

RESEARCH

Open Access



# Inertial proximal alternating minimization for nonconvex and nonsmooth problems

Yaxuan Zhang\*  and Songnian He

\*Correspondence:  
bunnyxuan@tju.edu.cn  
College of Science, Civil Aviation  
University of China, Tianjin, 300300,  
China

## Abstract

In this paper, we study the minimization problem of the type  $L(x, y) = f(x) + R(x, y) + g(y)$ , where  $f$  and  $g$  are both nonconvex nonsmooth functions, and  $R$  is a smooth function we can choose. We present a proximal alternating minimization algorithm with inertial effect. We obtain the convergence by constructing a key function  $H$  that guarantees a sufficient decrease property of the iterates. In fact, we prove that if  $H$  satisfies the Kurdyka-Lojasiewicz inequality, then every bounded sequence generated by the algorithm converges strongly to a critical point of  $L$ .

**Keywords:** nonconvex nonsmooth optimization; proximal alternating minimization; inertial; Kurdyka-Lojasiewicz inequality; convergence

## 1 Introduction

Nonconvex and nonsmooth optimization problems are extremely useful in many applied sciences, including statistics, machine learning, regression, classification, and so on. One of the most practical and classical optimization problems is of the form

$$\min_{x \in \mathbb{R}^n} [f(x) + g(x)]. \quad (1)$$

In this paper, we study the problem in the nonconvex and nonsmooth setting, where  $f, g : \mathbb{R}^n \rightarrow (-\infty, \infty]$  are proper lower semicontinuous functions. We aim at finding the critical points of

$$L(x, y) = f(x) + R(x, y) + g(y) \quad (2)$$

(with  $R$  being smooth) and possibly solving the corresponding minimization problem (1). This can be seen by setting

$$L(x, y) = f(x) + \frac{\rho}{2} \|x - y\|^2 + g(y),$$

where  $\rho > 0$  is a relaxation parameter.

For problem (2), we introduce a proximal alternating minimization algorithm with inertial effect and investigate the convergence of the generated iterates. Inertial proximal methods go back to [1, 2], where it has been noticed that the discretization of a differential

system of second order in time gives rise to a generalization of the classical proximal-point algorithm. The main feature of the inertial proximal algorithm is that the next iterate is defined by using the last two iterates. Recently, there has been an increasing interest in algorithms with inertial effect; see [3–12].

Generally, we consider the problem

$$\min L(x, y) = \min \{f(x) + R(x, y) + g(y)\}$$

with  $x \in \mathbb{R}^n$  and  $y \in \mathbb{R}^m$ .

In [13], the authors proposed the alternating minimization algorithm

$$\begin{cases} x_{k+1} \in \operatorname{argmin}\{L(u, y_k) + \frac{1}{2\lambda_k} \|u - x_k\|^2 : u \in \mathbb{R}^n\}, \\ y_{k+1} \in \operatorname{argmin}\{L(x_{k+1}, v) + \frac{1}{2\mu_k} \|v - y_k\|^2 : v \in \mathbb{R}^m\}, \end{cases}$$

which can be viewed as a proximal regularization of a two-block Gauss-Seidel method for minimizing  $L$ ,

$$\begin{cases} x_{k+1} \in \operatorname{argmin}\{L(u, y_k) : u \in \mathbb{R}^n\}, \\ y_{k+1} \in \operatorname{argmin}\{L(x_{k+1}, v) : v \in \mathbb{R}^m\}. \end{cases}$$

Inspired by [13], we propose the algorithm

$$x_{k+1} \in \operatorname{argmin}_{x \in \mathbb{R}^n} \left\{ L(x, y_k) + \frac{1}{2\lambda_k} \|x - x_k\|^2 + \alpha_k \langle x, x_k - x_{k-1} \rangle \right\}, \tag{3}$$

$$y_{k+1} \in \operatorname{argmin}_{y \in \mathbb{R}^m} \left\{ L(x_{k+1}, y) + \frac{1}{2\mu_k} \|y - y_k\|^2 + \beta_k \langle y, y_k - y_{k-1} \rangle \right\}. \tag{4}$$

We need the following assumptions on the functions and parameters.

- (H1)  $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$  and  $g : \mathbb{R}^m \rightarrow (-\infty, \infty]$  are proper lower semicontinuous functions;
- (H2)  $R : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$  is a continuously differentiable function;
- (H3)  $\nabla R$  is Lipschitz continuous on bounded subsets of  $\mathbb{R}^n \times \mathbb{R}^m$ ;
- (H4)  $\inf L > -\infty$ ;
- (H5)  $0 < \mu_- \leq \mu_k \leq \mu_+, 0 < \lambda_- \leq \lambda_k \leq \lambda_+, 0 \leq \alpha_k \leq \alpha, 0 \leq \beta_k \leq \beta$ ;
- (H6)  $\sigma > \max\{\alpha, \beta\} \cdot \max\{\lambda_+, \mu_+\} \cdot (\sigma^2 + 1)$ .

To prove the convergence of the algorithm under these assumptions, we construct a key function  $H$ , which is defined as in (11). Based on  $H$ , we can obtain a sufficient decrease property of the iterates, the existence of a subgradient lower bound for the iterate gap, and some important analytic features of the objective function. Finally, we can prove that every bounded sequence generated by the algorithm converges to a critical point of  $L$  if  $H$  satisfies the Kurdyka-Lojasiewicz inequality.

The rest of the paper is arranged as follows. In Section 2, we recall some elementary notions and facts of nonsmooth nonconvex analysis. In Section 3, we present a detailed proof of the convergence of the algorithm. In Section 4, we give a brief conclusion.

## 2 Preliminaries

In this section, we recall some definitions and results. Let  $\mathbb{N}$  be the set of nonnegative integers. For  $m \geq 1$ , the Euclidean scalar product and induced norm on  $\mathbb{R}^m$  are denoted by  $\langle \cdot, \cdot \rangle$  and  $\| \cdot \|$ , respectively.

The *domain* of a function  $f : \mathbb{R}^m \rightarrow (-\infty, \infty]$  is defined by  $\text{dom} f = \{x \in \mathbb{R}^m : f(x) < \infty\}$ . We say that  $f$  is *proper* if  $\text{dom} f \neq \emptyset$ . For the following generalized subdifferential notions and their basic properties we refer to [14] and [15]. Let  $f : \mathbb{R}^m \rightarrow (-\infty, \infty]$  be a proper lower semicontinuous function. For  $x \in \text{dom} f$ , we consider the *Frechet (viscosity) subdifferential* of  $f$  at  $x$  defined by the set

$$\hat{\partial}f(x) = \left\{ v \in \mathbb{R}^m : \liminf_{y \rightarrow x} \frac{f(y) - f(x) - \langle v, y - x \rangle}{\|y - x\|} \geq 0 \right\}. \tag{5}$$

For  $x \notin \text{dom} f$ , we set  $\hat{\partial}f(x) := \emptyset$ . The *limiting (Mordukhovich) subdifferential* of  $f$  at  $x \in \text{dom} f$  is defined by

$$\partial f(x) = \left\{ v \in \mathbb{R}^m : \exists x_n \rightarrow x, f(x_n) \rightarrow f(x) \text{ and } \exists v_n \in \hat{\partial}f(x_n), v_n \rightarrow v \text{ as } n \rightarrow \infty \right\}, \tag{6}$$

whereas for  $x \notin \text{dom} f$ , we take  $\partial f(x) := \emptyset$ .

It is known that both notions of subdifferentials coincide with the *convex subdifferential* if  $f$  is convex, that is,  $\hat{\partial}f(x) = \partial f(x) = \{v \in \mathbb{R}^m : f(y) \geq f(x) + \langle v, y - x \rangle, \forall y \in \mathbb{R}^m\}$ . Notice that if  $f$  is continuously differentiable around  $x \in \mathbb{R}^m$ , then we have  $\partial f(x) = \{\nabla f(x)\}$ . Generally, the inclusion  $\hat{\partial}f(x) \subset \partial f(x)$  holds for each  $x \in \mathbb{R}^m$ .

The *Fermat rule* reads in this nonsmooth setting as follows: if  $x \in \mathbb{R}^m$  is a local minimizer of  $f$ , then

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

Denote by

$$\text{crit} f = \{x \in \mathbb{R}^m : 0 \in \partial f(x)\}$$

the set of (*limiting*) *critical points* of  $f$ . Let us mention also the following subdifferential rule: if  $f : \mathbb{R}^m \rightarrow (-\infty, \infty]$  is proper lower semicontinuous and  $g : \mathbb{R}^m \rightarrow \mathbb{R}$  is a continuously differentiable function, then

$$\partial(f + g)(x) = \partial f(x) + \nabla g(x)$$

for all  $x \in \mathbb{R}^m$ .

We also denote  $\text{dist}(x, \Omega) = \inf_{y \in \Omega} \|x - y\|$  for  $x \in \mathbb{R}^m$  and  $\Omega \subset \mathbb{R}^m$ .

Now let us recall the *Kurdyka-Lojasiexicz property*, which plays an important role in the proof of the convergence of our algorithm.

**Definition 2.1** (Kurdyka-Lojasiexicz property; see [13, 16]) Let  $f : \mathbb{R}^m \rightarrow (-\infty, \infty]$  be a proper lower semicontinuous function. We say that  $f$  satisfies the *Kurdyka-Lojasiexicz (KL) property* at  $\bar{x} \in \text{dom} \partial f = \{x \in \mathbb{R}^m : \partial f(x) \neq \emptyset\}$  if there exist  $\eta > 0$ , a neighborhood  $U$  of  $\bar{x}$ , and a continuous concave function  $\varphi : [0, \eta) \rightarrow [0, \infty)$  such that

- (i)  $\varphi(0) = 0$ ;
- (ii)  $\varphi$  is continuously differentiable on  $(0, \eta)$  and continuous at 0;
- (iii)  $\varphi'(s) > 0$  for all  $s \in (0, \eta)$ ;
- (iv) for all  $x \in U \cap \{x \in \mathbb{R}^m : f(\bar{x}) < f(x) < f(\bar{x}) + \eta\}$ , we have the *KL inequality*:

$$\varphi'(f(x) - f(\bar{x})) \operatorname{dist}(0, \partial f(x)) \geq 1. \tag{7}$$

If  $f$  satisfies the KL property at each point in  $\operatorname{dom} \partial f$ , then we call  $f$  a *KL function*.

It is worth mentioning that many functions in applied science are the KL functions (see [16]). In fact, semialgebraic functions, real subanalytic functions, semiconvex functions, and uniformly convex functions are all KL functions.

The following result (see [16], Lemma 6) is crucial to our convergence analysis.

**Lemma 2.1** *Let  $\Omega \subset \mathbb{R}^m$  be a compact set, and let  $f : \mathbb{R}^m \rightarrow (-\infty, \infty]$  be a proper lower semicontinuous function. Assume that  $f$  is constant on  $\Omega$  and  $f$  satisfies the KL property at each point of  $\Omega$ . Then there exist  $\epsilon > 0, \eta > 0$ , and a continuous concave function  $\varphi$  such that*

- (i)  $\varphi(0) = 0$ ;
- (ii)  $\varphi$  is continuously differentiable on  $(0, \eta)$  and continuous at 0;
- (iii)  $\varphi'(s) > 0$  for all  $s \in (0, \eta)$ ;
- (iv) for all  $\bar{x} \in \Omega$  and all  $x \in \{x \in \mathbb{R}^m : \operatorname{dist}(x, \Omega) < \epsilon\} \cap \{x \in \mathbb{R}^m : f(\bar{x}) < f(x) < f(\bar{x}) + \eta\}$ , we have the *KL inequality*:

$$\varphi'(f(x) - f(\bar{x})) \operatorname{dist}(0, \partial f(x)) \geq 1. \tag{8}$$

We need the following two lemmas. The first one was often used in the context of Fejer monotonicity techniques for proving convergence results of classical algorithms for convex optimization problems or, more generally, for monotone inclusion problems (see [17]). The second one is easy to verify (see [12]).

**Lemma 2.2** *Let  $\{a_n\}_{n \in \mathbb{N}}$  and  $\{b_n\}_{n \in \mathbb{N}}$  be real sequences such that  $b_n \geq 0$  for all  $n \in \mathbb{N}$ ,  $\{a_n\}_{n \in \mathbb{N}}$  is bounded below, and  $a_{n+1} + b_n \leq a_n$  for all  $n \in \mathbb{N}$ . Then  $\{a_n\}_{n \in \mathbb{N}}$  is a monotonically decreasing and convergent sequence, and  $\sum_{n \in \mathbb{N}} b_n < +\infty$ .*

**Lemma 2.3** *Let  $\{a_n\}_{n \in \mathbb{N}}$  and  $\{b_n\}_{n \in \mathbb{N}}$  be nonnegative real sequences such that  $\sum_{n \in \mathbb{N}} b_n < \infty$  and  $a_{n+1} \leq a \cdot a_n + b \cdot a_{n-1} + b_n$  for all  $n \geq 1$ , where  $a \in \mathbb{R}, b \geq 0$ , and  $a + b < 1$ . Then  $\sum_{n \in \mathbb{N}} a_n < \infty$ .*

### 3 The convergence of the algorithm

In this section, we prove the convergence of our algorithm. Motivated by [11] and [13], we divide the proof into three main steps, which are listed in the following three subsections, respectively.

We always use  $\{(x_k, y_k)\}$  as the sequence generated by (3)-(4).

### 3.1 A sufficient decrease property of the iterates

In this subsection, we construct the key function  $H$  and prove that the iterates have a sufficient decrease property.

**Lemma 3.1** *Under assumptions (H1)-(H6), the sequence  $\{(x_k, y_k)\}$  is well defined, and  $\{L(x_k, y_k)\}$  is decreasing. More precisely, there exist two positive constants  $m_2 > m_1 > 0$  such that*

$$\begin{aligned} L(x_{k+1}, y_{k+1}) + m_2(\|x_{k+1} - x_k\|^2 + \|y_{k+1} - y_k\|^2) \\ \leq L(x_k, y_k) + m_1(\|x_k - x_{k-1}\|^2 + \|y_k - y_{k-1}\|^2). \end{aligned} \tag{9}$$

*Proof* Assumption (H4) indicates that, for any  $r_1, r_2 > 0$  and  $(\bar{x}, \bar{y}), (\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ , the functions

$$x \rightarrow L(x, \bar{y}) + \frac{1}{2r_1} \|x - \bar{x}\|^2 + r_2 \langle x, \bar{x} - \hat{x} \rangle$$

and

$$y \rightarrow L(\bar{x}, y) + \frac{1}{2r_1} \|y - \bar{y}\|^2 + r_2 \langle y, \bar{y} - \hat{y} \rangle$$

are coercive. Thus the sequence  $\{(x_k, y_k)\}$  is well defined.

Now we prove (9). Using the definition of  $x_{k+1}$  and  $y_{k+1}$  in (3) and (4), we have

$$\begin{aligned} L(x_{k+1}, y_{k+1}) + \frac{1}{2\mu_k} \|y_{k+1} - y_k\|^2 + \beta_k \langle y_{k+1}, y_k - y_{k-1} \rangle \\ \leq L(x_{k+1}, y_k) + \beta_k \langle y_k, y_k - y_{k-1} \rangle \\ = L(x_{k+1}, y_k) + \frac{1}{2\lambda_k} \|x_{k+1} - x_k\|^2 + \alpha_k \langle x_{k+1}, x_k - x_{k-1} \rangle + \beta_k \langle y_k, y_k - y_{k-1} \rangle \\ - \frac{1}{2\lambda_k} \|x_{k+1} - x_k\|^2 - \alpha_k \langle x_{k+1}, x_k - x_{k-1} \rangle \\ \leq L(x_k, y_k) + \alpha_k \langle x_k, x_k - x_{k-1} \rangle + \beta_k \langle y_k, y_k - y_{k-1} \rangle \\ - \frac{1}{2\lambda_k} \|x_{k+1} - x_k\|^2 - \alpha_k \langle x_{k+1}, x_k - x_{k-1} \rangle. \end{aligned}$$

This leads to

$$\begin{aligned} L(x_{k+1}, y_{k+1}) + \frac{1}{2\lambda_k} \|x_{k+1} - x_k\|^2 + \frac{1}{2\mu_k} \|y_{k+1} - y_k\|^2 \\ \leq L(x_k, y_k) + \alpha_k \langle x_k - x_{k+1}, x_k - x_{k-1} \rangle + \beta_k \langle y_k - y_{k+1}, y_k - y_{k-1} \rangle \\ \leq L(x_k, y_k) + \alpha_k \|x_k - x_{k+1}\| \|x_k - x_{k-1}\| + \beta_k \|y_k - y_{k+1}\| \|y_k - y_{k-1}\| \\ \leq L(x_k, y_k) + \frac{\sigma\alpha_k}{2} \|x_k - x_{k+1}\|^2 + \frac{\alpha_k}{2\sigma} \|x_k - x_{k-1}\|^2 + \frac{\sigma\beta_k}{2} \|y_k - y_{k+1}\|^2 \\ + \frac{\beta_k}{2\sigma} \|y_k - y_{k-1}\|^2 \end{aligned}$$

for any  $\sigma > 0$ . Thus it yields

$$\begin{aligned} L(x_{k+1}, y_{k+1}) &+ \frac{1}{2} \left( \frac{1}{\lambda_k} - \sigma \alpha_k \right) \|x_{k+1} - x_k\|^2 + \frac{1}{2} \left( \frac{1}{\mu_k} - \sigma \beta_k \right) \|y_{k+1} - y_k\|^2 \\ &\leq L(x_k, y_k) + \frac{1}{2} \frac{\alpha_k}{\sigma} \|x_k - x_{k-1}\|^2 + \frac{1}{2} \frac{\beta_k}{\sigma} \|y_k - y_{k-1}\|^2. \end{aligned}$$

Clearly assumption (H5) implies that

$$\frac{\alpha_k}{\sigma} \leq \frac{\alpha}{\sigma} := 2M_1, \quad \frac{1}{\lambda_k} - \sigma \alpha_k \geq \frac{1}{\lambda_+} - \sigma \alpha := 2M_3$$

and

$$\frac{\beta_k}{\sigma} \leq \frac{\beta}{\sigma} := 2M_2, \quad \frac{1}{\mu_k} - \sigma \beta_k \geq \frac{1}{\mu_+} - \sigma \beta := 2M_4.$$

Thus

$$\begin{aligned} L(x_{k+1}, y_{k+1}) &+ M_3 \|x_{k+1} - x_k\|^2 + M_4 \|y_{k+1} - y_k\|^2 \\ &\leq L(x_k, y_k) + M_1 \|x_k - x_{k-1}\|^2 + M_2 \|y_k - y_{k-1}\|^2. \end{aligned}$$

Set  $m_1 = \max\{M_1, M_2\}$ ,  $m_2 = \min\{M_3, M_4\}$ . Then

$$\begin{aligned} L(x_{k+1}, y_{k+1}) &+ m_2 (\|x_{k+1} - x_k\|^2 + \|y_{k+1} - y_k\|^2) \\ &\leq L(x_k, y_k) + m_1 (\|x_k - x_{k-1}\|^2 + \|y_k - y_{k-1}\|^2). \end{aligned} \tag{10}$$

An elementary verification shows that  $m_2 > m_1 > 0$  under assumption (H6). □

**Remark 3.1** Based on Lemma 3.1, we can define the new function

$$H(z, w) = L(x, y) + m_1 \|z - w\|^2, \tag{11}$$

where  $z = (x, y)$ ,  $w = (u, v)$ , and  $\|z - w\|^2 = \|x - u\|^2 + \|y - v\|^2$ . Set  $z_k = (x_k, y_k)$ . Then Lemma 3.1 implies that the sequence  $\{H(z_{k+1}, z_k)\}$  is decreasing. The decrease property of the iterates  $\{x_k, y_k\}$  showed in Lemma 3.2 is of vital importance for the convergence proof. Thus, we call  $H(x, y)$  the *key function*.

More precisely, we have the following lemma.

**Lemma 3.2** *Let  $H(z, w)$  be defined as in (11). Then under assumptions (H1)-(H6), we have*

$$H(z_{k+1}, z_k) + m \|z_{k+1} - z_k\|^2 \leq H(z_k, z_{k-1}), \tag{12}$$

where  $z_k = (x_k, y_k)$ , that is, the sequence  $\{H(z_{k+1}, z_k)\}$  is decreasing.

*Proof* Set  $m := m_2 - m_1 > 0$ . Then the result follows directly from (9) or (10). □

### 3.2 Norm estimate of the subdifferential of $H$

In this subsection, we prove that there exists a subgradient lower bound for the iterate gap. First, we estimate the norm of the subdifferential of  $L$ .

**Lemma 3.3** *Define*

$$p_{k+1} := \nabla_x R(x_{k+1}, y_{k+1}) - \nabla_x R(x_{k+1}, y_k) - \frac{1}{\lambda_k}(x_{k+1} - x_k) - \alpha_k(x_k - x_{k-1}), \tag{13}$$

$$q_{k+1} := -\frac{1}{\mu_k}(y_{k+1} - y_k) + \beta_k(y_k - y_{k-1}). \tag{14}$$

Then, under assumptions (H1)-(H6),  $(p_{k+1}, q_{k+1}) \in \partial L(x_{k+1}, y_{k+1})$ . Moreover, if  $\{(x_k, y_k)\}$  is bounded, then there exists a positive constant  $C_1 > 0$  such that

$$\begin{aligned} & \| (p_{k+1}, q_{k+1}) \| \\ & \leq C_1 (\|x_{k+1} - x_k\| + \|x_k - x_{k-1}\| + \|y_{k+1} - y_k\| + \|y_k - y_{k-1}\|). \end{aligned}$$

*Proof* According to the definition of  $x_{k+1}$  and  $y_{k+1}$  and the Fermat rule, we get

$$\begin{aligned} 0 & \in \partial_x L(x_{k+1}, y_k) + \frac{1}{\lambda_k}(x_{k+1} - x_k) + \alpha_k(x_k - x_{k-1}) \\ & = \partial f(x_{k+1}) + \nabla_x R(x_{k+1}, y_k) + \frac{1}{\lambda_k}(x_{k+1} - x_k) + \alpha_k(x_k - x_{k-1}), \\ 0 & \in \partial_y L(x_{k+1}, y_{k+1}) + \frac{1}{\mu_k}(y_{k+1} - y_k) + \beta_k(y_k - y_{k-1}) \\ & = \nabla_y R(x_{k+1}, y_{k+1}) + \partial g(y_{k+1}) + \frac{1}{\mu_k}(y_{k+1} - y_k) + \beta_k(y_k - y_{k-1}). \end{aligned}$$

Thus

$$\begin{aligned} p_{k+1} & := -\frac{1}{\lambda_k}(x_{k+1} - x_k) - \alpha_k(x_k - x_{k-1}) \\ & \quad - \nabla_x R(x_{k+1}, y_k) + \nabla_x R(x_{k+1}, y_{k+1}) \\ & \in \partial f(x_{k+1}) + \nabla_x R(x_{k+1}, y_{k+1}) = \partial_x L(x_{k+1}, y_{k+1}) \end{aligned}$$

and

$$\begin{aligned} q_{k+1} & := -\frac{1}{\mu_k}(y_{k+1} - y_k) - \beta_k(y_k - y_{k-1}) \\ & \in \nabla_y R(x_{k+1}, y_{k+1}) + \partial g(y_{k+1}) = \partial_y L(x_{k+1}, y_{k+1}). \end{aligned}$$

Using assumption (H3), we obtain that

$$\|q_{k+1}\| \leq \frac{1}{\mu_k} \|y_{k+1} - y_k\| + \beta_k \|y_k - y_{k-1}\|$$

and

$$\|p_{k+1}\| \leq \frac{1}{\lambda_k} \|x_{k+1} - x_k\| + \alpha_k \|x_k - x_{k-1}\| + \ell \|y_{k+1} - y_k\|,$$

where  $\ell$  is the Lipschitz constant of  $\nabla R(x, y)$  on the bounded set  $\{(x_k, y_k)\}$ .

Hence the norm estimate can be immediately derived. □

The norm estimate of the subdifferential of  $H$  is a direct consequence of Lemma 3.3.

**Lemma 3.4** *For all  $k \in \mathbb{N}$ ,  $H(z, w)$  has a subdifferential at  $(z_{k+1}, z_k)$  of the form*

$$\omega_{k+1} := \begin{pmatrix} p_{k+1} + 2m_1(x_{k+1} - x_k) \\ q_{k+1} + 2m_1(y_{k+1} - y_k) \\ -2m_1(x_{k+1} - x_k) \\ -2m_1(y_{k+1} - y_k) \end{pmatrix}. \tag{15}$$

Moreover, there exists a positive constant  $C_2 > 0$  such that

$$\|\omega_{k+1}\| \leq C_2 (\|z_{k+1} - z_k\| + \|z_k - z_{k-1}\|). \tag{16}$$

*Proof* According to the definition of  $H(z, w)$ , we get

$$\partial H(z, w) = \begin{pmatrix} \partial L(x, y) + \begin{pmatrix} 2m_1(x-u) \\ 2m_1(y-v) \end{pmatrix} \\ 2m_1(u - x) \\ 2m_1(v - y) \end{pmatrix}. \tag{17}$$

The rest is immediately obtained. □

The norm estimate, together with the closeness of the limiting subdifferential, is used to obtain the following convergence of the subsequence of  $\{x_k, y_k\}$ .

**Lemma 3.5** (Preconvergence result) *Under assumptions (H1)-(H6), we have the following statements:*

- (i)  $\sum_{k=1}^{\infty} \|z_{k+1} - z_k\|^2 < \infty$ ; particularly,  $\|x_{k+1} - x_k\| \rightarrow 0, \|y_{k+1} - y_k\| \rightarrow 0, k \rightarrow \infty$ ;
- (ii) the sequence  $\{L(x_k, y_k)\}$  is convergent;
- (iii) the sequence  $\{H(z_{k+1}, z_k)\}$  is convergent;
- (iv) if  $\{(x_k, y_k)\}$  has a cluster point  $(x^*, y^*)$ , then  $(x^*, y^*) \in \text{crit } L$ .

*Proof* Set  $a_k := L(x_k, y_k) + m_1(\|x_k - x_{k-1}\|^2 + \|y_k - y_{k-1}\|^2)$  and  $b_k = (m_2 - m_1)(\|x_{k+1} - x_k\|^2 + \|y_{k+1} - y_k\|^2)$ . Then Lemma 3.2 gives  $a_{k+1} + b_k \leq a_k$ . Then assumption (H4) ensures that  $a_n$  is bounded below. Thus Lemma 2.2 implies (i) and (ii). Moreover, the definition of  $H(z, w)$  yields that

$$H(z_{k+1}, z_k) = L(x_k, y_k) + m_1(\|x_{k+1} - x_k\|^2 + \|y_{k+1} - y_k\|^2).$$

Thus (iii) is derived from (i) and (ii).



Now let  $\{(x_{k_j}, y_{k_j})\}$  be a subsequence of  $\{(x_k, y_k)\}$  such that  $\{(x_{k_j}, y_{k_j})\} \rightarrow (x^*, y^*), j \rightarrow \infty$ . Since  $f$  is lower semicontinuous, we have

$$\liminf_{j \rightarrow \infty} f(x_{k_j}) \geq f(x^*).$$

On the other hand, the definition of  $x_{k+1}$  shows that

$$\begin{aligned} f(x_{k+1}) + R(x_{k+1}, y_k) + g(y_k) + \frac{1}{2\lambda_k} \|x_{k+1} - x_k\|^2 + \alpha_k \langle x_{k+1}, x_k - x_{k-1} \rangle \\ \leq f(x^*) + R(x^*, y_k) + g(y_k) + \frac{1}{2\lambda_k} \|x^* - x_k\|^2 + \alpha_k \langle x^*, x_k - x_{k-1} \rangle, \end{aligned}$$

from which we get

$$\begin{aligned} R(x_{k+1}, y_k) - R(x^*, y_k) + \frac{1}{2\lambda_k} [\|x_{k+1} - x_k\|^2 - \|x^* - x_k\|^2] \\ + \alpha_k \langle x_{k+1} - x^*, x_k - x_{k-1} \rangle + f(x_{k+1}) \\ \leq f(x^*). \end{aligned}$$

Hence

$$\begin{aligned} R(x_{k_j+1}, y_{k_j}) - R(x^*, y_{k_j}) + \frac{1}{2\lambda_{k_j}} [\|x_{k_j+1} - x_{k_j}\|^2 - \|x^* - x_{k_j}\|^2] \\ + \alpha_{k_j} \langle x_{k_j+1} - x^*, x_{k_j} - x_{k_j-1} \rangle + f(x_{k_j+1}) \\ \leq f(x^*), \end{aligned}$$

where we have used assumption (H5) and replaced  $x_k, y_k$  by  $x_{k_j}, y_{k_j}$ .

Due to the fact that  $\|x_{k+1} - x_k\| \rightarrow 0$  from (i), we have  $\|x_{k_j+1} - x_{k_j}\| \rightarrow 0$ . This, together with  $\|x_{k_j} - x^*\| \rightarrow 0$ , yields  $\|x_{k_j+1} - x^*\| \rightarrow 0$ . Using the continuity of  $R(x, y)$  by assumption (H2), the last inequality yields

$$\limsup_{j \rightarrow \infty} f(x_{k_j}) \leq f(x^*).$$

Therefore

$$\lim_{j \rightarrow \infty} f(x_{k_j}) = f(x^*).$$

In a similar way, we can prove that  $\lim_{j \rightarrow \infty} g(y_{k_j}) = g(y^*)$ . Combining with the continuity of  $R(x, y)$ , we immediately obtain that

$$L(x_{k_j}, y_{k_j}) \rightarrow L(x^*, y^*), \quad j \rightarrow \infty.$$

On the other hand, Lemma 3.5(i) and Lemma 3.3 give  $(p_{k_j+1}, q_{k_j+1}) \in \partial L(x_{k_j+1}, y_{k_j+1})$  and  $(p_{k_j+1}, q_{k_j+1}) \rightarrow 0, j \rightarrow \infty$ . Thus the closeness of the limiting subdifferential (see (6)) indicates that  $0 \in \partial L(x^*, y^*)$ . □

### 3.3 Analytic property of the key function $H$

Denote by  $\Omega$  the set of the cluster points of the sequence  $\{(z_{k+1}, z_k)\}$ .

**Lemma 3.6** *Suppose that the sequence  $\{(x_k, y_k)\}$  is bounded. Under assumptions (H1)-(H6), we have that*

- (i)  $\Omega$  is nonempty, compact, and connected. Moreover,  $\text{dist}((z_{k+1}, z_k), \Omega) \rightarrow 0$  as  $k \rightarrow \infty$ .
- (ii)  $\Omega \subset \text{crit } H = \{(z^*, z^*) : z^* = (x^*, y^*) \in \text{crit } L\}$ .
- (iii)  $H$  is finite and constant on  $\Omega$ .

*Proof* (i) It is easy to check by some elementary tools (see, e.g., [16]).

(ii) According to Lemma 3.5(i),  $\sum_{k=1}^{\infty} \|z_{k+1} - z_k\|^2 < \infty$ , and hence  $\|z_{k+1} - z_k\| \rightarrow 0, k \rightarrow \infty$ . Note that  $z_k = (x_k, y_k)$ , so  $\Omega \subset \{(z^*, z^*) : z^* = (x^*, y^*) \in \text{crit } L\}$ . On the other hand, from (15) we see that

$$0 \in \partial H(z, w) \Leftrightarrow u = x, v = y \text{ and } 0 \in \partial L(x, y).$$

Thus  $\text{crit } H = \{(z^*, z^*) : z^* = (x^*, y^*) \in \text{crit } L\}$ , and hence  $\Omega \subset \text{crit } H$ .

(iii) Notice that  $\{L(x_k, y_k)\}$  is convergent by Lemma 3.5(ii). Let  $L^* = \lim_{k \rightarrow \infty} L(x_k, y_k)$  be a constant. For any  $(z^*, z^*) \in \Omega$ , we have  $z^* = (x^*, y^*) \in \text{crit } L$ , and there exists a subsequence  $\{(x_{k_j}, y_{k_j})\}$  of  $\{(x_k, y_k)\}$  such that  $\{(x_{k_j}, y_{k_j})\} \rightarrow (x^*, y^*)$ . So

$$H(z_{k_j}, z_{k_{j-1}}) \rightarrow H(z^*, z^*), \quad j \rightarrow \infty.$$

Thus

$$H(z^*, z^*) = \lim_{j \rightarrow \infty} H(z_{k_j}, z_{k_{j-1}}) = \lim_{j \rightarrow \infty} L(x_{k_j}, y_{k_j}) = L^*. \quad \square$$

**Theorem 3.1** (Convergence) *Assume that  $H(z, w)$  is a KL function and that the sequence  $\{(x_k, y_k)\}$  is bounded. Then, under assumptions (H1)-(H6), we have*

- (i)  $\sum_{k=1}^{\infty} \|z_k - z_{k-1}\| < \infty$ , that is,  $\sum_{k=1}^{\infty} (\|x_k - x_{k-1}\| + \|y_k - y_{k-1}\|) < \infty$ ;
- (ii)  $\{(x_k, y_k)\}$  converges to a critical point  $(x^*, y^*)$  of  $L(x, y)$ .

*Proof* According to Lemma 3.6, we consider an element  $(x^*, y^*) \in \text{crit } L(x, y)$  such that  $(z^*, z^*) \in \Omega$ , where  $z^* = (x^*, y^*)$ . From the previous proof we can easily obtain that  $\lim_{k \rightarrow \infty} H(z_{k+1}, z_k) = H(z^*, z^*)$ . Next, we prove the theorem in two cases.

*Case 1.* There exists a positive integer  $k_0$  such that  $H(z_{k_0+1}, z_{k_0}) = H(z^*, z^*)$ .

Since  $\{H(z_{k+1}, z_k)\}$  is decreasing, we know that  $H(z_{k+1}, z_k) = H(z^*, z^*)$  for all  $k \geq k_0$ . This, together with the definition of  $H$ , shows that  $z_k = z_{k_0}$  for all  $k \geq k_0$ , and the desired results follow.

*Case 2.*  $H(z_{k+1}, z_k) > H(z^*, z^*)$  for all  $k \in \mathbb{N}$ .

Since  $H$  satisfies the KL property, Lemma 2.1 says that there exist  $\epsilon, \eta > 0$  and a concave function  $\varphi$  such that

- (i)  $\varphi(0) = 0$ ;
- (ii)  $\varphi$  is continuously differentiable on  $(0, \eta)$  and continuous at 0;
- (iii)  $\varphi'(s) > 0$  for all  $s \in (0, \eta)$ ;

(iv) for all

$$\begin{aligned} (z, w) \in & \{(z, w) \in \mathbb{R}^n \times \mathbb{R}^m : \text{dist}((z, w), \Omega) < \epsilon\} \\ & \cap \{(z, w) \in \mathbb{R}^n \times \mathbb{R}^m : H(z^*, z^*) < H(z, w) < H(z^*, z^*) + \eta\}, \end{aligned} \tag{18}$$

we have

$$\varphi'(H(z, w) - H(z^*, z^*)) \text{dist}(0, \partial H(z, w)) \geq 1.$$

Notice that  $H(z_{k+1}, z_k) \rightarrow H(z^*, z^*)$ ,  $k \rightarrow \infty$ , and  $H(z_{k+1}, z_k) > H(z^*, z^*)$ . Let  $k_1$  be such that  $H(z^*, z^*) < H(z_{k+1}, z_k) < H(z^*, z^*) + \eta$  for all  $k \geq k_1$ . By Lemma 3.6(i) there exists  $k_2$  such that  $\text{dist}((z_{k+1}, z_k), \Omega) < \epsilon$  for all  $k \geq k_2$ . Take  $k_3 = \max\{k_1, k_2\}$ . Then, for  $k \geq k_3$ ,  $\{(z_{k+1}, z_k)\}$  belongs to the intersection in (18). Hence

$$\varphi'(H(z_{k+1}, z_k) - H(z^*, z^*)) \text{dist}(0, \partial H(z_{k+1}, z_k)) \geq 1, \quad \forall k \geq k_3.$$

Due to the concavity of  $\varphi$ ,

$$\begin{aligned} & \varphi(H(z_k, z_{k-1}) - H(z^*, z^*)) - \varphi(H(z_{k+1}, z_k) - H(z^*, z^*)) \\ & \geq \varphi'(H(z_k, z_{k-1}) - H(z^*, z^*)) (H(z_k, z_{k-1}) - H(z_{k+1}, z_k)) \\ & \geq \frac{H(z_k, z_{k-1}) - H(z_{k+1}, z_k)}{\text{dist}(0, \partial H(z_k, z_{k-1}))}, \quad k > k_3. \end{aligned}$$

By Lemma 3.4 there exist a point  $\omega_{k+1} \in \partial H(z_{k+1}, z_k)$  defined as in (15) and a positive constant  $C_2 > 0$  such that

$$\|\omega_{k+1}\| \leq C_2 (\|z_{k+1} - z_k\| + \|z_k - z_{k-1}\|).$$

Thus

$$\begin{aligned} & \varphi(H(z_k, z_{k-1}) - H(z^*, z^*)) - \varphi(H(z_{k+1}, z_k) - H(z^*, z^*)) \\ & \geq \frac{H(z_k, z_{k-1}) - H(z_{k+1}, z_k)}{C_2 (\|z_k - z_{k-1}\| + \|z_{k-1} - z_{k-2}\|)}, \quad k > k_3. \end{aligned}$$

From Lemma 3.2 we have  $H(z_k, z_{k-1}) - H(z_{k+1}, z_k) \geq m \|z_{k+1} - z_k\|^2$ . Thus

$$\begin{aligned} & \varphi(H(z_k, z_{k-1}) - H(z^*, z^*)) - \varphi(H(z_{k+1}, z_k) - H(z^*, z^*)) \\ & \geq \frac{m \|z_{k+1} - z_k\|^2}{C_2 (\|z_k - z_{k-1}\| + \|z_{k-1} - z_{k-2}\|)}, \quad k > k_3. \end{aligned} \tag{19}$$

Set  $b_k = \frac{C_2}{m} (\varphi(H(z_k, z_{k-1}) - H(z^*, z^*)) - \varphi(H(z_{k+1}, z_k) - H(z^*, z^*))) \geq 0$ ,  $a_k = \|z_k - z_{k-1}\| \geq 0$ . Then (19) can be equivalently rewritten as

$$b_k \geq \frac{a_{k+1}^2}{a_k + a_{k-1}}, \quad k > k_3. \tag{20}$$

Since  $\varphi \geq 0$ , we know that

$$\sum_{k=1}^N b_k \leq \frac{C_2}{m} \varphi(H(z_1, z_0) - H(z^*, z^*)), \quad \forall N \in \mathbb{N},$$

and hence  $\sum_{k=1}^\infty b_k < \infty$ . Note that from (20) we have

$$a_{k+1} \leq \sqrt{b_k(a_k + a_{k-1})} \leq \frac{1}{4}(a_k + a_{k-1}) + b_k, \quad k > k_3.$$

So Lemma 2.3 gives that  $\sum_{k=1}^\infty a_k < \infty$ , that is,  $\sum_{k=1}^\infty \|z_k - z_{k-1}\| < \infty$ , which is equivalent to  $\sum_{k=1}^\infty (\|x_k - x_{k-1}\| + \|y_k - y_{k-1}\|) < \infty$ . This indicates that  $\{z_k\}$  is a Cauchy sequence. So  $\{z_k\} = \{(x_k, y_k)\}$  is convergent. Let  $(x_k, y_k) \rightarrow (x^*, y^*), k \rightarrow \infty$ . According to Lemma 3.5(iv), it is clear that  $(x^*, y^*)$  is a critical point of  $H$ . □

#### 4 Conclusion

In this paper, we present a proximal alternating minimization algorithm with inertial effect for the minimization problem of the type  $L(x, y) = f(x) + R(x, y) + g(y)$ , where  $f$  and  $g$  are both nonconvex nonsmooth functions, and  $R$  is a smooth function. We prove that every bounded sequence generated by the algorithm converges to a critical point of  $L$ . The key point is to construct a function  $H$  (see (11)) that satisfies the Kurdyka-Lojasiewicz inequality. It is worth mentioning that assumption (H6) requires

$$\max\{\alpha, \beta\} \cdot \max\{\lambda_+, \mu_+\} < \frac{1}{2},$$

which can be achieved by appropriate choice of the parameters.

#### Acknowledgements

The authors would like to express great thanks to the referees for their valuable comments, which notably improved the presentation of this manuscript. The authors also thank Professor Qiaoli Dong for her helpful advice.

#### Funding

This research was supported by National Natural Science Foundation of China (No. 61503385), and the Science Research Foundation in CAUC (No. 2011QD025).

#### Competing interests

We confirm that we have read SpringerOpen's guidance on competing interests and none of the authors has any financial and nonfinancial competing interests in the manuscript.

#### Authors' contributions

All the authors contributed, read, and approved this manuscript.

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 25 April 2017 Accepted: 8 September 2017 Published online: 20 September 2017

#### References

1. Alvarez, F: On the minimizing property of a second order dissipative system in Hilbert spaces. *SIAM J. Control Optim.* **38**(4), 1102-1119 (2000)
2. Alvarez, F, Attouch, H: An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. *Set-Valued Anal.* **9**, 3-11 (2001)
3. Mainge, PE, Noudafi, A: Convergence of new inertial proximal methods for DC programming. *SIAM J. Optim.* **19**(1), 397-413 (2008)
4. Beck, A, Teboulle, M: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2**, 183-202 (2009)

5. Ochs, P, Chen, Y, Brox, T, Pock, T: iPiano: inertial proximal algorithm for non-convex optimization. *SIAM J. Imaging Sci.* **7**, 1388-1419 (2014)
6. Bot, RI, Csetnek, ER, Hendrich, C: Inertial Douglas-Rachford splitting for monotone inclusion problems. *Appl. Math. Comput.* **256**, 472-487 (2015)
7. Bot, RI, Csetnek, ER: In inertial alternating direction method of multipliers. *Minimax Theory Appl.* **1**, 29-49 (2016)
8. Chambolle, A, Dossal, C: On the convergence of the iterates of the 'fast iterative shrinkage/thresholding algorithm'. *J. Optim. Theory Appl.* **166**, 968-982 (2016)
9. Chen, C, Ma, S, Yang, J: A general inertial proximal for mixed variational inequality problem. *SIAM J. Optim.* **25**, 2120-2142 (2015)
10. Dong, QL, Lu, YY, Yang, JF: The extragradient algorithm with inertial effects for solving the variational inequality. *Optimization* **65**(12), 2217-2226 (2016)
11. Bot, RI, Csetnek, ET, Laszlo, SC: An inertial forward-backward algorithm for the minimization of the sum of two nonconvex functions. *EURO J. Comput. Optim.* **4**(1), 1-23 (2016)
12. Bot, RI, Csetnek, ER: An inertial Tseng's type proximal algorithm for nonsmooth and nonconvex optimization problems. *J. Optim. Theory Appl.* **171**(2), 600-616 (2016)
13. Attouch, H, Bolte, J, Redont, P, Soubeyran, A: Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Lojasiewicz inequality. *Math. Oper. Res.* **35**(2), 428-457 (2010)
14. Mordukhovich, B: *Variational Analysis and Generalized Differentiation, I: Basic Theory, II: Applications*. Springer, Berlin (2006)
15. Rochafellar, RT, Wets, RJ-B: *Variational Analysis. Fundamental Principles of Mathematical Sciences*, vol. 317. Springer, Berlin (1998)
16. Bolte, J, Sabach, S, Teboulle, M: Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Math. Program., Ser. A* **146**(1-2), 459-494 (2014)
17. Bauschke, HH, Combettes, PL: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Springer, New York (2011)

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)

---