as (4) and

$$\eta_k := \max_{y \in G} \{\|(U^k)^T y\|\}. \tag{6}$$

If $v^k = 0$ then stop, otherwise compute

$$x^{k+1} = x^k + v^k. \tag{7}$$

We define the orthogonal projection of x onto C, denoted by $P_C(x)$, as the unique point in C, such that $\|P_C(x) - y\| \leq \|x - y\|$ for all $y \in C$. Let us consider Algorithm A in the constrained scalar case, i.e., $G = \{1\}$. In this case $U^k = u^k \in \partial f(x^k)$, and (5) becomes

$$\begin{aligned} v^k &= \arg\min_{w \in C_k} \left\{ \frac{1}{2}\|w\|^2 + \frac{\beta_k}{\eta_k} (u^k)^T w \right\} \\ &= \arg\min_{w \in C_k} \left\{ \frac{1}{2}\|w\|^2 + \frac{\beta_k}{\eta_k} (u^k)^T w + \frac{\beta_k^2}{2\eta_k^2}\|u^k\|^2 \right\} \\ &= \arg\min_{w \in C_k} \left\{ \frac{1}{2}\|w + \frac{\beta_k}{\eta_k} u^k\|^2 \right\} = P_{C_k}\left(-\frac{\beta_k}{\eta_k} u^k \right). \end{aligned} \tag{8}$$

Using (8) and (7), we have

$$x^{k+1} = P_C\left(x^k - \frac{\beta_k}{\eta_k} u^k \right).$$

This is the classical iteration of the Projected Subgradient Method; see [26, 27, 29, 2, 1]. The projected subgradient method has been used widely in practical applications and given that it is a simple method, it has several useful advantages. Primarily, it is easy to implement (especially for optimization problems with relatively simple constraints). The method uses little storage and readily exploits any sparsity or separable structure of ∂f or C. Furthermore, it is able to drop or add active constraints during the iterations. Thus, the projected subgradient method has been strongly used for solving several special cases of the scalar convex problem. Its extensions and some modifications were studied for the generalized convex case in [20, 30, 5, 3, 21].

A popular strategy for solving vector optimization problems is the scalarization approach. The most widely used scalarization technique is the weighting method. Basically, one minimizes a linear combination of the objectives, where the vector of "weights" is not known a priori. This procedure may lead to unbounded numerical problems, which therefore may lack minimizers; see [7, 14, 19, 24, 23]. Another disadvantage of this approach is that the choice of the parameters is not known in advance, leaving the modeler and the decision-maker with the burden of choosing them; see [6]. In our method we do not used scalarization, a popular strategy for solving vector optimization problems, exploring strongly the structure of the vector problem.

3 Preliminary

In this section, we present some definitions and results that are needed for the convergence analysis of Algorithm A. The following two results are important for the proof of the convergence results, especially for the proof of Theorem 1.

Lemma 1. *If f is K-convex and $x \in \mathbb{R}^n$, then $\partial f(x)$ is a nonempty convex compact set.*

Proof. See Theorem 4.12 of [25]. $\square$

Lemma 2. *Let $\{x^k\}$ be a bounded sequence and f a K-convex function. Then, any sequence $\{U^k\}$ with $U^k \in \partial f(x^k)$ is bounded.*

Proof. Since $\{x^k\}$ is bounded, we can define ρ as the radius of the closed ball centered at 0, which contains the sequence $\{x^k\}$, i.e., $\{x^k\} \subset B[0;\rho]$. By Theorem 3.1 of [25] f is continuous. Then, there exists $M > 0$ such that $\|f(y)\| < M$ for all $y \in B[0;\rho+1]$. Assume by contradiction that $\|U^k\| > k$. Hence, for $k > 2M$ we can take a sequence $\{y^k\} \subset B[0;1]$ such that $\|U^k(y^k)\| \geq \|U^k\| > k$. Define $z^k := y^k + x^k$ and $\bar{z}^k := -y^k + x^k$, which belong to $B[0;\rho+1]$ and hence $\|f(z^k)\| < M$ and $\|f(\bar{z}^k)\| < M$ for all $k > 2M$. Since $U^k \in \partial f(x^k)$,

$$f(z) - f(x^k) - U^k(z - x^k) \in K,$$

for all $z \in \mathbb{R}^n$. Taking in the above inclusion $z = z^k$ and $z = \bar{z}^k$, we obtain

$$f(z^k) - f(x^k) - U^k(y^k) \in K$$

and

$$f(\bar{z}^k) - f(x^k) + U^k(y^k) \in K,$$

respectively. Define

$$w^k := \frac{f(z^k) - f(x^k) - U^k(y^k)}{\|f(z^k) - f(x^k) - U^k(y^k)\|}$$

and

$$\bar{w}^k := \frac{f(\bar{z}^k) - f(x^k) + U^k(y^k)}{\|f(\bar{z}^k) - f(x^k) + U^k(y^k)\|},$$

for $k > 2M$. Clearly, $w^k, \bar{w}^k \in K$ and $\|w^k\| = \|\bar{w}^k\| = 1$. Without loss of generality, we may assume that $\{w^k\}$, $\{\bar{w}^k\}$ converge to some unit vectors w, $\bar{w}$, respectively. Then,

$$w,\ \bar{w} \in K, \tag{9}$$

and

$$\lim_{k\to\infty} (w^k + \bar{w}^k) = w + \bar{w}. \tag{10}$$

For all $k > 2M$,

$$\|f(z^k) - f(x^k) - U^k(y^k)\| \geq \|U^k(y^k)\| - \|f(z^k) - f(x^k)\| \geq k - (\|f(z^k)\| + \|f(x^k)\|) \geq k - 2M, \tag{11}$$

$$\frac{\|f(\bar{z}^k) - f(x^k) + U^k(y^k)\|}{\|U^k\|} \geq \frac{\|U^k(y^k)\|}{\|U^k\|} - \frac{\|f(x^k) - f(\bar{z}^k)\|}{\|U^k(y^k)\|} \geq 1 - \frac{(\|f(x^k)\| + \|f(\bar{z}^k)\|)}{\|U^k(y^k)\|} \geq k - \frac{2M}{k}, \tag{12}$$

and

$$\begin{aligned}
\|w^k+\bar{w}^k\| &= \left\|\frac{f(z^k)-f(x^k)-U^k(y^k)}{\|f(z^k)-f(x^k)-U^k(y^k)\|}+\frac{f(\bar{z}^k)-f(x^k)+U^k(y^k)}{\|f(\bar{z}^k)-f(x^k)+U^k(y^k)\|}\right\| \\
&\leq \left\|\frac{f(z^k)-f(x^k)}{\|f(z^k)-f(x^k)-U^k(y^k)\|}+\frac{f(\bar{z}^k)-f(x^k)}{\|f(\bar{z}^k)-f(x^k)+U^k(y^k)\|}\right\| \\
&\quad +\|U^k(y^k)\|\left|\frac{1}{\|f(\bar{z}^k)-f(x^k)+U^k(y^k)\|}-\frac{1}{\|f(z^k)-f(x^k)-U^k(y^k)\|}\right| \\
&\leq \frac{4M}{k-2M}+\|U^k(y^k)\|\left|\frac{\|f(z^k)-f(x^k)-U^k(y^k)\|-\|f(\bar{z}^k)-f(x^k)+U^k(y^k)\|}{\|f(\bar{z}^k)-f(x^k)+U^k(y^k)\|\|f(z^k)-f(x^k)-U^k(y^k)\|}\right| \\
&\leq \frac{4M}{k-2M}+\|U^k(y^k)\|\frac{\|f(z^k)-2f(x^k)+f(\bar{z}^k)\|}{\|f(\bar{z}^k)-f(x^k)+U^k(y^k)\|\|f(z^k)-f(x^k)-U^k(y^k)\|} \\
&\leq \frac{4M}{k-2M}+\frac{4M}{\left(1-\frac{2M}{k}\right)(k-2M)},
\end{aligned}$$

using in the last inequality (11) and (12). Taking limits and using (10), we get $w+\bar{w}=0$, which together with (9) contradicting the pointedness of K. □

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

Definition 1. *Let S be a nonempty subset of $\mathbb{R}^n$. A sequence $\{x^k\}$ is said to be quasi-Fejér convergent to S if and only if for all $x \in S$, there exists $\bar{k}$ and a summable sequence $\{\delta_k\} \subset \mathbb{R}_+$ such that*

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

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

This definition originates in [9] and has been further elaborated in [16]. A useful result on quasi-Fejér sequences is the following.

Lemma 3. *If $\{x^k\}$ is quasi-Fejér convergent to S then, $\{x^k\}$ is bounded.*

Proof. See Lemma 6 in [9]. □

The technical results in this section are similar to those presented in [13]. Given $x \in dom(\partial f)$, $U \in \partial f(x)$, $w \in C-x$ and α, $\eta > 0$, we define the functions: $\theta_x \colon C-x \to \mathbb{R}$, by

$$\theta_x(w) := \max_{y\in G}\left\{y^T U w\right\};$$

$\psi_x \colon C-x \to \mathbb{R}$, by

$$\psi_x(w) := \frac{1}{2}\|w\|^2+\frac{\alpha}{\eta}\theta_x(w) = \frac{1}{2}\|w\|^2+\max_{y\in G}\left\{y^T U w\right\};$$

$\sigma\colon\ C \to \mathbb{R}$, by

$$\sigma(x) := \min_{w\in C-x}\{\psi_x(w)\};$$

and $v\colon C \to \mathbb{R}^n$, by

$$v(x) := \arg\min_{w\in C-x}\{\psi_x(w)\} = \arg\min_{w\in C-x}\left\{\frac{1}{2}\|w\|^2+\frac{\alpha}{\eta}\max_{y\in G}\left\{y^T U w\right\}\right\}.$$

The function θ_x is well-defined because G is compact and it is convex because it is the maximum of linear functions. Since ψ_x is strongly convex, we obtain that the functions σ and v are well-defined. Our algorithm can be written using the above functions, taking $\alpha = \beta_k$, $\eta = \eta_k$, and $U = U^k \in \partial f(x^k)$, as

$$x^{k+1} = x^k + v(x^k),$$

with

$$\begin{aligned} v(x^k) &= \arg \min_{w \in C - x^k} \{\psi_{x^k}(w)\} = \arg \min_{w \in C - x^k} \left\{ \frac{1}{2}\|w\|^2 + \frac{\beta_k}{\eta_k} \theta_{x^k}(w) \right\} \\ &= \arg \min_{w \in C - x^k} \left\{ \frac{1}{2}\|w\|^2 + \frac{\beta_k}{\eta_k} \max_{y \in G} \left\{ y^T U^k w \right\} \right\}. \end{aligned}$$

4 Convergence analysis

We attempt to establish convergence of the generated sequence to a point in the solutions set. From now on we denote by $\{x^k\}$ the sequence generated by our algorithm starting from some $x^0 \in C$. First, we establish the feasibility of $\{x^k\}$.

Proposition 1. *The sequences $\{x^k\}$ generated by Algorithm A belong to C.*

Proof. By induction. The initial iterate x^0 belongs to C by the Initialization step of Algorithm A. Assuming $x^k \in C$, since $v^k \in C_k = C - x^k$ and using (7), we conclude that x^{k+1} belongs to C. □

The next proposition establishes the validity of the stoping criterion.

Proposition 2. *Let $\{x^k\}$ and $\{v^k\}$ be the sequences generated by Algorithm A. If $v^k = 0$, then $x^k \in S^*$.*

Proof. Assume that $x^k \notin S^*$. Then, there exists $z \in C$ such that $f(z) \prec f(x^k)$. Take $v = z - x^k$. It is true that $v \neq 0$. By the K-convexity of f, we have

$$f(z) \succcurlyeq f(x^k) + U^k v,$$

implying that $U^k v \in -\text{int}(K)$, and therefore $\max_{y \in G}\{y^T U^k v\} < 0$. Considering $z(\lambda) = \lambda z + (1-\lambda)x^k$, for $\lambda \in [0,1]$, we get

$$\begin{aligned} \frac{\|v^k\|^2}{2} + \frac{\beta_k}{\eta_k} \max_{y \in G}\{y^T U^k v^k\} &\leq \frac{\|z(\lambda) - x^k\|^2}{2} + \frac{\beta_k}{\eta_k} \max_{y \in G} \left\{ y^T U^k [z(\lambda) - x^k] \right\} \\ &= \lambda \left[\lambda \frac{\|v\|^2}{2} + \frac{\beta_k}{\eta_k} \max_{y \in G}\{y^T U^k v\} \right] \end{aligned}$$

for all $\lambda \in [0,1]$. Then

$$\frac{\|v^k\|^2}{2} + \frac{\beta_k}{\eta_k} \max_{y \in G}\{y^T U^k v^k\} < 0,$$

and henceforth $v^k \neq 0$. □

We continue the convergence analysis with this auxiliary result.

Proposition 3. *Let $\{x^k\}$ and $\{v^k\}$ be the sequences generated by Algorithm A and β_k defined in (4). For all k, $\|v^k\| \leq \beta_k$ and $\|x^{k+1} - x^k\| \leq \beta_k$.*

Proof. In case of $v^k = 0$, both assertions are trivially valid. So we assume that $v^k \neq 0$. The first order optimality conditions for $\min_{w\in C_k} \psi_{x^k}(w)$, where

$$\psi_{x^k}(w) = \frac{1}{2}\|w\|^2 + \frac{\beta_k}{\eta_k}\max_{y\in G}\{y^T U^k w\},$$

with β_k and η_k defined in (4) and (6) respectively. In view of the convexity of ψ_{x^k} these optimality conditions are necessary and sufficient, implying the existence of $u^k \in \partial\psi_{x^k}(v^k)$ such that

$$\langle u^k, z - v^k\rangle \geq 0 \ \ \forall z \in C_k. \tag{13}$$

Since $x^k \in C$, we obtain that $0 \in C_k$, and therefore, taking $z = 0$ in (13), we have

$$\langle u^k, v^k\rangle \leq 0. \tag{14}$$

Using the formula for the subdifferential of a maximum of convex functions, and (5), there exists a positive integer $q(k)$ and $y_i^k \in G$ and $\lambda_i^k > 0$ with $1 \leq i \leq q(k)$ such that

$$\sum_{i=1}^{q(k)} \lambda_i^k = 1, \tag{15}$$

$$(y_i^k)^T U^k v^k = \max_{y\in G}\{y^T U^k v^k\} \qquad (1 \leq i \leq q(k)), \tag{16}$$

and

$$u^k = v^k + \frac{\beta_k}{\eta_k}\sum_{i=1}^{q(k)} \lambda_i^k (U^k)^T y_i^k. \tag{17}$$

In view of (14)-(17), we get

$$\|v^k\|^2 \leq -\frac{\beta_k}{\eta_k}\sum_{i=1}^{q(k)} \lambda_i^k (y_i^k)^T U^k v^k = -\frac{\beta_k}{\eta_k}\max_{y\in G}\{y^T U^k v^k\}. \tag{18}$$

Since $\psi_{x^k}(v^k) = \frac{1}{2}\|v^k\|^2 + \frac{\beta_k}{\eta_k}\max_{y\in G}\{y^T U^k v^k\} \leq \psi_{x^k}(0) = 0$, we have

$$\frac{\beta_k}{\eta_k}\max_{y\in G}\{y^T U^k v^k\} \leq -\frac{1}{2}\|v^k\|^2 \leq 0. \tag{19}$$

Combining (18)-(19), we obtain

$$\|v^k\|^2 \leq \frac{\beta_k}{\eta_k}|\max_{y\in G}\{y^T U^k v^k\}| = \frac{\beta_k}{\eta_k}(\bar{y}^k)^T U^k v^k, \tag{20}$$

where $\bar{y}^k \in G$ realizes maximum in the second term of (20). Now using the fact that $\eta_k = \max_{y\in G}\{\|(U^k)^T y\|\}$ and (20), we get

$$\|v^k\|^2 \leq \frac{\beta_k}{\eta_k}|\max_{y\in G}\{y^T U^k v^k\}| \leq \frac{\beta_k}{\eta_k}\|(U^k)^T \bar{y}^k\|\|v^k\| \leq \beta_k\|v^k\|.$$

Since $\|v^k\| \neq 0$,

$$\|v^k\| \leq \beta_k. \tag{21}$$

Finally, from (21) and (7), we get

$$\|x^{k+1} - x^k\| \leq \beta_k.$$

□

The next result establishes a fundamental inequality.

Proposition 4. *Let $\{x^k\}$ be the sequence generated by Algorithm A, β_k and η_k defined in* (4) *and* (6)*, respectively. For all $x \in C$,*

$$3\beta_k^2 + \|x - x^k\|^2 - \|x - x^{k+1}\|^2 \geq 2\frac{\beta_k}{\eta_k} \min_{y \in G}\{(f(x^k) - f(x))^T y\}.$$

Proof. Using (7) and Proposition 3, we get

$$\begin{aligned} \beta_k^2 + \|x - x^k\|^2 - \|x - x^{k+1}\|^2 &\geq \|x^{k+1} - x^k\|^2 + \|x - x^k\|^2 - \|x - x^{k+1}\|^2 \\ &= 2\langle x^k - x^{k+1}, x^k - x\rangle = 2\langle v^k, x - x^k\rangle. \end{aligned} \tag{22}$$

By (17), $v^k = u^k - \frac{\beta_k}{\eta_k}\sum_{i=1}^{q(k)} \lambda_i^k (U^k)^T y_i^k$, and by (13), $\langle u^k, (x - x^k) - v^k\rangle \geq 0$, for all $x \in C$. Then

$$\begin{aligned} \langle v^k, x - x^k\rangle &= \langle u^k, x - x^k\rangle + \frac{\beta_k}{\eta_k}\sum_{i=1}^{q(k)} \lambda_i^k \langle (U^k)^T y_i^k, x^k - x\rangle \\ &\geq \langle u^k, v^k\rangle + \frac{\beta_k}{\eta_k}\sum_{i=1}^{q(k)} \lambda_i^k (y_i^k)^T (U^k)^T (x^k - x) \\ &= \langle v^k, v^k + \frac{\beta_k}{\eta_k}\sum_{i=1}^{q(k)} \lambda_i^k (U^k)^T y_i^k\rangle + \frac{\beta_k}{\eta_k}\sum_{i=1}^{q(k)} \lambda_i^k (y_i^k)^T U^k (x^k - x) \\ &= \|v^k\|^2 + \frac{\beta_k}{\eta_k}\sum_{i=1}^{q(k)} \lambda_i^k (y_i^k)^T U^k v^k + \frac{\beta_k}{\eta_k}\sum_{i=1}^{q(k)} \lambda_i^k (y_i^k)^T U^k (x^k - x) \\ &= \|v^k\|^2 + \frac{\beta_k}{\eta_k}\max_{y \in G}\{y^T U^k v^k\} + \frac{\beta_k}{\eta_k}\sum_{i=1}^{q(k)} \lambda_i^k (y_i^k)^T U^k (x^k - x) \\ &\geq \frac{\beta_k}{\eta_k}(\bar{y}^k)^T U v^k + \frac{\beta_k}{\eta_k}\sum_{i=1}^{q(k)} \lambda_i^k (y_i^k)^T U^k (x^k - x), \end{aligned} \tag{23}$$

with $\bar{y}^k$ as in the proof of Proposition 3. Now combining (22) and (23), we obtain

$$\beta_k^2 + \|x - x^k\|^2 - \|x - x^{k+1}\|^2 \geq 2\frac{\beta_k}{\eta_k}(\bar{y}^k)^T U^k v^k + 2\frac{\beta_k}{\eta_k}\sum_{i=1}^{q(k)} \lambda_i^k (y_i^k)^T U^k (x^k - x). \tag{24}$$

By (15), the definition of η_k, Proposition 3 and (19), we get from (24) that

$$\begin{aligned} \beta_k^2 + \|x - x^k\|^2 - \|x - x^{k+1}\|^2 &\geq -2\frac{\beta_k}{\eta_k}\|(U^k)^T \bar{y}^k\|\|v^k\| + 2\frac{\beta_k}{\eta_k}\sum_{i=1}^{q(k)} \lambda_i^k (y_i^k)^T U^k (x^k - x) \\ &\geq -2\beta_k\|v^k\| + 2\frac{\beta_k}{\eta_k}\sum_{i=1}^{q(k)} \lambda_i^k (y_i^k)^T U^k (x^k - x) \\ &\geq -2\beta_k^2 + 2\frac{\beta_k}{\eta_k}\sum_{i=1}^{q(k)} \lambda_i^k (y_i^k)^T U^k (x^k - x). \end{aligned} \tag{25}$$

Using the gradient-like inequality for K-convex functions, given in (3), we get

$$U^k(x^k - x) \succeq f(x^k) - f(x).$$

Since $y_i^k \in G \subset K^*$, we get

$$(y_i^k)^T U^k (x^k - x) \geq (y_i^k)^T (f(x^k) - f(x)).$$

Using the above inequality in (25), we obtain

$$\begin{aligned} \beta_k^2 + \|x - x^k\|^2 - \|x - x^{k+1}\|^2 &\geq -2\beta_k^2 + 2\frac{\beta_k}{\eta_k} \sum_{i=1}^{q(k)} \lambda_i^k (y_i^k)^T (f(x^k) - f(x)) \\ &\geq -2\beta_k^2 + 2\frac{\beta_k}{\eta_k} \min_{y \in G} \{(f(x^k) - f(x))^T y\}, \end{aligned}$$

establishing the proposition. □

Define the auxiliary set T as

$$T := \left\{ x \in C \colon \exists \bar{k} \text{ such that } \min_{y \in G} \{(f(x^k) - f(x))^T y\} \geq 0, \ \forall\, k \geq \bar{k} \right\}.$$

From now on we assume that T is nonempty. The assumption that $T \neq \emptyset$ was used in [13, 12, 6] for proving the convergence in the smooth convex and quasi convex cases. This assumption has a relation with the completeness of the image of f, namely that all non-increasing sequences in the image of f have a lower bound. It is important to say that completeness is a standard assumption for ensuring existence of efficient points [24].

Proposition 5. *The sequence $\{x^k\}$ is bounded.*

Proof. Take $x \in T$. By Proposition 4,

$$\|x^{k+1} - x\|^2 \leq \|x^k - x\|^2 + 3\beta_k^2,$$

for all $k \geq \bar{k}$. Using Definition 1, we conclude, in view of the fact that β_k satisfies (4) that the sequence $\{x^k\}$ is quasi-Fejér convergent to T. The result follows from Proposition 3. □

Theorem 1. *The sequences $\{x^k\}$ generated by Algorithm A has a cluster point, x^*, belonging to S^*.*

Proof. Using Proposition 4, for each $x \in C$, we have that

$$2\frac{\beta_k}{\eta_k} \min_{y \in G} \{(f(x^k) - f(x))^T y\} \leq \|x^k - x\|^2 - \|x^{k+1} - x\|^2 + 3\beta_k^2, \tag{26}$$

for all k. Since $\{x^k\}$ is bounded by Proposition 5 and then U^k is bounded in virtue that Lemma 2, establishing that $\eta_k \leq \rho$ for all k. We define $\gamma_k := \gamma_k(x) = \min_{y \in G} \{(f(x^k) - f(x))^T y\}$ and rewrite (26) as

$$2\frac{\beta_k}{\rho} \gamma_k \leq \|x^k - x\|^2 - \|x^{k+1} - x\|^2 + 3\beta_k^2. \tag{27}$$

Summing (27), with k between 0 and m,

$$\begin{aligned} \frac{2}{\rho} \sum_{k=0}^{m} \beta_k \gamma_k &\leq \sum_{k=0}^{m} \left(\|x^k - x\|^2 - \|x^{k+1} - x\|^2 \right) + 3 \sum_{k=0}^{m} \beta_k^2 \\ &= \|x^0 - x\|^2 - \|x^{m+1} - x\|^2 + 3 \sum_{k=0}^{m} \beta_k^2 \\ &\leq \|x^0 - x\|^2 + 3 \sum_{k=0}^{\infty} \beta_k^2 < \infty. \end{aligned} \tag{28}$$

Taking limits in (28), with $m \to \infty$, we get

$$\sum_{k=0}^{\infty} \beta_k \gamma_k < \infty. \tag{29}$$

We claim that there exists a subsequence $\{\gamma_{i_k}\}$ of γ_k such that $\lim_{k\to\infty} \gamma_{i_k} \le 0$. If the claim does not hold then there exists $\sigma > 0$ and $k \ge \tilde{k}$ such that $\sum_{k=\tilde{k}}^{\infty} \beta_k \gamma_k \ge \sigma \sum_{k=\tilde{k}}^{\infty} \beta_k$ for all $k \ge \tilde{k}$, in contradiction with $\sum_{k=\tilde{k}}^{\infty} \beta_k = \infty$. This establishes the claim.

Let x^* be a cluster point of $\{x^k\}$ associated to the subsequence $\{x^{i_k}\}$. Follows from Proposition 1 that $x^* \in C$. Suppose that $x^* \notin S^*$, so there exists $\hat{x} \in C$ such that $f(\hat{x}) \prec f(x^*)$, implying that

$$(f(x^*) - f(\hat{x}))^T y > 0, \tag{30}$$

for all $y \in G$.

By compactness of G there exists a point $\bar{y}^k$ realizing the minimum in G of $(f(x^k) - f(\hat{x}))^T y$, and also we can assume that $\{\bar{y}^{i_k}\}$ converges to $\bar{y} \in G$. Then

$$\begin{aligned} 0 &\ge \lim_{k\to\infty} \left\{ \min_{y\in G} \{(f(x^{i_k}) - f(\hat{x}))^T y\} \right\} = \lim_{k\to\infty} \left\{ (f(x^{i_k}) - f(\hat{x}))^T \bar{y}^{i_k} \right\} \\ &= \lim_{k\to\infty} \left\{ (f(x^{i_k}) - f(\hat{x}))^T (\bar{y}^{i_k} - \bar{y}) + (f(x^{i_k}) - f(\hat{x}))^T \bar{y} \right\} \\ &= \lim_{k\to\infty} \left\{ (f(x^{i_k}) - f(\hat{x}))^T \bar{y} \right\} \\ &\ge (f(x^*) - f(\hat{x}))^T \bar{y}. \end{aligned}$$

Summarizing, there exists a cluster point x^* of $\{x^k\}$ such that

$$(f(x^*) - f(\hat{x}))^T \bar{y} \le 0, \tag{31}$$

for some $\bar{y} = \bar{y}(\hat{x}) \in G$. The above inequality contradicts (30), so x^* belongs to S^*. □

Finally, we prove that all cluster points of $\{x^k\}$ belong to S^*.

Theorem 2. *All cluster points of* $\{x^k\}$ *solve Problem* (1)-(2).

Proof. By using the proof of Theorem 1 is sufficient to show that all cluster points of $\gamma_k = \gamma_k(x) := \min_{y\in G}\{(f(x^k) - f(x))^T y\}$ is nonpositive, for any $x \in C$.
First we prove that there exists $\rho > 0$ such that $\gamma_k - \gamma_{k+1} \le \rho\beta_k$ for all k.

$$\begin{aligned} \gamma_k - \gamma_{k+1} &= \min_{y\in G}\{(f(x^k) - f(x))^T y\} - \min_{y\in G}\{(f(x^{k+1}) - f(x))^T y\} \\ &\le \min_{y\in G}\{(f(x^k) - f(x))^T y\} - \min_{y\in G}\{(f(x^{k+1}) - f(x^k))^T y\} - \min_{y\in G}\{(f(x^k) - f(x))^T y\} \\ &= -\min_{y\in G}\{(f(x^{k+1}) - f(x^k))^T y\} = \max_{y\in G}\{(f(x^k) - f(x^{k+1}))^T y\} \\ &\le \max_{y\in G}\{\left(U^k(x^k - x^{k+1})\right)^T y\} \le \max_{y\in G} \|(U^k)^T y\| \|x^k - x^{k+1}\| \\ &\le \rho\beta_k, \end{aligned} \tag{32}$$

where ρ is an upper bound of $\max_{y\in G} \|(U^k)^T y\|$, which is finite in view of Lemma 2.
Next we claim that all cluster points of γ_k are nonpositive.

From Theorem 1 there exists a subsequence γ_{i_k} of γ_k such that $\lim_{k\to\infty}\gamma_{i_k} \le 0$. If the claim does not hold then there exists some $\delta > 0$ and another subsequence γ_{ℓ_k} of γ_k such that $\gamma_{\ell_k} \ge \delta$ for all k. Thus, we can construct a third subsequence γ_{j_k} of γ_k, where the indices j_k are chosen in the following way:

$$j_0 := \min_{m\ge 0}\{\gamma_m \ge \delta\},$$

$$j_{2k+1} := \min_{m\ge j_{2k}}\{\gamma_m \le \delta/2\},$$

$$j_{2k+2} := \min_{m\ge j_{2k+1}}\{\gamma_m \ge \delta\}.$$

The existence of the subsequences $\gamma_{i_k}, \gamma_{\ell_k}$ of $\{x^k\}$ guarantees that the subsequence γ_{j_k} of $\{x^k\}$ is well defined for all $k \ge 0$. Follows from the definition of j_k that

$$\gamma_m \ge \delta/2 \quad \text{for} \quad j_{2k} \le m \le j_{2k+1} - 1. \tag{33}$$

In view of (29), we have

$$\begin{aligned}
\infty \quad > \quad & \sum_{k=0}^{\infty}\beta_k\gamma_k \ge \sum_{k=0}^{\infty}\sum_{m=j_{2k}}^{j_{2k+1}-1}\beta_m\gamma_m \ge \delta/2\sum_{k=0}^{\infty}\sum_{m=j_{2k}}^{j_{2k+1}-1}\beta_m \\
= \quad & \frac{\delta}{2\rho}\sum_{k=0}^{\infty}\sum_{m=j_{2k}}^{j_{2k+1}-1}\rho\beta_m \ge \frac{\delta}{2\rho}\sum_{k=0}^{\infty}\sum_{m=j_{2k}}^{j_{2k+1}-1}(\gamma_m - \gamma_{m+1}) \\
= \quad & \frac{\delta}{2\rho}\sum_{k=0}^{\infty}(\gamma_{j_{2k}} - \gamma_{j_{2k+1}}) \ge \frac{\delta}{2\rho}\sum_{k=0}^{\infty}\delta/2 = \infty,
\end{aligned}$$

using (32) in the third inequality and (33) in the last inequality. The above contradiction establishes the claim. Thus, all cluster points of $\{x^k\}$ belong to S^*.

□

In vector optimization, the main goal is to reconstruct the set of solution. So, if the generated sequence has many accumulation points, a clustering technique such as K-means; see [22], can be used to identify them. Furthermore, we summarize in the following corollary the convergence sequence properties of Algorithm A, which are direct consequences of Theorems 1 and 2.

Corollary 1. *The sequence $\{x^k\}$ generated by Algorithm A is bounded, and all cluster points of $\{x^k\}$ belong to S^*. If Problem* (1)-(2) *has a unique solution then the whole sequence $\{x^k\}$ converges to it.*

In this paper, we do not study the subgradient algorithm for vector optimization for its practical value: in our opinion, it is suitable for "real life" vector optimization problems, only when compared to the classical subgradient method for scalar optimization problems. However, we start from the ideas of the classical subgradient method, in an attempt to deal with more efficient methods; for instance ϵ-subgradient methods, bundle methods and cutting-plane algorithms for vector optimization, expecting that it will be a first step toward more efficient methods to be developed in the future. However, how to extend these efficient procedures, remains an open question. We foresee further progress along this path in the future.

Acknowledgment

Research for this paper was partially supported by PROCAD-nf - UFG/UnB/IMPA research, PRONEX - CNPq-FAPERJ - Optimization research and by the Project CAPES-MES-CUBA 226/2012, "MODELOS DE OTIMIZAÇÃO E APLICAÇÕES". The author would like to extend your gratitude towards anonymous referees whose suggestions helped us to improve the presentation of this paper.

References

[1] Alber, Ya.I., Iusem, A.N. Extension of subgradient techniques for nonsmooth optimization in Banach spaces. *Set-Valued Analysis* **9** (2001) 315-335.

[2] Alber, Ya.I., Iusem, A.N., Solodov, M.V. On the projected subgradient method for nonsmooth convex optimization in a Hilbert space. *Mathematical Programming* **81** (1998) 23-37.

[3] Bello Cruz, J.Y., Iusem, A.N. A strongly convergent method for nonsmooth convex minimization in Hilbert spaces. *Numerical Functional Analysis and Optimization* **32** (2011) 1009 -1018.

[4] Bello Cruz, J.Y., Iusem, A.N. Convergence of direct methods for paramonotone variational inequalities. *Computational Optimization and Application* **46** (2010) 247-263.

[5] Bello Cruz, J.Y., Lucambio Pérez, L.R. Convergence of a projected gradient method variant for quasiconvex objectives. *Nonlinear Analysis* **9** (2010) 2917-2922.

[6] Bello Cruz, J.Y., Lucambio Pérez, L.R., Melo, J.G. Convergence of the projected gradient method for quasiconvex multiobjective optimization. *Nonlinear Analysis* **74** (2011) 5268-5273.

[7] Bolintinéanu, S. Approximate efficiency and scalar stationarity in unbounded nonsmooth convex vector optimization problems. *Journal of Optimization Theory and Applications* **106** (2000) 265-296.

[8] Bonnel, H., Iusem, A.N., Svaiter, B.F. Proximal methods in vector optimization. *SIAM Journal on Optimization* **15** (2005) 953-970.

[9] Browder, F.E. Convergence theorems for sequences of nonlinear operators in Banach spaces. *Mathematische Zeitschrift* **100** (1967) 201-225.

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

[11] Fliege, J., Svaiter, B.F. Steepest descent methods for multicriteria optimimization. *Mathematical Methods of Operations Research* **51** (2000) 479-494.

[12] Fukuda, E.H., Graña Drummond, L.M. On the convergence of the projected gradient method for vector optimization, Optimization **60** (2011) 1009-1021.

[13] Graña Drummond, L.M., Iusem, A.N. A projected gradient method for vector optimization problems. *Computational Optimization and Applications* **28** (2004) 5-30.

[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** (2008) 201-216.

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

[16] Iusem, A.N., Svaiter, B.F., Teboulle, M. Entropy-like proximal methods in convex programming. *Mathematics of Operations Research* **19** (1994) 790-814.

[17] Jahn, J. *Vector optimization - Theory, applications and extensions*. Springer, Erlangen (2003).

[18] Jahn, J. *Mathematical vector optimization in Partially Ordered Linear Spaces*. Verlag Peter D. Lang, Frankfurt (1986).

[19] Jahn, J. Scalarization in vector optimization. *Mathematical Programming* **29** (1984) 203-218.

[20] Kiwiel, K.C. Convergence and efficiency of subgradient methods for quasiconvex minimization. *Mathematical Programming* **90** (2001) 1-25.

[21] Kiwiel, K.C. The efficiency of subgradient projection methods for convex optimization. I. General level methods. *SIAM Journal on Control and Optimization* **34** (1996) 660-676.

[22] Kogan, J.: Introduction to clustering large and high-dimensional data. Cambridge University Press, Cambridge (2007).

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

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

[25] Luc, D.T., Tan, N.X., Tinh, P.N. Convex vector functions and their subdifferential. *Acta Mathematica Vietnamica* **23** (1998) 107-127.

[26] Polyak, B.T. A general method for solving extremal problems. *Soviet Mathematics Doklady* **8** (1967) 593-597.

[27] Polyak, B.T. Minimization of unsmooth functionals. *U.S.S.R. Computational Mathematics and Mathematical Physics* **9** (1969) 14-29.

[28] Polyak, B.T. *Introduction to optimization*. Optimization Software, New York (1987).

[29] Shor, N.Z. *Minimization methods for nondifferentiable functions*. Springer, Berlin (1985).

[30] Wang, C., Xiu, N. Convergence of the gradient projection method for generalized convex minimization. *Computational Optimization and Applications* **16** (2000) 111-120.