Relaxed-Inertial Proximal Point Type Algorithms for Quasiconvex Minimization

Sorin-Mihai Grad, Felipe Lara, Raul Tintaya Marcavillaca

Relaxed-Inertial Proximal Point Type Algorithms for Quasiconvex Minimization

S.-M. Grad* F. Lara† R. T. Marcavillaca ‡

August 22, 2022

Abstract We propose a relaxed-inertial proximal point type algorithm for solving optimization problems consisting in minimizing strongly quasiconvex functions whose variables lie in finitely dimensional linear subspaces. A relaxed version of the method where the constraint set is only closed and convex is also discussed, and so is the case of a quasiconvex objective function. Numerical experiments illustrate the theoretical results.

Keywords: Proximal point algorithms, Relaxed methods, Inertial methods, Generalized convexity, Strong quasiconvexity.

1 Introduction

Initially introduced by Martinet in [23] and afterwards extended (first by Rockafellar in [24]) to other contexts, the proximal point algorithm was quickly extended from simple unconstrained convex optimization problems to dealing with various classes of problems, including the minimization of structured convex and even nonconvex functions. Currently there exist various proximal point type algorithms for minimizing generalized convex functions, weakly convex, DC (difference of convex), Kurdyka-Łojasiewicz functions, and even ratios with certain properties (see [10, 12, 15, 20, 22] and the references therein).

Studied both for theoretical reasons and due to concrete applications, for instance in economic theory, financial theory and approximation theory, quasiconvex functions belong to the most investigated classes of generalized convex functions, see, for instance, [6, 13, 14, 26]. Although quasiconvex functions inherit some important features from convex functions, they have some stability properties that convex functions do not. For example, for any constant $\theta \in \mathbb{R}$,

*Department of Applied Mathematics, ENSTA Paris, Polytechnic Institute of Paris, F-91120 Palaiseau, France, and Corvinus Center for Operations Research, Corvinus Institute for Advanced Studies, Corvinus University of Budapest, H-1093 Budapest, Hungary. E-mail: sorin-mihai.grad@ensta-paris.fr, ORCID-ID: 0000-0002-1139-7504

†Departamento de Matemática, Facultad de Ciencias, Universidad de Tarapacá, Arica, Chile. E-mail: felipelaraobreque@gmail.com; flarao@uta.cl. Web: felipelara.cl

‡Departamento de Matemática, Facultad de Ciencias, Universidad de Tarapacá, Arica, Chile. E-mail: raultm.rt@gmail.com, ORCID-ID: 0000-0003-3748-0768

the truncation $\min\{h, \theta\}$ is quasiconvex when $h : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ is quasiconvex, however it may fail to be convex when h is convex. Hence the importance of the quasiconvex functions.

There are already several proximal point type algorithmic schemes for minimizing such functions in works like [12, 15, 20, 22]. An important subclass of quasiconvex functions is the strongly quasiconvex ones (not to be confused with the semistrictly quasiconvex functions, as it sometimes happened in the literature), see e.g., [19, 21, 25]. It has only very recently been shown in [21] that the classical proximal point algorithm converges under standard assumptions when employed for minimizing a strongly quasiconvex function.

The extension of proximal point algorithms from convex functions to generalized convex functions (in particular quasiconvex and strongly quasiconvex) may provide new tools for dealing with applications in machine learning and stochastic optimization problems beyond convexity (see for instance [7, 18] for strongly convex functions). Further, such investigations could as well serve for developing splitting methods for minimizing sums involving quasiconvex and/or strongly quasiconvex functions.

Motivated by works like [1, 5, 15, 22], where inertial effects and relaxations of the iterative steps were added to classical proximal point algorithms for solving convex optimization problems in order to improve them in terms of velocity and efficiency, we augment in this paper the study from [21] by considering a relaxed-inertial proximal point type algorithm from [5] for solving optimization problems consisting in minimizing strongly quasiconvex functions whose variables lie in finitely dimensional linear subspaces. Moreover, leaving aside the inertial steps and taking convenient relaxation parameters, the constraint set can be taken only closed and convex, and the algorithm remains well-defined. We provide convergence statements for the iterative sequences generated by these algorithms and, in the more general case of a quasiconvex objective function, for the values of the objective function at these points. Moreover, we discuss different conditions which guarantee the fulfillment of the hypotheses of the convergence statements. Numerical experiments illustrate the theoretical results, as the computational results show that the relaxed-inertial proximal point type algorithm considered in this work minimizes a strongly quasiconvex function faster than the classical proximal point algorithm investigated in [21].

The structure of the paper is as follows. Section 2 contains notations, preliminaries and basic definitions concerning generalized convexity functions and proximal point algorithms. In Section 3 we present a relaxed-inertial proximal point algorithm for solving nonconvex minimization problems, in particular, we study the convergence of the generated sequence for minimizing strongly quasiconvex functions and the general case of quasiconvex functions is revisited. In Section 4 we present different sufficient conditions for ensuring the convergence of the considered algorithm. Finally, in Section 5, numerical implementations are presented in order to show the advantages of the considered method with respect to its "pure" proximal point counterpart.

2 Preliminaries and Basic Definitions

The *inner product* of $\mathbb{R}^n$ and the *Euclidean norm* are denoted by $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$, respectively. Given a convex and closed set $K \subseteq \mathbb{R}^n$, the *projection* of $x \in \mathbb{R}^n$ on K is denoted by $P_K(x)$, and the *indicator function* on K by $\iota_K : \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$.

Given any $x, y, z \in \mathbb{R}^n$ and any $\beta \in \mathbb{R}$, the following relations hold:

$$\langle x - z, y - x \rangle = \frac{1}{2}\|z - y\|^2 - \frac{1}{2}\|x - z\|^2 - \frac{1}{2}\|y - x\|^2, \tag{2.1}$$

$$\|\beta x + (1-\beta)y\|^2 = \beta\|x\|^2 + (1-\beta)\|y\|^2 - \beta(1-\beta)\|x - y\|^2. \tag{2.2}$$

Given any extended real-valued function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$, the *effective domain* of h is defined by $\mathrm{dom}\, h := \{x \in \mathbb{R}^n : h(x) < +\infty\}$. It is said that h is *proper* if $\mathrm{dom}\, h$ is nonempty and $h(x) > -\infty$ for all $x \in \mathbb{R}^n$. We denote by $\mathrm{epi}\, h := \{(x,t) \in \mathbb{R}^n \times \mathbb{R} : h(x) \leq t\}$ the *epigraph* of h, by $S_\lambda(h) := \{x \in \mathbb{R}^n : h(x) \leq \lambda\}$ the *sublevel set of h at the height* $\lambda \in \mathbb{R}$ and by $\arg\min_{\mathbb{R}^n} h$ the *set of minimal points of h*. The function h is *lower semicontinuous* at $\overline{x} \in \mathbb{R}^n$ if for any sequence $\{x_k\}_k \in \mathbb{R}^n$ with $x_k \to \overline{x}$, $h(\overline{x}) \leq \liminf_{k \to +\infty} h(x_k)$. Furthermore, the usual convention $\sup_{\emptyset} h := -\infty$ and $\inf_{\emptyset} h := +\infty$ is adopted.

A function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ with convex domain is said to be

(a) *convex* if, given any $x, y \in \mathrm{dom}\, h$, then

$$h(\lambda x + (1-\lambda)y) \leq \lambda h(x) + (1-\lambda)h(y),\ \forall\ \lambda \in [0,1]; \tag{2.3}$$

(b) *semistrictly quasiconvex* if, given any $x, y \in \mathrm{dom}\, h$, with $h(x) \neq h(y)$, then

$$h(\lambda x + (1-\lambda)y) < \max\{h(x), h(y)\},\ \forall\ \lambda \in]0,1[; \tag{2.4}$$

(c) *quasiconvex* if, given any $x, y \in \mathrm{dom}\, h$, then

$$h(\lambda x + (1-\lambda)y) \leq \max\{h(x), h(y)\},\ \forall\ \lambda \in [0,1]. \tag{2.5}$$

It is said that h is *strictly convex* (resp. *strictly quasiconvex*) if the inequality in (2.3) (resp. (2.5)) is strict whenever $x \neq y$.

Every (stricly) convex function is (stricly) quasiconvex and semistrictly quasiconvex, and every semistrictly quasiconvex and lower semicontinuous function is quasiconvex (see [13, Theorem 2.3.2]). The function $h : \mathbb{R} \to \mathbb{R}$, with $h(x) := \min\{|x|, 1\}$, is quasiconvex without being semistrictly quasiconvex. Recall that

$$h \text{ is convex} \Longleftrightarrow \mathrm{epi}\, h \text{ is a convex set};$$
$$h \text{ is quasiconvex} \Longleftrightarrow S_\lambda(h) \text{ is a convex set for all } \lambda \in \mathbb{R}.$$

For algorithmic purposes, the following notions from [25] are useful.

A function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ with a convex domain is said to be:

(*a*) *strongly convex* if there exists $\gamma \in]0,+\infty[$ such that for all $x, y \in \mathrm{dom}\, h$ and all $\lambda \in [0,1]$, we have

$$h(\lambda y + (1-\lambda)x) \leq \lambda h(y) + (1-\lambda)h(x) - \lambda(1-\lambda)\frac{\gamma}{2}\|x-y\|^2; \tag{2.6}$$

(*b*) *strongly quasiconvex* if there exists $\gamma \in]0,+\infty[$ such that for all $x, y \in \mathrm{dom}\, h$ and all $\lambda \in [0,1]$, we have

$$h(\lambda y + (1-\lambda)x) \leq \max\{h(y), h(x)\} - \lambda(1-\lambda)\frac{\gamma}{2}\|x-y\|^2. \tag{2.7}$$

In these cases, it is said that h is *strongly convex* (resp. *strongly quasiconvex*) *with modulus* $\gamma > 0$. When relations (2.6) or (2.7) hold only for all $x, y \in K \subset \mathrm{dom}\, h$ it is said that h is *strongly convex (quasiconvex) on* K.

Note that every strongly convex function is strongly quasiconvex, and every strongly quasiconvex function is strictly quasiconvex. The Euclidean norm $\|\cdot\|$ is strongly quasiconvex without being strongly convex on any bounded convex set $K \subseteq \mathbb{R}^n$ (see [19, Theorem 2]) and the function $x \mapsto x^3$ is strictly quasiconvex without being strongly quasiconvex on $\mathbb{R}$.

Summarizing, we have the following implications between the (generalized) convexity notions introduced above (quasiconvex is denoted by qcx in the following scheme)

$$\begin{array}{ccccccc}
\text{strongly convex} & \Longrightarrow & \text{strictly convex} & \Longrightarrow & \text{convex} & \Longrightarrow & \text{qcx} \\
\Downarrow & & \Downarrow & & \Downarrow & & \\
\text{strongly qcx} & \Longrightarrow & \text{strictly qcx} & \Longrightarrow & \text{semistrictly qcx} & & \\
 & & \Downarrow & & & & \\
 & & \text{qcx} & & & &
\end{array}$$

Remark 2.1. *There is no relationship in general between convexity and strong quasiconvexity of functions. Indeed, the function* $h : \mathbb{R}^n \to \mathbb{R}$ *given by* $h(x) = \sqrt{\|x\|}$ *is strongly quasiconvex on any bounded and convex set* K *in* $\mathbb{R}^n$ *without being convex, while the function* $h(x) \equiv 1$ *is convex without being strongly quasiconvex. However, strongly convex functions are both convex and strongly quasiconvex.*

A proper function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ is said to be *coercive*, if $S_\lambda(h)$ is bounded for all $\lambda \in \mathbb{R}$, or equivalently, if

$$\lim_{\|x\| \to +\infty} h(x) = +\infty. \tag{2.8}$$

The following existence result is the starting point of our investigations.

Lemma 2.1. *(cf. [21, Corollary 3]) Let* $K \subseteq \mathbb{R}^n$ *be a closed and convex set and* $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ *be a proper, lower semicontinuous, and strongly quasiconvex function such that* $K \subseteq \mathrm{dom}\, h$*. Then,* $\arg\min_K h$ *is a singleton.*

The above result ensures that every lower semicontinuous and strongly quasiconvex function has exactly one minimizer on every closed and convex subset K of $\mathbb{R}^n$. Therefore, Lemma 2.1 is useful for analyzing proximal point algorithms for classes of quasiconvex functions (see [21]).

Let K be a closed and convex set in $\mathbb{R}^n$ and $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper function with $K \cap \mathrm{dom}\, h \neq \emptyset$. The *proximity operator on K of parameter $\beta > 0$ of* h at $x \in \mathbb{R}^n$ is defined as $\mathrm{Prox}_{\beta h} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ where

$$\mathrm{Prox}_{\beta h}(K, x) = \arg\min_{y \in K} \left\{ h(y) + \frac{1}{2\beta} \|y - x\|^2 \right\}, \tag{2.9}$$

whenever the minimum in (2.9) exists. If $K = \mathbb{R}^n$, then we simply write $\mathrm{Prox}_{\beta h}(\mathbb{R}^n, x) = \mathrm{Prox}_{\beta h}(x)$. Consequently, $\mathrm{Prox}_{\beta h}(K, \cdot) = \mathrm{Prox}_{\beta h + \iota_K}(\cdot)$ for all closed and convex $K \subseteq \mathbb{R}^n$. The *fixed points set* of $\mathrm{Prox}_{\beta h}(K, \cdot)$ is denoted by $\mathrm{Fix}(\mathrm{Prox}_{\beta h}(K, \cdot))$. When h is proper, lower semicontinuous and convex, $\mathrm{Prox}_{\beta h}$ turns out to be a single-valued operator (see, for instance, [8, Proposition 12.15]). Although in most of the literature the proximity operator of a function is considered (in the spirit of "full splitting") on the whole space, there are works like [10, 16, 17] where the employed functions are not split from the corresponding sets, as defined above.

The following results will be useful in the sequel. Note that quasiconvexity can be seen as strong quasiconvexity with modulus $\gamma = 0$.

Lemma 2.2. *(cf. [21, Proposition 7]) Let $K \subseteq \mathbb{R}^n$ be a closed and convex set, $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper, lower semicontinuous, strongly quasiconvex function with modulus $\gamma \geq 0$ and such that $K \subseteq \mathrm{dom}\, h$, $\beta > 0$ and $x \in K$. If $\overline{x} \in \mathrm{Prox}_{\beta h}(K, x)$, then*

$$h(\overline{x}) - \max\{h(y), h(\overline{x})\} \leq \frac{\lambda}{\beta}\langle \overline{x} - x, y - \overline{x}\rangle + \frac{\lambda}{2}\left(\frac{\lambda}{\beta} - \gamma + \lambda\gamma\right)\|y - \overline{x}\|^2,$$
$$\forall\ y \in K,\ \forall\ \lambda \in [0, 1]. \tag{2.10}$$

Lemma 2.3. *(cf. [21, Proposition 9]) Let $K \subseteq \mathbb{R}^n$ be a closed and convex set, $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper, lower semicontinuous and strongly quasiconvex function with modulus $\gamma > 0$ such that $K \subseteq \mathrm{dom}\, h$, and $\beta > 0$. Then*

$$\mathrm{Fix}\left(\mathrm{Prox}_{\beta h}(K, \cdot)\right) = \arg\min_K h. \tag{2.11}$$

The following lemma was essentially proven in [2, Theorem 2.1] (see also [4, Lemma A.4] for a short and direct proof) and will be used in the next section to analyze the convergence of the proposed algorithm.

Lemma 2.4. *Let the sequences $\{\eta_k\}_k$, $\{s_k\}_k$, $\{\alpha_k\}_k$ and $\{\delta_k\}_k$ in $[0, +\infty[$ and let $\alpha \in \mathbb{R}$ be such that $\eta_0 = \eta_{-1}$, $0 \leq \alpha_k \leq \alpha < 1$ and*

$$\eta_{k+1} - \eta_k + s_{k+1} \leq \alpha_k(\eta_k - \eta_{k-1}) + \delta_k,\ \forall\ k \geq 0. \tag{2.12}$$

Then the following assertions hold.

(*a*) *For all* $k \geq 1$,

$$\eta_k + \sum_{j=1}^{k} s_j \leq \eta_0 + \frac{1}{1-\alpha} \sum_{j=0}^{k-1} \delta_j. \tag{2.13}$$

(*b*) *If* $\sum_{k=0}^{\infty} \delta_k < +\infty$, *then* $\lim_{k\to\infty} \eta_k$ *exists, i.e., the sequence* $\{\eta_k\}_k$ *converges to some element in* $[0, +\infty[$.

Finally, we also consider the following lemmata for our convergence analysis.

Lemma 2.5. *Consider an arbitrary parameter* $\rho \in]0, 2[$.

(*a*) *Let* $\phi_1 :]\rho/(1+\rho), 2[\to]0, 1[$ *be a real function given by*

$$\phi_1(x) = \frac{2\rho\,(2-x)}{2\rho + 2x - \rho x + \sqrt{4x^2 - 4x^2\rho - 7x^2\rho^2 + 8x\rho + 20x\rho^2 - 12\rho^2}}.$$

Then its inverse function $\phi_1^{-1} :]0, 1[\to]\rho/(1+\rho), 2[$ *is given by*

$$\phi_1^{-1}(x) = \frac{2\rho(x^2 - x + 1)}{2\rho x^2 + (2-\rho)x + \rho}.$$

(*b*) *Let* $\phi_2 :]2\rho/(2+\rho), 2[\to]0, 1[$ *be a real function given by*

$$\phi_2(x) = \frac{2(2-x)}{(\frac{4}{\rho} - 1)x + \sqrt{\left[(\frac{4}{\rho} - 1)^2 - 8\right] x^2 + 24x - 16}}.$$

Then its inverse function $\phi_2^{-1} :]0, 1[\to]2\rho/(2+\rho), 2[$ *is given by*

$$\phi_2^{-1}(x) = \frac{2\rho(x^2 + 1)}{2\rho x^2 + (4-\rho)x + \rho}.$$

(*c*) *Let* $\phi_3 :]0, 2[\to]0, 1[$ *be a real fuction given by*

$$\phi_3(x) = \frac{2(2-x)}{4 - x + \sqrt{16x - 7x^2}}.$$

Then its inverse $\phi_3^{-1} :]0, 1[\to]0, 2[$ *is given by*

$$\phi_3^{-1}(x) = \frac{2(x-1)^2}{2(x-1)^2 + 3x - 1}.$$

A proof for item (c) of the previous lemma can be found in [3, Lemma A.2]. The other two statements can be shown in the same lines.

Lemma 2.6. *(cf. [3, Lemma A.3]) Let $q : \mathbb{R} \to \mathbb{R}$ be given by $q(x) := ax^2 - bx + c$. Assume that $b, c > 0$, $b^2 - 4ac > 0$ and define*

$$\beta := \frac{2c}{b + \sqrt{b^2 - 4ac}} > 0. \tag{2.14}$$

(*a*) *If $a = 0$, then $q(\cdot)$ is a decreasing affine function and $\beta > 0$ is its unique root.*

(*b*) *If $a > 0$ (resp. $a < 0$), then $q(\cdot)$ is a convex (resp. concave) quadratic function and $\beta > 0$ is its smallest (resp. largest) root.*

In both cases (a) and (b), $\beta > 0$ is a root of $q(\cdot)$, and $q(\cdot)$ is decreasing on the interval $[0, \beta]$.

For a further study regarding strong quasiconvexity and generalized convexity we refer to [13, 19, 21, 25], while for applications in economics and financial theory of quasiconvex and strongly quasiconvex functions we refer to [6, 14] among others.

3 A Relaxed Inertial Proximal Point Algorithm

Let K be a closed and convex set in $\mathbb{R}^n$, and $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper function such that $K \cap \text{dom } h \neq \emptyset$. We consider the constrained optimization problem

$$\min_{x \in K} h(x). \tag{COP}$$

In order to provide a *relaxed-inertial proximal point algorithm* (RIPPA henceforth) for solving problem (COP) when it has a strongly quasiconvex or a quasiconvex objective function, we take $K \subseteq \mathbb{R}^n$ to be a linear subspace, and consider the following assumptions on h

(A) $h : K \to \mathbb{R}$ is a continuous and strongly quasiconvex function with modulus $\gamma > 0$;

($B1$) $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a proper, continuous and quasiconvex function on $K \subseteq \text{dom } h$;

($B2$) h is 2-weakly coercive on K, that is,

$$\liminf_{x \in K, \, \|x\| \to +\infty} \frac{h(x)}{\|x\|^2} \geq 0. \tag{3.1}$$

Remark 3.1. *We need to take $K \subseteq \mathbb{R}^n$ to be a linear subspace because of the extrapolation step in the algorithm proposed below. We opted to work in the most general framework (i.e. by taking K to be a linear subspace instead of the whole space $\mathbb{R}^n$ like in the literature) because if the extrapolation step is omitted and convenient relaxation parameters are considered the resulting* *algorithm is capable of minimizing a strongly quasiconvex function over a closed convex set. Moreover, this setting might turn out to be relevant at some point to some application we are not yet aware of, as, for instance, quasiconvex functions over linear subspaces were considered in works like [26].*

In order to make assumption (A) comply with the general situation considered in this work we can take $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $K = \mathrm{dom}\, h$. Before continuing, we note the following.

Remark 3.2. *(i) Clearly, (A) implies $(B1)$ and $(B2)$ (see [21, Theorem 1]), while the converse implications do not hold even for continuous convex functions. Indeed, the function $h(x) = \alpha$, with $\alpha \in \mathbb{R}$, satisfies assumptions $(B1)$ and $(B2)$, but does not satisfy (A).*

(ii) If h satisfies assumptions $(B1)$ and $(B2)$, then $\mathrm{Prox}_{\beta h}(K, \cdot)$ is nonempty since $h(\cdot) + (1/(2\beta))\|\cdot\|^2$ is lower semicontinuous and coercive on K for all $\beta > 0$. If in addition, given $w \in K$ and $c > 0$, the function $h(\cdot) + (1/(2c))\|w - \cdot\|^2$ is strongly quasiconvex on K, then $\mathrm{Prox}_{ch}(K, w)$ is a singleton by Lemma 2.1.

(iii) Assumption (3.1) is not too restrictive. Indeed, every convex function satisfies assumption (3.1) as it has an affine minorant. Furthermore, every coercive function satisfies assumption (3.1).

(iv) The continuous quasiconvex function $h : \mathbb{R} \to \mathbb{R}$ given by $h(x) = \sqrt{|x|}$ for $x < 0$, and $h(x) = -\sqrt{x}$ for $x \geq 0$, satisfies assumption (3.1) and is not bounded from below by an affine minorant.

The algorithm that we propose can be found in the convex case in [5] and it is based on the ones proposed in [1, 15, 22]. Note that it uses at each iteration an extrapolation, a proximal step and a relaxation step.

The precise statement of Algorithm 1 is given below.

Algorithm 1 RIPPA for Strongly Quasiconvex Functions (RIPPA-SQ)

Step 0. (Initialization). Let $x^0 = x^{-1} \in K$, $\alpha \in [0,1[$, $0 < \rho' \le \rho'' < 2$, $\{c_k\}_{k\in\mathbb{N}}$ be a sequence of positive numbers and $k = 0$.

Step 1. Choose $\alpha_k \in [0,\alpha]$ and set

$$y^k = x^k + \alpha_k(x^k - x^{k-1}), \qquad \text{[extrapolation step]} \tag{3.2}$$

and compute

$$z^k \in \mathrm{Prox}_{c_k h}(K, y^k) \qquad \text{[proximal step]}. \tag{3.3}$$

Step 2. If $z^k = y^k$, then Stop, and $y^k \in \arg\min_K h$. Otherwise, choose $\rho_k \in [\rho', \rho'']$ and update

$$x^{k+1} = (1-\rho_k)y^k + \rho_k z^k \qquad \text{[relaxation step]}. \tag{3.4}$$

Step 3. Let $k = k+1$ and go to Step 1.

Before continuing, we make some remarks regarding Algorithm 1.

Remark 3.3. (*i*) *The iterative steps are well-defined by Lemma 2.1 and the stopping criterion follows from Lemma 2.3, because $z^k = y^k$ in relation* (3.3) *implies that y^k is a fixed point of the proximity operator, and hence an optimal solution to* (COP).

(*ii*) *The extrapolation step in* (3.2)*, which is controlled by the parameter α_k, introduces inertial effects in Algorithm 1, while the parameter ρ_k at step* (3.4) *represents the relaxation parameter.*

(*iii*) *In virtue of relation* (3.2) *and of* (3.4) *when $\rho_k \in [1,2]$, K should be assumed as a linear subspace and not only as a closed and convex set. We note that the authors in [1–3, 15, 22] considered their algorithms on the whole space.*

(*iv*) *Note that if $z^k \neq y^k$, then $h(z^k) < h(y^k)$. Indeed, if h satisfies (A) ($\gamma > 0$) or if h satisifies (B1) and (B2) ($\gamma = 0$), then by* (3.3) *and Lemma 2.2, we have (for $\gamma \ge 0$)*

$$h(z^k) - \max\{h(y), h(z^k)\} \le \frac{\lambda}{c_k}\langle z^k - y^k, y - z^k\rangle + \frac{\lambda}{2}\left(\frac{\lambda}{c_k} - \gamma + \lambda\gamma\right)\|y - z^k\|^2, \ \forall\, y \in K, \ \forall\, \lambda \in [0,1].$$

Take $y = y^k \in K$. Then,

$$h(z^k) - \max\{h(y^k), h(z^k)\} \le \frac{\lambda}{2c_k}\left(\lambda - 2 + \gamma c_k(\lambda - 1)\right)\|y^k - z^k\|^2, \ \forall\, \lambda \in [0,1].$$

Since $y^k \neq z^k$, by taking $\lambda = \frac{1}{2}$, we have

$$h(z^k) - \max\{h(y^k), h(z^k)\} \leq \frac{1}{4c_k}\left(-\frac{3}{2} - \frac{\gamma c_k}{2}\right)\|y^k - z^k\|^2 < 0$$
$$\implies h(z^k) < \max\{h(y^k), h(z^k)\} = h(y^k).$$

Therefore, if $z^k \neq y^k$, then $h(z^k) < h(y^k)$.

Let us consider the set

$$\Omega := \{x \in K : \, h(x) \leq h(z^k), \ \forall \, k \in \mathbb{N}\}. \tag{3.5}$$

A similar construction was used in [9] where the proximal point algorithm was extended for vector optimization problems. Note that under assumption (A), $\arg\min_K h$ is a singleton by Lemma 2.1, hence $\Omega \neq \emptyset$. Therefore, in order to encompass both cases in our analysis, quasiconvexity and strong quasiconvexity separately, we consider the following assumptions on h

$(C1)$ h satisfies assumption (A).

$(C2)$ h satisfies assumptions $(B1)$, $(B2)$ and $\Omega \neq \emptyset$.

We start the convergence analysis of Algorithm 1 by proving the following result.

Proposition 3.1. *Let $K \subseteq \mathbb{R}^n$ be a linear subspace, $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function such that assumption $(C1)$ or $(C2)$ holds, $0 < \rho' \leq \rho'' < 2$, $\{\rho_k\}_k \subseteq [\rho', \rho'']$ and $\{x^k\}_k$, $\{y^k\}_k$ and $\{z^k\}_k$ be the sequences generated by Algorithm 1. Then, for every $x^* \in \Omega$, we have*

$$||x^{k+1} - x^*||^2 + \rho_k(2-\rho_k)||y^k - z^k||^2 \leq ||y^k - x^*||^2, \ \forall \, k \in \mathbb{N}. \tag{3.6}$$

Proof. From the relaxation step (3.4) of Algorithm 1 we have

$$x^{k+1} - x^* = (1-\rho_k)(y^k - x^*) + \rho_k(z^k - x^*).$$

Now, using identity (2.2), we have

$$||x^{k+1} - x^*||^2 = (1-\rho_k)||y^k - x^*||^2 + \rho_k||z^k - x^*||^2 - \rho_k(1-\rho_k)||y^k - z^k||^2,$$

or equivalently,

$$\begin{aligned}\|x^{k+1} - x^*\|^2 - \|y^k - x^*\|^2 = {} & \rho_k(\|z^k - x^*\|^2 - \|y^k - x^*\|^2) \\ & - \rho_k(1-\rho_k)\|y^k - z^k\|^2.\end{aligned} \tag{3.7}$$

Furthermore, it follows from relation (2.1) that

$$\|z^k - x^*\|^2 - \|y^k - x^*\|^2 = \|z^k - y^k\|^2 + 2\langle z^k - y^k, y^k - x^*\rangle. \tag{3.8}$$

On the other hand, from the proximal step (3.3) and Lemma 2.2, we have for every $k \geq 0$ that

$$\begin{aligned} h(z^k) - \max\{h(y), h(z^k)\} \leq \ & \frac{\lambda}{c_k}\langle z^k - y^k, y - z^k\rangle \\ & + \frac{\lambda}{2}\left(\frac{\lambda}{c_k} - \gamma + \lambda\gamma\right)\|y - z^k\|^2,\ \forall\, y \in K,\ \forall\, \lambda \in [0,1]. \end{aligned} \tag{3.9}$$

Take $y = x^* \in \Omega$ in relation (3.9). Then,

$$\begin{aligned} 0 &\leq 2\langle z^k - y^k, x^* - z^k\rangle + c_k\left(\frac{\lambda}{c_k} - \gamma + \lambda\gamma\right)\|x^* - z^k\|^2 \\ &= 2\langle z^k - y^k, x^* - y^k\rangle - 2\|z^k - y^k\|^2 + c_k\left(\frac{\lambda}{c_k} - \gamma + \lambda\gamma\right)\|x^* - z^k\|^2. \end{aligned} \tag{3.10}$$

We have two cases.

(*i*) If h satisfies $(C1)$, then $\gamma > 0$, thus by taking λ such that $0 < \lambda < \gamma c_k/(1 + \gamma c_k)$ in (3.10), we have

$$\begin{aligned} & 2\langle z^k - y^k, y^k - x^*\rangle \leq -2\|z^k - y^k\|^2 \\ \overset{(3.8)}{\Longrightarrow}\ & \|z^k - x^*\|^2 - \|y^k - x^*\|^2 \leq -\|z^k - y^k\|^2. \end{aligned} \tag{3.11}$$

Using inequality (3.11) in relation (3.7), we obtain (3.6).

(*ii*) If h satisfies $(C2)$, then $\gamma = 0$, thus we obtain relation (3.10) with $\gamma = 0$. Then, by taking $\lambda \downarrow 0$ in (3.10), we have

$$\begin{aligned} & 2\langle z^k - y^k, y^k - x^*\rangle \leq -2\|z^k - y^k\|^2 \\ \overset{(3.8)}{\Longrightarrow}\ & \|z^k - x^*\|^2 - \|y^k - x^*\|^2 \leq -\|z^k - y^k\|^2. \end{aligned} \tag{3.12}$$

Similarly to (i), using inequality (3.12) in relation (3.7), we obtain (3.6).

Therefore, in both cases, we obtain (3.6) and the proof is complete. □

As a consequence of the previous result, we have the following.

Proposition 3.2. *Let $K \subseteq \mathbb{R}^n$ be a linear subspace, $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ be such that one of conditions $(C1)$ and $(C2)$ holds, $0 < \rho' \leq \rho'' < 2$, $\{\rho_k\}_k \subseteq [\rho', \rho'']$, $\alpha \in [0,1[$, $\{\alpha_k\}_k \subseteq [0,\alpha]$, and $\{x^k\}_k$, $\{y^k\}_k$ and $\{z^k\}_k$ be the sequences generated by Algorithm 1. Given any $x^* \in \Omega$, we set $\eta_k := ||x^k - x^*||^2$ for all $k \geq 0$. Then,*

$$\begin{aligned} \eta_{k+1} - \eta_k - \alpha_k(\eta_k - \eta_{k-1}) + \frac{2 - \rho_k}{\rho_k}||x^{k+1} - y^k||^2 & \\ \leq (\alpha_k^2 + \alpha_k)||x^k - x^{k-1}||^2,\ \forall\ k \in \mathbb{N}. & \end{aligned} \tag{3.13}$$

Proof. Observe that, from (3.2), for every $x^* \in \Omega$, we have

$$x^k - x^* = \frac{1}{1+\alpha_k}(y^k - x^*) + \frac{\alpha_k}{1+\alpha_k}(x^{k-1} - x^*), \tag{3.14}$$

$$y^k - x^{k-1} = (1+\alpha_k)(x^k - x^{k-1}). \tag{3.15}$$

Using identities (2.2) and (3.14), we obtain

$$||x^k - x^*||^2 = \frac{||y^k - x^*||^2}{1+\alpha_k} + \frac{\alpha_k}{1+\alpha_k}||x^{k-1} - x^*||^2 - \frac{\alpha_k}{(1+\alpha_k)^2}||y^k - x^{k-1}||^2,$$

which combined with (3.15) gives

$$\begin{aligned} ||y^k - x^*||^2 &= (1+\alpha_k)||x^k - x^*||^2 - \alpha_k||x^{k-1} - x^*||^2 + \alpha_k(1+\alpha_k)||x^k - x^{k-1}||^2 \\ &= \eta_k + \alpha_k(\eta_k - \eta_{k-1}) + (\alpha_k^2 + \alpha_k)||x^k - x^{k-1}||^2. \end{aligned}$$

Hence, relation (3.13) follows immediately from Proposition 3.1 and the last equality. □

We give now our first main result, which shows that the sequences $\{x^k\}_k$, $\{y^k\}_k$ and $\{z^k\}_k$, generated by Algorithm 1 converge to an optimal solution to problem (COP) considering that h satisfies assumption $(C1)$.

Theorem 3.1. *Let $K \subseteq \mathbb{R}^n$ be a linear subspace, $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function such that assumption $(C1)$ holds, $0 < \rho' \leq \rho'' < 2$, $\{\rho_k\}_k \subseteq [\rho', \rho'']$, $\alpha \in [0, 1[$, $\{\alpha_k\}_k \subseteq [0, \alpha]$ and $\{x^k\}_k$, $\{y^k\}_k$ and $\{z^k\}_k$, be the sequences generated by Algorithm 1. If*

$$\sum_{k=0}^{\infty} \alpha_k ||x^k - x^{k-1}||^2 < +\infty, \tag{3.16}$$

then the following assertions hold.

- (a) *For every $x^* \in \Omega$, the limit $\lim_{k\to\infty}\|x^k - x^*\|$ exists and*

$$\lim_{k\to+\infty}\|x^{k+1} - y^k\| = \lim_{k\to+\infty}\|z^k - y^k\| = 0. \tag{3.17}$$

- (b) *If in addition $c_k \geq c' > 0$ for every $k \geq 0$, then the sequence $\{x^k\}_k$ converges to $\overline{x} = \arg\min_K h$ and $\lim_{k\to+\infty} h(x^k) = \min_K h$. Moreover, the sequences $\{y^k\}_k$ and $\{z^k\}_k$ converge both to $\overline{x}$, too.*

Proof. (a): Let $x^* \in \Omega$. Defining $\delta_k = (\alpha_k^2 + \alpha_k)||x^k - x^{k-1}||^2$, by Proposition 3.2 we observe that condition (2.12) in Lemma 2.4 is fulfilled with $s_{k+1} = ((2-\rho_k)/\rho_k)||x^{k+1} - y^k||^2$. Hence, using assumption (3.16) and Lemma 2.4(b), we conclude that the limit $\lim_{k\to\infty}||x^k - x^*||$ exists. In particular, $\{x^k\}_k$ is bounded. Moreover, from assumption (3.16) and Lemma 2.4(a), we conclude that $\sum_{k=0}^{\infty} s_{k+1} < +\infty$, and so, $s_{k+1} \to 0$ as $k \to +\infty$, thus due to the boundedness of $\{\rho_k\}_k$ it holds $\lim_{k\to+\infty}\|x^{k+1} - y^k\| = 0$.

Finally, since $x^{k+1} - y^k = \rho_k(z^k - y^k)$ by relation (3.4) and $0 < \rho' \leq \rho_k < 2$, it follows that $\lim_{k \to +\infty} \|z^k - y^k\| = 0$.

(b): By part (a), the sequence $\{x^k\}_k$ is bounded, i.e., it has cluster points. Let $\widehat{x} \in K$ be a cluster point of $\{x^k\}_k$. Then, there exists a subsequence $\{x^{k_l}\}_l$ of $\{x^k\}_k$ such that $x^{k_l} \to \widehat{x}$ as $l \to +\infty$. Applying (3.17) to the subsequence $\{x^{k_l}\}_l$ and since $\lim_{l \to +\infty} x^{k_l} = \widehat{x}$, we conclude that

$$\lim_{l \to +\infty} y^{k_l} = \lim_{l \to +\infty} z^{k_l} = \widehat{x}. \tag{3.18}$$

Now, by the construction of the sequence, (3.3) and Lemma 2.2, we have

$$\begin{aligned} h(z^k) - \max\{h(y), h(z^k)\} \leq & \ \frac{\lambda}{c_k} \langle z^k - y^k, y - z^k \rangle \\ & + \frac{\lambda}{2} \left(\frac{\lambda}{c_k} - \gamma + \lambda\gamma \right) \|y - z^k\|^2, \ \forall \, y \in K, \ \forall \, \lambda \in [0, 1]. \end{aligned} \tag{3.19}$$

Replace k by k_l in the previous inequality. Then, by taking $\liminf_{l \to +\infty}$, relation (3.18), $x^{k_l} \to \widehat{x}$ and the continuity of h, we obtain

$$h(\widehat{x}) - \max\{h(y), h(\widehat{x})\} \leq \frac{\lambda}{2} \left(\frac{\lambda}{c'} - \gamma + \lambda\gamma \right) \|y - \widehat{x}\|^2, \ \forall \, y \in K, \ \forall \, \lambda \in [0, 1].$$

Take $\lambda < (\gamma c')/(1 + \gamma c') < 1$, thus

$$h(\widehat{x}) - \max\{h(y), h(\widehat{x})\} < 0, \ \ \forall \ y \in K \backslash \{\widehat{x}\}.$$

Therefore, $\widehat{x} \in \arg\min_K h$ and every cluster point of the sequence $\{x^k\}_k$ belongs to $\arg\min_K h$. Since h is strongly quasiconvex, $\arg\min_K h = \{\overline{x}\}$, that is, $\widehat{x} = \overline{x}$ and the whole sequence $\{x^k\}_k$ converges to $\overline{x} = \arg\min_K h$ and $\lim_{k \to +\infty} h(x^k) = \min_K h$.

Finally, since $x^k \to \overline{x} = \arg\min_K h$, it follows from relation (3.17) that $y^k \to \overline{x}$ and $z^k \to \overline{x}$. $\square$

By taking $\alpha = 0$, we obtain $\alpha_k = 0$ for every $k \geq 0$. Hence, as a consequence of Theorem 3.1, we have the following convergence statement involving a *relaxed proximal point algorithm* (RPPA henceforth) for solving (COP). Note that its convergence statement does not require K to be a linear subspace when $\rho_k \in]0, 1]$ for every $k \geq 0$.

Corollary 3.1. *Let $K \subseteq \mathbb{R}^n$ be closed and convex, h be a function such that assumption ($C1$) holds, $\{c_k\}_{k \in \mathbb{N}}$ be a sequence of positive numbers and $\rho_k \in]0, 1]$ for every $k \geq 0$. Then, for any sequence $\{x^k\}_k$ generated by*

$$\text{(RPPA)} \qquad \left\{ \begin{array}{l} k = k + 1 \\ z^k \in \mathrm{Prox}_{c_k h}(K, x^k) \\ x^{k+1} = (1 - \rho_k) x^k + \rho_k z^k, \end{array} \right. \tag{3.20}$$

we have

(*a*) $\displaystyle\sum_{k=1}^{+\infty} \frac{2-\rho_k}{\rho_k}\|x^{k+1}-x^k\|^2 < +\infty$;

(*b*) *for every* $x^* \in \Omega$, *the limit* $\lim_{k\to+\infty}\|x^k - x^*\|$ *exists, and hence* $\{x^k\}$ *is bounded;*

(*c*) *if in addition* $c_k \geq c' > 0$ *for every* $k \geq 0$, *then the sequence* $\{x^k\}_k$, *generated by* $(RPPA)$, *converges to* $\{\overline{x}\} = \arg\min_K h$ *and* $\lim_{k\to+\infty} h(x^k) = \min_K h$.

The statement of Corollary 3.1 remains valid when $\rho_k \in\,]0,2[$ for every $k \geq 0$ if $K \subseteq \mathbb{R}^n$ is a linear subspace. If in addition to $\alpha = 0$, we consider $\rho_k = 1$ for all $k \geq 0$ in Algorithm 1, we obtain a variant of [21, Theorem 10] for the strongly quasiconvex minimization problem (COP).

Corollary 3.2. *Let* $K \subseteq \mathbb{R}^n$ *be a closed and convex set,* h *be a function such that assumption* $(C1)$ *holds, and* $c_k \geq c' > 0$. *Then the sequence* $\{x^k\}_k$, *generated by*

$$x^{k+1} \in \mathrm{Prox}_{c_k h}(K, x^k), \tag{3.21}$$

is a minimizing sequence of h, *i.e.,* $h(x^k) \downarrow \min_{x\in K} h(x)$.

Proof. Since $x^{k+1} \in \mathrm{Prox}_{c_k h}(K, x^k)$, by Lemma 2.2, we have

$$\begin{aligned} h(x^{k+1}) - \max\{h(y), h(x^{k+1})\} \leq\ & \frac{\lambda}{c_k}\langle x^{k+1}-x^k, y - x^{k+1}\rangle \\ & + \frac{\lambda}{2}\left(\frac{\lambda}{c_k} - \gamma + \lambda\gamma\right)\|y - x^{k+1}\|^2,\ \forall\, y \in K,\ \forall\, \lambda \in [0,1]. \end{aligned} \tag{3.22}$$

Take $y = x^k$ in (3.22). Then, for all $\lambda \in [0,1]$,

$$h(x^{k+1}) - \max\{h(x^k), h(x^{k+1})\} \leq \frac{\lambda}{2}\left(\frac{\lambda}{c_k} - \gamma + \lambda\gamma - \frac{2}{c_k}\right)\|x^k - x^{k+1}\|^2.$$

Take $\lambda = 1/2$. Then

$$\frac{1}{4}\left(\frac{1}{2c_k} - \gamma + \frac{\gamma}{2} - \frac{2}{c_k}\right) = \frac{1}{8c_k}(-3 - \gamma c_k) < 0,$$

thus

$$h(x^{k+1}) - \max\{h(x^k), h(x^{k+1})\} < 0 \iff h(x^{k+1}) < h(x^k),\ \forall\ k \in \mathbb{N}.$$

Hence, $\{h(x^k)\}_{k\in\mathbb{N}}$ is a decreasing sequence. The rest of the proof follows from Theorem 3.1. $\square$

Remark 3.4. *A similar result to Corollary 3.2 is provided in [21, Theorem 10], where* h *is assumed to be only lower semicontinuous, instead of continuous. Inspecting the proof of Corollary 3.2 reveals that the mentioned weaker hypothesis on* h *is sufficient to guarantee the assertion. However, for the convergence of*

Algorithm 1 it seems to be necessary to take h continuous (and also K a linear subspace in order to ensure that the algorithm is well-defined), these additional assumptions being the price to pay for the accelerations and flexibility provided by this method in comparison with the classical proximal point algorithm considered in [21] for solving the same optimization problem.

For the general quasiconvex case, we have the following statement.

Proposition 3.3. *Let $K \subseteq \mathbb{R}^n$ be a linear subspace, $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function such that assumption $(C2)$ holds, $0 < \rho' \leq \rho'' < 2$, $\{\rho_k\}_k \subseteq [\rho', \rho'']$, $\alpha \in [0, 1[$, $\{\alpha_k\}_k \subseteq [0, \alpha]$, $\{x^k\}_k$, $\{y^k\}_k$ and $\{z^k\}_k$ be the sequences generated by Algorithm 1. Suppose that $\Omega \neq \emptyset$. Then the following assertions hold.*

(a) *If condition (3.16) hold, then for every $x^* \in \Omega$, the limit $\lim_{k\to\infty} ||x^k - x^*||$ exists and*

$$\lim_{k\to+\infty} \|x^{k+1} - y^k\| = \lim_{k\to+\infty} \|z^k - y^k\| = 0. \tag{3.23}$$

(b) *If in addition h is bounded from below and $c_k \geq c' > 0$ for any k, then the sequence $\{h(x^k)\}_k$ is convergent.*

Proof. (a): Since $\Omega \neq \emptyset$ by assumption, we take $x^* \in \Omega$ and $\eta_k := \|x^k - x^*\|^2$ for all $k \geq 0$. Then we simply repeat the proof of Theorem 3.1(a).

(b): See the proof of [22, Lemma 3.2$(j1)$]. □

Remark 3.5. *Supposing that $\Omega \neq \emptyset$ and assumption $(B1)$ holds, and if in addition h is bounded from below, Algorithm 1 recovers all the properties from [22, Lemma 3.2 and Theorem 3.1] with the same proofs, hence even the convergence of the iterative sequence $\{x^k\}_k$ towards the minimal point of h can be achieved under additional hypotheses.*

4 Sufficient Conditions

In this section, we provide sufficient conditions for ensuring the fulfillment of assumption (3.16). To that end, we prove the following result.

Proposition 4.1. *Let $K \subseteq \mathbb{R}^n$ be a linear subspace, $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ be such that assumption $(C1)$ or $(C2)$ holds, $0 < \rho' \leq \rho'' < 2$, $\{\rho_k\}_k \subseteq [\rho', \rho'']$, $\alpha \in [0, 1[$, $\{\alpha_k\}_k \subseteq [0, \alpha]$, and $\{x^k\}_k$, $\{y^k\}_k$ and $\{z^k\}_k$ be the sequences generated by Algorithm 1. Given any $x^* \in \Omega$, we set $\eta_k := \|x^k - x^*\|^2$ for all $k \geq 0$. Then,*

$$\begin{aligned}\eta_{k+1} - \eta_k - \alpha_k(\eta_k - \eta_{k-1}) - \left(2\alpha_k^2\left(1 - \frac{1}{\rho_k}\right) + \frac{2}{\rho_k}\alpha_k\right)\|x^k - x^{k-1}\|^2 \\ \leq \frac{(2-\rho_k)}{\rho_k}(\alpha_k - 1)||x^{k+1} - x^k||^2, \ \forall\, k \in \mathbb{N}. \quad (4.1)\end{aligned}$$

As a consequence,

$$\eta_{k+1} - \eta_k - \alpha_k(\eta_k - \eta_{k-1}) - \nu_k\|x^k - x^{k-1}\|^2 \\ \leq \frac{2-\rho''}{\rho''}(\alpha_k - 1)||x^{k+1} - x^k||^2, \ \forall\ k \in \mathbb{N}, \quad (4.2)$$

where $\nu_k := 2\left(1 - \frac{1}{\rho''}\right)\alpha_k^2 + \frac{2}{\rho'}\alpha_k$ *for all* $k \geq 0$.

Proof. To show (4.1), we investigate the boundedness of $||x^{k+1} - y^k||^2$ from below in relation (3.13) of Proposition 3.2. A direct calculation using definition of y^k and the Cauchy-Schwartz inequality yield

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

From this and the well-known inequality $2pq \leq p^2 + q^2$ with $p = ||x^{k+1} - x^k||$ and $q = ||x^k - x^{k-1}||$, we conclude

$$||x^{k+1} - y^k||^2 \geq (1 - \alpha_k)||x^{k+1} - x^k||^2 + (\alpha_k^2 - \alpha_k)||x^k - x^{k-1}||^2. \quad (4.3)$$

By assumption $(C1)$ or $(C2)$, $\Omega \neq \emptyset$. Then, for every $x^* \in \Omega$, we set $\eta_k := \|x^k - x^*\|^2$ for all $k \geq 0$. Hence, by replacing (4.3) in relation (3.13), we have

$$\eta_{k+1} - \eta_k - \alpha_k(\eta_k - \eta_{k-1}) \leq \left(2\alpha_k^2\left(1 - \frac{1}{\rho_k}\right) + \frac{2}{\rho_k}\alpha_k\right)\|x^k - x^{k-1}\|^2 \\ + \frac{(2-\rho_k)}{\rho_k}(\alpha_k - 1)||x^{k+1} - x^k||^2, \ \forall\ k \in \mathbb{N},$$

which proves relation (4.1).

Finally, since $\alpha \in [0, 1[$, $\{\alpha_k\}_k \subseteq [0, \alpha]$, $\{\rho_k\}_k \subseteq [\rho', \rho'']$ and $\rho'' < 2$, from relation (4.1) we obtain (4.2). □

The following result is a variant of [1, Proposition 2.5] which was proved for the relaxed-inertial proximal point algorithm for solving monotone inclusion problems. We emphasize that this result is valid under assumption $(C1)$.

Theorem 4.1. *Let* $K \subseteq \mathbb{R}^n$ *be a linear subspace,* $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ *be a function such that assumption* $(C1)$ *holds,* $0 < \rho' \leq \rho'' < 2$, $\{\rho_k\}_k \subseteq [\rho', \rho'']$, $\alpha \in [0, 1[$, $\{\alpha_k\}_k \subseteq [0, \alpha]$ *and* $\{\alpha_k\}_k$ *is nondecreasing satisfying the following (for some* $\beta > 0$*)*

$$0 \leq \alpha_k \leq \alpha_{k+1} \leq \alpha < \beta < 1, \ \forall\ k \geq 0, \quad (4.4)$$

and

$$\rho'' = \rho''(\beta, \rho') := \frac{2\rho'(\beta^2 - \beta + 1)}{2\rho'\beta^2 + (2 - \rho')\beta + \rho'}. \quad (4.5)$$

If $\{x^k\}_k$ is the sequence generated by Algorithm 1, then

$$\sum_{k=1}^{\infty} ||x^k - x^{k-1}||^2 < +\infty. \tag{4.6}$$

As a consequence, if $c_k \geq c' > 0$ for every $k \geq 0$, then $\{x^k\}_k$ converges to $\overline{x} = \arg\min_K h$ and $\lim_{k\to+\infty} h(x^k) = \min_K h$.

Proof. Since assumption $(C1)$ holds, $\arg\min_K h \neq \emptyset$, i.e., $\Omega \neq \emptyset$ too. Then, for any $x^* \in \Omega$, we set $\eta_k := \|x^k - x^*\|^2$ for all $k \geq 0$. Define

$$\mu_k := \eta_k - \alpha_k \eta_{k-1} + \nu_k ||x^k - x^{k-1}||^2, \ \forall\ k \geq 0, \tag{4.7}$$

Since $\eta_k \geq 0$, $\alpha_k \leq \alpha_{k+1}$ for all $k \geq 0$ and identity (4.2) of Proposition 4.1, we have

$$\begin{aligned}
\mu_{k+1} - \mu_k &= \big(\eta_{k+1} - \eta_k - \alpha_k(\eta_k - \eta_{k-1}) - \nu_k \|x^k - x^{k-1}\|^2\big) + \nu_{k+1}||x^{k+1} - x^k||^2 \\
&\leq \left((\frac{2}{\rho'} - 1)\alpha_k - \frac{2}{\rho''} + 1 + \nu_{k+1} \right) ||x^{k+1} - x^k||^2 \\
&\leq \left((\frac{2}{\rho'} - 1)\alpha_{k+1} - \frac{2}{\rho''} + 1 + \nu_{k+1} \right) ||x^{k+1} - x^k||^2 \\
&= -\left[2\left(\frac{1}{\rho''} - 1\right)\alpha_{k+1}^2 - \left(\frac{2}{\rho'} + \frac{2}{\rho''} - 1\right)\alpha_{k+1} + \frac{2-\rho''}{\rho''} \right] ||x^{k+1} - x^k||^2 \\
&= -q(\alpha_{k+1})||x^{k+1} - x^k||^2,
\end{aligned} \tag{4.8}$$

where $q : \mathbb{R} \to \mathbb{R}$ is a quadratic function defined by

$$q(x) := 2\left(\frac{1}{\rho''} - 1\right)x^2 - \left(\frac{2}{\rho'} + \frac{2}{\rho''} - 1\right)x + \left(\frac{2}{\rho''} - 1\right). \tag{4.9}$$

Note that $q(\alpha_{k+1})$ admits a uniform positive lower bound. Indeed, from (4.5) and Lemma 2.5(a) with $\rho = \rho'$ and $x = \rho''$, after some arrangements we have

$$\beta = \frac{2\left(\frac{2}{\rho''} - 1\right)}{\frac{2}{\rho'} + \frac{2}{\rho''} - 1 + \sqrt{\left(\frac{2}{\rho'} + \frac{2}{\rho''} - 1\right)^2 - 8\left(\frac{1}{\rho''} - 1\right)\left(\frac{2}{\rho''} - 1\right)}}. \tag{4.10}$$

From this and Lemma 2.6 with with $q(\cdot)$ as in (4.9), $a = 2(1/\rho'' - 1)$, $b = 2/\rho' + 2/\rho'' - 1$ and $c = 2/\rho'' - 1$, we conclude that $q(\beta) = 0$ and $q(\cdot)$ is decreasing on $[0, \beta]$. Then in view of (4.4), we conclude

$$q(\alpha_{k+1}) \geq q(\alpha) > q(\beta) = 0.$$

Combining this with (4.8), we deduce

$$\mu_{k+1} - \mu_k \leq -q(\alpha_{k+1})||x^{k+1} - x^k||^2 \leq -q(\alpha)||x^{k+1} - x^k||^2 \leq 0. \tag{4.11}$$

Hence, $\mu_{k+1} \leq \mu_k$ for all $k \geq 0$, so the sequence $\{\mu_k\}_k$ is non increasing and bounded from above by $\mu_0 = (1-\alpha_0)\eta_0$. So, equation (4.7) and the monotonicity of $\{\mu_k\}_k$ yield $\eta_k - \alpha_k \eta_{k-1} \leq \mu_k \leq \mu_0$ for all $k \geq 0$. From the latter and (4.4) we recursively obtain

$$\eta_k \leq \alpha\eta_{k-1} + \mu_0 \leq \cdots \leq \alpha^k \eta_0 + \mu_0 \sum_{j=0}^{k-1} \alpha^j \leq \alpha^k \eta_0 + \frac{\mu_0}{1-\alpha}.$$

By using relation (4.11), the above inequality, and taking into account that $\mu_{k+1} \geq -\alpha\eta_k$, we have

$$\begin{aligned}
\sum_{j=0}^{k} ||x^{j+1} - x^j||^2 &\leq \frac{1}{q(\alpha)}(\mu_0 - \mu_{k+1}) \leq \frac{1}{q(\alpha)}(\mu_0 + \alpha\eta_k) \\
&\leq \frac{1}{q(\alpha)}\left(\mu_0 + \alpha\left(\alpha^k \eta_0 + \frac{\mu_0}{1-\alpha}\right)\right) \\
&= \frac{1}{q(\alpha)}\left(\frac{\mu_0}{1-\alpha} + \alpha^{k+1}\eta_0\right).
\end{aligned}$$

Taking $k \to +\infty$ in the previous inequality, we obtain

$$\sum_{k=0}^{+\infty} ||x^{k+1} - x^k||^2 \leq \frac{1}{q(\alpha)} \frac{\mu_0}{1-\alpha} < +\infty.$$

The remainder follows directly from Theorem 3.1. $\square$

Let us now particularize Theorem 4.1 to the case $\rho_k = 1$ for every $k \geq 0$, which corresponds to the absence of relaxation effects in Algorithm 1. In this case, when replacing $\rho_k = \rho' = \rho'' = 1$ into (4.5), we obtain $\beta = 1/3$, recovering the condition on the inertial parameters in [2, Proposition 2.1] in our *inertial proximal point method* for solving the strongly quasiconvex minimization problem (COP) proposed below.

Corollary 4.1. *Let $K \subseteq \mathbb{R}^n$ be a linear subspace, $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function such that assumption $(C1)$ holds. Assume that $\alpha \in [0, 1[$, and $\{\alpha_k\}_k$ is nondecreasing satisfying $0 \leq \alpha_k \leq \alpha_{k+1} \leq \alpha < 1/3$, and $\{c_k\}_{k\in\mathbb{N}}$ is a sequence of positive numbers. Then for any sequence $\{x^k\}_k$ generated by*

$$\text{(IPPA)} \qquad \begin{cases} k \leftarrow k+1 \\ y^k = x^k + \alpha_k(x^k - x^{k-1}) \\ x^{k+1} \in \operatorname{Prox}_{c_k h}(K, y^k), \end{cases} \tag{4.12}$$

we have

(a) $\displaystyle\sum_{k=1}^{+\infty} \|x^k - x^{k-1}\|^2 < +\infty$;

(b) *for* $\{\overline{x}\} = \arg\min_K h$*, the limit* $\lim_{k\to+\infty}\|x^k - \overline{x}\|$ *exists, and hence* $\{x^k\}$ *is bounded;*

(c) *if in addition* $c_k \geq c' > 0$ *for every* $k \geq 0$*, then the sequence* $\{x^k\}_k$*, generated by* (IPPA)*, converges to* $\overline{x} = \arg\min_K h$ *and* $\lim_{k\to+\infty} h(x^k) = \min_K h$*.*

Remark 4.1. *Note that if we consider* $\alpha = 0$ *in Corollary 4.1, we rediscover Corollary 3.2.*

Next we propose alternatives for the choice of the sequence $\{\rho_k\}_k$. To that end, we consider different upper bounds in Proposition 3.2 (see equation (3.13)). We resume this in the following remark.

Remark 4.2. (i) *Since* $(2-\rho_k)/\rho_k(\alpha_k - 1) \leq (2/\rho' - 1)\,\alpha_k - 2/\rho'' + 1$*, from relation* (3.13)*, by following the proof of Theorem 4.1, we obtain in* (4.9) *that* $q(\cdot)$ *is given by*

$$q(x) := 2\left(\frac{1}{\rho''} - 1\right)x^2 - \left(\frac{4}{\rho'} - 1\right)x + \left(\frac{2}{\rho''} - 1\right). \tag{4.13}$$

Hence, relation (4.10) *should be replaced by*

$$\beta = \frac{2\,(2-\rho'')}{\left(\frac{4}{\rho'} - 1\right)\rho'' + \sqrt{\left[\left(\frac{4}{\rho'} - 1\right)^2 - 8\right]\rho''^2 + 24\rho'' - 16}}. \tag{4.14}$$

Therefore, condition (4.5) (*the inverse of* (4.10) *by Lemma 2.5(b)*) *becomes*

$$\rho'' = \rho''(\beta, \rho') := \frac{2\rho'(\beta^2 + 1)}{2\rho'\beta^2 + (4-\rho')\beta + \rho'}. \tag{4.15}$$

A quadratic function similar to $q(\cdot)$*, as defined in* (4.13)*, was also considered by Alvarez in [1], where for the fulfillment of a condition like* (4.6) *it is sufficient that* $\{\alpha_k\}_k$ *is nondecreasing and* $q(\alpha) > 0$*.*

(ii) *Since* $(2-\rho_k)/\rho_k \geq 2/\rho'' - 1$ *from relation* (3.13)*, we have* (*in Proposition 4.1*) $\nu_k = 2\,(1 - 1/\rho'')\,\alpha_k^2 + 2\alpha_k/\rho''$ *for all* $k \geq 0$*. Hence, by following the proof of Theorem 4.1, we obtain in* (4.9) *that* $q(\cdot)$ *is given by*

$$q(x) := 2\left(\frac{1}{\rho''} - 1\right)x^2 - \left(\frac{4}{\rho''} - 1\right)x + \left(\frac{2}{\rho''} - 1\right). \tag{4.16}$$

Hence, relation (4.10) *is replaced by*

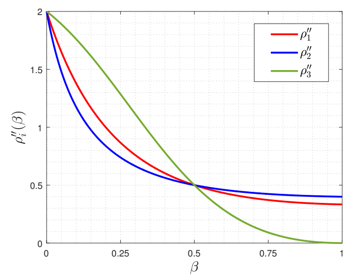
$$\beta = \frac{2(2-\rho'')}{4 - \rho'' + \sqrt{16\rho'' - 7\rho''^2}}. \tag{4.17}$$

Therefore, condition (4.5) *(the inverse of* (4.10) *by Lemma 2.5(c)) is*

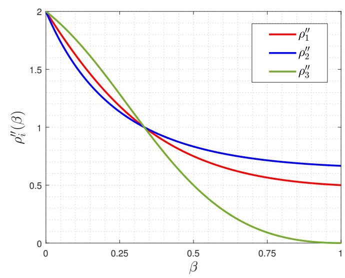
$$\rho'' = \rho''(\beta) := \frac{2(\beta - 1)^2}{2(\beta - 1)^2 + 3\beta - 1}. \tag{4.18}$$

A similar approach considered here has been addressed in [4, Theorem 3.5] (see also [3, Theorem 2] for aplications to splitting methods in the convex setting).

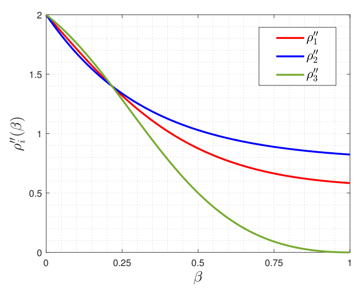
(iii) *To summarize, we have considered three different types of upper bounds of the relaxation parameters as a function of the upper bound of the inertial parameters as shown in Figure 1 below.*



(a) $\rho' = 0.6$



(b) $\rho' = 1$



(c) $\rho' = 1.4$

Figure 1: Upper bound of the relaxation parameters $\rho_i''(\beta)$ $(i = 1, 2, 3)$ given in (4.5), (4.15) and (4.18) as functions of inertial step upper bound $\beta > 0$.

Finally we provide the following result for the general quasiconvex case.

Proposition 4.2. *Let $K \subseteq \mathbb{R}^n$ be a linear subspace, $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function such that assumption $(C2)$ holds, and $\{x^k\}_k$ and $\{z^k\}_k$ be sequences generated by Algorithm 1. Suppose that $\alpha \in [0, 1[$, $0 < \rho' \leq \rho'' < 2$ and $\{\alpha_k\}_k$ is nondecreasing satisfying the following (for some $\beta > 0$):*

$$0 \leq \alpha_k \leq \alpha_{k+1} \leq \alpha < \beta < 1, \ \forall \ k \geq 0, \tag{4.19}$$

and

$$\rho'' = \rho''(\beta, \rho') := \frac{2\rho'(\beta^2 - \beta + 1)}{2\rho'\beta^2 + (2 - \rho')\beta + \rho'}. \tag{4.20}$$

Then,

$$\sum_{k=1}^{\infty} ||x^k - x^{k-1}||^2 < +\infty. \tag{4.21}$$

If in addition $c_k \geq c' > 0$ for every $k \geq 0$ and h is bounded from below, then the sequence $\{h(x^k)\}_k$, generated by Algorithm 1, is convergent.

Proof. Since assumption $(C2)$ holds, for any $x^* \in \Omega$, we set $\eta_k := \|x^k - x^*\|^2$ for all $k \geq 0$. Then we repeat the proof of Theorem 4.1 invoking at the end Proposition 3.3 instead of Theorem 3.1. $\square$

Remark 4.3. *The above proposition is complementary to [22, Theorem 3.2] in the finite dimensional setting and under more flexible assumptions (without both differentiability and Lipschitz continuity (of the gradient) of the objective function).*

5 Numerical Experiments

In the following we present some computational results obtained in MATLAB 2019b-Win64 on a Lenovo Yoga 260 Laptop with Windows 10 and an Intel Core i7 6500U CPU with 2.59 GHz and 16GB RAM by implementing Algorithm 1 and, for comparison, [21, Algorithm 1]. The example treated below does not stem from a concrete application and is merely meant to present a situation where the relaxed-inertial proximal point algorithm has a superior performance to the one of its "simple" proximal point counterpart. As stopping criterion of the proposed algorithm we considered the situation when the norm of the differences between the generated sequences $\{y^k\}_k$ and $\{z^k\}_k$ is not larger than an a priori given error $\varepsilon > 0$, i. e., $\|z^k - y^k\| < \varepsilon$. The strongly quasiconvex function we minimized is introduced below.

Example 5.1. *Let $q \in \mathbb{N}$ and $h_1, h_2 : \mathbb{R}^n \to \mathbb{R}$, $h_1(x) = \sqrt{\|x\|}$ and $h_2(x) = \|x\|^2 - q$. Both h_1 and h_2 are continuous functions. It is known that h_2 is strongly convex, hence also strongly quasiconvex, while according to [21, Theorem 16] the function h_1 is strongly quasiconvex on any convex and bounded set $K \subseteq \mathbb{R}^n$.*

Then the function $h : \mathbb{R}^n \to \mathbb{R}$ defined by $h(x) := \max\{h_1(x), h_2(x)\}$ is continuous and strongly quasiconvex (as the maximum of two strongly quasiconvex functions), without being convex. Note also that $\arg\min_{\mathbb{R}^n} h = \{0\}$.

Consider the optimization problem (COP) with the objective function h introduced in Example 5.1 for various values of $q \in \mathbb{N}$, and $K = \mathbb{R}^n$. The parameter sequences $\{\alpha_k\}_k$, $\{\rho_k\}_k$ and $\{c_k\}_k$ were taken in order to comply with the theoretical results.

We implemented both the relaxed-inertial proximal point method (Algorithm 1) and the basic proximal point one [21, Algorithm 1] for the same constellations of initial values and parameters. In each case the proximal step (that occurs in both methods) was computed by employing the MATLAB function `fmincon` because, as it is unfortunately often the case when dealing with nonconvex functions, a closed form of the proximity operator of h (even when $K = \mathbb{R}^n$) is not available yet. Possible ways for determining it could be by direct calculation or by adapting the method considered in [11] for providing the one of the root function.

In order to verify the convergence of Algorithm 1 and to compare its performance with the one of [21, Algorithm 1] we considered several constellations as follows. When $n = 3$ the differences between the best performance of Algorithm 1 and [21, Algorithm 1] were mostly minimal, the first one turning for instance to deliver slightly faster an approximate optimal solution to (COP) for $\alpha < \beta \approx 0.065$ (in order to comply with (4.4) and (4.10), and one could also use the upper bounds given in Remark 4.2), $\{\alpha_k\}_k$ constant, $\rho' = 0.8$, $\rho'' = 1.7$ and $\rho_k = (1 - 1/k)\rho' + (1/k)\rho''$, $k \geq 0$.

Increasing the dimension to $n = 5$, considering $\varepsilon = 0.0000001$, $q = 133$, $x^0 = [7, -8, 5, 2, 55]^\top$, $\rho' = 0.9$, $\rho'' = 1.5$, $\alpha = 0.125$ (in order to comply with (4.4) and (4.10)), as $\beta \approx 0.126$), $c_1 = 1$, $c_{k+1} = 100/k^2 + c_k$, $\alpha_{k+1} = \alpha_k + 1/(900(k+1)^2)$ and $\rho_k = (1 - 1/k)\rho' + (1/k)\rho''$, $k \geq 0$, Algorithm 1 stopped when $k = 11$, after 0.9306 seconds, by delivering an approximate optimal solution to (COP), while [21, Algorithm 1] required 43 iterations and 0.9885 seconds for a similar output.

In a similar constellation for $n = 50$ and x^0 randomly chosen, Algorithm 1 delivered an approximate optimal solution to (COP) after 13 iterations and 1.1360 seconds, while the basic proximal point algorithm stopped when $k = 46$, after 1.5866 seconds.

Increasing the dimension to $n = 500$, in a similar constellation, when $\varepsilon = 0.0001$ Algorithm 1 stopped after 23 iterations and 4.9289 seconds, while [21, Algorithm 1] needed 43 iterative steps and 9.4276 seconds. Improving the accuracy of solving (COP) to $\varepsilon = 0.00001$, our algorithm delivered an approximate optimal solution at $k = 63$ after 15.5450 seconds, in contrast to $k = 93$ and a time of 24.7519 required by the basic proximal point algorithm for the same purpose. The improvements brought by the newly proposed algorithm are best noticed when the accuracy increases to $\varepsilon = 0.000001$, as it required only 68 iterations and 16.3472 seconds for solving the problem, while the classical proximal point method stopped only when $k = 1105$ after 294.0988 seconds.

Shrinking the bounded set where h is not convex by taking $q = 25$, when $n = 500$ and $\varepsilon = 0.00001$ Algorithm 1 provided the sought approximate optimal solution to (COP) in 14 iterative steps after 7.8103 seconds, while [21, Algorithm 1] iterated 59 times during 14.6288 seconds to the same end. Tables 1 and 2 summarize the performance of Algorithm 1 compared with [21, Algorithm 1] as was described above.

Table 1: Running time (in seconds) and number of iterations performed by Algorithm 1 and [21, Algorithm 1] to reach the stopping criterion $\|z^k - y^k\| < \varepsilon$, for $n \in \{5, 50, 500\}$, where $x^0 = [7, -8, 5, 2, 55]^\top$ for $n = 5$ and randomly chosen for the other two cases.

	$n = 5,\ \varepsilon = 10^{-6}$		$n = 50,\ \varepsilon = 10^{-6}$		$n = 500,\ \varepsilon = 10^{-3}$	
$q = 133$	Alg. 1	[21, Alg. 1]	Alg. 1	[21, Alg. 1]	Alg. 1	[21, Alg. 1]
Time (s)	0.9306	0.9885	1.1360	9.4276	4.9289	9.4276
Iterations	11	43	13	46	23	43

Table 2: (Table 1 continued.) Here, we show the comparison of the performance between Alg. 1 and [21, Algorithm 1] while reaching the stopping criterion $\|z^k - y^k\| < \varepsilon$ for $n = 500$, where x^0 is randomly chosen.

	$\varepsilon = 10^{-4}$		$\varepsilon = 10^{-5}$		$\varepsilon = 10^{-4} (q = 25)$	
$n = 500$	Alg. 1	[21, Alg. 1]	Alg. 1	[21, Alg. 1]	Alg. 1	[21, Alg. 1]
Time (s)	15.5450	24.7519	16.3472	294.0988	7.8103	14.6288
Iterations	63	93	68	1105	14	59

Given the flexibility provided by the (sequences of) parameters involved in Algorithm 1, better results (in terms of both number of iterations and time) than the one presented above are safe to be expected by fine tuning it, the role of our experiments being merely to show that this method is indeed faster and cheaper (with respect to the number of iterations) than the basic proximal point method. Note also that taking $\{\alpha_k\}_k$ as indicated in [5] for the convex case did not improve the velocity of Algorithm 1 in comparison to the outcomes listed above. Last but not least, since the optimal solution to (COP) is actually the origin of the considered Euclidean space, we repeated some of these numerical experiments by taking as an alternate stopping criterion the situation when the norm of x^k was not larger than ε, i.e., $\|x^k\| < \varepsilon$. The results were only slightly different to the ones presented above, exhibiting the same tendency, that is, Algorithm 1 worked cheaper and faster than [21, Algorithm 1].

6 Conclusions

We have shown that the relaxed-inertial proximal point algorithm proposed by Attouch and Cabot for solving convex optimization problems remains convergent when employed for minimizing strongly quasiconvex functions whose variables lie in finitely dimensional linear subspaces. Numerical experiments show that this method solves such a problem faster and cheaper than the basic proximal point algorithm. We also discuss some special cases of the considered algorithm, where some connections to the existing literature are noted, and provide some preliminary results on the convergence of the method in the case of a quasiconvex objective function, that is planned to be approached in a subsequent work (in order to complement the existing literature on inertial type proximal point algorithms for minimizing quasiconvex functions). Another direction worth exploring concerns splitting proximal point methods for minimizing structured (strongly) quasiconvex functions.

7 Declarations

7.1 Funding

This research was partially supported by ANID–Chile under project Fondecyt Regular 1220379 (Lara), and by a CIAS Senior Research Fellow Grant of the Corvinus Institute for Advanced Studies (Grad).

7.2 Conflicts of interest/Competing interests

There are no conflicts of interest or competing interests related to this manuscript.

7.3 Availability of data and material

The datasets generated during and/or analysed during the current study are available from S.-M. Grad on reasonable request.

7.4 Code availability

Not applicable.

7.5 Authors' contributions

The three authors contributed equally to the study conception and design.

References

[1] F. Alvarez, Weak convergence of a relaxed and inertial hybrid projection-proximal point algorithm for maximal monotone operators in Hilbert space, *SIAM J. Optim.*, **14**, 773–782, (2003).

[2] F. Alvarez and H. Attouch, An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping, *Set-Valued Var. Anal.*, **9(1-2)**, 3–11, (2001).

[3] M.M. Alves, J. Eckstein, M. Geremia and J.G. Melo, Relative-error inertial-relaxed inexact versions of Douglas-Rachford and ADMM splitting algorithms, *Comput. Optim. Appl.*, **75**, 389–422, (2020).

[4] M.M. Alves and R.T. Marcavillaca, On inexact relative-error hybrid proximal extragradient, forward-backward and Tsengs modified forward-backward methods with inertial effects, *Set-Valued Var. Anal.*, **28**, 301–325, (2020).

[5] H. Attouch and A. Cabot, Convergence rate of a relaxed inertial proximal algorithm for convex minimization, *O*ptimization, **6**9, 1281–1312, (2020).

[6] M. Avriel, W. E. Diewert, S. Schaible and I. Zang, "Generalized Concavity". SIAM, Classics in Applied Mathematics, Philadelphia, (2010).

[7] N.S. Aybat, A. Fallah, M. Gürbüzbalaban and A. Ozdaglar, Robust accelerated gradient methods for smooth strongly convex functions, *SIAM J. Optim.*, **30**, 717–751, (2020).

[8] H.H. Bauschke and P.L. Combettes, "Convex Analysis and Monotone Operators Theory in Hilbert Spaces". *CMS Books in Mathematics.* Springer-Verlag, second edition, (2017).

[9] H. Bonnel, A.N. Iusem and B.F. Svaiter, Proximal methods in vector optimization, *SIAM J. Optim.*, **15**, 953–970, (2005).

[10] R.I. Boţ and E.R. Csetnek, Proximal-gradient algorithms for fractional programming, *Optimization*, **66**, 1383–1396, (2017).

[11] K. Bredies and D. Lorenz, Iterated hard shrinkage for minimization problems with sparsity constraints, *SIAM J. Sci Comput*, **30**, 657–683, (2008).

[12] A.S. Brito, J.X. Da Cruz Neto, J.O. Lopes and P.R. Oliveira, Interior proximal algorithm for quasiconvex programming problems and variational inequalities with linear constraints, *J. Optim. Theory Appl.*, **154**, 217–234, (2012).

[13] A. Cambini and L. Martein. "Generalized Convexity and Optimization". Springer-Verlag, Berlin-Heidelberg, (2009).

[14] W. Ginsberg, Concavity and quasiconcavity in economics, *J. Econ. Theory*, **6**, 596–605, (1973).

[15] X. Goudou and J. Munier, The gradient and heavy ball with friction dynamical systems: the quasiconvex case, *Math. Program. Ser. B*, **116**, 173–191, (2009).

[16] S.-M. Grad and F. Lara, An extension of the proximal point algorithm beyond convexity, *J. Global Optim.*, **82**, 313–329, (2022).

[17] S.-M. Grad and F. Lara, Solving mixed variational inequalities beyond convexity , *J. Optim. Theory Appl.*, **190**, 565–580, (2021).

[18] E. Hazan and S. Kale, Beyond the regret minimization barrier: optimal algorithms for stochastic strongly-convex optimization, *J. Mach. Learn. Res.*, **15**, 2489–2512, (2014).

[19] M. Jovanović, A note on strongly convex and quasiconvex functions, *Math. Notes*, **60**, 584–585, (1996).

[20] A. Kaplan and R. Tichatschke, Proximal point methods and nonconvex optimization, *J. Global Optim.*, **13**, 389–406, (1998).

[21] F. Lara, On strongly quasiconvex functions: existence results and proximal point algorithms, *J. Optim. Theory Appl.*, **192**, 891–911, (2022).

[22] P.E. Maingé, Asymptotic convergence of an inertial proximal method for unconstrained quasiconvex minimization, *J. Global Optim.*, **45**, 631–644, (2009).

[23] B. Martinet, Regularisation d'inequations variationelles par approximations successives, *Rev. Francaise Inf. Rech. Oper.*, 154–159, (1970).

[24] R.T. Rockafellar, Monotone operators and proximal point algorithms, *SIAM J. Control Optim.*, **14**, 877–898, (1976).

[25] J.P. Vial, Strong and weak convexity of sets and functions, *Math. Oper. Res.*, **8**, 231–259, (1983).

[26] K. Zhang, Quasi-convex functions on subspaces and boundaries of quasiconvex sets, *Proc. R. Soc. Edinb., Sect. A, Math.*, **134**, 783–799, (2004).