

RESEARCH

Open Access



An inexact proximal gradient algorithm with extrapolation for a class of nonconvex nonsmooth optimization problems

Zehui Jia¹, Zhongming Wu^{2*} and Xiaomei Dong³

*Correspondence:
wuzm@seu.edu.cn

²School of Economics and Management, Southeast University, Nanjing, P.R. China
Full list of author information is available at the end of the article

Abstract

In this paper, we propose an inexact version of proximal gradient algorithm with extrapolation for solving a class of nonconvex nonsmooth optimization problems. Specifically, the subproblem in proximal gradient algorithm with extrapolation is allowed to be solved inexactly by certain relative error criterion, in the sense that the criterion can be updated adaptively in each iteration. Under the assumption that an auxiliary function satisfies the Kurdyka–Łojasiewicz (KL) inequality, we prove that the iterative sequence generated by the inexact proximal gradient algorithm with extrapolation converges to a stationary point of the considered problem. Furthermore, the convergence rate of the proposed algorithm can be established when the KL exponent is known. Moreover, we illustrate the advantage by applying the algorithm to solve a nonconvex optimization problem.

Keywords: Nonconvex minimization; Proximal gradient algorithm; Relative error criterion; Extrapolation; Global convergence

1 Introduction

In this paper, we consider the following structured optimization problem:

$$\min_{x \in \mathcal{R}^n} F(x) := f(x) + g(x), \quad (1)$$

where $g: \mathcal{R}^n \rightarrow \mathcal{R} \cup \{+\infty\}$ is a proper closed convex function, and $f: \mathcal{R}^n \rightarrow \mathcal{R}$ is a possibly nonconvex function which has a Lipschitz continuous gradient. We assume that the optimal value of (1) is finite and is attained. Problem (1) arises in many applications such as compressed sensing [1, 2] and image processing [3]. Due to the special structure and properties, the first-order methods, especially the proximal gradient algorithm, are widely used for solving problem (1).

The proximal gradient algorithm, also known as forward-backward splitting method [4], takes full advantage of the property of the problem whose objective is the sum of a smooth function and a nonsmooth function. In each iteration, the algorithm executes a gradient step for the smooth part and a proximal step for the nonsmooth part. In the convex case, i.e., both functions f and g are convex, this method has been widely studied, see, e.g., [4, 5]. However, the proximal gradient algorithm, in its original form, usually performs

slowly in practice [6]. To accelerate the convergence speed of the proximal gradient algorithm, many strategies have been proposed in the last decades. One of the most efficient strategies is to incorporate extrapolation, where a momentum term based on the previous iterations is introduced to update the current iteration. The concrete iterative scheme takes the following form:

$$\begin{cases} y^k = x^k + \beta_k(x^k - x^{k-1}), \\ x^{k+1} = \text{prox}_{\mu g}(y^k - \mu \nabla f(y^k)), \end{cases} \tag{2}$$

where $\mu > 0$, $\beta_k \in [0, 1)$ and $\text{prox}_{\mu g}(\cdot)$ denotes the proximal mapping [7] which is defined by

$$\text{prox}_{\mu g}(v) = \arg \min_{x \in \mathcal{R}^n} \left\{ g(x) + \frac{1}{2\mu} \|x - v\|^2 \right\}$$

for any $v \in \mathcal{R}^n$. Indeed, many accelerated methods can be contained in the framework of scheme (2). A famous example is the fast iterative shrinkage-thresholding algorithm (FISTA) proposed by Beck and Teboulle [8], where they require β_k satisfying a certain recurrence relation. In [8], it has been shown that FISTA possesses $O(1/k^2)$ convergence rate for the convex case which is faster than the original proximal gradient algorithm, where k counts the iteration number. For the nonconvex case, there are also some works considering the proximal gradient method with or without acceleration, see, e.g., [9–15]. In [13], Wen et al. proved the linear convergence of proximal gradient algorithm with extrapolation for nonconvex optimization problem (1), based on the error bound condition, while the works [9–12, 14, 15] studied the proximal gradient algorithm or its variants under the Kurdyka–Łojasiewicz (KL) framework for the nonconvex case, in which they usually require some potential functions satisfying the KL property (see Definition 2.1).

However, the efficiency of the proximal gradient method or its accelerated versions largely relies on the solving difficulty of the subproblem in (2). In many applications, the proximal mapping $\text{prox}_{\mu g}(\cdot)$ is not easy to evaluate and does not possess closed-form solution. Therefore, in practice, one prefers to solve the subproblem in (2) inexactly with some tolerance initially, and then tighten the solution as the iteration goes, instead of solving it with high accuracy. Such an idea is reasonable, as it can help avoid spending too much effort at the beginning of the iterations for an exact minimizer. In order to achieve the inexact solving, many inexact criteria for the proximal-based methods have been proposed recently. The oldest one for the proximal point algorithm is the absolute summable error criterion [16], which involves a sequence of error tolerance parameters $\epsilon_k \subset [0, \infty)$ with $\sum_{k=1}^{\infty} \epsilon_k < \infty$. However, such an absolute error criterion does not provide any guidance on selecting the value of ϵ_k in the implementation. To overcome this drawback, a class of relative error criteria is proposed for approximating the proximal point algorithm [17–19], which only involves a single scalar parameter to control the ratio of the residual of the subproblem and other quantities in algorithm. The advantages of such criteria include that they do not need to determine the value of a sequence and can be adjusted adaptively in each iteration. Based on these works, a natural question is whether we can design an inexact proximal gradient method with a relative error criterion for solving the nonconvex optimization problems considered in this paper.

Along this line, we propose an inexact version of proximal gradient algorithm with extrapolation for solving the nonconvex nonsmooth optimization problem (1). In particular, the subproblem in proximal gradient algorithm with extrapolation is allowed to be solved inexactly under a certain relative error criterion, the scheme is as follows:

$$\begin{cases} y^k = x^k + \beta_k(x^k - x^{k-1}), \\ x^{k+1} \approx x_{\text{exact}}^{k+1} = \text{prox}_{\mu g}(y^k - \mu \nabla f(y^k)). \end{cases} \tag{3}$$

Evidently, if we set $x^{k+1} = x_{\text{exact}}^{k+1}$ for each $k \geq 0$, scheme (3) reduces to the proximal gradient algorithm with extrapolation (2), which is actually well studied in [13]. By introducing a reasonable relative inexact criterion, we analyze the global convergence of the sequence generated by this inexact algorithm (3) based on the KL property. Besides, the convergence rate of the proposed method can be established if the KL exponent is known. It is worth noting that Li et al. [20] have proved that if the error bound condition and the assumption of the separability of stationary values hold, the potential function of the proximal gradient algorithm with extrapolation for optimization problem (1) satisfies the KL property with an exponent of 1/2. With the later condition (a milder one), in this paper, we can prove the linear convergence of the inexact algorithm (3) (which contains (2) as a special case). This indicates that our work can get the same linear convergence result under a weaker condition compared with [13].

The rest of this paper is organized as follows. Section 2 presents some basic notations and preliminary materials. In Sect. 3, we present the inexact proximal gradient algorithm with extrapolation under relative error criteria concretely. Under the Kurdyka–Łojasiewicz framework, we establish the convergence properties and convergence rate of the iterates generated by the proposed method. In Sect. 4, we perform a numerical experiment to illustrate the feasibility and advantage of the proposed method. Finally, we make some conclusions in Sect. 5.

2 Preliminaries

In this section, we summarize some notations and preliminaries which will be used in further analysis.

Throughout this paper, we use \mathcal{R}^n to denote the n -dimensional Euclidean space, with its standard inner product denoted by $\langle \cdot, \cdot \rangle$. The Euclidean norm is denoted by $\| \cdot \|$. For a matrix $A \in \mathcal{R}^{m \times n}$, we use A^T to denote its transpose. For any subset $\Omega \subset \mathcal{R}^n$ and any point $x \in \mathcal{R}^n$, the distance from x to Ω , denoted by $\text{dist}(x, \Omega)$, is defined as

$$\text{dist}(x, \Omega) = \inf_{y \in \Omega} \|y - x\|.$$

When Ω is closed and convex, we use $P_\Omega(x)$ to denote the projection of x onto Ω .

For an extended-real-valued function $g : \mathcal{R}^n \rightarrow \mathcal{R} \cup \{+\infty\}$, the domain of g is defined as

$$\text{dom } g = \{x \in \mathcal{R}^n | g(x) < +\infty\}.$$

We say that g is proper if it never equals $-\infty$ and $\text{dom } g \neq \emptyset$. Such a function is closed if it is lower semicontinuous. For a proper closed convex function $g : \mathcal{R}^n \rightarrow \mathcal{R} \cup \{+\infty\}$, the

subdifferential of g at $x \in \text{dom } g$ is given by

$$\partial g(x) = \{ \xi \in \mathcal{R}^n \mid g(u) - g(x) - \langle \xi, u - x \rangle \geq 0, \forall u \in \mathcal{R}^n \}.$$

A necessary condition for $x \in \mathcal{R}^n$ to be a minimizer of the sum of a differentiable function f and a closed convex function g is

$$0 \in \nabla f(x) + \partial g(x). \tag{4}$$

Throughout the paper, a point which satisfies (4) is called critical point or stationary point of problem (1), and the set of all critical points satisfying (4) is denoted by $\text{crit } F$.

Next, we recall the definitions of the KL property, KL function, and KL exponent from [10].

Definition 2.1 (KL property) Let $f : \mathcal{R}^n \rightarrow \mathcal{R} \cup \{+\infty\}$ be a proper lower semicontinuous function. For $-\infty < \eta_1 < \eta_2 \leq +\infty$, set $[\eta_1 < f < \eta_2] = \{x \in \mathcal{R}^n : \eta_1 < f(x) < \eta_2\}$. We say that f has the KL property at $x^* \in \text{dom } \partial f$ if there exist $\eta \in (0, +\infty]$, a neighborhood U of x^* , and a continuous concave function $\phi : [0, \eta) \rightarrow \mathcal{R}_+$ such that

- (i) $\phi(0) = 0$ and ϕ is continuously differentiable on $(0, \eta)$ with $\phi'(s) > 0, \forall s \in (0, \eta)$;
- (ii) for all x in $U \cap [f(x^*) < f < f(x^*) + \eta]$, the following KL inequality holds:

$$\phi'(f(x) - f(x^*))d(0, \partial f(x)) \geq 1.$$

Definition 2.2 (KL function) If f satisfies the KL property at each point of $\text{dom } \partial f$, then f is called a KL function.

Definition 2.3 (KL exponent) If the function ϕ can be chosen as $\phi(s) = cs^{1-\theta}, \theta \in [0, 1), c > 0$, i.e., there exists $\eta > 0$, so that

$$d(0, \partial f(x)) \geq c(f(x) - f(x^*))^\theta$$

for all x in $U \cap [f(x^*) < f < f(x^*) + \eta]$, then we say that f has the KL property at x^* with an exponent of θ .

Remark 2.4 One can easily check that the KL property is automatically satisfied at any noncritical point $x^* \in \text{dom } f$, see, e.g., [14, Lemma 2.1]. Besides, a big class of functions that have the KL property is given by real semialgebraic functions [10], which include most of the convex functions and some other classes of nonconvex functions.

Below we recall an important property for the KL functions, whose proof can be found in [21].

Lemma 2.5 (Uniformized KL property [21]) Denote by Φ_η the set of functions which satisfy Definition 2.1(i), and let Ω be a compact set, and let $f : \mathcal{R}^n \rightarrow \mathcal{R} \cup \{+\infty\}$ be a proper lower semicontinuous function. Assume that f is constant on Ω and satisfies the KL property at

each point of Ω . Then there exist $\epsilon > 0, \eta > 0$, and $\phi \in \Phi_\eta$ such that, for all $\bar{x} \in \Omega$ and for all x in the following intersection:

$$\{x \in \mathcal{R}^n : \text{dist}(x, \Omega) < \epsilon\} \cap [f(\bar{x}) < f < f(\bar{x}) + \eta],$$

one has

$$\phi'(f(x) - f(\bar{x})) \text{dist}(0, \partial f(x)) \geq 1.$$

The following descent lemma for a smooth function is useful for the convergence analysis.

Lemma 2.6 ([22]) *Let $f : \mathcal{R}^n \rightarrow \mathcal{R}$ be a continuous differentiable function, and the gradient ∇f is Lipschitz continuous with modulus $L_f > 0$, then for any $x, y \in \mathcal{R}^n$, we have*

$$|f(y) - f(x) - \langle \nabla f(x), y - x \rangle| \leq \frac{L_f}{2} \|y - x\|^2.$$

3 Algorithm and convergence analysis

In this section, we first propose an inexact proximal gradient algorithm with extrapolation for solving the possibly nonconvex nonsmooth optimization problems. Then, based on the KL property, we establish the global convergence and convergence rate of the proposed method.

3.1 An inexact proximal gradient algorithm with extrapolation

In this subsection, we propose an inexact proximal gradient algorithm with extrapolation under the relative error criterion. The concrete algorithmic framework is presented in Algorithm 1, where only one nonnegative constant σ and the subgradient information are needed to control the error tolerance and obtain a candidate solution. Note that Algorithm 1 reduces to the proximal gradient algorithm with extrapolation (2) if we take $\sigma = 0$.

Remark 3.1 In Algorithm 1, the d^{k+1} in Step 2 can be chosen as follows:

$$d^{k+1} = \xi^{k+1} - \eta^{k+1},$$

where $\eta^{k+1} = -L(x^{k+1} - y^k) - \nabla f(y^k)$ and $\xi^{k+1} = P_{\partial g(x^{k+1})}(\eta^{k+1})$.

Algorithm 1 An inexact proximal gradient algorithm with extrapolation and relative error criteria

Let $\sigma \in [0, \frac{1}{2}]$, $\beta_k \in [0, \sqrt{\frac{L-2\sigma}{L+1}}]$, and $\epsilon > 0$. Choose $x^0 \in \text{dom } g$, and set $x^{-1} = x^0$. For $k = 0, 1, \dots$

Step 1. $y^k = x^k + \beta_k(x^k - x^{k-1})$

Step 2. Compute $x^{k+1} \approx \arg \min_{x \in \mathcal{R}^n} \{\langle \nabla f(y^k), x \rangle + \frac{1}{2} \|x - y^k\|^2 + g(x)\}$ such that

$$|\langle d^{k+1}, x^k - x^{k+1} \rangle| + \|d^{k+1}\|^2 \leq \sigma \|x^k - x^{k+1}\|^2, \tag{5}$$

with $d^{k+1} \in \partial g(x^{k+1}) + L(x^{k+1} - y^k) + \nabla f(y^k)$.

Step 3. If $\|x^{k+1} - x^k\| \leq \epsilon$, stop; otherwise set $k := k + 1$ and go to **Step 1**.

Before analyzing the convergence of Algorithm 1, we firstly define an auxiliary function

$$H_\alpha(x, w) = f(x) + g(x) + \alpha \|x - w\|^2,$$

where α is a fixed non-negative constant. By the definition of a critical point, we know that (x^*, w^*) is a critical point of the function H_α if it satisfies

$$\begin{cases} 0 \in \nabla f(x^*) + \partial g(x^*) + 2\alpha(x^* - w^*), \\ 0 = w^* - x^*. \end{cases}$$

The critical points set of H_α is denoted by $\text{crit } H_\alpha$. Indeed, it is easy to verify that if $(x^*, w^*) \in \text{crit } H_\alpha$, then x^* is a critical point of problem (1), i.e., $x^* \in \text{crit } F$. In this paper, we assume that there is at least a critical point of problem (1).

Correspondingly, an auxiliary function sequence is given as follows:

$$H_{k+1,\alpha} := H_\alpha(x^{k+1}, w^{k+1}) = f(x^{k+1}) + g(x^{k+1}) + \alpha \|x^{k+1} - w^{k+1}\|^2$$

for fixed $\alpha \in (\frac{L+l}{2}\bar{\beta}^2, \frac{l}{2} - \sigma)$ with $\bar{\beta} = \sup_k \beta_k$, where $\{x^k\}$ is generated by Algorithm 1, and $\{w^k | w^k := x^{k-1}\}$. Through studying the nonincreasing property and the convergence of $H_{k,\alpha}$, we can obtain that $\{(x^k, w^k)\}$ converges to a critical point of H_α , and thus $\{x^k\}$ converges to a stationary point of problem (1).

3.2 Convergence analysis

In this subsection, we analyze the convergence and the convergence rate of the sequence generated by the inexact proximal gradient algorithm with extrapolation for solving (1). Invoking the optimality condition of the subproblem in Algorithm 1, we have

$$\begin{cases} y^k = x^k + \beta_k(x^k - x^{k-1}), \\ d^{k+1} \in \partial g(x^{k+1}) + L(x^{k+1} - y^k) + \nabla f(y^k). \end{cases} \tag{6}$$

As discussed in [13], any function f with Lipschitz continuous gradient can be decomposed to the difference of two convex and differentiable functions, and their gradients are Lipschitz continuous. In other words, there exist convex and differentiable functions f_1 and f_2 with Lipschitz continuous gradients such that

$$f = f_1 - f_2.$$

For instance, one can decompose f so that $f_1(x) = f(x) + \frac{\tau}{2} \|x\|^2$ and $f_2(x) = \frac{\tau}{2} \|x\|^2$ with $\tau \geq L_f$, where L_f denotes the Lipschitz constant of ∇f . Without loss of generality, we suppose that $f = f_1 - f_2$ for some convex functions f_1 and f_2 with Lipschitz continuous gradients in the following analysis. We also denote the Lipschitz continuity moduli of ∇f_1 and ∇f_2 by $L > 0$ and $l \geq 0$, respectively. Furthermore, by taking larger L if necessary, we assume that $L \geq l$. Then it is not hard to show that ∇f is Lipschitz continuous with a modulus $L_f = L$. Therefore, it holds that

$$f_1(y^k) + \langle \nabla f_1(y^k), z - y^k \rangle \leq f_1(z), \tag{7}$$

and

$$f_2(z) - f_2(y^k) - \langle \nabla f_2(y^k), z - y^k \rangle \leq \frac{l}{2} \|z - y^k\|^2. \tag{8}$$

Now we begin our analysis with the following lemma.

Lemma 3.2 *Suppose that $\sigma \in [0, \frac{l}{2})$ and $\bar{\beta} \in (0, \sqrt{\frac{l-2\sigma}{L+l}})$. Let $\{x^k\}$ be the sequence generated by Algorithm 1, and $\{w^k | w^k := x^{k-1}\}$. Then $H_{k,\alpha}$ is monotonically nonincreasing. In particular, it holds that*

$$H_{k+1,\alpha} \leq H_{k,\alpha} - \delta \|u^{k+1} - u^k\|^2, \tag{9}$$

where $u^k := (x^k, w^k), \forall k > 0$, and δ is a positive constant.

Proof Fix any k and $z \in \text{dom}g$. Due to the convexity of g , for any $\xi^{k+1} \in \partial g(x^{k+1})$, we have

$$g(z) \geq g(x^{k+1}) + \langle \xi^{k+1}, z - x^{k+1} \rangle.$$

From the second relation in (6), we can set $\xi^{k+1} = d^{k+1} - L(x^{k+1} - y^k) - \nabla f(y^k)$, then the above inequality can be written as

$$\begin{aligned} g(z) - g(x^{k+1}) &\geq \langle d^{k+1} - L(x^{k+1} - y^k) - \nabla f(y^k), z - x^{k+1} \rangle \\ &= \langle d^{k+1}, z - x^{k+1} \rangle - \langle \nabla f(y^k), z - x^{k+1} \rangle - \langle L(x^{k+1} - y^k), z - x^{k+1} \rangle. \end{aligned}$$

Rearranging the above inequality, we obtain

$$\begin{aligned} g(x^{k+1}) &\leq g(z) + \langle d^{k+1}, x^{k+1} - z \rangle + \langle \nabla f(y^k), z - x^{k+1} \rangle + \langle L(x^{k+1} - y^k), z - x^{k+1} \rangle \\ &= g(z) + \langle d^{k+1}, x^{k+1} - z \rangle + \langle \nabla f(y^k), z - x^{k+1} \rangle \\ &\quad + \frac{L}{2} (\|y^k - z\|^2 - \|y^k - x^{k+1}\|^2 - \|x^{k+1} - z\|^2). \end{aligned} \tag{10}$$

Since ∇f is Lipschitz continuous with modulus L , it follows from Lemma 2.6 that

$$f(x^{k+1}) \leq f(y^k) + \langle \nabla f(y^k), x^{k+1} - y^k \rangle + \frac{L}{2} \|x^{k+1} - y^k\|^2. \tag{11}$$

Combining (10) and (11), we get

$$\begin{aligned} f(x^{k+1}) + g(x^{k+1}) &\leq f(y^k) + g(z) + \langle d^{k+1}, x^{k+1} - z \rangle + \langle \nabla f(y^k), z - y^k \rangle \\ &\quad + \frac{L}{2} (\|y^k - z\|^2 - \|x^{k+1} - z\|^2). \end{aligned} \tag{12}$$

Together with inequalities (7) and (8), we obtain

$$f(y^k) + \langle \nabla f(y^k), z - y^k \rangle \leq f(z) + \frac{l}{2} \|z - y^k\|^2.$$

Substituting the above inequality into (12), we get

$$\begin{aligned}
 f(x^{k+1}) + g(x^{k+1}) &\leq f(z) + g(z) + \langle d^{k+1}, x^{k+1} - z \rangle \\
 &\quad + \frac{L+l}{2} \|y^k - z\|^2 - \frac{L}{2} \|x^{k+1} - z\|^2.
 \end{aligned}
 \tag{13}$$

Setting $z := x^k$ in (13), we obtain

$$\begin{aligned}
 F(x^{k+1}) - F(x^k) &\leq \langle d^{k+1}, x^{k+1} - x^k \rangle + \frac{L+l}{2} \|y^k - x^k\|^2 - \frac{L}{2} \|x^{k+1} - x^k\|^2 \\
 &= \langle d^{k+1}, x^{k+1} - x^k \rangle + \frac{L+l}{2} \beta_k^2 \|x^k - x^{k-1}\|^2 - \frac{L}{2} \|x^{k+1} - x^k\|^2,
 \end{aligned}$$

where the equality follows from (6). Rearranging this inequality, we can deduce

$$\begin{aligned}
 F(x^{k+1}) + \alpha \|x^{k+1} - x^k\|^2 &\leq F(x^k) + \alpha \|x^k - x^{k-1}\|^2 + \langle d^{k+1}, x^{k+1} - x^k \rangle \\
 &\quad - \left(\alpha - \frac{L+l}{2} \beta_k^2 \right) \|x^k - x^{k-1}\|^2 - \left(\frac{L}{2} - \alpha \right) \|x^{k+1} - x^k\|^2 \\
 &\leq F(x^k) + \alpha \|x^k - x^{k-1}\|^2 \\
 &\quad - \left(\alpha - \frac{L+l}{2} \bar{\beta}^2 \right) \|w^{k+1} - w^k\|^2 - \left(\frac{L}{2} - \alpha - \sigma \right) \|x^{k+1} - x^k\|^2,
 \end{aligned}
 \tag{14}$$

where α is a non-negative constant, and the second inequality follows from (5). Recalling that $\bar{\beta} < \sqrt{\frac{L-2\sigma}{L+l}}$, one can choose $\alpha \in (\frac{L+l}{2} \bar{\beta}^2, \frac{L}{2} - \sigma)$, and then (14) becomes

$$H_{k+1,\alpha} \leq H_{k,\alpha} - \delta_1 \|w^{k+1} - w^k\|^2 - \delta_2 \|x^{k+1} - x^k\|^2,$$

where δ_1, δ_2 are two positive constants. Let $\delta = \min\{\delta_1, \delta_2\} > 0$, and then assertion (9) follows immediately. This evidently implies $\{H_{k,\alpha}\}$ is nonincreasing. This completes the proof. \square

Lemma 3.3 *Let $\{x^k\}$ be the sequence generated by Algorithm 1, and $\{w^k | w^k := x^{k-1}\}$. Then there exists $\tilde{c} > 0$ such that*

$$\text{dist}(0, \partial H_\alpha(x^{k+1}, w^{k+1})) \leq \tilde{c} \|u^{k+1} - u^k\|,$$

where $u^k := (x^k, w^k), \forall k > 0$.

Proof From the definition of $H_\alpha(x^{k+1}, w^{k+1})$, it follows that

$$\begin{aligned}
 \partial_x H_\alpha(x^{k+1}, w^{k+1}) &= \nabla f(x^{k+1}) + \partial g(x^{k+1}) + 2\alpha(x^{k+1} - x^k), \\
 \nabla_w H_\alpha(x^{k+1}, w^{k+1}) &= 2\alpha(x^k - x^{k+1}).
 \end{aligned}$$

From this, together with (6), we obtain

$$\zeta^{k+1} := d^{k+1} - L(x^{k+1} - y^k) - \nabla f(y^k) + \nabla f(x^{k+1}) + 2\alpha(x^{k+1} - x^k) \in \partial_x H_\alpha(x^{k+1}, w^{k+1}).$$

Then, using the trigonometric inequality, we have

$$\begin{aligned}
 & \|d^{k+1} - L(x^{k+1} - y^k) - \nabla f(y^k) + \nabla f(x^{k+1}) + 2\alpha(x^{k+1} - x^k)\| \\
 & \leq \|d^{k+1}\| + \|-L(x^{k+1} - y^k) + 2\alpha(x^{k+1} - x^k)\| + \|\nabla f(x^{k+1}) - \nabla f(y^k)\| \\
 & \leq \|d^{k+1}\| + \|-L(x^{k+1} - y^k) + 2\alpha(x^{k+1} - x^k)\| + L\|x^{k+1} - y^k\| \\
 & \leq \sqrt{\sigma}\|x^{k+1} - x^k\| + \|(2\alpha - L)(x^{k+1} - x^k) + \beta_k L(x^k - x^{k-1})\| \\
 & \quad + L\|x^{k+1} - x^k - \beta_k(x^k - x^{k-1})\| \\
 & \leq (\sqrt{\sigma} + L - 2\alpha + L)\|x^{k+1} - x^k\| + 2\beta_k L\|x^k - x^{k-1}\| \\
 & = (\sqrt{\sigma} + 2L - 2\alpha)\|x^{k+1} - x^k\| + 2\beta_k L\|w^{k+1} - w^k\|,
 \end{aligned}$$

where the second inequality follows from the Lipschitz continuity of ∇f , and the third inequality is due to (5) and (6). Thus, there exist c_1 and $c_2 > 0$ such that

$$\begin{aligned}
 \text{dist}(0, \partial H_\alpha(x^{k+1}, w^{k+1})) & \leq \sqrt{\|\zeta^{k+1}\|^2 + \|2\alpha(x^k - x^{k+1})\|^2} \\
 & \leq \sqrt{c_1\|x^{k+1} - x^k\|^2 + c_2\|w^{k+1} - w^k\|^2} \\
 & \leq \tilde{c}\sqrt{(\|x^{k+1} - x^k\|^2 + \|w^{k+1} - w^k\|^2)} \\
 & = \tilde{c}\|u^{k+1} - u^k\|,
 \end{aligned}$$

where $\tilde{c} = \max\{\sqrt{c_1}, \sqrt{c_2}\}$. This completes the proof. □

In the following lemma, we present several properties of the limit point set of $\{(x^k, w^k)\}$.

Lemma 3.4 *Let $\{x^k\}$ be the sequence generated by Algorithm 1, which is assumed to be bounded, and let $\{w^k | w^k := x^{k-1}\}$. Let S denote the set of the limit points of $\{(x^k, w^k)\}$. Then*

- (i) S is a nonempty compact set, and $\text{dist}((x^k, w^k), S) \rightarrow 0$, as $k \rightarrow +\infty$;
- (ii) $S \subset \text{crit } H_\alpha$;
- (iii) H_α is finite and constant on S , equal to $\inf_{k \in \mathbb{N}} H_\alpha(x^k, w^k) = \lim_{k \rightarrow +\infty} H_\alpha(x^k, w^k)$.

Proof (i) By definition, it is trivial. (ii) Let $(x^*, w^*) \in S$, then there exists a subsequence $\{(x^{k_j}, w^{k_j})\}$ of $\{(x^k, w^k)\}$ converging to (x^*, w^*) . Note that Lemma 3.2 implies

$$\|x^{k+1} - x^k\|^2 \rightarrow 0, \quad \|w^{k+1} - w^k\|^2 \rightarrow 0,$$

which means that the sequences $\{(x^{k_j+1}, w^{k_j+1})\}$ and $\{(x^{k_j-1}, w^{k_j-1})\}$ also converge to (x^*, w^*) , and further $x^* = w^*$. Together with (5), we can also deduce $\|d^k\| \rightarrow 0$. Considering the continuity of ∇f and the closeness of ∂g , by taking the limit in (6) along the sequence (x^{k_j+1}, w^{k_j+1}) , we have

$$0 \in \partial g(x^*) + \nabla f(x^*).$$

Therefore, $\{(x^*, w^*)\}$ is a critical point of H_α , hence $\{(x^*, w^*)\} \in \text{crit } H_\alpha$.

(iii) In the following, we will consider the value of H_α on the set of accumulation points. Considering the convexity of g , we have

$$g(x^*) \geq g(x_{k_j+1}) + \xi_{k_j+1}^T(x^* - x_{k_j+1}), \quad \forall \xi_{k_j+1} \in \partial g(x_{k_j+1}).$$

According to the optimality condition (6), we can take $\xi_{k_j+1} = d^{k_j+1} - L(x^{k_j+1} - y^{k_j}) - \nabla f(y^{k_j})$. Thus we have

$$g(x^*) \geq g(x_{k_j+1}) + (d^{k_j+1} - L(x^{k_j+1} - y^{k_j}) - \nabla f(y^{k_j}))^T(x^* - x_{k_j+1}).$$

From this together with the continuity of $f(x)$ with respect to x , and the continuity of $\alpha\|x - w\|^2$ with respect to both x and w , we have

$$\begin{aligned} & f(x^*) + g(x^*) + \alpha\|x^* - w^*\|^2 \\ &= \lim_{j \rightarrow +\infty} f(x^*) + g(x^*) + \alpha\|x^* - w_{k_j+1}\|^2 \\ &\geq \limsup_{j \rightarrow +\infty} \{f(x_{k_j+1}) + g(x_{k_j+1}) + \alpha\|x_{k_j+1} - w_{k_j+1}\|^2 \\ &\quad + (d^{k_j+1} - L(x^{k_j+1} - y^{k_j}) - \nabla f(y^{k_j}))^T(x^* - x_{k_j+1})\}. \end{aligned}$$

Furthermore, from (5) and (6), we get $x^{k_j+1} - y^{k_j} \rightarrow 0$ and $\|d^{k_j+1}\| \rightarrow 0$. Combining the boundedness of sequence $\{x^k\}$, we further have

$$H_\alpha(x^*, w^*) \geq \limsup_{j \rightarrow +\infty} H_\alpha(x^{k_j+1}, w^{k_j+1}). \tag{15}$$

On the other hand, according to the lower semicontinuity of H_α , we have

$$H_\alpha(x^*, w^*) \leq \liminf_{j \rightarrow +\infty} H_\alpha(x^{k_j+1}, w^{k_j+1}). \tag{16}$$

Using inequalities (15), (16), and $\{H_\alpha(x^k, w^k)\}$ is nonincreasing, we obtain

$$H_\alpha(x^*, w^*) = \lim_{k \rightarrow +\infty} H_\alpha(x^k, w^k).$$

Therefore, $\{H_\alpha(x, w)\}$ is constant on S . Moreover, it holds that $\inf_{k \in N} H_\alpha(x^k, w^k) = \lim_{k \rightarrow +\infty} H_\alpha(x^k, w^k)$. □

We are now ready to prove the main result of this paper.

Theorem 3.5 *Let $\{x^k\}$ be the sequence generated by Algorithm 1, which is assumed to be bounded, and let $\{w^k | w^k := x^{k-1}\}$. Suppose that f and g are semi-algebraic functions, then*

$$\sum_{k=0}^{+\infty} \|x^{k+1} - x^k\| < +\infty, \quad \sum_{k=0}^{+\infty} \|w^{k+1} - w^k\| < +\infty,$$

and thus $\{(x^k, w^k)\}$ converges to a critical point of H_α .

Proof From the proof of Lemma 3.4, we know that $\lim_{k \rightarrow +\infty} H_\alpha(x^k, w^k) = H_\alpha(x^*, w^*)$ for all $(x^*, w^*) \in S$. Then there are two cases we need to consider.

Case I: There exists an integer k_0 such that $H_\alpha(x^{k_0}, w^{k_0}) = H_\alpha(x^*, w^*)$. Rearranging terms of inequality (9), and using that $H_\alpha(x^k, w^k)$ is nonincreasing, for any $k \geq k_0$, we have

$$\delta \|x^{k+1} - x^k\|^2 + \delta \|w^{k+1} - w^k\|^2 \leq H_{k,\alpha} - H_{k+1,\alpha} \leq H_{k_0,\alpha} - H_\alpha(x^*, w^*) = 0.$$

Thus, for $\forall k \geq k_0$, we have $x^{k+1} = x^k$ and $w^{k+1} = w^k$, the assertion holds.

Case II: Now we assume that $H_\alpha(x^k, w^k) > H_\alpha(x^*, w^*), \forall k$. Since $\text{dist}((x^k, w^k), S) \rightarrow 0$, it follows that for all $\epsilon > 0$, there exists $K_1 > 0$ such that, for any $k > K_1$, $\text{dist}((x^k, w^k), S) < \epsilon$. Considering that $\lim_{k \rightarrow +\infty} H_\alpha(x^k, w^k) = H_\alpha(x^*, w^*)$, then for any given $\eta > 0$, there exists $K_2 > 0$ such that $H_\alpha(x^k, w^k) < H_\alpha(x^*, w^*) + \eta, \forall k > K_2$. Therefore, for any $\epsilon, \eta > 0$, we have

$$\text{dist}((x^k, w^k), S) < \epsilon, \quad H_\alpha(x^k, w^k) < H_\alpha(x^*, w^*) + \eta, \quad \forall k > \tilde{K},$$

where $\tilde{K} = \max\{k_1, k_2\}$. Since S is a nonempty compact set and H_α is constant on S , applying Lemma 2.5 with $\Omega = S$, we deduce

$$\phi'(H_\alpha(x^k, w^k) - H_\alpha(x^*, w^*)) \text{dist}(0, \partial H_\alpha(x^k, w^k)) \geq 1, \quad \forall k > \tilde{K}.$$

Recalling the concavity of ϕ and $H_{k,\alpha} - H_{k+1,\alpha} = (H_{k,\alpha} - H_\alpha(x^*, w^*)) - (H_{k+1,\alpha} - H_\alpha(x^*, w^*))$, we get

$$\phi(H_{k,\alpha} - H_\alpha(x^*, w^*)) - \phi(H_{k+1,\alpha} - H_\alpha(x^*, w^*)) \geq \phi'(H_{k,\alpha} - H_\alpha(x^*, w^*))(H_{k,\alpha} - H_{k+1,\alpha}).$$

From Lemma 3.3, we know $\text{dist}(0, \partial H_\alpha(x^{k+1}, w^{k+1})) \leq \tilde{c} \|u^{k+1} - u^k\|$. Together with $\phi'(H_{k,\alpha} - H_\alpha(x^*, w^*)) > 0$, we obtain

$$\begin{aligned} & H_{k,\alpha} - H_{k+1,\alpha} \\ & \leq \frac{\phi(H_{k,\alpha} - H_\alpha(x^*, w^*)) - \phi(H_{k+1,\alpha} - H_\alpha(x^*, w^*))}{\phi'(H_{k,\alpha} - H_\alpha(x^*, w^*))} \\ & \leq \text{dist}(0, \partial H_\alpha(x^k, w^k)) [\phi(H_{k,\alpha} - H_\alpha(x^*, w^*)) - \phi(H_{k+1,\alpha} - H_\alpha(x^*, w^*))] \\ & \leq \tilde{c} \|u^k - u^{k-1}\| [\phi(H_{k,\alpha} - H_\alpha(x^*, w^*)) - \phi(H_{k+1,\alpha} - H_\alpha(x^*, w^*))], \quad \forall k > \tilde{K}. \end{aligned}$$

For notational convenience, we denote $\Delta_{k,k+1} = \phi(H_{k,\alpha} - H_\alpha(x^*, w^*)) - \phi(H_{k+1,\alpha} - H_\alpha(x^*, w^*))$. Combining Lemma 3.2 with the above inequality yields that, for all $k > \tilde{K}$,

$$\delta \|u^{k+1} - u^k\|^2 \leq \tilde{c} \|u^k - u^{k-1}\| \Delta_{k,k+1},$$

and hence

$$\|u^{k+1} - u^k\| \leq \sqrt{\|u^k - u^{k-1}\| \left(\frac{\tilde{c}}{\delta} \Delta_{k,k+1} \right)}.$$

Using the fact that $2\sqrt{ab} \leq a + b$ for any $a, b > 0$, we obtain

$$2\|u^{k+1} - u^k\| \leq \|u^k - u^{k-1}\| + \frac{\tilde{c}}{\delta} \Delta_{k,k+1}.$$

Summing up the equation above for $k = \tilde{K} + 1, \dots, m$ yields

$$2 \sum_{k=\tilde{K}+1}^m \|u^{k+1} - u^k\| \leq \sum_{k=\tilde{K}+1}^m \|u^k - u^{k-1}\| + \frac{\tilde{c}}{\delta} \Delta_{\tilde{K}+1,m+1}.$$

By rearranging the terms of the above inequality, we can write the above inequality as follows:

$$\begin{aligned} \sum_{k=\tilde{K}+1}^m \|u^{k+1} - u^k\| &\leq \|u^{\tilde{K}+1} - u^{\tilde{K}}\| - \|u^{m+1} - u^m\| \\ &\quad + \frac{\tilde{c}}{\delta} (\phi(H_{\tilde{K}+1,\alpha} - H_\alpha(x^*, w^*)) - \phi(H_{m+1,\alpha} - H_\alpha(x^*, w^*))). \end{aligned}$$

Notice that $\phi(H_{m+1,\alpha} - H_\alpha(x^*, w^*)) > 0$ and $\|u^{m+1} - u^m\| \geq 0$, we get

$$\sum_{k=\tilde{K}+1}^m \|u^{k+1} - u^k\| \leq \|u^{\tilde{K}+1} - u^{\tilde{K}}\| + \frac{\tilde{c}}{\delta} \phi(H_{\tilde{K}+1,\alpha} - H_\alpha(x^*, w^*)).$$

Letting $m \rightarrow +\infty$ in the above inequality, we obtain

$$\sum_{k=\tilde{K}+1}^{+\infty} \|u^{k+1} - u^k\| \leq \|u^{\tilde{K}+1} - u^{\tilde{K}}\| + \frac{\tilde{c}}{\delta} \phi(H_{\tilde{K}+1,\alpha} - H_\alpha(x^*, w^*)), \tag{17}$$

which implies that

$$\sum_{k=0}^{+\infty} \|u^{k+1} - u^k\| \leq +\infty.$$

Due to $u^k := (x^k, w^k), \forall k > 0$, we know that

$$\sum_{k=0}^{+\infty} \|x^{k+1} - x^k\| \leq +\infty \quad \text{and} \quad \sum_{k=0}^{+\infty} \|w^{k+1} - w^k\| \leq +\infty.$$

Therefore, $\{u^k := (x^k, w^k)\}$ is a Cauchy sequence and thus is convergent. The assertion then follows immediately from Lemma 3.4. □

We now give another main result about the convergence rate for Algorithm 1. Consider the KL property has been applied to analyzing local convergence rate of various first-order methods by many researchers [10, 23, 24].

Theorem 3.6 (Convergence rate) *Let $\{x^k\}$ be the sequence generated by Algorithm 1 and converging to x^* , and $\{w^k | w^k := x^{k-1}\}$. Suppose that H_α has the KL property at (x^*, w^*) with $\phi(s) = cs^{1-\theta}, \theta \in [0, 1), c > 0$. Then the following results hold:*

- (i) If $\theta = 0$, then the sequence $\{x^k, w^k\}$ converges finitely.
- (ii) If $\theta \in (0, \frac{1}{2}]$, then there exist $\mu > 0$ and $\tau \in [0, 1)$ such that

$$\|(x^k, w^k) - (x^*, w^*)\| \leq \mu \tau^k.$$

- (iii) If $\theta \in (\frac{1}{2}, 1)$, then there exists $\mu > 0$ such that

$$\|(x^k, w^k) - (x^*, w^*)\| \leq \mu k^{(\theta-1)/(2\theta-1)}.$$

Proof We first consider the case of $\theta = 0$. In this case, $\phi(s) = cs$ and $\phi'(s) = c$. If $\{(x^k, w^k)\}$ does not converge in a finite number of iterations, then the KL property at (x^*, w^*) yields, for any k sufficiently large, $c \cdot \text{dist}(0, \partial H_\alpha(x^k, w^k)) \geq 1$, a contradiction to Lemma 3.3.

Now, we consider the case of $\theta > 0$. Here we set $\Delta_k = \sum_{i=k}^{+\infty} \|u^{i+1} - u^i\|, k \geq 0$, then inequality (17) becomes

$$\Delta_{\tilde{k}+1} \leq (\Delta_{\tilde{k}} - \Delta_{\tilde{k}+1}) + \frac{\tilde{c}}{\delta} \phi(H_\alpha(x^{\tilde{k}+1}, w^{\tilde{k}+1}) - H_\alpha(x^*, w^*)). \tag{18}$$

Together with the KL property at (x^*, w^*) , we get

$$\phi'(H_\alpha(x^{\tilde{k}+1}, w^{\tilde{k}+1}) - H_\alpha(x^*, w^*)) \text{dist}(0, \partial H_\alpha(x^{\tilde{k}+1}, w^{\tilde{k}+1})) \geq 1,$$

which is equivalent to

$$(H_\alpha(x^{\tilde{k}+1}, w^{\tilde{k}+1}) - H_\alpha(x^*, w^*))^\theta \leq c(1 - \theta) \text{dist}(0, \partial H_\alpha(x^{\tilde{k}+1}, w^{\tilde{k}+1})). \tag{19}$$

Using Lemma 3.3, we deduce

$$\text{dist}(0, \partial H_\alpha(x^{\tilde{k}+1}, w^{\tilde{k}+1})) \leq \tilde{c} \|u^{\tilde{k}+1} - u^{\tilde{k}}\| = \tilde{c}(\Delta_{\tilde{k}} - \Delta_{\tilde{k}+1}). \tag{20}$$

Combining (19) and (20), we obtain that there exists $\gamma > 0$ such that

$$\begin{aligned} \phi(H_\alpha(x^{\tilde{k}+1}, w^{\tilde{k}+1}) - H_\alpha(x^*, w^*)) &= c(H_\alpha(x^{\tilde{k}+1}, w^{\tilde{k}+1}) - H_\alpha(x^*, w^*))^{1-\theta} \\ &\leq \gamma(\Delta_{\tilde{k}-1} - \Delta_{\tilde{k}+1})^{\frac{1-\theta}{\theta}}, \end{aligned}$$

where $\gamma = c^{\frac{1}{\theta}}(\tilde{c}(1 - \theta))^{\frac{1-\theta}{\theta}}$. Then inequality (18) becomes

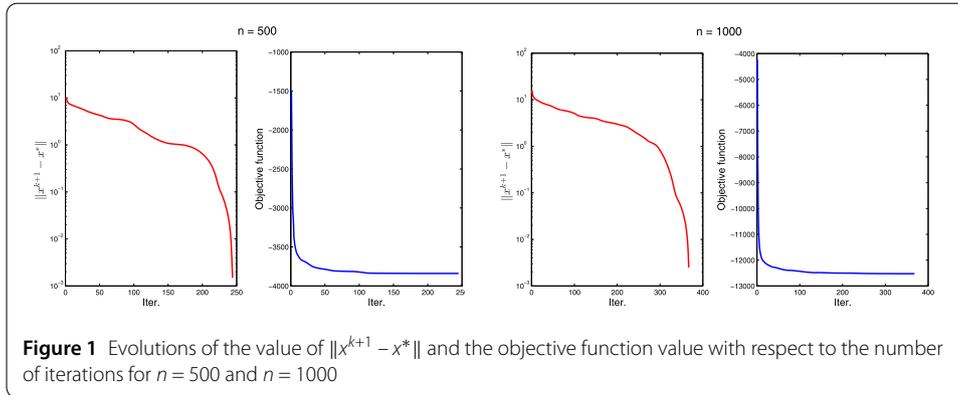
$$\Delta_{\tilde{k}+1} \leq (\Delta_{\tilde{k}} - \Delta_{\tilde{k}+1}) + \frac{\tilde{c}}{\delta} \gamma (\Delta_{\tilde{k}-1} - \Delta_{\tilde{k}+1})^{\frac{1-\theta}{\theta}}.$$

Sequences satisfying the above inequality have been considered in [9]. Therefore, as the proof in [9], it follows that if $\theta \in (0, \frac{1}{2}]$, there exist $\mu > 0$ and $\tau \in [0, 1)$ such that

$$\|(x^k, w^k) - (x^*, w^*)\| \leq \mu \tau^k,$$

and if $\theta \in (\frac{1}{2}, 1)$, there exist $\mu > 0$ and $\tau \in [0, 1)$ such that

$$\|(x^k, w^k) - (x^*, w^*)\| \leq \mu k^{(\theta-1)/(2\theta-1)}. \tag{□}$$



4 Numerical experiment

In this section, we apply the proposed method to solve a nonconvex optimization problem which arises in portfolio selection [25], neural network [26, 27], and compressed sensing [28]. Some preliminary numerical results are reported to demonstrate the feasibility and advantage of the method. All numerical experiments are performed in MATLAB 2014b on a 64-bit PC with an Intel Core i7-7500U CPU (2.70 GHz) and 16 GB of RAM.

We consider the following optimization problem:

$$\begin{aligned} \min & \frac{1}{2}x^T Ax - b^T x + \|x\|_1 \\ \text{s.t. } & x \in S, \end{aligned} \tag{21}$$

where $A \in \mathcal{R}^{n \times n}$ is a symmetric matrix that is not necessarily positive semidefinite, $b \in \mathcal{R}^n$ is a vector, S is a polyhedral set in \mathcal{R}^n . Problem (21) is obviously nonconvex. We also assume that the optimal value of (21) is finite and can be attained. Notice that one can write (21) equivalently as the following separable optimization problem:

$$\min f(x) + g(x) \tag{22}$$

by setting $f(x) := \frac{1}{2}x^T Ax - b^T x$ and $g(x) := \|x\|_1 + \delta(x|S)$. In this setting, f is a possibly nonconvex function and ∇f is Lipschitz continuous with modulus $L > 0$, g is a proper closed convex function, and $f + g$ is level bounded. The parameter $L = \max\{\lambda_{\max}(A), |\lambda_{\min}(A)|\}$, where $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ are the largest and smallest eigenvalues of A respectively. Furthermore, the objective function in (21) satisfies the KL property as discussed in [14]. Therefore, we can apply the inexact proximal gradient algorithm with extrapolation in Algorithm (1) to solve the equivalent model (22).

In the numerical experiments, the test data of problem (21) is generated by the following way. We set the symmetric matrix $A = D + D^T \in \mathcal{R}^{n \times n}$, where D is a matrix generated with i.i.d. standard Gaussian entries; the polyhedral set $S = [0, 1]^n$; the vector b generated with i.i.d. standard Gaussian entries. We initialize the $x^0 \in \text{dom } g$ randomly and set $x^{-1} = x^0$. Besides, the test method is terminated when $\|x^{k+1} - x^k\| \leq 10^{-4}$ and the maximal iteration number is set as 5000.

For this nonconvex problem, we apply the inexact proximal gradient algorithm with extrapolation to solve it. In Fig. 1, we first plot the evolution curves of the value $\|x^{k+1} - x^*\|$ and the objective function value with respect to the iteration number, where the sequence

Table 1 Iteration number and CPU computing time of Algorithm 1 with different dimensions

n	100	300	500	700	900	1000	1500
Iter.	57	110	245	317	357	368	564
Initer.	936	1110	2074	2243	2320	2391	2434
Time (s)	0.04	0.08	0.29	0.64	1.07	1.80	4.59

$\{x^k\}$ is generated by the proposed inexact algorithm and x^* denotes the approximate solution obtained at termination of the algorithm. From Fig. 1, we can see that the sequence $\|x^{k+1} - x^*\|$ converges to zero, and the objective function value decreases as the iteration increases, which conforms with our theory. Furthermore, the number of iterations, inner iterations (the iteration needed to calculate the subproblem in the algorithm inexactly), and the cpu time (in seconds) required by the algorithm are reported in Table 1 for different dimensions of the problem, and are denoted as “Iter.,” “Initer.,” and “Time (s)” respectively. The results also indicate the feasibility and effectiveness of the proposed method.

5 Conclusions

In this paper, we proposed an inexact proximal gradient algorithm with extrapolation for solving a class of nonconvex optimization problems. The convergence of this inexact algorithm was established under the assumption that an auxiliary function satisfies the KL property. We proved that the iterative sequence generated by the proposed method converges to a stationary point of the problem. Furthermore, the convergence rate result was obtained by the means of KL exponent.

Funding

This work was supported by the National Natural Science Foundation of China (Grant No. 11801279), the Natural Science Foundation of Jiangsu Province (Grant No. BK20180782), and the Startup Foundation for Introducing Talent of NUIST (Grant No. 2017r059).

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

All authors contributed to drafting this manuscript. All authors read and approved the final manuscript.

Author details

¹Department of Information and Computing Science, School of Mathematics and Statistics, Nanjing University of Information Science and Technology, Nanjing, P.R. China. ²School of Economics and Management, Southeast University, Nanjing, P.R. China. ³School of Mathematical Sciences, Nanjing Normal University, Nanjing, P.R. China.

Publisher's Note

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

Received: 15 March 2019 Accepted: 23 April 2019 Published online: 08 May 2019

References

- Candès, E.J., Tao, T.: Decoding by linear programming. *IEEE Trans. Inf. Theory* **51**, 4203–4215 (2005)
- Donoho, D.L.: Compressed sensing. *IEEE Trans. Inf. Theory* **52**, 1289–1306 (2006)
- Chambolle, A.: An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.* **20**, 89–97 (2004)
- Bauschke, H.H., Combettes, P.L.: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, New York (2011)
- Jia, Z.H., Cai, X.J.: A relaxation of the parameter in the forward-backward splitting method. *Pac. J. Optim.* **13**, 665–681 (2017)
- O'Donoghue, B., Candès, E.J.: Adaptive restart for accelerated gradient schemes. *Found. Comput. Math.* **15**, 715–732 (2015)
- Parikh, N., Boyd, S.: Proximal algorithms. *Found. Trends Optim.* **1**, 127–239 (2014)

8. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2**, 183–202 (2009)
9. Attouch, H., Bolte, J.: On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. *Math. Program.* **116**, 5–16 (2009)
10. 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**, 438–457 (2010)
11. Frankel, P., Garrigos, G., Peyrouquet, J.: Splitting methods with variable metric for Kurdyka–Lojasiewicz functions and general convergence rates. *J. Optim. Theory Appl.* **165**, 874–900 (2015)
12. Ochs, P., Chen, Y., Brox, T., Pock, T.: iPiano: inertial proximal algorithm for non-convex optimization. *SIAM J. Imaging Sci.* **7**, 1388–1419 (2014)
13. Wen, B., Chen, X.J., Pong, T.K.: Linear convergence of proximal gradient algorithm with extrapolation for a class of nonconvex nonsmooth minimization problems. *SIAM J. Optim.* **27**, 124–145 (2017)
14. Attouch, H., Bolte, J., Svaiter, B.F., Soubeyran, A.: Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss–Seidel methods. *Math. Program.* **137**, 91–129 (2013)
15. Wu, Z.M., Li, M.: General inertial proximal gradient method for a class of nonconvex nonsmooth optimization problems. *Comput. Optim. Appl.* (2019). <https://doi.org/10.1007/s10589-019-00073-1>
16. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.* **14**, 877–898 (1976)
17. Solodov, M.V., Svaiter, B.F.: A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Var. Anal.* **7**, 323–345 (1999)
18. Solodov, M.V., Svaiter, B.F.: A hybrid projection-proximal point algorithm. *J. Convex Anal.* **6**, 59–70 (1999)
19. Solodov, M.V., Svaiter, B.F.: An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions. *Math. Oper. Res.* **25**, 214–230 (2000)
20. Li, G.Y., Pong, T.K.: Calculus of the exponent of Kurdyka–Lojasiewicz inequality and its applications to linear convergence of first-order methods. *Found. Comput. Math.* **18**, 1199–1232 (2018)
21. Bolte, J., Sabach, S., Teboulle, M.: Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Math. Program.* **146**, 459–494 (2014)
22. Nesterov, Y.: *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer Academic Publishers, Boston (2004)
23. Li, G.Y., Pong, T.K.: Douglas–Rachford splitting for nonconvex optimization with application to nonconvex feasibility problems. *Math. Program.* **159**, 371–401 (2016)
24. Xu, Y.Y., Yin, W.T.: A block coordinate descent method for regularized multi-convex optimization with applications to nonnegative tensor factorization and completion. *SIAM J. Imaging Sci.* **6**, 1758–1789 (2013)
25. Markowitz, H.: Portfolio selection. *J. Finance* **7**, 77–91 (1952)
26. Liu, Q.S., Dang, C.Y., Cao, J.D.: A novel recurrent neural network with one neuron and finite-time convergence for k -winners-take-all operation. *IEEE Trans. Neural Netw.* **21**, 1140–1148 (2010)
27. Liu, Q.S., Cao, J.D., Chen, G.R.: A novel recurrent neural network with finite-time convergence for linear programming. *Neural Comput.* **22**, 2962–2978 (2010)
28. Chen, X., Peng, J.M., Zhang, S.Z.: Sparse solutions to random standard quadratic optimization problems. *Math. Program.* **141**, 273–293 (2013)

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](https://www.springeropen.com)
