A Proximal Gradient Splitting Method for Solving Convex Vector Optimization Problems

Yunier Bello-Cruz* Jefferson G. Melo† Ray V.G. Serra‡

July 20, 2020

Abstract

In this paper, a proximal gradient splitting method for solving nondifferentiable vector optimization problems is proposed. The convergence analysis is carried out when the objective function is the sum of two convex functions where one of them is assumed to be continuously differentiable. The proposed splitting method exhibits full convergence to a weakly efficient solution without assuming the Lipschitz continuity of the Jacobian of the differentiable component. To carry this analysis, the popular Beck–Teboulle's line-search procedure is extended to the vectorial setting under mild assumptions. It is also shown that the proposed scheme obtains an ε-approximate solution to the vector optimization problem in at most $\mathcal{O}(1/\varepsilon)$ iterations.

2010 Mathematics Subject Classification: 49J52, 90C25, 90C52, 65K05

Keywords: Convex programming, Full convergence, Proximal Gradient method, Vector optimization.

1 Introduction

Recently, there has been a growing interest in the extension and analysis of classical optimization algorithms for solving vector optimization problems, i.e., given a function $F : \mathbb{R}^n \to \mathbb{R}^m$, the goal is to obtain a point $\bar{x} \in \mathbb{R}^n$ which can not be strictly improved. The vector optimization problem is as follows

$$\min_{\mathcal{K}} \; F(x) \qquad \text{subject to} \quad x \in \mathbb{R}^n, \tag{1}$$

where the partial order is given by a pointed, convex and closed cone $\mathcal{K} \subset \mathbb{R}^m$ with a nonempty interior. Hence, the *weakly efficient* solutions are the points $\bar{x}$ such that there is no $x \in \mathbb{R}^n$ satisfying $F(x) \prec F(\bar{x})$ (or equivalently $F(\bar{x}) - F(x) \in \operatorname{int}(\mathcal{K})$). The above problem becomes the so-called

*Department of Mathematical Sciences, Northern Illinois University. Watson Hall 366, DeKalb, IL, USA - 60115. E-mail: yunierbello@niu.edu.

†Institute of Mathematics and Statistics, Federal University of Goias, Campus II- Caixa Postal 131, CEP 74001-970, Goiânia-GO, Brazil. E-mail: jefferson@ufg.br

‡Institute of Mathematics and Statistics, Federal University of Goias, Campus II- Caixa Postal 131, CEP 74001-970, Goiânia-GO, Brazil. E-mail: rvictorgs@gmail.com

multiobjective optimization problem when $\mathcal{K}$ is $\mathbb{R}^m_+ := \{x \in \mathbb{R}^n \mid x_i \geq 0, \ \forall i = 1, 2, \ldots, m\}$ and weakly efficient solutions are usually called weak *Pareto* points, i.e., there is no $x \in \mathbb{R}^n$ such $F_i(x) < F_i(\bar{x}), \forall i = 1, \ldots, m$.

A plethora of methods has been proposed for solving problem (1), most of them coming from the *scalar* optimization, i.e., $m = 1$ and $\mathcal{K} = \mathbb{R}_+$. In [19], the authors proposed an extension of the classical steepest descent method for minimizing continuously differentiable multiobjective functions. A generalized version of [19] was introduced in [23] for solving vector optimization problems. The latter method was further extended to solve constrained vector optimization problems in [21]. The full convergence of the projected gradient method for solving quasi-convex multiobjective optimization problems was studied in [8]. In [11], the authors generalized the gradient method of [19] to solve multiobjective optimization problems over a Riemannian manifold. The proximal point method was first introduced in [13] for solving vector optimization problems. A multiobjective proximal point method based on variable scalarizations was proposed in [10] to solve quasi-convex multiobjective problems. The authors also showed that if the weak Pareto optimum set satisfies a special property, then the proposed scheme possesses finite termination. Variants of the vector proximal point method for solving some classes of nonconvex multiobjective optimization problems were considered in [9, 12]. The subgradient method was extended to the multiobjective setting in [16] and more generally for solving vector optimization problems in [3]. The Newton method was first introduced in the multiobjective setting in [18] and further extended to solve vector optimization problems in [22]. Recently, [26] proposed and analyzed a conjugate gradient method for solving vector optimization problems using different line-search procedures. Finally, several methods for solving other vector-type optimization problems related to (1) have been studied in [4–6].

Consider the following structured vector optimization problem

$$\min_{\mathcal{K}} F(x) := G(x) + H(x) \qquad \text{subject to} \quad x \in \mathbb{R}^n, \tag{2}$$

where $G : \mathbb{R}^n \to \mathbb{R}^m$ is a continuously differentiable $\mathcal{K}$-convex function, and $H : \mathbb{R}^n \to \overline{\mathbb{R}}^m := \mathbb{R}^m \cup \{+\infty_{\mathcal{K}}\}$ is a proper $\mathcal{K}$-convex function which is possibly nonsmooth. Here, $y \prec +\infty_{\mathcal{K}}$ for any $y \in \mathbb{R}^m$. One of the most studied methods for solving the scalar version of this problem is the proximal gradient method (PGM) which has been intensively investigated; e.g., [2] and references therein. This classical splitting iteration is shown to be an effective and simple scheme for solving large scale problems and ℓ_1-regularized problems; see, for instance, [1, 7, 28]. Each iteration of this scheme basically consists of applying a gradient step for the differentiable component of the objective function followed by a proximal step of the nonsmooth part. Its applicability is mainly directly associated with the capability of solving the proximal subproblem, which can be achieved in some important cases. For instance, when each component h_i of H is such that $h_i = \| \cdot \|_1$ or h_i is the indicator function of a "simple" convex set. In the latter case, this scheme reduces to the classical projected gradient method. The extension of the PGM for solving composite multiobjective optimization problems was recently proposed in [27], where it was proved that every accumulation point, if any, satisfies a necessary optimality condition associated to problem (2) with $\mathcal{K} = \mathbb{R}^m_+$. The latter reference reformulated a robust multiobjective optimization problem [20] in the setting of (2) and some numerical experiments were considered for solving this problem. An inertial forward-backward algorithm was proposed in [14] for solving the vector optimization problem (2), where the Jacobian of the differentiable component is assumed to be Lipschitz. This scheme combines a gradient-like step and the proximal point iteration [13]. It is worth emphasizing that this algorithm requires a constraint set in the computation of the proximal step which makes, in general, the evaluation of the proximal operator more complicated. Moreover, in the latter algorithm the knowledge of the aforementioned Lipschitz constant is required in its fomulation.

The main goal of this paper is to analyze a variant of the PGM for solving the convex composite vector optimization problem (2). We extend to the vectorial setting the celebrated Beck–Teboulle's line-search, which makes possible to relax the Lipschitz continuity assumption and to prove the full convergence of the generated sequence to a weakly efficient solution of the problem. Additionally, an iteration complexity bound to obtain an approximate weakly efficient solution of problem (2) is established. Specifically, it is shown that, for a given tolerance $\varepsilon > 0$, an "ε-approximate solution" for problem (2) is obtained after at most $\mathcal{O}(1/\varepsilon)$ iterations.

Organization of the paper. Section 2 contains notations, definitions and some basic results. The concept of $\mathcal{K}$–convexity of a vector function and its properties are presented. A vector proximal scheme and the extension of Beck–Teboulle's line-search are discussed in Section 3. Section 4 formally describes the vector proximal gradient method for solving the structured vector optimization problem (2), and contains the convergence analysis of the algorithm. The last section is devoted to some final remarks and considerations.

2 Notation and Basic Results

In this section, following [1, 25], we formally state some basic definitions, results, and notations used in the paper.

In this paper, $\mathbb{N}$ and $\mathbb{R}^p$ denote the nonnegative integers $\{0, 1, 2, \ldots\}$ and a p-dimensional real vector space, respectively.

Let $\mathcal{K}$ be a closed, convex, pointed (i.e., $\mathcal{K} \cap (-\mathcal{K}) = \{0\}$) cone in $\mathbb{R}^m$ with nonempty interior. The partial order $\preceq$ $(\prec)$ induced in $\mathbb{R}^m$ by $\mathcal{K}$ is defined as $x \preceq y$ (or $x \prec y$) if and only if $y - x \in \mathcal{K}$ (or $y - x \in \operatorname{int}(\mathcal{K})$). The partial order $\succeq$ $(\succ)$ is defined in a similar way.

Similarly to [13, 14], we consider the extended $\mathbb{R}^m$, denoted by $\overline{\mathbb{R}}^m := \mathbb{R}^m \cup \{+\infty_{\mathcal{K}}\}$, where $+\infty_{\mathcal{K}}$ is assumed to be such that $y \prec +\infty_{\mathcal{K}}$, for every $y \in \mathbb{R}^m$. We also assume that

$$y + (+\infty_{\mathcal{K}}) = (+\infty_{\mathcal{K}}) + y = +\infty_{\mathcal{K}}, \quad t \cdot (+\infty_{\mathcal{K}}) = +\infty_{\mathcal{K}}, \qquad y \in \overline{\mathbb{R}}^m, t > 0.$$

The image of a function $F : \mathbb{R}^n \to \overline{\mathbb{R}}^m$ is the set $\operatorname{Im}(F) := \{y \in \mathbb{R}^m \mid y = F(x) \mid x \in \mathbb{R}^n\}$. Moreover, F is said to be proper if its domain defined by $\operatorname{dom}(F) := \{x \in \mathbb{R}^n \mid F(x) \prec +\infty_{\mathcal{K}}\}$ is nonempty.

Let $\langle \cdot, \cdot \rangle$ be the inner product and $\| \cdot \|$ be its induced norm. The positive dual cone $\mathcal{K}^*$ of $\mathcal{K}$ is defined by $\mathcal{K}^* := \{w \in \mathbb{R}^m \mid \langle x, w \rangle \geq 0, \forall x \in \mathcal{K}\}$. From now on, we assume that there exists a compact set $C \subset \mathcal{K}^* \setminus \{0\}$ such that $\operatorname{cone}(\operatorname{conv}(C)) = \mathcal{K}^*$ and $0 \notin C$. Following [13, 14], we assume by convention that $\langle +\infty_{\mathcal{K}}, w \rangle = +\infty$ for any $w \in C$. Moreover, we assume without loss of generality that $\|w\| = 1$ for every $w \in C$. Since $\mathcal{K} = \mathcal{K}^{**}$ (see [25, Theorem 14.1]), we have

$$-\mathcal{K} = \{u \in \mathbb{R}^m \mid \langle u, w \rangle \leq 0, \ \ \forall w \in C\} \tag{3}$$

$$-\operatorname{int}(\mathcal{K}) = \{u \in \mathbb{R}^m \mid \langle u, w \rangle < 0, \ \ \forall w \in C\}. \tag{4}$$

For a general cone $\mathcal{K}$, the generator set C as above can be chosen as $C = \{w \in \mathcal{K}^* \mid \|w\| = 1\}$. Frequently, it is possible to take much smaller sets for C, e.g. If $\mathcal{K}$ is a polyhedral cone then $\mathcal{K}^*$ is also polyhedral, hence C can be chosen as a normalized finite set of its extreme rays. For example, in scalar optimization, $\mathcal{K} = \mathbb{R}_+$ and we may take $C = \{1\}$. For multiobjective optimization, $\mathcal{K}$ and $\mathcal{K}^*$ are the positive orthant of $\mathbb{R}^m$, defined as $\mathbb{R}^m_+ := \{y \in \mathbb{R}^m \mid y_i \geq 0, \forall i = 1, \ldots, m\}$, and we may take C as the canonical basis of $\mathbb{R}^m$.

A vector function $F : \mathbb{R}^n \to \overline{\mathbb{R}}^m$ is called $\mathcal{K}$–convex if and only if for all $x, y \in \mathbb{R}^n$ and $\gamma \in [0, 1]$,

$$F(\gamma x + (1-\gamma)y) \preceq \gamma F(x) + (1-\gamma)F(y).$$

We consider $\partial F : \mathbb{R}^n \rightrightarrows \mathbb{R}^{m\times n}$ defined as

$$\partial F(x) = \{U \in \mathbb{R}^{m\times n} \mid F(y) \succeq F(x) + U(y-x),\ \forall y \in \mathbb{R}^n\}, \quad x \in \mathbb{R}^n.$$

If $G : \mathbb{R}^n \to \mathbb{R}^m$ is $\mathcal{K}-$convex and differentiable, the natural extension of the classical gradient inequality to the vectorial setting is:

$$G(y) \succeq G(x) + J_G(x)(y-x), \tag{5}$$

for any $x, y \in \mathbb{R}^n$, where $J_G(x)$ denotes the Jacobian matrix of G at x; see [25, Lemma 5.2].

Next we recall some definitions and properties for scalar convex functions which are easily derived from the vector convex functions properties presented before (i.e., taking $m = 1$). A scalar convex function $h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} := \overline{\mathbb{R}}$ is proper, if the domain of h, $\mathrm{dom}(h) := \{x \in \mathbb{R}^n \mid h(x) < +\infty\}$ is nonempty. The subdifferential of h at $x \in \mathbb{R}^n$ is defined by

$$\partial h(x) := \{u \in \mathbb{R}^n \mid h(y) \geq h(x) + \langle u, y-x\rangle,\ \forall y \in \mathbb{R}^n\}. \tag{6}$$

Moreover, the graph of $\partial h : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is given by

$$\mathrm{Gph}(\partial h) := \{(x, u) \in \mathbb{R}^n \times \mathbb{R}^n \mid u \in \partial h(x)\}.$$

The subdifferential ∂h is a monotone operator, i.e., for any $(x, u), (y, v) \in \mathrm{Gph}(\partial h)$, we have

$$\langle v-u, y-x\rangle \geq 0. \tag{7}$$

Next, we formally present the concept of weakly efficient solution.

Definition 2.1 *An element $x^* \in \mathbb{R}^n$ is a weakly efficient solution of problem* (2) *if there is no $x \in \mathbb{R}^n$ such that $F(x) \prec F(x^*)$.*

The next result, whose proof can be easily verified, presents a necessary and sufficient condition for a point to be weakly efficient.

Lemma 2.2 *A point $x \in \mathrm{dom}(F)$ is a weakly efficient solution of problem* (2) *if and only if*

$$\max_{w\in C} \langle F(u) - F(x), w\rangle \geq 0, \quad \forall u \in \mathbb{R}^n. \tag{8}$$

Let us end the section by recalling the well-known concept of Fejér convergence. The definition originates in [17] and has been elaborated further in [15, 24].

Definition 2.3 *Let S be a nonempty subset of $\mathbb{R}^n$. A sequence $\{x^k\}_{k\in\mathbb{N}}$ in $\mathbb{R}^n$ is said to be Fejér convergent to S if and only if for every $x \in S$, there holds $\|x^{k+1} - x\| \leq \|x^k - x\|$ for all $k \in \mathbb{N}$.*

Lemma 2.4 ([24, Theorem 4.1]) *If $\{x^k\}_{k\in\mathbb{N}}$ is Fejér convergent to S, then the following statements hold:*

(i) *the sequence $\{x^k\}_{k\in\mathbb{N}}$ is bounded;*

(ii) *if an accumulation point of $\{x^k\}_{k\in\mathbb{N}}$ belongs to S then $\{x^k\}_{k\in\mathbb{N}}$ converges to a point in S.*

A function $Q : \mathbb{R}^n \to \overline{\mathbb{R}}^m$ will be called positively lower semicontinuous if, for every $w \in C$, the scalar function $\langle Q(\cdot), w\rangle$ is lower semicontinuous.

Lemma 2.5 *Let $Q : \mathbb{R}^n \to \overline{\mathbb{R}}^m$ be a proper, $\mathcal{K}$-convex and positively lower semicontinuous function. Let $\{(y^k, U^k)\}_{k\in\mathbb{N}}$ be a bounded sequence in $\mathrm{Gph}(\partial Q)$ such that $\{y^k\}_{k\in\mathbb{N}}$ converges to $\bar{y}$. Then the following statements hold:*

(i) *the sequence $\{Q(y^k)\}_{k\in\mathbb{N}}$ is bounded;*

(ii) *if $\{w^k\}_{k\in\mathbb{N}} \subset C$ converges to $\bar{w}$, then $\liminf_{k\to+\infty} \langle Q(y^k), w^k\rangle \geq \langle Q(\bar{y}), \bar{w}\rangle$.*

Proof. (i) Suppose by contradiction that $\{Q(y^k)\}_{k\in\mathbb{N}}$ has an unbounded subsequence. Without loss of generality, let us assume that the whole sequence satisfies $\lim_{k\to+\infty} \|Q(y^k)\| = +\infty$. Let $z \in \mathrm{int}(\mathcal{K})$ (which is assumed to be nonempty). Then, there exists $t > 0$ sufficiently small such that $z^k := z + t\frac{Q(y^k)}{\|Q(y^k)\|} \in \mathcal{K}$. Since $\|z^k\| \leq \|z\| + t < +\infty$, we obtain by using the Cauchy-Schwarz inequality that for any $x \in \mathbb{R}^n$

$$(\|z\| + t)\|Q(x)\| \geq \langle Q(x), z^k\rangle. \tag{9}$$

Moreover, since $\{(y^k, U^k)\}_{k\in\mathbb{N}} \subset \mathrm{Gph}(\partial Q)$, we have

$$Q(x) \succeq Q(y^k) + U^k(x - y^k), \qquad \forall x \in \mathbb{R}^n, \forall k \in \mathbb{N},$$

which in turn implies that

$$\langle Q(x), z^k\rangle \geq \langle Q(y^k), z^k\rangle + \langle U^k(x - y^k), z^k\rangle, \qquad \forall x \in \mathbb{R}^n, \forall k \in \mathbb{N}.$$

In view of the definition of z^k, it follows by combining the last inequality with (9) and by using the Cauchy-Schwarz inequality that

$$\begin{aligned}(\|z\| + t)\|Q(x)\| &\geq \langle Q(y^k), z^k\rangle + \langle U^k(x - y^k), z^k\rangle \\ &\geq \langle Q(y^k), z\rangle + t\|Q(y^k)\| - \|U^k(x - y^k)\|\|z^k\|,\end{aligned}$$

for any $x \in \mathbb{R}^n, k \in \mathbb{N}$. Hence, since $\langle Q(\cdot), z\rangle$ is lower semicontinuous, we conclude that

$$\begin{aligned}+\infty = \lim_{k\to+\infty} t\|Q(y^k)\| &\leq (\|z\| + t)\|Q(x)\| - \liminf_{k\to+\infty} \langle Q(y^k), z\rangle + \liminf_{k\to+\infty} \|U^k\|\|x - y^k\|\|z^k\| \\ &\leq (\|z\| + t)\|Q(x)\| - \langle Q(\bar{y}), z\rangle + \|\bar{U}\|\|x - \bar{y}\|(\|z\| + t) < +\infty,\end{aligned}$$

which is a contradiction.

(ii) Let $\{w^k\}_{k\in\mathbb{N}} \subset C$ converging to $\bar{w}$. Clearly $\bar{w} \in C$, since C is closed. Then, it follows by using the Cauchy-Schwarz inequality, the lower semicontinuity of $\langle Q(\cdot), \bar{w}\rangle$, the boundedness of $\{Q(y^k)\}_{k\in\mathbb{N}}$ proved in (i), and the fact that $\{\|w^k - \bar{w}\|\}_{k\in\mathbb{N}}$ converges to zero, that

$$\liminf_{k\to+\infty} \langle Q(y^k), w^k\rangle \geq \liminf_{k\to+\infty} \langle Q(y^k), w^k - \bar{w}\rangle + \liminf_{k\to+\infty} \langle Q(y^k), \bar{w}\rangle = \langle Q(\bar{y}), \bar{w}\rangle,$$

concluding the proof. ■

3 A Proximal Regularization and the Line-search Procedure

This section is devoted to analyze a proximal regularization and a line-search procedure, which will be essential to introduce the proximal gradient method for solving composite vector optimization problems.

Let $x \in \mathrm{dom}(H)$ and $\alpha > 0$ be given. Consider the following function $\theta_{x,\alpha} : \mathbb{R}^n \to \overline{\mathbb{R}}$ defined as

$$\theta_{x,\alpha}(u) := \psi_x(u) + \frac{1}{2\alpha}\|u - x\|^2, \qquad \forall u \in \mathbb{R}^n, \tag{10}$$

where $\psi_x : \mathbb{R}^n \to \overline{\mathbb{R}}$ is given by

$$\psi_x(u) := \max_{w \in C} \langle J_G(x)(u - x) + H(u) - H(x), w\rangle, \qquad \forall u \in \mathbb{R}^n. \tag{11}$$

Moreover, define $p(x, \alpha)$ as

$$p(x,\alpha) := \arg\min_{u \in \mathbb{R}^n} \theta_{x,\alpha}(u). \tag{12}$$

Since ψ_x is convex, $\theta_{x,\alpha}$ is strongly convex. Hence, $p(x, \alpha)$ is uniquely determined and belongs to $\mathrm{dom}(H)$ (note that $\mathrm{dom}(F) = \mathrm{dom}(H) = \mathrm{dom}(\theta_{x,\alpha}) = \mathrm{dom}(\psi_x)$). It is worth noting that when $m = 1$ in problem (2), $p(x, \alpha)$ is directly related to the classical *forward-backward* operator evaluated at x.

In the following, we present some basic properties associating the above elements with weakly efficient solutions of problem (2). First, let us fix a point $\zeta \in \mathrm{int}(\mathcal{K})$ such that

$$0 < \delta_1 := \min_{w \in C} \langle \zeta, w\rangle \le \max_{w \in C} \langle \zeta, w\rangle \le 1. \tag{13}$$

Since $\mathrm{int}(\mathcal{K}) \neq \emptyset$ and C is a compact set satisfying (4), it is easy to show that there always exists a point $\zeta \in \mathrm{int}(\mathcal{K})$ satisfying the above condition. For multiobjective optimization where $\mathcal{K} = \mathbb{R}^m_+$, for example, we can choose $\zeta = (1, \dots, 1)$.

Lemma 3.1 *Let $x \in \mathrm{dom}(F)$ and $\alpha > 0$ be given, and let $p(x, \alpha)$ be computed as in* (12)*. Then, the following statements hold:*

(i) *for every $z \in \mathrm{dom}(F)$ and $\gamma \in (0, 1)$, we have*

$$\theta_{x,\alpha}(p(x,\alpha)) \le \gamma \max_{w \in C} \langle F(z) - F(x), w\rangle + \frac{\gamma^2}{2\alpha}\|z - x\|^2; \tag{14}$$

(ii) *$\theta_{x,\alpha}(p(x, \alpha)) \le 0$, and $\theta_{x,\alpha}(p(x, \alpha)) = 0$ if and only if $p(x, \alpha) = x$;*

(iii) *if $p(x, \alpha) = x$ then x is a weakly efficient solution of problem* (2)*;*

(iv) *let ζ be as in* (13) *and assume that*

$$G(p(x,\alpha)) - G(x) \preceq J_G(x)(p(x,\alpha) - x) + \frac{1}{2\alpha}\|p(x,\alpha) - x\|^2\zeta. \tag{15}$$

Then

$$\max_{w \in C} \langle F(p(x,\alpha)) - F(x), w\rangle \le \theta_{x,\alpha}(p(x,\alpha)). \tag{16}$$

As a consequence, if x is a weakly efficient solution of problem (2) *then $p(x, \alpha) = x$.*

Proof. (i) Let $z \in \mathrm{dom}(F)$ and $\gamma \in (0,1)$. It follows from (10)-(12) that

$$\begin{aligned}\theta_{x,\alpha}(p(x,\alpha)) \leq& \theta_{x,\alpha}(x+\gamma(z-x)) \\ =& \max_{w\in C}\langle J_G(x)\gamma(z-x) + H(x+\gamma(z-x)) - H(x), w\rangle + \frac{1}{2\alpha}\|\gamma(z-x)\|^2.\end{aligned}$$

Since G is differentiable and $\mathcal{K}$-convex, we have from (3) and (5) that for every $w \in C$, holds

$$\gamma\langle J_G(x)(z-x), w\rangle \leq \langle G(x+\gamma(z-x)) - G(x), w\rangle.$$

Combining the last two inequalities, we conclude that

$$\theta_{x,\alpha}(p(x,\alpha)) \leq \max_{w\in C}\langle G(x+\gamma(z-x)) - G(x) + H(x+\gamma(z-x)) - H(x), w\rangle + \frac{\gamma^2}{2\alpha}\|z-x\|^2.$$

Hence, using the $\mathcal{K}$-convexity of G and H, we obtain

$$\theta_{x,\alpha}(p(x,\alpha)) \leq \gamma\max_{w\in C}\langle G(z) - G(x) + H(z) - H(x), w\rangle + \frac{\gamma^2}{2\alpha}\|z-x\|^2.$$

The result follows trivially from the definition of F, given in problem (2).

(ii) This statement follows immediately from the fact that $\theta_{x,\alpha}(\cdot)$ is strongly convex, $p(x,\alpha)$ is uniquely determined and $\theta_{x,\alpha}(p(x,\alpha)) \leq \theta_{x,\alpha}(x) = 0$.

(iii) Assume that $p(x,\alpha) = x$. In view of (ii), we have $\theta_{x,\alpha}(p(x,\alpha)) = 0$. Then, it follows from (i) that

$$0 = \theta_{x,\alpha}(p(x,\alpha)) \leq \gamma\max_{w\in C}\langle F(z) - F(x), w\rangle + \frac{\gamma^2}{2\alpha}\|z-x\|^2, \qquad \forall z \in \mathrm{dom}(F), \forall\gamma \in (0,1).$$

Hence, dividing by γ and letting $\gamma \downarrow 0$, we have $\max_{w\in C}\langle F(z) - F(x), w\rangle \geq 0$ for every $z \in \mathrm{dom}(F)$. By convention $\langle +\infty_{\mathcal{K}}, w\rangle = +\infty$ for any $w \in C$, hence the latter inequality trivially holds for $z \notin \mathrm{dom}(F)$. Using Lemma 2.2, we conclude that x is a weakly efficient solution of (2).

(iv) Using (11), (15), the decomposition $F = G + H$, and the last inequality in (13), we obtain that, for any $w \in C$,

$$\begin{aligned}\langle F(p(x,\alpha)), w\rangle &= \langle F(x) + G(p(x,\alpha)) - G(x) + H(p(x,\alpha)) - H(x), w\rangle \\ &\leq \langle F(x) + J_G(x)(p(x,\alpha)-x) + H(p(x,\alpha)) - H(x) + \frac{1}{2\alpha}\|p(x,\alpha)-x\|^2\zeta, w\rangle \\ &\leq \langle F(x), w\rangle + \psi_x(p(x,\alpha)) + \frac{1}{2\alpha}\|p(x,\alpha)-x\|^2,\end{aligned}$$

which clearly proves (16), in view of the definition of $\theta_{x,\alpha}$ given in (10). Now in order to prove the last statement in (iv), assume by contradiction that $p(x,\alpha) \neq x$. Then, combining (ii) and (16), we obtain $\langle F(p(x,\alpha)) - F(x), w\rangle < 0$ for any $w \in C$. The latter conclusion implies that x is not a weakly efficient solution of problem (2), see (4), which is a contradiction. ■

Next we present a monotone property satisfied by $p(x,\alpha)$ associated with the proximal gradient iteration, which is essential to our analysis. Although the proof is similar to the one presented in Lemma 2.4 of [7] for the scalar case, we present its proof for the sake of completeness.

Lemma 3.2 *For any $x \in \mathrm{dom}(F)$ and $\alpha_2 \geq \alpha_1 > 0$, we have*

$$\frac{\alpha_2}{\alpha_1}\|x - p(x,\alpha_1)\| \geq \|x - p(x,\alpha_2)\| \geq \|x - p(x,\alpha_1)\|. \tag{17}$$

Proof. Given $x \in \mathrm{dom}(F)$, it follows from the first-order optimality condition of (12) that, for every $\alpha > 0$,

$$\frac{x - p(x,\alpha)}{\alpha} \in \partial\psi_x(p(x,\alpha)).$$

Let arbitrary $\alpha_2 \geq \alpha_1 > 0$ be given. From the latter inclusion and the monotonicity of $\partial\psi_x$ given in (7), we have

$$\begin{aligned} 0 &\leq \left\langle \frac{x - p(x,\alpha_2)}{\alpha_2} - \frac{x - p(x,\alpha_1)}{\alpha_1}, p(x,\alpha_2) - p(x,\alpha_1) \right\rangle \\ &= \left\langle \frac{x - p(x,\alpha_2)}{\alpha_2} - \frac{x - p(x,\alpha_1)}{\alpha_1}, (x - p(x,\alpha_1)) - (x - p(x,\alpha_2)) \right\rangle \\ &= -\frac{\|x - p(x,\alpha_2)\|^2}{\alpha_2} - \frac{\|x - p(x,\alpha_1)\|^2}{\alpha_1} + \left(\frac{1}{\alpha_1} + \frac{1}{\alpha_2}\right) \langle x - p(x,\alpha_2), x - p(x,\alpha_1)\rangle. \end{aligned}$$

Hence, multiplying the above inequality by α_2 and using the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned} 0 &\leq -\|x - p(x,\alpha_2)\|^2 - \frac{\alpha_2}{\alpha_1}\|x - p(x,\alpha_1)\|^2 + \left(\frac{\alpha_2}{\alpha_1} + 1\right) \|x - p(x,\alpha_2)\| \cdot \|x - p(x,\alpha_1)\| \\ &= (\|x - p(x,\alpha_2)\| - \|x - p(x,\alpha_1)\|) \cdot \left(\frac{\alpha_2}{\alpha_1}\|x - p(x,\alpha_1)\| - \|x - p(x,\alpha_2)\|\right). \end{aligned}$$

Since $\frac{\alpha_2}{\alpha_1} \geq 1$, both above factors are nonnegative which clearly implies (17). ■

Next we present a useful result.

Lemma 3.3 *Let $\{y^k\}_{k\in\mathbb{N}} \subset \mathrm{dom}(F)$ be a bounded sequence. Then $\{p(y^k,\beta)\}_{k\in\mathbb{N}}$ is also bounded for any $\beta > 0$.*

Proof. Let $p^k := p(y^k,\beta)$ and note that the optimality condition of (12) with $x = y^k$, $\alpha = \beta$ yields

$$\frac{y^k - p^k}{\beta} \in \partial\psi_{y^k}(p^k), \qquad \forall k \in \mathbb{N}.$$

Hence, using the subgradient inequality (6), we have

$$\psi_{y^k}(u) \geq \psi_{y^k}(p^k) + \frac{1}{\beta}\langle y^k - p^k, u - p^k\rangle, \qquad \forall u \in \mathbb{R}^n, \forall k \in \mathbb{N}.$$

Using the above inequality twice, firstly with $u = p^0$ and then with $k = 0$ and $u = p^k$, and adding the resulting inequalities, we obtain

$$\begin{aligned} \psi_{y^k}(p^0) + \psi_{y^0}(p^k) &\geq \psi_{y^k}(p^k) + \psi_{y^0}(p^0) + \frac{1}{\beta}\langle y^k - p^k, p^0 - p^k\rangle + \frac{1}{\beta}\langle y^0 - p^0, p^k - p^0\rangle \\ &= \psi_{y^k}(p^k) + \psi_{y^0}(p^0) + \frac{1}{\beta}\langle y^0 - y^k, p^k - p^0\rangle + \frac{1}{\beta}\|p^k - p^0\|^2. \end{aligned} \tag{18}$$

Note that in view of (11) and the compactness of C, there exists a bounded sequence $\{w^k\} \subset C$ such that

$$\psi_{y^0}(p^k) = \tilde{\psi}^k_{y^0}(p^k) + \langle H(p^k), w^k\rangle, \qquad \forall k \in \mathbb{N}, \tag{19}$$

where, for every $y \in \text{dom}(F)$,

$$\tilde{\psi}^k_y(p) = \langle J_G(y)(p-y) - H(y), w^k \rangle, \qquad \forall k \in \mathbb{N}. \tag{20}$$

Since (11) implies that $\psi_{y^k}(p^k) \geq \tilde{\psi}^k_{y^k}(p^k) + \langle H(p^k), w^k \rangle$, we obtain from (18) and (19) that

$$\begin{aligned} &\psi_{y^k}(p^0) + \tilde{\psi}^k_{y^0}(p^k) + \langle H(p^k), w^k \rangle \\ &\geq \tilde{\psi}^k_{y^k}(p^k) + \langle H(p^k), w^k \rangle + \psi_{y^0}(p^0) + \frac{1}{\beta}\langle y^0 - y^k, p^k - p^0 \rangle + \frac{1}{\beta}\|p^k - p^0\|^2. \end{aligned}$$

Simplifying $\langle H(p^k), w^k \rangle$ and rewriting the above inequality, we get

$$\frac{1}{\beta}\|p^k - p^0\|^2 \leq \psi_{y^k}(p^0) + \tilde{\psi}^k_{y^0}(p^k) - \psi_{y^0}(p^0) - \tilde{\psi}^k_{y^k}(p^k) - \frac{1}{\beta}\langle y^0 - y^k, p^k - p^0 \rangle.$$

Note that $\tilde{\psi}^k_{y^k}(p^k)$ does not contain the term $H(p^k)$. Moreover, since $\{y^k\}_{k\in\mathbb{N}}$ and $\{w^k\}_{k\in\mathbb{N}}$ are bounded, we easily see that the left hand side of the above inequality is $\mathcal{O}(\|p^k - p^0)\|^2)$ whereas the right hand side is $\mathcal{O}(\|p^k - p^0)\|)$, in view of the definitions of ψ and $\tilde{\psi}^k$ given in (11) and (20), respectively. This observation clearly implies that $\{\|p^k - p^0\|\}_{k\in\mathbb{N}}$ is bounded, proving the lemma.

■

Next we present an extension of Beck-Teboulle's backtracking line-search [2] to the vectorial optimization setting.

Line-search Procedure

Step 0. Given ζ as in (13) and $\eta \in (0,1)$. Input $(x, \sigma) \in \text{dom}(F) \times \mathbb{R}_{++}$;

Step 1. Let $p(x, \alpha)$ be computed as in (12) with $\alpha = \sigma$;

Step 2. If

$$G(p(x,\alpha)) - G(x) \preceq J_G(x)(p(x,\alpha) - x) + \frac{1}{2\alpha}\|p(x,\alpha) - x\|^2\zeta, \tag{21}$$

stop and output $(p(x,\alpha), \alpha)$. Otherwise, set $\alpha := \eta\alpha$ and return to Step 1.

end

Some remarks about the above line-search procedure follows. First, if J_G is Lipschitz continuous with Lipschitz constant L, then any $\alpha \leq 1/L$ satisfies the vector inequality in (21). Second, as it will be shown in the next proposition, the above backtracking procedure stops after a finite number of steps even if J_G fails to be Lipschitz continuous. Note that, the Lipschitz continuity is required for proving the well-definition of Beck-Teboulle's line-search in [2]. Third, it is worth mentioning that this procedure may also be useful as a subroutine of forward-backward type algorithms for solving problems in which the Lipschitz constant of J_G exists but is not known or hard to estimate. Finally, for convenience, we use the notation $(p(x,\alpha), \alpha) = LS(x, \sigma)$ to refer to the output of the above line-search procedure with input (x, σ).

In the following, we present a technical lemma regarding the violation of the vector inequality (21). This result will be useful to prove the finite termination of the line-search procedure and to conclude that our vector proximal gradient algorithm converges to a weakly efficient solution of problem (2).

Lemma 3.4 *Let $\{\beta_k\}_{k\in\mathbb{N}} \subset (0,\sigma]$ be a sequence converging to zero and let $\{y^k\}_{k\in\mathbb{N}} \subset \operatorname{dom}(F)$ be a bounded sequence such that the inequality in* (21) *fails with $(x,\alpha) = (y^k, \beta_k)$, for every $k \in \mathbb{N}$. Then,*

$$\lim_{k\to+\infty} \frac{\|y^k - p(y^k,\beta_k)\|}{\beta_k} = 0.$$

Proof. First, for convenience, denote $\hat{y}^k := p(y^k,\beta_k)$. Since, for every $k \in \mathbb{N}$, the inequality in (21) fails with $(x,\alpha) = (y^k,\beta_k)$, it follows that there exists $w^k \in C$ such that

$$\begin{aligned}\langle J_G(y^k)(\hat{y}^k - y^k), w^k\rangle + \frac{1}{2\beta_k}\|y^k - \hat{y}^k\|^2\langle \zeta, w^k\rangle &< \langle G(\hat{y}^k) - G(y^k), w^k\rangle \\ &\le \langle J_G(\hat{y}^k)(\hat{y}^k - y^k), w^k\rangle,\end{aligned}$$

where the last inequality follows by the $\mathcal{K}$–convexity of G and (5). Hence, rewriting the above inequality and using the Cauchy-Schwarz inequality and the fact that $\langle w^k, \zeta\rangle \ge \delta_1 > 0$, we obtain

$$\begin{aligned}\frac{\delta_1}{2\beta_k}\left\|\hat{y}^k - y^k\right\|^2 &< \left\langle \left[J_G(\hat{y}^k) - J_G(y^k)\right](\hat{y}^k - y^k), w^k\right\rangle \\ &\le \|w^k\| \left\|J_G(\hat{y}^k) - J_G(y^k)\right\| \cdot \left\|\hat{y}^k - y^k\right\|.\end{aligned}$$

The above inequalities together with that fact that $\|w^k\| = 1$ imply that $\hat{y}^k \neq y^k$ and

$$\left\|\hat{y}^k - y^k\right\| < 2\frac{\beta_k}{\delta_1}\left\|J_G(\hat{y}^k) - J_G(y^k)\right\|, \quad \forall k \in \mathbb{N}. \tag{22}$$

Since $\beta_k \le \sigma$, using the triangle inequality and Lemma 3.2, we obtain

$$\begin{aligned}\|\hat{y}^k\| \le \|y^k - \hat{y}^k\| + \|y^k\| &\le \|y^k - p(y^k,\sigma)\| + \|y^k\| \\ &\le \|y^k - p(y^0,\sigma)\| + \|p(y^0,\sigma) - p(y^k,\sigma)\| + \|y^k\|.\end{aligned}$$

It follows from the above inequalities combined with boundedness of $\{y^k\}_{k\in\mathbb{N}}$ and Lemma 3.3 that $\{\hat{y}^k\}_{k\in\mathbb{N}}$ is bounded, which in turn, in view of the continuity of J_G, implies that $\{\|J_G(\hat{y}^k)\|\}_{k\in\mathbb{N}}$ is also bounded. Hence, since $\{\beta_k\}_{k\in\mathbb{N}}$ converges to zero, it follows from (22) that $\{\|y^k - \hat{y}^k\|\}_{k\in\mathbb{N}}$ converges to zero. It follows then by using again the continuity of J_G and (22) that

$$\lim_{k\to+\infty} \frac{\|y^k - \hat{y}^k\|}{\beta_k} = 0, \tag{23}$$

which proves the lemma. ∎

Next we show that the above line-search procedure provides an output after a finite number of steps, without assuming J_G to be Lipschitz continuous.

Proposition 3.5 *Let $x \in \operatorname{dom}(F)$. If x is not a weakly efficient solution of problem* (2)*, then the line-search procedure stops after a finite number of steps.*

Proof. Let $x \in \text{dom}(F)$ and assume that it is not a weakly efficient solution of problem (2). Suppose by contradiction that the line-search procedure does not stop after a finite number of steps. Hence, for every $k \in \mathbb{N}$, the vector inequality in (21) is violated with $\alpha = \alpha_k := \sigma\eta^k$. It follows from Lemma 3.4 with $(y^k, \beta_k) = (x, \alpha_k)$ that

$$\lim_{k\to+\infty} \frac{\|x - p(x, \alpha_k)\|}{\alpha_k} = 0. \tag{24}$$

In particular, $\{p(x, \alpha_k)\}_{k\in\mathbb{N}}$ converges to x, in view of $\{\alpha_k\}_{k\in\mathbb{N}} \downarrow 0$. On the other hand, we have from the optimality condition of (12) that

$$\frac{x - p(x, \alpha_k)}{\alpha_k} \in \partial\psi_x(p(x, \alpha_k)). \tag{25}$$

Hence, since the graph of $\partial\psi_x$ is closed and $\{p(x, \alpha_k)\}_{k\in\mathbb{N}}$ converges to x, we obtain

$$0 \in \partial\psi_x(x).$$

Then, in view of the convexity of ψ_x, we conclude that x minimizes ψ_x. Therefore, in view of the definition of ψ_x, we have, for every $u \in \mathbb{R}^n$,

$$\begin{aligned} 0 =&\psi_x(x) \le \psi_x(u) = \max_{w\in C}\langle J_G(x)(u - x) + H(u) - H(x), w\rangle \\ \le& \max_{w\in C}\langle G(u) - G(x) + H(u) - H(x), w\rangle \\ =& \max_{w\in C}\langle F(u) - F(x), w\rangle, \end{aligned}$$

where the second inequality is due the $\mathcal{K}$–convexity of G and (5). Hence, Lemma 2.2 implies that x is a weakly efficient solution of problem (2), which contradicts the assumption of the lemma. ■

4 Proximal Gradient Method and Convergence Analysis

This section states the vector proximal gradient method and analyzes its convergence properties. It is shown that, under some standard assumptions, the whole sequence generated by the proposed scheme converges to a weakly efficient solution of (2).

The proximal gradient method is stated as below.

Vector Proximal Gradient Method (VPGM) with Line-search

Step 0. Let $(x^0, \sigma) \in \text{dom}(F) \times \mathbb{R}_{++}$ be given, set $k = 0$;

Step 1. If x^k is a weakly efficient solution of (2), stop and output x^k;

Step 2. Use the line-search procedure to compute $(p(x^k, \alpha), \alpha) = LS(x^k, \sigma)$ and set

$$\alpha_k = \alpha, \quad x^{k+1} = p(x^k, \alpha_k);$$

set $k \leftarrow k + 1$ and return to step 1.

end

Some comments are in order. First, the VPGM is a natural extension to the vectorial setting of the well known proximal gradient algorithm [2, 7] for solving composite convex optimization problems. Second, the main iterative step consists of checking if the current point is a weakly efficient solution of problem (2) and if it is not, then the line-search procedure of Section 3 is invoked in order to compute the next iterate. Third, the line-search procedure returns the next iterate after a finite number of steps, in view of the stopping criterion and Proposition 3.5. Moreover, if x^k is not a weakly efficient solution of problem (2), then in view of (21) and the definition of x^{k+1}, we have

$$G(x^{k+1}) - G(x^k) \preceq J_G(x^k)(x^{k+1} - x^k) + \frac{1}{2\alpha_k}\|x^{k+1} - x^k\|^2\zeta, \qquad \forall k \in \mathbb{N}. \tag{26}$$

As a consequence, in view of the definition of x^{k+1} and Lemma 3.1(ii) and (iv), we obtain

$$F(x^{k+1}) \preceq F(x^k), \tag{27}$$

showing that the VPGM is a decreasing scheme in the order given by $\mathcal{K}$.

Note that if the VPGM stops at some iteration k, then x^k is a weakly efficient solution of problem (2). Hence, in our analysis, we assume that the VPGM generates an infinite sequence.

Next we present a key inequality satisfied by the sequence generated by the VPGM, which is essential to prove its full convergence.

Proposition 4.1 *Let $\{x^k\}_{k\in\mathbb{N}}$ be generated by the VPGM. Then, for every $x \in \operatorname{dom}(F)$, the following inequality holds*

$$\|x^{k+1} - x\|^2 \leq \|x^k - x\|^2 + 2\alpha_k \max_{w\in C}\left\langle F(x) - F(x^{k+1}), w\right\rangle, \quad \forall\, k \in \mathbb{N}.$$

Proof. Since $x^{k+1} = p(x^k, \alpha_k)$, it follows from (10)–(12) that

$$v^k := \frac{x^k - x^{k+1}}{\alpha_k} \in \partial\psi_{x^k}(x^{k+1}). \tag{28}$$

Hence, the subgradient inequality and the definition of $\psi_{x^k}(\cdot)$ given in (11) imply that

$$\begin{aligned}\psi_{x^k}(x) &\geq \psi_{x^k}(x^{k+1}) + \langle v^k, x - x^{k+1}\rangle \\ &= \psi_{x^k}(x^{k+1}) + \langle v^k, x - x^k\rangle + \frac{1}{\alpha_k}\|x^k - x^{k+1}\|^2.\end{aligned} \tag{29}$$

On the other hand, using the decomposition $F = G + H$, (11), the $\mathcal{K}$–convexity of G, and (5), we have

$$\begin{aligned}\max_{w\in C}\langle F(x) - F(x^k), w\rangle &= \max_{w\in C}\langle G(x) - G(x^k) + H(x) - H(x^k), w\rangle \\ &\geq \max_{w\in C}\langle J_G(x^k)(x - x^k) + H(x) - H(x^k), w\rangle = \psi_{x^k}(x).\end{aligned}$$

In view of step 2 of VPGM and (21), we have

$$\begin{aligned}\psi_{x^k}(x^{k+1}) &= \max_{w\in C}\langle J_G(x^k)(x^{k+1} - x^k) + H(x^{k+1}) - H(x^k), w\rangle \\ &\geq \max_{w\in C}\langle G(x^{k+1}) - G(x^k) - \frac{1}{2\alpha_k}\|x^{k+1} - x^k\|^2\zeta + H(x^{k+1}) - H(x^k), w\rangle \\ &= \max_{w\in C}\langle F(x^{k+1}) - F(x^k) - \frac{1}{2\alpha_k}\|x^{k+1} - x^k\|^2\zeta, w\rangle.\end{aligned}$$

Combining the latter two inequalities with (29), we obtain for every $w \in C$:

$$\begin{aligned}\max_{w\in C}\langle F(x)-F(x^k),w\rangle &\geq \psi_{x^k}(x^{k+1})+\langle v^k,x-x^k\rangle+\frac{1}{\alpha_k}\|x^k-x^{k+1}\|^2\\ &\geq \langle F(x^{k+1})-F(x^k),w\rangle-\frac{\langle\zeta,w\rangle}{2\alpha_k}\|x^{k+1}-x^k\|^2+\langle v^k,x-x^k\rangle+\frac{1}{\alpha_k}\|x^k-x^{k+1}\|^2\\ &\geq \langle F(x^{k+1})-F(x^k),w\rangle+\langle v^k,x-x^k\rangle+\frac{1}{2\alpha_k}\|x^{k+1}-x^k\|^2,\end{aligned}$$

where the last inequality is due to the fact that $\langle\zeta,w\rangle\leq 1$, in view of (13).

Hence, since C is compact, there exists $w^k \in C$ such that the above maximum is attained, and then rewriting the above inequality with $w = w^k$, we easily obtain

$$2\alpha_k\langle v^k,x-x^k\rangle+\|x^k-x^{k+1}\|^2\leq 2\alpha_k\langle F(x)-F(x^{k+1}),w^k\rangle\leq 2\alpha_k\max_{w\in C}\langle F(x)-F(x^{k+1}),w\rangle.$$

The result then follows by noting that

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

and $\alpha_k\langle v^k,x-x^k\rangle=\langle x^{k+1}-x^k,x^k-x\rangle$, in view of the definition of v^k in (28). ■

In order to prove the convergence of the sequence generated by the VPGM, the following standard assumption is required:

$$\Omega=\{x\in\mathbb{R}^n \mid F(x)\preceq F(x^k),\ \forall k\in\mathbb{N}\}\neq\emptyset.$$

This assumption has been considered in the vector/multiobjective optimization literature for the convergence analysis of several methods; see, for instance [8, 10, 13, 23]. In view of the decreasing property of the sequence $\{F(x^k)\}_{k\in\mathbb{N}}$ (see (27)), the above assumption holds automatically if Im(F) is complete. Recall that Im(F) is said to be complete if for every sequence $\{y^k\}_{k\in\mathbb{N}}$ satisfying $F(y^{k+1})\preceq F(y^k)$ for all $k\in\mathbb{N}$, there exists $y\in\mathbb{R}^n$ such that $F(y)\preceq F(y^k)$ for all $k\in\mathbb{N}$. The completeness of $\mathrm{Im}(F)$ ensures the existence of weakly efficient solutions for vector optimization problems (see [25, Section 3]). In scalar optimization, it is equivalent to the existence of solution.

Before proceeding, it will be useful to consider the following sequence

$$\theta_k:=\theta_{x^k,\alpha_k}(x^{k+1}),\qquad \forall k\in\mathbb{N}, \tag{30}$$

where $\theta_{x,\alpha}$ is as defined in (10). Note that Lemma 3.1(ii)-(iv) imply that $\theta_k\leq 0$ for every k, and $\theta_k=0$ if and only if $x^{k+1}=x^k$, in which case the VPGM stops obtaining a weakly efficient solution of problem (2).

Now note that Lemma 3.1(iv) combined with the definition of x^{k+1} yield

$$|\theta_k|=-\theta_k\leq\langle F(x^k)-F(x^{k+1}),w\rangle,\qquad \forall w\in C. \tag{31}$$

Next, we establish our main result, which shows that the whole sequence generated by the VPGM converges without assuming Lipschitz continuity on J_G.

Theorem 4.2 *The sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by the VPGM converges to a weakly efficient solution of problem* (2).

Proof. It follows from Proposition 4.1 that $\{x^k\}_{k\in\mathbb{N}}$ is Fejér convergent to Ω, which implies by Lemma 2.4 that $\{x^k\}_{k\in\mathbb{N}}$ is bounded. Let $\bar{x}$ be an accumulation point and consider a subsequence $\{x^{i_k}\}_{k\in\mathbb{N}}$ converging to $\bar{x}$. Since $\{F(x^k)\}_{k\in\mathbb{N}}$ is $\mathcal{K}$–decreasing (in view of (27)), we obtain that $\bar{x} \in \Omega$. In fact, for any $w \in C$,

$$\langle F(x^k), w\rangle \geq \langle F(x^{i_\ell}), w\rangle, \quad \forall\, k, \ell \in \mathbb{N} \text{ such that } k \leq i_\ell.$$

Hence, using that $\langle F(\cdot), w\rangle$ is lower semicontinuous for any $w \in C$, we get

$$\langle F(x^k), w\rangle \geq \liminf_{\ell\to+\infty} \langle F(x^{i_\ell}), w\rangle \geq \langle F(\bar{x}), w\rangle.$$

Thus, from Lemma 2.4 the whole sequence $\{x^k\}_{k\in\mathbb{N}}$ converges to a point $\bar{x} \in \Omega$. Let us now show that $\bar{x}$ is a weakly efficient solution of problem (2).

We split our analysis in two cases:

Case 1. There exists $\bar{\alpha} > 0$ such that $\bar{\alpha} \leq \alpha_k$ for all $k \in \mathbb{N}$.

Since $\{F(x^k)\}_{k\in\mathbb{N}}$ is $\mathcal{K}$–decreasing and $F(\bar{x}) \preceq F(x^k)$, we obtain that, for every $w \in C$, $\{\langle F(x^k), w\rangle\}_{k\in\mathbb{N}}$ is decreasing and bounded below by $\langle F(\bar{x}), w\rangle$. Hence, $\{\langle F(x^k), w\rangle\}_{k\in\mathbb{N}}$ converges, and then (31) implies that $\lim_{k\to+\infty} \theta_k = 0$. Now, since $\{\alpha_k\}_{k\in\mathbb{N}}$ is bounded from below by $\bar{\alpha}$, it follows from Lemma 3.1(i) that, for all $z \in \mathbb{R}^n$, $\gamma \in (0,1)$, and $k \in \mathbb{N}$,

$$\begin{aligned}\theta_k &\leq \gamma \max_{w\in C} \langle F(z) - F(x^k), w\rangle + \frac{\gamma^2}{2\alpha_k}\|z - x^k\|^2 \\ &\leq \gamma\langle F(z) - F(x^k), w^k\rangle + \frac{\gamma^2}{2\bar{\alpha}}\|z - x^k\|^2, \end{aligned} \tag{32}$$

where w^k is the point in C achieving the above maximum. Since C is compact, we may assume without loss of generality that $\{w^k\}_{k\in\mathbb{N}}$ converges to some $\bar{w}$. Hence, (32) together with Lemma 2.5(ii) and the fact that $\{\theta_k\}_{k\in\mathbb{N}}$ converges to zero imply that

$$0 \leq \gamma\langle F(z) - F(\bar{x}), \bar{w}\rangle + \frac{\gamma^2}{2\bar{\alpha}}\|z - \bar{x}\|^2, \quad \forall z \in \mathbb{R}^n.$$

Now, dividing both sides of the above inequality by γ and letting $\gamma \downarrow 0$, we obtain

$$0 \leq \langle F(z) - F(\hat{x}), \bar{w}\rangle \leq \max_{w\in C} \langle F(z) - F(\bar{x}), w\rangle, \quad \forall z \in \mathbb{R}^n,$$

which implies from Lemma 2.2 that $\bar{x}$ is a weakly efficient solution of problem (2).

Now, we consider the second case.

Case 2. There exists a subsequence of $\{\alpha_k\}_{k\in\mathbb{N}}$ converging to zero.

Suppose without loss of generality that $\{\alpha_k\}_{k\in\mathbb{N}}$ converges to zero. Define $\hat{\alpha}_k := \frac{\alpha_k}{\eta} > 0$ and $\hat{x}^k := p(x^k, \hat{\alpha}_k)$. Hence, it follows from step 2 of the VPGM and the line-search procedure that the vector inequality in (21) is violated with $(x, \alpha) = (\hat{x}^k, \hat{\alpha}_k)$. Hence, Lemma 3.4 with $(y^k, \beta_k) = (x^k, \hat{\alpha}_k)$ implies that

$$\lim_{k\to+\infty} \frac{\|x^k - \hat{x}^k\|}{\hat{\alpha}_k} = 0. \tag{33}$$

On the other hand, we have from the optimality condition for $\hat{x}^k$ (see (10)–(12)),

$$\hat{v}^k := \frac{x^k - \hat{x}^k}{\hat{\alpha}_k} \in \partial\psi_{x^k}(\hat{x}^k), \tag{34}$$

which implies that

$$\psi_{x^k}(x) \geq \psi_{x^k}(\hat{x}^k) + \langle \hat{v}^k, x - \hat{x}^k \rangle.$$

Consider $w^k \in C$ achieving the maximum in the expression of the function $\psi_{x^k}(x)$ given in (11), i.e., $\psi_{x^k}(x) = \langle J_G(x^k)(x - x^k) + H(x) - H(x^k), w^k \rangle$. Hence, the last inequality and (11) imply

$$\langle J_G(x^k)(x - x^k) + H(x) - H(x^k), w^k \rangle \geq \langle J_G(x^k)(\hat{x}^k - x^k) + H(\hat{x}^k) - H(x^k), w^k \rangle + \langle \hat{v}^k, x - \hat{x}^k \rangle.$$

Then,

$$\langle J_G(x^k)(x - x^k) + H(x), w^k \rangle \geq \langle J_G(x^k)(\hat{x}^k - x^k) + H(\hat{x}^k), w^k \rangle + \langle \hat{v}^k, x - \hat{x}^k \rangle, \tag{35}$$

which together with the $\mathcal{K}$-convexity of G and (5) yield

$$\langle G(x) - G(x^k) + H(x), w^k \rangle \geq \langle J_G(x^k)(\hat{x}^k - x^k), w^k \rangle + \langle H(\hat{x}^k), w^k \rangle + \langle \hat{v}^k, x - x^k \rangle.$$

Since $\{w^k\}_{k \in \mathbb{N}}$ is bounded, we can assume without loss of generality that it converges to some $\bar{w}$. Now, taking into account that $\{x^k\}_{k \in \mathbb{N}}$ converges to $\bar{x}$, it follows from (33)-(34) that $\{\hat{v}^k\}$ converges to zero and $\{\hat{x}^k\}_{k \in \mathbb{N}}$ converges to $\bar{x}$. Thus, by taking $\liminf$, as k goes to $+\infty$, in the above inequality and using Lemma 2.5(ii), we obtain that

$$\langle G(x) - G(\bar{x}) + H(x), \bar{w} \rangle \geq \langle H(\bar{x}), \bar{w} \rangle,$$

which implies that

$$\max_{w \in C} \langle G(x) - G(\bar{x}) + H(x) - H(\bar{x}), w \rangle \geq \langle G(x) - G(\bar{x}) + H(x) - H(\bar{x}), \bar{w} \rangle \geq 0, \quad \forall x \in \mathbb{R}^n.$$

Hence, since $F = G + H$, Lemma 2.2 implies that $\bar{x}$ is a weakly efficient solution of problem (2). ■

We end this section by establishing an iteration-complexity bound to obtain an approximate weakly efficient solution of problem (2).

For a given tolerance $\varepsilon > 0$, we are interested to estimate how many iterations of the VPGM is necessary to obtain a point x^k such that $|\theta_k| \leq \varepsilon$. In view of the remarks preceding Theorem 4.2, we may see such a point as an "ε-approximate" weakly efficient solution of problem (2).

Proposition 4.3 *The VPGM generates a point x^k such that $|\theta_k| \leq \varepsilon$ in at most $\mathcal{O}(1/\varepsilon)$ iterations, where θ_k is as defined in* (30).

Proof. Assume that none of the generated points x^k, $k = 0, \ldots, N$, is an ε-approximate weakly efficient solution of problem (2), i.e., $|\theta_k| > \varepsilon$. Hence, (31) implies that, for any $\bar{x} \in \Omega$ and $w \in C$,

$$(N+1)\varepsilon < (N+1) \min\{|\theta_k| : k = 0, \ldots, N\} \leq \sum_{k=0}^{N} |\theta_k| \leq \langle F(x^0) - F(x^{N+1}), w \rangle \leq \langle F(x^0) - F(\bar{x}), w \rangle,$$

which implies that

$$N + 1 < \frac{\max_{w \in C} \langle F(x^0) - F(\bar{x}), w \rangle}{\varepsilon} \leq \frac{\|F(x^0) - F(\bar{x})\|}{\varepsilon},$$

where the last inequality is due to Cauchy-Schwarz inequality and the fact that $\|w\| = 1$ for every $w \in C$. The above conclusion immediately proves the proposition. ■

5 Concluding Remarks

The proximal gradient method (PGM) is one of the most popular and efficient schemes for solving convex composite vector optimization problems. This method is well-known in the scalar case and was recently considered in the multiobjective setting in [27]. The main purpose here was to show that for problems in which the vector function is convex (without any Lipschitz assumption of the Jacobian of the differentiable component), the sequence generated by the PGM converges to a weakly efficient solution. An iteration-complexity result was also established in order to obtain an approximate weakly efficient solution of the vector problem under consideration.

Acknowledgements: We dedicate this paper in honor of the 70th birthday of Professor Alfredo N. Iusem. Yunier Bello-Cruz was partially supported by the National Science Foundation (NSF) Grant DMS - 1816449 and by Research & Artistry grant from NIU. Jefferson G. Melo was partially supported by CNPq grant 312559/2019-4 and FAPEG/GO.

References

[1] Bauschke HH, Combettes PL. Convex analysis and monotone operator theory in Hilbert spaces. CMS Books in Mathematics/Ouvrages de mathématiques de la SMC. New York(NY): Springer; 2017.

[2] Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. 2009; 2(1):183–202.

[3] Bello-Cruz JY. A subgradient method for vector optimization problems. SIAM J Optim. 2013; 23(4):2169-2182.

[4] Bello-Cruz JY, Bouza Allende G. A steepest descent-like method for variable order vector optimization problems. J Optim Theory Appl. 2014; 162(2): 371–391.

[5] Bello-Cruz JY, Bouza Allende G, Lucambio Pérez LR. Subgradient algorithms for solving variable inequalities. Appl Math Comput. 2014; 247, 1052–1063.

[6] Bello-Cruz JY, Lucambio Pérez LR. A subgradient-like algorithm for solving vector convex inequalities. J Optim Theory Appl 2014; 162(2):392-404.

[7] Bello-Cruz JY, Nghia TTA. On the convergence of the forward–backward splitting method with linesearches. Optim Methods Softw. 2016; 31(6):1209–1238.

[8] Bello-Cruz JY, Lucambio Pérez LR, Melo JG. Convergence of the projected gradient method for quasiconvex multiobjective optimization. Nonlinear Anal. 2011; 74(16):5268-5273.

[9] Bento GC, Cruz-Neto JX, López G, Soubeyran A, Souza JCO. The proximal point method for locally Lipschitz functions in multiobjective optimization with application to the compromise problem. SIAM J Optim. 2018; 28(2):1104–1120.

[10] Bento GC, Cruz-Neto JX, Soubeyran A. A proximal point-type method for multicriteria optimization. Set-Valued Var Anal. 2014; 22(3):557-573.

[11] Bento GC, Ferreira OP, Oliveira PR. Unconstrained steepest descent method for multicriteria optimization on Riemannian manifolds. J Optim Theory Appl. 2012; 154(1):88-107.

[12] Bento GC, Ferreira OP, Sousa Junior VL. Proximal point method for a special class of nonconvex multiobjective optimization functions. Optim Lett. 2018; 12(2):311-320.

[13] Bonnel H, Iusem AN, Svaiter BF. Proximal methods in vector optimization. SIAM J Optim. 2005; 15(4):953-970.

[14] Boţ RI, Grad SM. Inertial forward-backward methods for solving vector optimization problems. Optimization. 2018; 67(7):959-974.

[15] Combettes PL. Quasi-Fejérian analysis of some optimization algorithms. *Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications*. *Studies in Computational Mathematics* **8**, 115–152 North-Holland, Amsterdam; 2001.

[16] Cruz-Neto JX, Silva GJP, Ferreira OP, Lopes JO. A subgradient method for multiobjective optimization. Comput Optim Appl. 2013; 54(3):461-472.

[17] Ermoliev Yu. M. On the method of generalized stochastic gradients and quasi-Fejér sequences. *Cybernetics*. 1969; **5**: 208-220.

[18] Fliege J, Graña-Drummond LM, Svaiter BF. Newton's method for multiobjective optimization. SIAM J Optim. 2009; 20(2):602-626.

[19] Fliege J, Svaiter BF. Steepest descent methods for multicriteria optimization. Math Methods Oper Res. 2000; 51(3):479-494.

[20] Fliege J, Werner R. Robust multiobjective optimization and applications in portfolio optimization. Eur J Oper Res. 2014; 234(2):422-433.

[21] Graña-Drummond LM, Iusem AN. A projected gradient method for vector optimization problems. Comput Optim Appl. 2004; 28(1):5-29.

[22] Graña-Drummond LM, Raupp FMP, Svaiter BF. A quadratically convergent Newton method for vector optimization. Optimization. 2014; 63(5):661-677.

[23] Graña-Drummond LM, Svaiter BF. A steepest descent method for vector optimization. J Comput Appl Math. 2005; 175(2):395-414.

[24] Iusem AN, Svaiter BF, Teboulle M. Entropy-like proximal methods in convex programming. Math. Oper. Res. 1994; 19(4): 790-814.

[25] Luc DT. Theory of vector optimization. Berlin: Springer; 1989.

[26] Lucambio Pérez LR, Prudente LF. Nonlinear conjugate gradient methods for vector optimization. SIAM J Optim. 2018; 28(3):2690-2720.

[27] Tanabe H, Fukuda EH, Yamashita N. Proximal gradient methods for multiobjective optimization and their applications. Comput Optim Appl. 2019; 72(2): 339-361.

[28] Tseng P. Approximation accuracy, gradient methods, and error bound for structured convex optimization. Math Program. 2010; 125(2, Ser. B):263-295.