

Optimization

A Journal of Mathematical Programming and Operations Research

ISSN: 0233-1934 (Print) 1029-4945 (Online) Journal homepage: <https://www.tandfonline.com/loi/gopt20>

Inertial forward-backward methods for solving vector optimization problems

Radu Ioan Boţ & Sorin-Mihai Grad

To cite this article: Radu Ioan Boţ & Sorin-Mihai Grad (2018) Inertial forward-backward methods for solving vector optimization problems, Optimization, 67:7, 959-974, DOI: [10.1080/02331934.2018.1440553](https://doi.org/10.1080/02331934.2018.1440553)

To link to this article: <https://doi.org/10.1080/02331934.2018.1440553>



© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 20 Feb 2018.



Submit your article to this journal [↗](#)



Article views: 1055



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 2 View citing articles [↗](#)

Inertial forward–backward methods for solving vector optimization problems

Radu Ioan Boţ^{a,b}  and Sorin-Mihai Grad^{c,d} 

^aFaculty of Mathematics, University of Vienna, Vienna, Austria; ^bFaculty of Mathematics and Computer Science, Babeş-Bolyai University, Cluj-Napoca, Romania; ^cFaculty of Mathematics and Computer Science, University of Leipzig, Leipzig, Germany; ^dFaculty of Mathematics, Chemnitz University of Technology, Chemnitz, Germany

ABSTRACT

We propose two forward–backward proximal point type algorithms with inertial/memory effects for determining weakly efficient solutions to a vector optimization problem consisting in vector-minimizing with respect to a given closed convex pointed cone the sum of a proper cone-convex vector function with a cone-convex differentiable one, both mapping from a Hilbert space to a Banach one. Inexact versions of the algorithms, more suitable for implementation, are provided as well, while as a byproduct one can also derive a forward–backward method for solving the mentioned problem. Numerical experiments with the proposed methods are carried out in the context of solving a portfolio optimization problem.

ARTICLE HISTORY

Received 12 September 2017
Accepted 9 February 2018

KEYWORDS

Vector optimization problems; inertial proximal algorithms; forward–backward algorithms; weakly efficient solutions

1. Introduction and preliminaries

With this paper we propose the first, to the best of our knowledge, iterative methods for determining weakly efficient solutions to vector optimization problems consisting in vector-minimizing with respect to a convex cone the sum of two vector functions. The numerical methods we propose rely on the classical proximal point algorithm due to Martinet extended for vector optimization problems by Bonnel, Iusem and Svaiter. In order to treat the involved functions separately, we propose a forward–backward splitting scheme, to which inertial/memory effects are added. The inertial methods with memory effects were inspired from heavy ball with friction dynamical systems and have as a characteristic feature the fact that an iteration variable depends on the previous two elements of the same sequence, not only on its predecessor as it is usually the case for many algorithmic approaches. The first inertial proximal point type algorithm has been proposed by Alvarez and Attouch (cf. [1,2]) for the minimization of a proper, convex and lower semicontinuous function. To the best of our knowledge, the only other inertial type proximal method for solving vector optimization problems proposed so far in the literature is the one in [3], which follows a different path than our contribution as it is employed for determining ideally efficient solutions. The vector optimization problems we solve with our methods consist in vector-minimizing with respect to a given closed convex pointed cone the sum of a proper cone-convex vector function, evaluated in a backward step, with a cone-convex differentiable one that is evaluated in a forward step, both mapping from a Hilbert space to a Banach one. The usual way to approach vector optimization problems is by scalarizing them, but this

CONTACT Sorin-Mihai Grad  grad@math.tu-chemnitz.de
Dedicated to Johannes Jahn on the occasion of his 65th birthday.



can often lead to unbounded problems (see, for instance [4, Remark 1]), hence the need to address the vector optimization problems directly, especially when it comes to numerically solving them. One can find some results on the choice of scalarizing parameters in order to guarantee the existence of optimal solutions of the scalarized problems in the literature, but the imposed conditions are quite restrictive (see [5,6]) and their verification may prove to be too expensive from a computational point of view. This has motivated research on iterative methods for directly solving multiobjective or vector optimization problems consisting in vector-minimizing a vector function, sometimes subject to geometric constraints, and in the recent literature several contributions in this sense can be found in both smooth (cf. [7,8]) and convex nonsmooth (cf. [4,9]) cases. Different to these works, in this paper we vector-minimize with respect to a convex cone the sum of two cone-convex vector functions, that are handled separately in each iteration via a forward-backward scheme, covering thus a larger category of problems than in the existing literature. Moreover, we added inertial/memory effects to the proposed scheme that have a positive contribution to its robustness and speed. The proximal point scheme proposed in [4] is recovered as a special case of our method when the inertial steps and the smooth objective term are removed. Moreover, unlike the mentioned papers where iterative methods for solving vector optimization problems were proposed, but their implementation was left for later due to the difficulty of solving the employed subproblems, we present a concrete application as well that is solved in MATLAB.

For implementation purposes, we provide also versions of the inertial forward-backward algorithms which do not scalarize the original problem, but some approximations of it, another one at each iteration, as done for instance also in [4,9]. We have opted for the linear scalarization of the intermediate vector optimization problems instead of other existing alternatives (see, for instance, [10, Section 4.4]) for both simplicity and computational reasons. The construction of the algorithms guarantees the existence of an optimal solution to each of the considered (linearly) scalarized optimization problems, a feature usually mentioned as an advantage of other scalarization techniques in comparison to the linear one. Moreover, the linear scalarization is more flexible than its counterparts, allowing modifications of its parameters at each iteration.

Let X be a Hilbert space and Y a separable Banach space that is partially ordered by a pointed closed convex cone $C \subseteq Y$. The partial ordering induced by C on Y is denoted by ' \leq_C ' (i.e. it holds $x \leq_C y$ when $y - x \in C$, where $x, y \in Y$) and we write $x \leq_C y$ if $x \leq_C y$ and $x \neq y$. A greatest element with respect to ' \leq_C ' denoted by ∞_C which does not belong to Y is attached to this space, and let $Y^\bullet = Y \cup \{\infty_C\}$. Then for any $y \in Y$ one has $y \leq_C \infty_C$ and we consider on Y^\bullet the operations $y + \infty_C = \infty_C + y = \infty_C$ for all $y \in Y^\bullet$ and $t \cdot \infty_C = \infty_C$ for all $t \geq 0$. By $\langle y^*, y \rangle$ we denote the value at $y \in Y$ of the linear continuous functional $y^* \in Y^*$ and by convention we take $\langle y^*, \infty_C \rangle = +\infty$ for all $y^* \in C^*$, where $C^* = \{y^* \in Y^* : \langle y^*, y \rangle \geq 0 \forall y \in C\}$ is the *dual cone* to C . Given a subset U of X , by $\text{cl}U$, $\text{int}U$ and δ_U we denote its *closure*, *interior* and *indicator function*, respectively.

When $f : X \rightarrow \overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ is *proper* (i.e. is nowhere equal to $-\infty$ and has at least a real value) and $\varepsilon \geq 0$, if $f(x) \in \mathbb{R}$ the (convex) ε -subdifferential of f at x is $\partial_\varepsilon f(x) = \{x^* \in X^* : f(y) - f(x) \geq \langle x^*, y - x \rangle - \varepsilon \forall y \in X\}$, while if $f(x) = +\infty$ we take by convention $\partial_\varepsilon f(x) = \emptyset$. The ε -subdifferential of f becomes in case $\varepsilon = 0$ its classical (convex) subdifferential denoted by ∂f . Then $\bar{x} \in X$ is a minimum of f if and only if $0 \in \partial f(\bar{x})$. Denote also by $[t]_+ = \max\{t, 0\}$ for any $t \in \mathbb{R}$.

A vector function $F : X \rightarrow Y^\bullet = Y \cup \{\infty_C\}$ is said to be *proper* if its domain $\text{dom}F = \{x \in X : F(x) \in Y\}$ is nonempty, C -convex if $F(tx + (1-t)y) \leq_C tF(x) + (1-t)F(y)$ for all $x, y \in X$ and all $t \in [0, 1]$ and *positively C -lower semicontinuous* (in the literature also *star C -lower semicontinuous*) when the function $x \mapsto \langle z^*, F(x) \rangle$, further denoted by $(z^*F) : X \rightarrow \overline{\mathbb{R}}$, is lower semicontinuous for all $z^* \in C^* \setminus \{0\}$.

Assume further that $\text{int}C \neq \emptyset$. Consider the vector optimization problem

$$(VP) \quad \text{WMin}_{x \in X} [F(x) + G(x)],$$

where $F : X \rightarrow Y$ is a Fréchet differentiable vector function with an L -Lipschitz continuous gradient ∇F and $G : X \rightarrow Y^\bullet$ is a proper vector function. It is the aim of this paper to provide a proximal inertial forward–backward algorithm for determining the weakly efficient solutions to (VP). An element $\bar{x} \in \text{dom}G$ is said to be an *efficient solution* to (VP) if there is no $x \in X$ such that $F(x) + G(x) \leq_C F(\bar{x}) + G(\bar{x})$ and a *weakly efficient solution* to (VP) if $(F(\bar{x}) + G(\bar{x}) - \text{int}C) \cap (F + G)(\text{dom}G) = \emptyset$, respectively. We denote by $\mathcal{E}(VP)$ the set of all efficient solutions to (VP) and by $\mathcal{WE}(VP)$ the one of all weakly efficient ones. From [10, Corollary 2.4.26] one has the following characterization of the weakly efficient solutions to (VP) by means of a linear scalarization.

Lemma 1.1: *If $F + G$ is C -convex, then $\bar{x} \in \mathcal{WE}(VP)$ if and only if*

$$\exists z^* \in C^* \setminus \{0\} : \langle z^*, F(\bar{x}) + G(\bar{x}) \rangle \leq \langle z^*, F(x) + G(x) \rangle \quad \forall x \in X.$$

Remark 1: A sufficient hypothesis for guaranteeing the C -convexity of $F + G$ is to take both F and G to be C -convex.

Lemma 1.2: *When F is C -convex, then for any $z^* \in C^*$ the function $(z^*F) : Y \rightarrow \mathbb{R}$ is convex and Fréchet differentiable with an $L\|z^*\|$ -Lipschitz continuous gradient.*

Proof: Let $z^* \in C^*$. The convexity and continuous differentiability of (z^*F) were already proven in the literature, so we focus on the Lipschitz continuity of its gradient, that, by the chain rule, coincides at any $x \in X$ with the functional $z^* \circ \nabla F(x)$ defined by $w \in X \mapsto \langle z^*, \nabla F(x)(w) \rangle$.

For $x, y \in X$, one has, keeping in mind the linearity of z^* and of the gradient,

$$\begin{aligned} \|\nabla(z^*F)(x) - \nabla(z^*F)(y)\| &= \|z^* \circ \nabla F(x) - z^* \circ \nabla F(y)\| = \|z^* \circ (\nabla F(x) - \nabla F(y))\| \\ &\leq \|z^*\| \|\nabla F(x) - \nabla F(y)\| \leq L\|z^*\| \|x - y\|, \end{aligned}$$

therefore the gradient of (z^*F) is $L\|z^*\|$ -Lipschitz continuous. □

A result which is very useful in proving the convergence of numerical algorithms is the celebrated Opial's Lemma (cf. [11]).

Lemma 1.3: *Let $(x_n)_n \subseteq X$ be a sequence and $S \subseteq X$ a nonempty set such that*

- (a) $\lim_{n \rightarrow +\infty} \|x_n - x\|$ exists for every $x \in S$;
- (b) if $x_{n_j} \rightharpoonup z$ for a subsequence $n_j \rightarrow +\infty$, then $z \in S$,

where ' \rightharpoonup ' denotes the convergence in the weak topology. Then, there exists an $\bar{x} \in S$ such that $x_n \rightharpoonup \bar{x}$ when $n \rightarrow +\infty$.

2. Inertial forward–backward algorithm

We propose below an exact proximal inertial forward–backward iterative method for determining the weakly efficient solutions to (VP). It generates a sequence $(x_n)_n \subseteq X$ that, as seen later, converges under suitable but not very demanding hypotheses towards a weakly efficient solution to (VP).

Algorithm 1: *Choose the starting points $x_0, x_1 \in X$ and the sequences $(\beta_n)_n \subseteq [0, \beta]$, $(z_n^*)_n \subseteq C^* \setminus \{0\}$ and $(e_n)_n \subseteq \text{int}C$ such that $(\beta_n)_n$ is nondecreasing, $\beta < 1/9$, $\|z_n^*\| = 1$ and $\langle z_n^*, e_n \rangle = 1$ for all $n \geq 1$. Consider the following iterative steps*

- (1) let $n = 1$;
- (2) if $x_n \in \mathcal{WE}(VP)$: STOP;

- (3) find $x_{n+1} \in \mathcal{WE} \left\{ G(x) + \frac{L}{2} \left\| x - (x_n + \beta_n(x_n - x_{n-1}) - \frac{1}{L} \nabla(z_n^* F)(x_n)) \right\|^2 e_n : x \in \Omega_n \right\}$, where $\Omega_n = \{x \in X : (F + G)(x) \leq_C (F + G)(x_n)\}$;
 (4) let $n := n + 1$ and go to Step 2.

Remark 2: When $F \equiv 0$, Algorithm 1 becomes an inertial proximal point method for solving vector optimization problems, which by additionally taking $\beta_n = 0$ for all $n \geq 1$ collapses into the proximal point method for vector-minimizing a nonsmooth vector function introduced in [4]. On the other hand, when $Y = \mathbb{R}$ and $C = \mathbb{R}_+$ (i.e. in the scalar case), Algorithm 1 becomes the inertial proximal-gradient method for scalar optimization problems, that can be derived from the algorithm for finding zeros of maximally monotone operators proposed in [12]. When, furthermore, $F \equiv 0$, it collapses into the one from [1], while when $\beta_n = 0$ for all $n \geq 1$ it becomes the celebrated ISTA method (see, for instance, [13]), however in a more general framework.

Remark 3: Analyzing Algorithm 1 one can notice that at every iteration a different vector optimization problem is addressed, each of them having a smaller feasible set than its predecessor.

Before formulating the convergence statement concerning the sequence $(x_n)_n$ generated by Algorithm 1, it is necessary to introduce a new notion, considered in most of the papers dealing with proximal methods for vector optimization problems (see [4,9]).

Definition 2.1: Given $x_0 \in X$, the set $F(X) \cap (F(x_0) - C)$ is said to be *C-complete* when for all sequences $(a_n)_n \subseteq X$ with $a_0 = x_0$ such that $F(a_{n+1}) \leq_C F(a_n)$ for all $n \geq 1$ there exists an $a \in X$ such that $F(a) \leq_C F(a_n)$ for all $n \geq 1$.

Theorem 2.1: Let F be *C-convex* and G be *C-convex* and positively *C-lower semicontinuous* and assume that $(F + G)(X) \cap (F(x_0) + G(x_0) - C)$ is *C-complete*. If Algorithm 1 does not stop in finitely many steps, then any sequence $(x_n)_n$ generated by it converges weakly towards a weakly efficient solution to (VP).

Proof: We show first that the algorithm is well defined, i.e. if the stopping rule is not activated the next iteration exists. Assuming that we have obtained an x_n , where $n \geq 1$, we have to secure the existence of x_{n+1} .

The hypotheses imposed on F and G guarantee that the set Ω_n is convex and closed, since G is *C-epi closed* and F is Fréchet differentiable. Thus, δ_{Ω_n} is convex and lower semicontinuous. As the function

$$x \mapsto \left\langle z_n^*, G(x) + \frac{L}{2} \left\| x - x_n - \beta_n(x_n - x_{n-1}) + \frac{1}{L} \nabla(z_n^* F)(x_n) \right\|^2 e_n \right\rangle + \delta_{\Omega_n}(x)$$

is lower semicontinuous, being a sum of lower semicontinuous and continuous functions, and strongly convex, as the sum of some convex functions and a squared norm, it has exactly one minimum. By Lemma 1.1 this minimum is a weakly efficient solution to the vector optimization problem in Step 3 of Algorithm 1 and we denote it by x_{n+1} .

Thus one has

$$0 \in \partial \left(\left\langle z_n^*, G(\cdot) + \frac{L}{2} \left\| \cdot - x_n - \beta_n(x_n - x_{n-1}) + \frac{1}{L} \nabla(z_n^* F)(x_n) \right\|^2 e_n \right\rangle + \delta_{\Omega_n}(\cdot) \right) (x_{n+1}),$$

which, due to the continuity of the norm, turns into (cf. [10, Theorem 3.5.5])

$$0 \in \partial \left(\langle z_n^*, G(\cdot) + \delta_{\Omega_n}(\cdot) \rangle \right) (x_{n+1}) \\ + L \partial \left(\frac{1}{2} \left\| \cdot - x_n - \beta_n(x_n - x_{n-1}) + \frac{1}{L} \nabla(z_n^* F)(x_n) \right\|^2 \right) (x_{n+1}).$$

or, equivalently,

$$-L \left(x_{n+1} - x_n - \beta_n(x_n - x_{n-1}) + \frac{1}{L} \nabla(z_n^* F)(x_n) \right) \in \partial(\langle z_n^*, G(\cdot) \rangle + \delta_{\Omega_n}(\cdot))(x_{n+1}).$$

Thus, for any $x \in \Omega_n$, one has

$$\langle z_n^*, G(x) - G(x_{n+1}) \rangle \geq \langle L(x_{n+1} - x_n - \beta_n(x_n - x_{n-1})) + \nabla(z_n^* F)(x_n), x_{n+1} - x \rangle. \quad (1)$$

The set $\Omega := \bigcap_{k \geq 1} \Omega_k = \{x \in \mathbb{R}^n : (F + G)(x) \leq_C (F + G)(x_k) \ \forall k \geq 1\}$ is nonempty because of the C -completeness hypothesis. Let $\tilde{x} \in \Omega$. Then, using that $\tilde{x} \in \Omega_n$ and $z_n^* \in C^*$, (1) yields

$$\begin{aligned} 0 &\geq \langle z_n^*, (F + G)(\tilde{x}) - (F + G)(x_{n+1}) \rangle \\ &\geq \langle L(x_{n+1} - x_n - \beta_n(x_n - x_{n-1})) + \nabla(z_n^* F)(x_n), x_{n+1} - \tilde{x} \rangle + \langle z_n^*, F(\tilde{x}) - F(x_{n+1}) \rangle. \end{aligned}$$

Employing the descent lemma (e.g. [14, Theorem 18.15(iii)]) for $\langle z_n^*, F \rangle$, which is convex and Fréchet differentiable with $L\|z_n^*\| = L$ -Lipschitz continuous gradient, this implies

$$\langle L(x_{n+1} - x_n - \beta_n(x_n - x_{n-1})) + \nabla(z_n^* F)(x_n) - \nabla(z_n^* F)(\tilde{x}), x_{n+1} - \tilde{x} \rangle \leq \frac{L}{2} \|\tilde{x} - x_{n+1}\|^2.$$

The $\frac{1}{L}$ -cocoercivity of $\nabla(z_n^* F)$ (cf. [14, Theorem 18.15(v)]) gives

$$\begin{aligned} &\langle \nabla(z_n^* F)(x_n) - \nabla(z_n^* F)(\tilde{x}), x_{n+1} - \tilde{x} \rangle \\ &= \langle \nabla(z_n^* F)(x_n) - \nabla(z_n^* F)(\tilde{x}), x_n - \tilde{x} \rangle + \langle \nabla(z_n^* F)(x_n) - \nabla(z_n^* F)(\tilde{x}), x_{n+1} - x_n \rangle \\ &\geq \frac{1}{L} \|\nabla(z_n^* F)(x_n) - \nabla(z_n^* F)(\tilde{x})\|^2 + \langle \nabla(z_n^* F)(x_n) - \nabla(z_n^* F)(\tilde{x}), x_{n+1} - x_n \rangle \\ &= \frac{1}{L} \left\| \nabla(z_n^* F)(x_n) - \nabla(z_n^* F)(\tilde{x}) + \frac{L}{2}(x_{n+1} - x_n) \right\|^2 - \frac{L}{4} \|x_{n+1} - x_n\|^2 \\ &\geq -\frac{L}{4} \|x_{n+1} - x_n\|^2, \end{aligned}$$

therefore

$$\langle L(x_{n+1} - x_n - \beta_n(x_n - x_{n-1})), x_{n+1} - \tilde{x} \rangle - \frac{L}{2} \|\tilde{x} - x_{n+1}\|^2 - \frac{L}{4} \|x_{n+1} - x_n\|^2 \leq 0.$$

For each $k \geq 1$ denote $\varphi_k = (1/2)\|x_k - \tilde{x}\|^2$. As $\varphi_n - \varphi_{n+1} = \frac{1}{2}\|x_{n+1} - x_n\|^2 + \langle x_{n+1} - x_n - \beta_n(x_n - x_{n-1}), \tilde{x} - x_{n+1} \rangle + \beta_n \langle x_n - x_{n-1}, \tilde{x} - x_{n+1} \rangle$, the previous inequality can be rewritten, after dividing with L , as

$$\varphi_{n+1} - \varphi_n + \frac{1}{4} \|x_{n+1} - x_n\|^2 - \frac{1}{2} \|\tilde{x} - x_{n+1}\|^2 - \beta_n \langle x_n - x_{n-1}, x_{n+1} - \tilde{x} \rangle \leq 0,$$

and, since $\langle x_n - x_{n-1}, x_{n+1} - \tilde{x} \rangle = \varphi_n - \varphi_{n-1} + (1/2)\|x_n - x_{n-1}\|^2 + \langle x_n - x_{n-1}, x_{n+1} - x_n \rangle$, it turns into

$$\varphi_{n+1} - \varphi_n - \beta_n(\varphi_n - \varphi_{n-1}) \leq \frac{\beta_n}{2} \|x_n - x_{n-1}\|^2 + \beta_n \langle x_n - x_{n-1}, x_{n+1} - x_n \rangle - \frac{1}{4} \|x_{n+1} - x_n\|^2.$$

The right-hand side of the above inequality can be rewritten as



$$\frac{\beta_n}{2} \|x_n - x_{n-1}\|^2 + \beta_n^2 \|x_n - x_{n-1}\|^2 - \frac{1}{4} \|x_{n+1} - x_n - 2\beta_n(x_n - x_{n-1})\|^2,$$

that, since $\beta_n \in [0, \beta[$, is less than or equal to

$$\beta_n \|x_n - x_{n-1}\|^2 - \frac{1}{4} \|x_{n+1} - x_n - 2\beta_n(x_n - x_{n-1})\|^2,$$

and, taking also in consideration that $\beta_n < \beta < 1/9 < 1/8$, even to

$$\frac{9}{8} \beta_n \|x_n - x_{n-1}\|^2 - \frac{1}{8} \|x_{n+1} - x_n\|^2.$$

Therefore, one gets

$$\varphi_{n+1} - \varphi_n - \beta_n(\varphi_n - \varphi_{n-1}) \leq \frac{9}{8} \beta_n \|x_n - x_{n-1}\|^2 - \frac{1}{8} \|x_{n+1} - x_n\|^2. \quad (2)$$

Denoting for all $k \geq 1$

$$\mu_k := \varphi_k - \beta_k \varphi_{k-1} + \frac{9}{8} \beta_k \|x_k - x_{k-1}\|^2,$$

it follows that

$$\mu_{k+1} - \mu_k \leq \frac{9\beta - 1}{8} \|x_{k+1} - x_k\|^2 \leq 0, \quad (3)$$

thus the sequence $(\mu_k)_k$ is nonincreasing, as $n \geq 1$ was arbitrarily chosen. Then, for all $k \geq 1$,

$$\varphi_k - \beta_k \varphi_{k-1} \leq \mu_k \leq \mu_1,$$

hence

$$\varphi_k \leq \beta^k \varphi_0 + \frac{1 - \beta^k}{1 - \beta} \mu_1 \leq \beta^k \varphi_0 + \frac{1}{1 - \beta} \mu_1,$$

and one also gets $\|x_{k+1} - x_k\|^2 \leq \frac{8}{9\beta - 1} (\mu_{k+1} - \mu_k)$. Consequently,

$$\begin{aligned} \sum_{k=1}^n \|x_{k+1} - x_k\|^2 &\leq \frac{8}{1 - 9\beta} (\mu_1 - \mu_{n+1}) \leq \frac{8}{1 - 9\beta} (\mu_1 + \beta \varphi_n) \\ &\leq \frac{8}{1 - 9\beta} \left(\frac{1}{1 - \beta} \mu_1 + \beta^{n+1} \varphi_0 \right) < +\infty, \end{aligned}$$

in particular

$$\sum_{k=1}^{+\infty} \|x_{k+1} - x_k\|^2 < +\infty. \quad (4)$$

Using the intermediate step towards (2) and denoting $\tau_{k+1} := x_{k+1} - x_k - 2\beta_k(x_k - x_{k-1})$, $\theta_k := \varphi_k - \varphi_{k-1}$ and $\delta_k := \beta_k \|x_k - x_{k-1}\|^2$ for all $k \geq 1$, one obtains

$$\theta_{k+1} - \beta_k \theta_k \leq \delta_k - \frac{1}{4} \|\tau_{k+1}\|^2 \leq \delta_k \quad \forall k \geq 1. \quad (5)$$

Then

$$[\theta_{k+1}]_+ \leq \frac{1}{9} [\theta_k]_+ + \delta_k \quad \forall k \geq 1,$$

which yields

$$[\theta_{n+1}]_+ \leq \frac{1}{9^n} [\theta_1]_+ + \sum_{k=0}^{n-1} \frac{\delta_{n-k}}{9^k}.$$

Hence

$$\sum_{k=0}^{+\infty} [\theta_{k+1}]_+ \leq \frac{9}{8} \left([\theta_1]_+ + \sum_{k=0}^{+\infty} \delta_k \right)$$

and, as the right-hand side of this inequality is finite due to (4), so is $\sum_{k=1}^{+\infty} [\theta_k]_+$, too. This yields that the sequence $(w_k)_k$ defined as $w_k = \varphi_k - \sum_{j=1}^k [\theta_j]_+$, for all $k \geq 1$, is bounded. Moreover, $w_{k+1} - w_k = \varphi_{k+1} - \varphi_k - [\varphi_{k+1} - \varphi_k]_+ \leq 0$ for all $k \geq 1$, thus $(w_k)_k$ is convergent. Consequently, $(\varphi_k)_k$ is convergent. Finally, $(\|x_k - \tilde{x}\|^2)_k$ is convergent, too, i.e. (a) in Lemma 1.3 with $S = \Omega$ is fulfilled.

The next step is to show that $(x_k)_k$ is weakly convergent. The convergence of $(\varphi_k)_k$ implies that $(x_k)_k$ is bounded, so it has weak cluster points. Let $\hat{x} \in X$ be one of them and $(x_{k_j})_j$ the subsequence that converges weakly towards it as $j \rightarrow +\infty$. Then, as $F+G$ is positively C -lower semicontinuous and C -convex, it follows that for any $z^* \in C^* \setminus \{0\}$ the function $\langle z^*, (F+G)(\cdot) \rangle$ is lower semicontinuous and convex, thus

$$\langle z^*, (F+G)(\hat{x}) \rangle \leq \liminf_{j \rightarrow +\infty} \langle z^*, (F+G)(x_{k_j}) \rangle = \inf_{k \geq 1} \langle z^*, (F+G)(x_k) \rangle, \quad (6)$$

with the last equality following from the fact that the sequence $((F+G)(x_k))_k$ is by construction C -nonincreasing. Assuming that there exists a $k \geq 1$ such that $(F+G)(\hat{x}) \not\leq_C (F+G)(x_k)$, there exists a $\tilde{z} \in C^* \setminus \{0\}$ such that $\langle \tilde{z}, (F+G)(\hat{x}) - (F+G)(x_k) \rangle > 0$, which contradicts (6), consequently $(F+G)(\hat{x}) \leq_C (F+G)(x_k)$ for all $k \geq 1$, i.e. $\hat{x} \in \Omega$, therefore one can employ Lemma 1.3 with $S = \Omega$ since its hypothesis (b) is fulfilled as well. This guarantees then the weak convergence of $(x_k)_k$ to a point $\bar{x} \in \Omega$.

The proof is not complete without showing that $\bar{x} \in \mathcal{WE}(VP)$. Assuming the opposite, there would exist an $x' \in X$ such that $(F+G)(x') \in (F+G)(\bar{x}) - \text{int}C$. This yields $x' \in \Omega$. Since $\|z_k^*\| = 1$ for all $k \geq 1$, the sequence $(z_k^*)_k$ has a weak* cluster point, say \bar{z}^* , that is the limit of a subsequence $(z_{k_j}^*)_j$. Because $z_k^* \in C^*$ for all $k \geq 1$ and C^* is weakly* closed, it follows that $\bar{z}^* \in C^*$. Moreover, from [15, Lemma 2.2] it follows that $\langle \bar{z}^*, c \rangle > 0$ for any $c \in \text{int}C$, thus $\bar{z}^* \neq 0$. Consequently, $\langle \bar{z}^*, (F+G)(x') - (F+G)(\bar{x}) \rangle < 0$. For any $j \geq 1$ it holds then by (1) and employing the descent lemma and the $\frac{1}{L}$ -cocoercivity of $\nabla(z_{k_j}^* F)$

$$\begin{aligned} \langle z_{k_j}^*, (F+G)(x') - (F+G)(\bar{x}) \rangle &\geq \langle z_{k_j}^*, (F+G)(x') - (F+G)(x_{k_j+1}) \rangle \\ &\geq -L \left\langle x_{k_j+1} - x_{k_j} - \beta_{k_j}(x_{k_j} - x_{k_j-1}) + \frac{1}{L} \nabla(z_{k_j}^* F)(x_{k_j}), x' - x_{k_j+1} \right\rangle + \langle z_{k_j}^*, F(x') - F(x_{k_j+1}) \rangle \\ &\geq -L \|x' - x_{k_j+1}\| (\|x_{k_j+1} - x_{k_j}\| + \beta_{k_j} \|x_{k_j} - x_{k_j-1}\|) + \langle \nabla(z_{k_j}^* F)(x_{k_j}) - \nabla(z_{k_j}^* F)(x'), x_{k_j+1} - x' \rangle \\ &\geq -L \|x' - x_{k_j+1}\| (\|x_{k_j+1} - x_{k_j}\| + \beta_{k_j} \|x_{k_j} - x_{k_j-1}\|) - \frac{L}{4} \|x_{k_j+1} - x_{k_j}\|^2. \end{aligned} \quad (7)$$

Because of (4) $(\|x_k - x_{k-1}\|)_k$ converges towards 0 for $k \rightarrow +\infty$ and so does the last expression in the inequality chain (7) when $j \rightarrow +\infty$ as well. Letting j converge towards $+\infty$, (7) yields $\langle \bar{z}^*, (F+G)(x') - (F+G)(\bar{x}) \rangle \geq 0$, contradicting the inequality obtained above. Consequently, $\bar{x} \in \mathcal{WE}(VP)$. \square

Remark 4: The conclusion of Theorem 2.1 remains valid when G is taken to be C -lower semicontinuous in the sense of [10, Definition 2.2.14] instead of positively C -lower semicontinuous.

Remark 5: As can be seen in the proof of Theorem 2.1, its conclusion remains valid if the sequence $(x_n)_n$ generated by Algorithm 1 fulfills the condition $\sum_{k=1}^{+\infty} \beta_k \|x_k - x_{k-1}\|^2 < +\infty$ (see (4), mentioned also in the literature, for instance in [1,12]), in which case $(\beta_n)_n$ needs not necessarily be nondecreasing and one can take $\beta \in [0, 1[$. However, this dynamic condition might be more difficult to verify since it involves the generated sequence $(x_n)_n$, while the static hypotheses considered in this paper can simply be imposed while defining the parameters β and $(\beta_n)_n$, respectively.

Remark 6: In the proof of Theorem 2.1, we have employed some ideas inspired from the ones of [4, Theorem 3.1], [12, Theorem 2.1] and [1, Theorem 2.1 and Proposition 2.1]. The difficulties encountered while adapting the techniques from the mentioned statements to our framework consisted mainly of the fact that here one has to deal, as mentioned in Remark 3, at each iteration with a different optimization problem, while in [1,12] the objective function of the considered problem is not modified as the algorithm advances.

Remark 7: Different to the inertial proximal methods proposed in the literature for solving scalar optimization problems or monotone inclusions (see, for instance, [1,12]), in our approach it is not necessary to assume the existence of a solution of the considered problem, i.e. a weakly efficient solution to (VP) , in order to prove the convergence of Algorithm 1. The role of such a hypothesis in showing the convergence of the method has been fully covered in the proof of Theorem 2.1 by the assumed C -completeness hypothesis. Considering the former instead of the latter, the role of Ω would be taken by $\mathcal{WE}(VP)$. However, then is the inclusion $\mathcal{WE}(VP) \subseteq \Omega_n$ for all $n \geq 1$ not guaranteed by construction and should be separately investigated. Note moreover that assuming the existence of some $\bar{x} \in \mathcal{WE}(VP)$ does not automatically deliver the corresponding scalarizing parameter \bar{z}^* that exists according to Lemma 1.1, which would probably be needed in formulating the algorithm under the mentioned hypothesis.

Remark 8: Any $z^* \in C^* \setminus \{0\}$ provides a suitable scalarization functional (whose existence is guaranteed by Lemma 1.1) for the vector optimization problems in Step 3 of Algorithm 1. This endows our method with additional flexibility properties that may prove to be useful when implementing it. Moreover, even if the function

$$x \mapsto \left\langle z^*, G(x) + \frac{L}{2} \left\| x - x_n - \beta_n(x_n - x_{n-1}) + \frac{1}{L} \nabla(z_n^* F)(x_n) \right\|^2 e_n \right\rangle + \delta_{\Omega_n}(x)$$

has, because it is lower semicontinuous and strongly convex, exactly one minimum that is x_{n+1} , the sequence $(x_n)_n$ is not uniquely determined because for each choice of $z^* \in C^* \setminus \{0\}$ one deals with a different such function. Note also that the sequence $(z_n^*)_n$ can be taken even constant, situation in which the intermediate vector optimization problems differ despite having the same objective vector function because their feasible sets become smaller at each iteration. This does not mean that the vector optimization problem (VP) is a priori scalarized by means of a linear continuous functional, because this scalarization is applied to the intermediate vector optimization problems not to (VP) .

Remark 9: For determining the optimal solutions of the scalarized optimization problems attached to the vector optimization problems in Step 3 of Algorithm 1 one can employ for instance a splitting type algorithm designed for finding the optimal solutions of optimization problems consisting in minimizing sums of convex functions, like the ones proposed in [16,17]. However, the processing of the functions δ_{Ω_n} , $n \geq 1$, may prove to be quite difficult, due to the special structure of the sets Ω_n , $n \geq 1$. A way to go round this nuisance is, as seen later in Section 4, by employing some other approaches for solving the intermediate scalar optimization problems, for instance one based on interior point methods.

Remark 10: Similarly to [4, Section 4], it can be shown that under additional hypotheses (for instance $\exists \delta > 0 : \{z^* \in Y^* : \langle z^*, y \rangle \geq \delta \|y\| \|z^*\| \text{ for all } y \in C\} \neq \emptyset$) Algorithm 1 can deliver efficient solutions to (VP) instead of weakly efficient. One can also modify Algorithm 1 in order to deliver *properly efficient solutions to (VP) with respect to linear scalarization* that are defined as those $\bar{x} \in \text{dom} G$ for which there exists a $z^* \in C^{*0} = \{y^* \in C^* : \langle y^*, y \rangle > 0 \ \forall y \in C \setminus \{0\}\}$ for which $\langle z^*, F(\bar{x}) + G(\bar{x}) \rangle \leq \langle z^*, F(x) + G(x) \rangle$ for all $x \in X$. For this, the interior of C needs not necessarily be nonempty and the sequence $(e_n)_n$ should lie in $C \setminus \{0\}$, while in Steps 2 and 3 of the algorithm one should consider properly efficient instead of weakly efficient solutions to (VP).

Remark 11: Vector optimization problems with the ordering cones of the image spaces having empty interiors, but nonempty generalized interiors can be found in finance mathematics (see, for instance, [18,19]) and other research fields. Motivated by them, the definition of the weakly efficient solutions to (VP) has been extended in some recent contributions (such as [19–21]) for the case when $\text{int}C = \emptyset$ by replacing this with the *quasi interior* of C (i.e. the set of all $y \in Y$ such that $\text{cl}(\text{cone}(V - y)) = Y$, where ‘cone’ denotes the *conical hull* of the corresponding set). In order to characterize these more general weakly efficient solutions to (VP) one can use [21, Corollary 9] instead of Lemma 1.1. The proof of the algorithm convergence statement Theorem 2.1 can be almost entirely adapted to the new framework, but, since [15, Lemma 2.2] does not hold in case $\text{int}C = \emptyset$, an alternative approach for guaranteeing that $\bar{z}^* \neq 0$ is necessary. A way to do this would be by taking the sequence $(z_n^*)_n$ constant with $z_n^* = \bar{z}^* \neq 0$. On the other hand, in finitely dimensional spaces so-called relatively weakly efficient solutions can be defined when C has an empty interior but a nonempty relative interior and characterized by means of linear scalarization (cf. [20]), while $\bar{z}^* \neq 0$ because of the coincidence of the corresponding weak and strong topologies.

One can simplify Algorithm 1 in order to become a ‘pure’ (i.e. non-inertial) forward-backward method, as follows.

Algorithm 2: Choose the starting point $x_0 \in X$ and the sequences $(z_n^*)_n \subseteq C^* \setminus \{0\}$ and $(e_n)_n \subseteq \text{int}C$ such that $\|z_n^*\| = 1$ and $\langle z_n^*, e_n \rangle = 1$ for all $n \geq 1$. Consider the following iterative steps

- (1) let $n = 1$;
- (2) if $x_n \in \mathcal{WE}(VP)$: STOP;
- (3) find $x_{n+1} \in \mathcal{WE} \left\{ G(x) + \frac{L}{2} \|x - (x_n - \frac{1}{L} \nabla(z_n^* F)(x_n))\|^2 e_n : x \in \Omega_n \right\}$, where $\Omega_n = \{x \in X : (F + G)(x) \leq_C (F + G)(x_n)\}$;
- (4) let $n := n + 1$ and go to 2.

This algorithm is interesting not only per se, but also because one can derive rates for its convergence, by employing some ideas from [13], when the sequence $(z_n^*)_n$ is constant. In order to prove the corresponding statement, an additional result is necessary (following [13, Lemma 2.3]).

Lemma 2.1: Let F be C -convex and G be C -convex and positively C -lower semicontinuous and denote, for fixed $z^* \in C^*$, $e \in \text{int}C$ with $\langle z^*, e \rangle = 1$ and $y \in X$,

$$z_y := \underset{x \in \Omega_y}{\operatorname{argmin}} \left\langle z^*, G(x) + \frac{L}{2} \left\| x - y + \frac{1}{L} \nabla(z^* F)(y) \right\|^2 e \right\rangle,$$

where $\Omega_y = \{x \in X : (F + G)(x) \leq_C (F + G)(y)\}$. One has

$$\langle z^*, F(x) + G(x) - F(z_y) - G(z_y) \rangle + \delta_{\Omega_y}(x) \geq \frac{L}{2} \|z_y - y\|^2 + L \langle z_y - y, y - x \rangle \quad \forall x \in X. \quad (8)$$

Proof: For a $y \in X$, by the definition of z_y one gets, taking into consideration the continuity of the norm (like in the proof of Theorem 2.1) and that $z_y \in \Omega_y$,

$$\langle z^*, G(x) - G(z_y) \rangle + \delta_{\Omega_y}(x) \geq \langle \nabla(z^*F)(y) + L(z_y - y), z_y - x \rangle \quad \forall x \in X.$$

Using also that $\langle z^*, F(x) - F(y) \rangle \geq \langle \nabla(z^*F)(y), x - y \rangle$ for all $x \in X$, one gets

$$\langle z^*, F(x) + G(x) - F(y) - G(z_y) \rangle + \delta_{\Omega_y}(x) \geq \langle \nabla(z^*F)(y) - L(x - z_y), z_y - y \rangle \quad \forall x \in X.$$

Employing the descent lemma, the above inequality yields

$$\langle z^*, F(x) + G(x) - F(z_y) - G(z_y) \rangle + \delta_{\Omega_y}(x) \geq -\frac{L}{2}\|z_y - y\|^2 - L\langle x - z_y, z_y - y \rangle \quad \forall x \in X,$$

and the right-hand side can be rewritten in order to deliver (8). \square

Theorem 2.2: Let F be C -convex and G be C -convex and positively C -lower semicontinuous and assume that $(F + G)(X) \cap (F(x_0) + G(x_0) - C)$ is C -complete. Consider the sequence $(x_n)_n$ generated by Algorithm 2, where one takes $z_n^* = z^* \in C^* \setminus \{0\}$, $n \geq 1$. Then for any $n \geq 1$ and $\tilde{x} \in \Omega$ one has

$$\langle z^*, F(x_n) + G(x_n) - F(\tilde{x}) - G(\tilde{x}) \rangle \leq \frac{L\|\tilde{x} - x_0\|^2}{2n}. \quad (9)$$

Proof: In order to prove the statement, one can follow the steps from the proof of [13, Theorem 3.1] by employing twice Lemma 2.1 for the functions (z^*F) and (z^*G) , respectively, first for $x = \tilde{x}$ and $y = x_n$, then for $x = y = x_n$ (for a fixed $n \geq 1$), taking also into consideration that $\tilde{x}, x_n \in \Omega_n$. \square

Remark 12: Note that the assertion of Theorem 2.2 is actually valid for all $\tilde{x} \in \Omega$, not only for the weakly efficient solution to (VP) obtained from the convergence statement Theorem 2.1. Moreover, when taking the sequence $(z_n^*)_n$ constant it is no longer necessary to take $\|z_n^*\| = 1$ for all $n \geq 1$. However, the constant z^* cannot be taken arbitrarily large (with respect to the ordering cone C^*) because it has to fulfill $\langle z^*, e_n \rangle = 1$ for all $n \geq 1$. Moreover, in this case one can consider the more general framework discussed in Remark 11. Without assuming that the sequence $(z_n^*)_n$ is constant, in order to show instead of (9) in a similar manner to the proof of Theorem 2.2 that

$$\langle z_n^*, F(x_n) + G(x_n) - F(\tilde{x}) - G(\tilde{x}) \rangle \leq \frac{L\|\tilde{x} - x_0\|^2}{2n} \quad \forall n \geq 1 \quad \forall \tilde{x} \in \Omega,$$

one needs additional assumptions that guarantee certain monotonicity properties for $(z_n^*)_n$, for instance

$$\langle z_n^* - z_{n+1}^*, F(x_{n+1}) + G(x_{n+1}) \rangle \geq 0 \geq \langle z_n^* - z_{n+1}^*, F(\tilde{x}) + G(\tilde{x}) \rangle.$$

Remark 13: For implementation purposes, one can provide an inexact version of Algorithm 1 as well, where Step 3 is replaced by

$$3' \quad \text{find } x_{n+1} \in X \text{ such that } 0 \in \partial_{\varepsilon_n}(\langle z_n^*, G(\cdot) \rangle + \frac{L}{2}\|\cdot - x_n - \beta_n(x_n - x_{n-1}) + \frac{1}{L}\nabla(z_n^*F)(x_n)\|^2 e_n) + \delta_{\Omega_n}(\cdot)(x_{n+1}),$$

where the additional sequence of tolerable nonnegative errors $(\varepsilon_n)_n$ fulfills some hypotheses, such as the ones considered in [4] or those from [2,12]. Employing the later, i.e. $\sum_{n \geq 1} \varepsilon_n < +\infty$, the converge statement obtained by correspondingly modifying Theorem 2.1 remains valid, only some minor adjustments in the proof (for instance one takes $\delta_k = \beta_k\|x_k - x_{k-1}\|^2 + \varepsilon_k$ for $k \geq 1$) being necessary.

3. Alternative inertial forward–backward algorithm

One can modify Algorithm 1 into another inertial forward–backward method, as follows.

Algorithm 3: Choose the starting points $x_0, x_1 \in X$ and the sequences $(\beta_n)_n \subseteq [0, \beta]$, $(z_n^*)_n \subseteq C^* \setminus \{0\}$ and $(e_n)_n \subseteq \text{int}C$ such that $(\beta_n)_n$ is nondecreasing, $\beta < 1/9$, $\|z_n^*\| = 1$ and $\langle z_n^*, e_n \rangle = 1$ for all $n \geq 1$. Consider the following iterative steps

- (1) let $n = 1$;
- (2) if $x_n \in \mathcal{WE}(VP)$: STOP;
- (3) find $x_{n+1} \in \mathcal{WE} \left\{ G(x) + \frac{1}{2} \left\| x - (x_n + \beta_n(x_n - x_{n-1})) - \frac{1}{L} \nabla(z_n^*F)(x_n + \beta_n(x_n - x_{n-1})) \right\|^2 e_n : x \in \Omega_n \right\}$;
- (4) let $n := n + 1$ and go to 2.

Remark 14: The difference between Algorithm 1 and Algorithm 3 resides in the point where the value of $\nabla(z_n^*F)$ is calculated, x_n vs. $x_n + \beta_n(x_n - x_{n-1})$. Thus, the comments from Remark 2 for $F \equiv 0$ remain valid for Algorithm 3 as well. On the other hand, when $Y = \mathbb{R}$ and $C = \mathbb{R}_+$ (i.e. in the scalar case), Algorithm 3 becomes a more general version of the celebrated FISTA method from [13], that can be recovered by taking $\beta_n = (t_n - 1)/t_{n+1}$, where $t_{n+1} = (1 + \sqrt{1 + 4t_n^2})/2$, $n \geq 1$, with $t_1 = 1$, and restricting the framework to finitely dimensional spaces.

The convergence of the sequence $(x_n)_n$ generated by Algorithm 3 can be investigated in an analogous manner to Theorem 2.1, hence the proof of the next statement contains only what is different compared to the mentioned statement.

Theorem 3.1: Let F be C -convex and G be C -convex and positively C -lower semicontinuous and assume that $(F + G)(X) \cap (F(x_0) + G(x_0) - C)$ is C -complete. If Algorithm 3 does not stop in finitely many steps, then any sequence $(x_n)_n$ generated by it converges weakly towards a weakly efficient solution to (VP) .

Proof: Let be $n \geq 1$. Like in the proof of Theorem 2.1 the algorithm is well defined and one gets

$$\begin{aligned} & \langle L(x_{n+1} - x_n - \beta_n(x_n - x_{n-1})) + \nabla(z_n^*F)(x_n + \beta_n(x_n - x_{n-1})) - \nabla(z_n^*F)(\tilde{x}), x_{n+1} - \tilde{x} \rangle \\ & \leq \frac{L}{2} \|\tilde{x} - x_{n+1}\|^2, \end{aligned}$$

which, by employing the $\frac{1}{L}$ -cocoercivity of $\nabla(z_n^*F)$, yields

$$\langle x_{n+1} - x_n - \beta_n(x_n - x_{n-1}), x_{n+1} - \tilde{x} \rangle - \frac{1}{2} \|\tilde{x} - x_{n+1}\|^2 - \frac{1}{4} \|x_{n+1} - x_n - \beta_n(x_n - x_{n-1})\|^2 \leq 0.$$

Denoting $\varphi_k = (1/2)\|x_k - \tilde{x}\|^2$, for $k \geq 1$, one gets, like in the proof of Theorem 2.1,

$$\begin{aligned} \varphi_{n+1} - \varphi_n - \beta_n(\varphi_n - \varphi_{n-1}) & \leq \frac{\beta_n}{2} \|x_n - x_{n-1}\|^2 + \beta_n \langle x_n - x_{n-1}, x_{n+1} - x_n \rangle \\ & \quad - \frac{1}{4} \|x_{n+1} - x_n - \beta_n(x_n - x_{n-1})\|^2. \end{aligned}$$

The right-hand side of the above inequality can be rewritten, taking into consideration that $\|x_{n+1} - x_n - \beta_n(x_n - x_{n-1})\|^2 = \|x_{n+1} - x_n\|^2 + \beta_n^2 \|x_n - x_{n-1}\|^2 - 2\beta_n \langle x_{n+1} - x_n, x_n - x_{n-1} \rangle$, as

$$\frac{\beta_n}{2} \|x_n - x_{n-1}\|^2 - \frac{\beta_n^2}{4} \|x_n - x_{n-1}\|^2 - \frac{1}{4} \|x_{n+1} - x_n\|^2 + \frac{3}{2} \beta_n \langle x_{n+1} - x_n, x_n - x_{n-1} \rangle,$$

being further equal to

$$\left(\frac{\beta_n}{2} + 2\beta_n^2\right) \|x_n - x_{n-1}\|^2 - \frac{1}{4} \|x_{n+1} - x_n - 3\beta_n(x_n - x_{n-1})\|^2.$$

Therefore, $\varphi_{n+1} - \varphi_n - \beta_n(\varphi_n - \varphi_{n-1})$ is less than or equal to

$$\frac{\beta_n}{2} (1 + 2\beta_n) \|x_n - x_{n-1}\|^2 - \frac{1}{4} \|x_{n+1} - x_n - 3\beta_n(x_n - x_{n-1})\|^2.$$

From

$$-\frac{1}{4} \|x_{n+1} - x_n - 3\beta_n(x_n - x_{n-1})\|^2 \leq \frac{9}{4} \beta_n^2 \|x_n - x_{n-1}\|^2 - \frac{1}{8} \|x_{n+1} - x_n\|^2,$$

one gets

$$\varphi_{n+1} - \varphi_n - \beta_n(\varphi_n - \varphi_{n-1}) \leq \beta_n \left(\frac{1}{2} + \frac{17}{4} \beta_n\right) \|x_n - x_{n-1}\|^2 - \frac{1}{8} \|x_{n+1} - x_n\|^2,$$

that, since $\beta_n < 1/8$, yields (2).

The rest of the proof follows analogously to the one of Theorem 2.1. \square

Remark 15: Making use of Lemma 2.1, one can provide, following [13, Theorem 4.4], a convergence rate statement for Algorithm 3 for a special choice of the parameters β_n , $n \geq 1$, and when the sequence $(z_n^*)_n$ is constant, that improves the assertion in the non-inertial case from Theorem 2.2.

As discussed for instance also in [4], one can consider in the presented algorithms a stopping rule that is easier to check than the original one. The following statement shows that if three consecutive iterations of the sequence $(x_n)_n$ generated by the inertial type algorithms we proposed coincide, they represent a weakly efficient solution to (VP), i.e. Step 2 of Algorithm 1 or Algorithm 3 can be replaced with

2' if $x_{n+1} = x_n = x_{n-1}$: STOP.

Proposition 3.1: Let F and G be C -convex and consider a sequence $(x_k)_k$ generated by Algorithm 1. If for some $n \geq 1$ one has

$$x_{n-1} = x_n \in \operatorname{argmin}_{x \in \Omega_n} \left\langle z_n^*, G(x) + \frac{L}{2} \left\| x - (x_n + \beta_n(x_n - x_{n-1})) - \frac{1}{L} \nabla(z_n^* F)(x_n) \right\|^2 e_n \right\rangle,$$

then $x_n \in \mathcal{WE}(VP)$.

Proof: Assuming that $x_n \notin \mathcal{WE}(VP)$, there exist $\tilde{x} \in X$ and $c \in \operatorname{int} C$ such that $F(\tilde{x}) + G(\tilde{x}) = F(x_n) + G(x_n) - c$. Then $\tilde{x} \in \Omega_n$. Denoting, for $t \in [0, 1]$, $x_t := tx_n + (1 - t)\tilde{x}$, the C -convexity of $F + G$ yields $F(x_t) + G(x_t) \leq_C F(x_n) + G(x_n) + (t - 1)c$, therefore $x_t \in \Omega_n$ as well.

Since

$$x_n \in \operatorname{argmin}_{x \in \Omega_n} \left\langle z_n^*, G(x) + \frac{L}{2} \left\| x - (x_n + \beta_n(x_n - x_{n-1})) - \frac{1}{L} \nabla(z_n^* F)(x_n) \right\|^2 e_n \right\rangle,$$

one gets then

$$\begin{aligned} 0 &\leq \langle z_n^*, G(x_t) - G(x_n) \rangle + \frac{L}{2} \left\| x_t - (x_n + \beta_n(x_n - x_{n-1})) - \frac{1}{L} \nabla(z_n^* F)(x_n) \right\|^2 \\ &\quad - \frac{L}{2} \left\| x_n - (x_n + \beta_n(x_n - x_{n-1})) - \frac{1}{L} \nabla(z_n^* F)(x_n) \right\|^2, \end{aligned}$$

which, employing the definition of x_t and the hypothesis $x_{n-1} = x_n$, yields

$$0 \leq \langle z_n^*, G(x_t) - G(x_n) \rangle + \frac{L}{2} \left\| (1-t)(\tilde{x} - x_n) + \frac{1}{L} \nabla(z_n^* F)(x_n) \right\|^2 - \frac{1}{2L} \|\nabla(z_n^* F)(x_n)\|^2. \quad (10)$$

As G is C -convex, one has $\langle z_n^*, G(x_t) - G(x_n) \rangle \leq (1-t)\langle z_n^*, G(\tilde{x}) - G(x_n) \rangle$, while

$$\begin{aligned} \left\| (1-t)(\tilde{x} - x_n) + \frac{1}{L} \nabla(z_n^* F)(x_n) \right\|^2 &= (1-t)^2 \|\tilde{x} - x_n\|^2 + \frac{1}{L^2} \|\nabla(z_n^* F)(x_n)\|^2 \\ &\quad + (1-t) \frac{2}{L} \langle \tilde{x} - x_n, \nabla(z_n^* F)(x_n) \rangle, \end{aligned}$$

consequently, (10) yields

$$0 \leq (1-t)\langle z_n^*, G(\tilde{x}) - G(x_n) \rangle + \frac{L}{2} (1-t)^2 \|\tilde{x} - x_n\|^2 + (1-t) \langle \tilde{x} - x_n, \nabla(z_n^* F)(x_n) \rangle.$$

Dividing with $1-t$ and using the convexity of $(z_n^* F)$, one gets

$$0 \leq \langle z_n^*, G(\tilde{x}) - G(x_n) \rangle + \frac{L}{2} (1-t) \|\tilde{x} - x_n\|^2 + \langle z_n^*, F(\tilde{x}) - F(x_n) \rangle,$$

followed by $\langle z_n^*, c \rangle \leq \frac{L}{2} (1-t) \|\tilde{x} - x_n\|^2$.

Letting t tend towards 1 and using that $z_n^* \in C^* \setminus \{0\}$ and $c \in \text{int}C$, the last inequality yields $0 < \langle z_n^*, c \rangle \leq 0$, which is a contradiction, consequently, $x_n \in \mathcal{WE}(VP)$. \square

Remark 16: Note that $x_{n-1} = x_n$ does not necessarily imply that x_{n+1} coincides with them, too, but the fact that it depends only on x_n and not on x_{n-1} . This can prove to be useful when starting the algorithm because one can begin with $x_0 = x_1$ without affecting the convergence of the method.

4. Numerical experiments

In order to verify the proposed methods, we present in the following an example where a multiobjective optimization problem is solved by implementing Algorithm 1 in MATLAB (version 9.0.0.341360/R2016a) on a Windows 7-PC with an Intel Core i5 processor with 3.40 GHz and 8 GB of RAM.

Consider the vector optimization problem

$$(EP) \quad \begin{aligned} &\text{Min} \\ &x=(x_1, \dots, x_d) \in \mathbb{R}_+^d, \quad \begin{pmatrix} -x^\top u \\ x^\top Vx \end{pmatrix}, \\ &\sum_{i=1}^d x_i = 1 \end{aligned}$$

where $u \in \mathbb{R}^d$ and $V \in \mathbb{R}^{d \times d}$ is a symmetric positive semidefinite matrix. Such problems can be found, for instance, in portfolio optimization, where x can be interpreted as the portfolio vector for d given assets having the proportions of the assets in the whole portfolio as components where the short sales are excluded, the first component of the objective vector function represents the negative of the expected return (that is to be maximized, therefore minimized with a leading minus), while the second is the variance of the portfolio, expressed by a quadratic function involving a symmetric positive semidefinite variance-covariance matrix $V \in \mathbb{R}^{d \times d}$, that quantifies the risk associated to the considered portfolio and should be concomitantly minimized.

The vector optimization problem (EP) can be recast as a special case of (VP) by taking $X = \mathbb{R}^d$, $Y = \mathbb{R}^2$, $C = \mathbb{R}_+^2$, $F(x) = (-x^\top u, x^\top Vx)^\top$ and $G(x) = (\delta_{\mathbb{R}_+^d \cap T}(x), \delta_{\mathbb{R}_+^d \cap T}(x))^\top$, where $T = \{x = (x_1, \dots, x_d) \in \mathbb{R}^d : \sum_{i=1}^d x_i = 1\}$. Note that F is proper, \mathbb{R}_+^2 -convex and Fréchet differentiable and has a Lipschitz continuous gradient $\nabla F(x) = (u, Vx)^\top$ and G is proper, \mathbb{R}_+^2 -convex and positively \mathbb{R}_+^2 -lower semicontinuous.

For the concrete implementation of the method, we use the real data considered in [22] that contains five stocks IBM, Microsoft, Apple, Quest Diagnostics and Bank of America, whose expected return and variance in the portfolio were calculated based on historical stock price and dividend payment from 1 February 2002 to 1 February 2007. Consider thus the problem (EP) with $d = 5$, $u = (0.4, 0.513, 4.085, 1.006, 1.236)^\top$ and

$$V = \begin{pmatrix} 0.006461 & 0.002983 & 0.00235487 & 0.00235487 & 0.00096889 \\ 0.002983 & 0.0039 & 0.00095937 & -0.0001987 & 0.00063459 \\ 0.002355 & 0.000959 & 0.01267778 & 0.00135712 & 0.00134481 \\ 0.002355 & -0.0002 & 0.00135712 & 0.00559836 & 0.00041942 \\ 0.000969 & 0.000635 & 0.00134481 & 0.00041942 & 0.0016229 \end{pmatrix}.$$

Take moreover $z_n^* = (1/\sqrt{2}, 1/\sqrt{2})^\top$ and $e_n = (1, 1)^\top$ for all $n \geq 1$, and as the starting points of the algorithm $x_0 = (0.25, 0.25, 0, 0.25, 0.25)^\top$ and $x_1 = (0.15, 0.25, 0.25, 0.2, 0.15)^\top$. Note that this choice of the scalarization function guarantees, in the light of Remark 10, that the iteratively generated sequence actually converges towards a properly efficient solution to (EP), that is, consequently, also efficient. In order to obtain only weakly efficient solutions to (EP), that are, as pointed by an anonymous reviewer, not quite relevant for the problem in discussion, one could take some components of the z_n^* to be 0. The intermediate problems scalarized with the corresponding z_n^* 's become

$$(SP_n) \quad \inf_{\substack{x=(x^1, x^2, x^3, x^4, x^5) \in \mathbb{R}_+^5, \\ x^1+x^2+x^3+x^4+x^5=1, \\ u^\top(x-x_n) \leq 0, \\ x^\top Vx - x_n^\top Vx_n \leq 0}} \frac{1}{2} \|x - x_n - \beta_n(x_n - x_{n-1}) + \frac{1}{\sqrt{2}}(u + Vx_n)\|^2, \quad n \geq 1.$$

For computational reasons, we consider an inexact version of the stopping rule 2', namely that $\|x_{n+1} - x_n\| \leq \varepsilon = 0.00001 \geq \|x_n - x_{n-1}\|$. The intermediate scalar problems (SP_n) are solved using the MATLAB function `fmincon`, the existing proximal-point methods being not employable because of the complicated constraint sets.

In the following tables, we present some of the achieved computational results. Taking the sequence $(\beta_n)_n$ constant, the program delivers the approximate properly efficient solution $\bar{x} = (0.00000015603, 0.0718, 0.3189, 0.1317, 0.4777)^\top$ to (EP) after 15.517354 seconds and 281 iterations when $\beta_n = 1/10$, $n \geq 1$. Although there is no certain rule, one can notice that when the value of β_n decreases the elapsed time and the number of iterations tend to increase. However, when the inertial step is omitted, i.e. $\beta_n = 0$ for all $n \geq 1$, and the method becomes a 'pure' forward-backward one, the algorithm needs 33.288016 seconds and 625 iterations until the stopping rule is activated. In all these cases `fmincon` delivers approximate global optimal solutions for the intermediate problems. On the other hand, when the sequence $(\beta_n)_n$ is nondecreasing but not constant, `fmincon` delivers for the intermediate problems mostly approximate local optimal solutions, however the elapsed time and the number of iterations decrease dramatically, less than a second being necessary to deliver the approximate properly efficient solution $\bar{x} = (0.000000046008, 0.00000072695, 0.3050, 0.0878, 0.6072)^\top$ to (EP) after 13 iterations when $\beta_n = 1/30 - 1/(n+30)$, $n \geq 1$.

β_n	Iterations	Time (s)
0	625	33.288016
$\frac{1}{10}$	281	15.517354
$\frac{1}{11}$	579	31.590278
$\frac{1}{12}$	707	38.557769
$\frac{1}{15}$	603	32.456625
$\frac{1}{20}$	507	28.100981
$\frac{1}{25}$	507	27.316595
$\frac{1}{30}$	571	30.952835
$\frac{1}{50}$	749	40.752453
$\frac{1}{100}$	813	43.925126
$\frac{1}{300}$	783	39.746160
$\frac{1}{500}$	1083	60.644596
$\frac{1}{10} - \frac{1}{10n}$	13	1.080089
$\frac{1}{10} - \frac{1}{n+10}$	117	11.592547
$\frac{1}{11} - \frac{1}{n+11}$	117	11.784833
$\frac{1}{15} - \frac{1}{n+15}$	117	11.692819
$\frac{1}{20} - \frac{1}{n+20}$	15	1.721906
$\frac{1}{25} - \frac{1}{n+25}$	13	1.038344
$\frac{1}{30} - \frac{1}{n+30}$	13	0.988701
$\frac{1}{50} - \frac{1}{n+50}$	15	1.245741
$\frac{1}{100} - \frac{1}{n+100}$	15	1.989177
$\frac{1}{300} - \frac{1}{n+300}$	15	1.960903
$\frac{1}{500} - \frac{1}{n+500}$	15	1.975370

5. Conclusions and further research

In this paper, we propose two forward–backward proximal point type algorithms with inertial/memory effects for determining weakly efficient solutions to a vector optimization problem consisting in vector-minimizing with respect to a given closed convex pointed cone the sum of a proper cone-convex vector function with a cone-convex differentiable one, the first ones with these characteristics in the literature, to the best of our knowledge. Among the ideas we consider for future work in this research direction we mention first the identification of possible ways to avoid using the constraint sets Ω_n , $n \geq 1$, without losing the convergence of the method. Likewise, we are also interested in finding alternative hypotheses to the C -completeness of $(F + G)(X) \cap (F(x_0) + G(x_0) - C)$ that are weaker than the ones mentioned in [4, Remark 3] as well as in providing convergence rates for our algorithm in more general frameworks than the one we give in the non-inertial case when the scalarizing sequence is constant. Moreover, we plan to investigate some ways to modify the proposed algorithms in order to encompass as a special case also the projected gradient method proposed in [8] for vector-minimizing a smooth cone-convex vector function. Extending our investigations by employing Bregman type distances instead of the classical one like in [9] is another idea worth exploring.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Research partially supported by FWF (Austrian Science Fund) [project number I 2419-N32]; DFG (German Research Foundation) [project number WA 922/8-1], [project number GR3367/4-1].



ORCID

Radu Ioan Boț <http://orcid.org/0000-0002-4469-314X>

Sorin-Mihai Grad <http://orcid.org/0000-0002-1139-7504>

References

- [1] Alvarez F, Attouch H. An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. *Set-Valued Anal.* **2001**;9:3–11.
- [2] Alvarez F. On the minimizing property of a second order dissipative system in Hilbert spaces. *SIAM J Control Optim.* **2000**;38:1102–1119.
- [3] Buong N. Inertial proximal point regularization algorithm for unconstrained vector convex optimization problems. *Ukrainian Math J.* **2008**;60:1483–1491.
- [4] Bonnel H, Iusem AN, Svaiter BF. Proximal methods in vector optimization. *SIAM J Optim.* **2005**;15:953–970.
- [5] Miglierina E, Molho E, Recchioni MC. Box-constrained multi-objective optimization: a gradient-like method without ‘a priori’ scalarization. *Eur J Oper Res.* **2008**;188:662–682.
- [6] Graña Drummond LM, Maculan N, Svaiter BF. On the choice of parameters for the weighting method in vector optimization. *Math Program.* **2008**;111:201–216.
- [7] Fliege J, Graña Drummond LM, Svaiter BF. Newton’s method for multiobjective optimization. *SIAM J Optim.* **2009**;20:602–626.
- [8] Graña Drummond LM, Iusem AN. A projected gradient method for vector optimization problems. *Comput Optim Appl.* **2004**;28:5–29.
- [9] Villacorta KDV, Oliveira PR. An interior proximal method in vector optimization. *Eur J Oper Res.* **2011**;214:485–492.
- [10] Boț RI, Grad S-M, Wanka G. *Duality in vector optimization*. Berlin: Springer-Verlag; **2009**.
- [11] Opial Z. Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bull Am Math Soc.* **1967**;73:591–597.
- [12] Moudafi A, Oliny M. Convergence of a splitting inertial proximal method for monotone operators. *J Comput Appl Math.* **2003**;155:447–454.
- [13] Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J Imaging Sci.* **2009**;2:183–202.
- [14] 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-Verlag; **2011**.
- [15] Bolintineanu Ș. Approximate efficiency and scalar stationarity in unbounded nonsmooth convex vector optimization problems. *J Optim Theory Appl.* **2000**;106:265–296.
- [16] Boț RI, Hendrich C. A Douglas-Rachford type primal-dual method for solving inclusions with mixtures of composite and parallel-sum type monotone operators. *SIAM J Optim.* **2013**;23:2541–2565.
- [17] Boț RI, Csetnek ER, Heinrich A. A primal-dual splitting algorithm for finding zeros of sums of maximal monotone operators. *SIAM J Optim.* **2013**;23:2011–2036.
- [18] Aliprantis CD, Florenzano M, Martins-da-Rocha VF, Tourky R. Equilibrium analysis in financial markets with countably many securities. *J Math Econ.* **2004**;40:683–699.
- [19] Gong X-H. Optimality conditions for Henig and globally proper efficient solutions with ordering cone has empty interior. *J Math Anal Appl.* **2005**;307:12–31.
- [20] Grad S-M. *Vector optimization and monotone operators via convex duality*. Cham: Springer-Verlag; **2015**.
- [21] Grad S-M, Pop EL. Vector duality for convex vector optimization problems by means of the quasi interior of the ordering cone. *Optimization.* **2014**;63:21–37.
- [22] Duan YC. A multi-objective approach to portfolio optimization. *Undergrad Math J.* **2007**;8(1):18p.