Journal of Computational Mathematics
Vol.xx, No.x, 2026, 1–23.
http://www.global-sci.org/jcm
doi:10.4208/jcm.2512-m2024-0107

AN ACCELERATED PRECONDITIONED PRIMAL-DUAL GRADIENT ALGORITHM FOR NONCONVEX COMPOSITE OPTIMIZATION PROBLEMS WITH APPLICATIONS*

Xian-Jun Long[1)]
School of Mathematics and Statistics, Chongqing Technology and Business University, Chongqing 400067, P.R. China;
Chongqing Key Laboratory of Statistical Intelligent Computing and Monitoring, Chongqing Technology and Business University, Chongqing 400067, P.R. China
Email: xianjunlong@ctbu.edu.cn
Jia-Lin Nie and Gao-Xi Li
School of Mathematics and Statistics, Chongqing Technology and Business University, Chongqing 400067, P.R. China
Emails: niejialin00@163.com, ligaoxicn@126.com
Zai-Yun Peng
School of Mathematics, Yunnan Normal University, Kunming 650092, P.R. China
Email: pengzaiyun@126.com

Abstract

In this paper, we consider a class of three-composite nonconvex optimization problems, in which the nonsmooth function is further composed with a linear operator. This problem has many applications such as sparse signal recovery, image processing and machine learning. Based on the conjugate duality theory, we present an accelerated preconditioned primal-dual gradient algorithm for this problem. Compared with the existing algorithms, our algorithm only needs to calculate the proximal mapping of the conjugate function h^* which is always convex and lower semicontinuous and it does not need to calculate the proximal mapping of nonconvex functions. This may significantly reduce the computation load. We prove that the sequence generated by the proposed algorithm globally converges to a critical point when the function satisfies the Kurdyka-Łojasiewicz property. We also obtain the convergence rate of the proposed algorithm. Finally, numerical results on sparse signal recovery and image processing illustrate the efficiency and competitiveness of the proposed algorithm.

Mathematics subject classification: 90C26, 90C30, 90C33.
Key words: Nonconvex composite optimization, Primal-dual algorithm, Convergence, Sparse signal recovery, Image processing.

1. Introduction

Let X and Y be finite-dimensional vector spaces, f and $g : X \to \mathbb{R}$ be two continuously differentiable functions, which are not necessarily convex, $h : Y \to \mathbb{R} \cup \{+\infty\}$ be a simple and possibly nonsmooth, nonconvex function, and $A : X \to Y$ be a linear operator. In this paper, we are concerned with the following nonconvex composite optimization problem:

$$\min_{x\in X} f(x) + h(Ax) + g(x). \tag{1.1}$$

* Received June 3, 2024 / Revised version received December 4, 2024 / Accepted December 22, 2025 / Published online March 19, 2026 /
1) Corresponding author

This problem has received much attention by many authors due to its extensive applications in many fields such as sparse signal recovery, image processing and low-rank minimization. We now provide three concrete examples of problem (1.1) as following.

Example 1.1 ([42]). Consider an elastic net problem which can be characterized as the following optimization problem:

$$\min_{x\in R^q} \frac{1}{2}\|Cx-b\|^2 + \lambda_1\|x\|_1 + \frac{\lambda_2}{2}\|x\|_2^2, \tag{1.2}$$

where $C \in R^{p\times q}$ is a matrix where its columns are features and $b \in R^p$ is the measurement vector, $\lambda_1 > 0$ and $\lambda_2 > 0$ are regularization parameters. Problem (1.2) can be expressed in the form of problem (1.1) with $f(x) = \|Cx-b\|^2/2$, $h(Ax) = \lambda_1\|x\|_1$ ($A = I$ is the identity operator) and $g(x) = \lambda_2\|x\|_2^2/2$.

Example 1.2 (Image Processing, [24]). Consider a weighted difference of anisotropic and isotropic total variation model for image processing as follows:

$$\min_{x\in R^q} \frac{1}{2}\|Cx-b\|_p^p + \lambda\|\mathbb{D}x\|_0, \tag{1.3}$$

where $\mathbb{D}$ is the well-known discrete gradient operator, C is a linear operator, b is a degraded image with noise, $\lambda > 0$ is a regularization parameter and α is a weighted parameter. Generally, we may specify $p = 2$ for Gaussian noise and $p = 1$ for salt and pepper noise, respectively. Problem (1.3) can be expressed in the form of problem (1.1) with $f(x) = \|Cx-b\|^2/2$, $h(Ax) = \lambda\|\mathbb{D}x\|_0$ ($Ax = \mathbb{D}x$) and $g(x) = 0$.

When the functions f, g and h are convex, there are many primal-dual algorithms for solving model (1.1). These include the Condat-Vu algorithm [14, 33], the primal-dual three-operator algorithm [39], the primal-dual Davis-Yin algorithm [30], the primal-dual fixed point algorithm [19], the primal-dual Bregman algorithm [12,20], the primal-dual hybrid gradient algorithm [11] (when $g = 0$), the golden ratio primal-dual algorithm (when $g = 0$) [13], the Arrow-Hurwicz algorithm [32] (when $h = 0$), and the Davis-Yin splitting algorithm [15] (when $A = I$). Some relevant results can be found in [12, 16, 41].

When the functions f, g and h are nonconvex and $A = I$ (I is an identity operator), Bian and Zhang [6] proposed a three-operator splitting algorithm for solving model (1.1) and established the convergence and the convergence rate of the algorithm when the involved functions are semialgebraic. When $g = 0$, Li and Pong [22] proposed the alternating direction method of multiplier to solve problem (1.1), where f and h are possibly nonconvex. Guo *et al.* [17] proposed a preconditioned primal-dual gradient algorithm to solve problem (1.1) with $g = 0$. They also established the convergence and the convergence rate of the algorithm under the Kurdyka-Łojasiewicz property. When $h = 0$, many authors discussed this problem, we refer to [1, 3, 21, 27, 31, 34–38, 40] and the references therein.

However, to the best of our knowledge, there does not exist any results about the fully nonconvex problem (1.1) in the literature. So, motivated by the works of [6, 17], we study the fully nonconvex problem (1.1). Our contributions in this paper are two aspects:

(a) We present the accelerated preconditioned primal-dual gradient algorithm to solve the nonconvex problem (1.1), where all three functions are nonconvex. We prove that the sequence generated by the proposed algorithm globally converges to a critical point when the function satisfies the Kurdyka-Łojasiewicz property. We also obtain the convergence rate of the proposed algorithm. Compared with the results obtained in [14, 19, 33, 39], we do not need the assumption of the convexity of functions.

(b) Compared with the algorithms introduced in [6, 23], our algorithm does not require the calculation of proximal mapping of nonconvex functions at each iteration. Moreover, our algorithm adopts an extrapolation technique which accelerates the convergence speed of the algorithm. The numerical experiments show that our algorithm is efficient and converges faster than the algorithms given in [6, 23].

The rest of the paper is organized as follows. In Section 2, we recall some basic definitions and results. In Section 3, we propose an accelerated preconditioned primal-dual gradient algorithm for solving problem (1.1) and analyze the convergence and the convergence rate of the algorithm. Some numerical experiments are given in Section 4 to illustrate the performance of the algorithm.

2. Preliminaries

Throughout this paper, let X, Y be two finite-dimensional real vector spaces equipped with standard inner product $\langle \cdot\,, \cdot \rangle$ and norm $\|\cdot\|$. Let X^*, Y^* be the dual spaces of X, Y respectively. The operator norm of a linear operator $A : X \to Y$ is

$$\|A\| := \max\{\|Ax\| : x \in X,\ \|x\| \leq 1\}.$$

Let $g : X \to \mathbb{R} \cup \{+\infty\}$ be a mapping, the domain of g is defined by

$$\mathrm{dom} g := \{x \in X : g(x) < +\infty\}.$$

We said that g is proper if $\mathrm{dom} g \neq \emptyset$ and is lower semicontinuous if, for any $x \in X$, $g(x) \leq \liminf_{z \to x} g(z)$. The conjugate function of g, denoted by g^*, is defined by

$$g^*(y) := \sup_{x \in X}\{\langle y, x\rangle - g(x)\}, \quad y \in X^*.$$

From [28, Theorem 4.3], the conjugate function g^* is always convex and lower semicontinuous. Let g be a convex function, the subdifferential of g at $x \in \mathbb{X}$ is defined by

$$\partial g(x) := \{\xi \in X : g(y) \geq g(x) + \langle \xi, y - x\rangle,\ \forall\, y \in X\}.$$

Let $C \subseteq X$ be a closed set, $x \in X$, the distance of x to C is defined by

$$\mathrm{dist}(x, C) := \inf\{\|y - x\| : y \in C\}.$$

Definition 2.1. *A continuously differentiable function $f : X \to \mathbb{R}$ is said to be L_f-smooth over X if there exist a constant $L_f > 0$ such that for any $x, z \in X$, one has*

$$\|\nabla f(x) - \nabla f(z)\| \leq L_f \|x - z\|.$$

Definition 2.2 ([28]). *A set-valued mapping $F : X \rightrightarrows Y$ is said to satisfy the outer semicontinuous at x, if for any $u \in Y$, there exist $x_n \to x$ and $u_n \to u$ with $u_n \in F(x_n)$ such that $u \in F(x)$.*

Remark 2.1. From [28, Theorem 24.4], if $g : X \to \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous, proper and convex, the set-valued mapping ∂g is outer semicontinuous.

Definition 2.3 ([4]). *The extended proximal mapping of g associated to a positive definite linear operator M is defined by*

$$\mathrm{Prox}_g^M(y) := \arg\min_{x \in X} \left\{ g(x) + \frac{1}{2}\|x - y\|_M^2 \right\}.$$

Here, $\|x\|_M^2 := \langle Mx, x \rangle$.

We next recall the Kurdyka-Łojasiewicz (KL) property which plays a important role in convergence analysis. For $\eta \in (0, +\infty]$, we denote by Θ_η the class of concave and continuous functions $\varphi : [0, \eta) \to [0, +\infty)$ such that $\varphi(0) = 0, \varphi$ is continuously differentiable on $(0, \eta)$ and $\varphi^{'}(s) > 0$ for all $s \in (0, \eta)$.

Definition 2.4 (Kurdyka-Łojasiewicz Property and KL Function, [2]). *Let $f : X \to (-\infty, +\infty]$ be a proper and lower semicontinuous function. f is said to have Kurdyka-Łojasiewicz property at $x_0 \in \mathrm{dom}(\partial f)$ if there exist $\eta \in (0, +\infty]$, a neighborhood U of x_0 and a function $\varphi \in \Theta_\eta$ such that for all*

$$x \in U \cap \{x \in X : f(x_0) < f(x) < f(x_0) + \eta\},$$

the following KL inequality holds:

$$\varphi^{'}\big(f(x) - f(x_0)\big) \cdot \mathrm{dist}\big(0, \partial f(x)\big) \geq 1.$$

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

Lemma 2.1 (Uniformized KL Property, [7, Lemma 6]). *Let $C \subseteq X$ be a compact set and let $f : X \to (-\infty, +\infty]$ be a proper and lower semicontinuous function. Assume that f is constant on C and f satisfies the KL property at each point of C. Then, there exist $\varepsilon > 0$, $\eta > 0$ and $\varphi \in \Theta_\eta$ such that for all $x_0 \in C$ and for all*

$$\{x \in X : \mathrm{dist}(0, C) < \varepsilon\} \cap \{x \in X : f(x_0) < f(x) < f(x_0) + \eta\},$$

the following inequality holds:

$$\varphi^{'}\big(f(x) - f(x_0)\big) \cdot \mathrm{dist}\big(0, \partial f(x)\big) \geq 1.$$

We also need the following assumption.

Assumption 2.1. The functions f and g are L_f-smooth and L_g-smooth on X, respectively.

Remark 2.2. From Assumption 2.1, we can deduce the following:

$$f(x_{n+1}) \leq f(x_n) + \langle \nabla f(x_n), x_{n+1} - x_n \rangle + \frac{L_f}{2}\|x_{n+1} - x_n\|^2, \tag{2.1}$$

$$g(x_{n+1}) \leq g(x_n) + \langle \nabla g(x_n), x_{n+1} - x_n \rangle + \frac{L_g}{2}\|x_{n+1} - x_n\|^2. \tag{2.2}$$

3. A Preconditioned Primal-dual Gradient Algorithm

In this section, we present a preconditioned primal-dual gradient algorithm for solving problem (1.1). We analyze the convergence and the convergence rate of the proposed algorithm.

Let

$$L(x,y) := f(x) + g(x) + \langle y, Ax\rangle - h^*(y).$$

By the conjugate duality theory presented in [9, Section 2.5.3], we have that the dual problem of (1.1) is

$$\max_{y\in Y^*}\{\inf_{x\in X} L(x,y)\}. \tag{3.1}$$

From [9, Theorem 2.158], if $\inf_{x\in X} L(x,y) > -\infty$ for any $y \in Y^*$, then x_0 and y_0 are optimal solutions of (1.1) and (3.1), respectively, if and only if the following hold:

$$\begin{cases} x_0 \in \arg\min_{x\in X} L(x, y_0), \\ 0 = h(Ax_0) + h^*(y_0) - \langle y_0, Ax_0\rangle. \end{cases} \tag{3.2}$$

By the definition of conjugate function, if (3.2) is satisfied, we have $0 \in \partial L(x_0, y_0)$, i.e.,

$$\begin{cases} 0 = \nabla f(x_0) + \nabla g(x_0) + A^\top y_0, \\ 0 \in -\partial h^*(y_0) + Ax_0, \end{cases} \tag{3.3}$$

where $A^\top$ is the adjoint operator of A. Denote by crit$L$ the set of critical points of L, that is

$$\mathrm{crit}L := \{(x_0, y_0) \in X \times Y^* : 0 \in \partial L(x_0, y_0)\}.$$

First, we propose the following algorithm.

Algorithm 3.1: Accelerated Preconditioned Primal-dual Gradient Algorithm.

Initialization: Choose parameters $\tau > 0, \theta > 0$ and a positive definite matrix M, for given $x_1 \in X$, $y_1 := y_0 \in Y^*$. Set $n = 1$.

Step 1. Update x_{n+1}, y_{n+1} as follows:

$$\begin{aligned} x_{n+1} &= x_n - \tau\big(A^\top y_n + \nabla f(x_n) + \nabla g(x_n)\big), \\ z_n &= y_n + \theta(y_n - y_{n-1}), \\ y_{n+1} &= \mathrm{Prox}_{h^*}^M\big(z_n + M^{-1}A\big(2x_{n+1} - x_n + \tau\nabla g(x_n) - \tau\nabla g(x_{n+1})\big)\big). \end{aligned} \tag{3.4}$$

Step 2. Set $n := n + 1$ and go to Step 1.

Remark 3.1. (i) We remark that (3.4) can be regarded as a proximal gradient step coupled with the preconditioning technique introduced in [26] for the relation $0 \in -\partial h^*(y_0) + Ax_0$. It is easy to see that (3.4) is equivalent to

$$y_{n+1} = \arg\min_{y\in Y^*}\left\{h^*(y) - \big\langle y, A\big(2x_{n+1} - x_n + \tau\nabla g(x_n) - \tau\nabla g(x_{n+1})\big)\big\rangle + \frac{1}{2}\|y - z_n\|_M^2\right\}.$$

Moreover, by the definition of $\mathrm{Prox}_{h^*}^M$, there exists a vector $s_{n+1} \in \partial h^*(y_{n+1})$ such that

$$s_{n+1} = -M(y_{n+1} - z_n) + A\big(2x_{n+1} - x_n + \tau\nabla g(x_n) - \tau\nabla g(x_{n+1})\big). \tag{3.5}$$

If the sequence $\{(x_n, y_n)\}$ converges to (x_0, y_0), then by Assumption 2.1, (3.5) and the outer semicontinuity of ∂h^*, one has $Ax_0 \in \partial h^*(y_0)$.

(ii) If $g \equiv 0$ and $\theta \equiv 0$, then Algorithm 3.1 reduces to the algorithm PPDG considered in [17].

Remark 3.2. From Remark 3.1(i), we know that Algorithm 3.1 only need to calculate the proximal mapping of the conjugate function h^* rather than calculating the proximal mapping of nonconvex functions. And hence, the computation load may be significantly reduced.

Remark 3.3. From Algorithm 3.1 and Assumption 2.1, we have

$$\begin{aligned} \|A^\top(y_{n+1} - y_n)\| &= \Bigg\| \left(\frac{x_{n+1} - x_{n+2}}{\tau} - \nabla f(x_{n+1}) - \nabla g(x_{n+1}) \right) \\ &\quad - \left(\frac{x_n - x_{n+1}}{\tau} - \nabla f(x_n) - \nabla g(x_n) \right) \Bigg\| \\ &\le \frac{1}{\tau}\|x_{n+1} - x_{n+2}\| + \left(\frac{1}{\tau} + L_f + L_g \right) \|x_{n+1} - x_n\|. \end{aligned} \tag{3.6}$$

To prove the main results of this paper, we also need the following assumption.

Assumption 3.1. (i) The linear operator A is surjective.

(ii) $\inf_{x \in X} L(x, y) > -\infty$ for any $y \in Y^*$.

(iii) The convex hull of h is proper.

Remark 3.4. (i) The linear operator A is surjective if and only if the matrix associated with $AA^\top$ is positive definite. Thus, Assumption 3.1(i) implies that a natural choice of M in Algorithm 3.1 is $\tau AA^\top$. Particularly, if A is the identity operator, then we can choose $M = \tau I$ and hence the extended proximal mapping $\mathrm{Prox}_{h^*}^M(\cdot)$ reduces to the classical proximal mapping $\mathrm{Prox}_{h^*/\tau}(\cdot)$. Moreover, for any $y \in Y^*$, we have

$$\hat{\lambda}\|y\| \le \|A^\top y\|, \tag{3.7}$$

where $\hat{\lambda} := \sqrt{\lambda_{\min}(AA^\top)}$ and $\lambda_{\min}(AA^\top)$ denotes the smallest eigenvalue of $AA^\top$.

(ii) Assumption 3.1(ii) guarantees that the sequence generated by Algorithm 3.1 is well-defined, and it is also essential for the analysis of convergence (see Theorem 3.1).

(iii) By [5, Theorem 4.3], without any assumption on h, the conjugate function h^* is lower semicontinuous and convex. However, in order to ensure that h^* is proper, an additional assumption is required; see Assumption 3.1(iii). Obviously, h^* is proper if Assumption 3.1(iii) is satisfied; see [29, Theorem 11.1].

Define the following Lyapunov function:

$$\begin{aligned} \mathcal{L}(x, y, u, v, w) &:= L(x, y) - a\|x - u\|^2 + b\|x - v\|^2 \\ &\quad + c\|x - w\|^2, \quad \forall\, x, u, v, w \in X, \quad y \in Y^*. \end{aligned}$$

Let

$$\begin{aligned} &a := \frac{3\theta}{\tau}, \quad b := \frac{13}{20\tau} - (L_g + L_f) - \frac{3\theta}{2\tau} - \frac{\tau L_f^2}{4\theta} - \frac{3\tau\theta(L_f + L_g + 1/\tau)^2}{2}, \\ &c := \tau\theta\left(L_f + L_g + \frac{1}{\tau}\right)^2, \end{aligned} \tag{3.8}$$

where θ is defined in Algorithm 3.1 satisfying $a > 0, b > 0$ and $c > 0$. Let

$$\begin{aligned} d &:= b - \left(\frac{2}{5\tau} - 2(L_f + L_g) + \frac{\tau L_f^2}{2\theta} + \frac{\tau\theta}{2} \times \frac{2}{\tau^2} \right), \\ e &= \frac{2}{25\tau} - \frac{2(L_f + L_g)}{5} - \frac{\tau\theta}{2} \times \frac{2}{\tau^2} - c. \end{aligned} \tag{3.9}$$

By an elementary calculation, if we choose θ correctly, which is sufficient to guarantee $d > 0$ and $e > 0$. So, in this paper, we assume that $d \geq 0$ and $e \geq 0$.

Let S denote the set of cluster points of the sequence $\{(x_n, y_n)\}$ generated by Algorithm 3.1. We next prove the following lemmas.

Lemma 3.1. *Let $x, u, v, w \in X, y \in Y^*$. Then, $(x, y, u, v, w) \in \mathrm{crit}\mathcal{L}$ is equivalent to $(x, y) \in \mathrm{crit}L$ and $u = v = w = x$.*

Proof. From the definition of $\mathcal{L}$ and $(x, y, u, v, w) \in \mathrm{crit}\mathcal{L}$, we have

$$\begin{aligned} &0 = \nabla_x \mathcal{L}(x, y, u, v, w) = \nabla_x L(x, y) - 2a(x - u) + 2b(x - v) + 2c(x - w), \\ &0 \in \partial_y \mathcal{L}(x, y, u, v) = \partial_y L(x, y), \\ &0 = \nabla_u \mathcal{L}(x, y, u, v, w) = 2a(x - u), \\ &0 = \nabla_v \mathcal{L}(x, y, u, v, w) = 2b(v - x), \\ &0 = \nabla_w \mathcal{L}(x, y, u, v, w) = 2c(w - x). \end{aligned}$$

The latter three relations imply that $u = v = w = x$. Using the first two relations, we obtain $0 \in \partial \mathcal{L}(x, y)$. It follows that $(x, y) \in \mathrm{crit}L$. The converse is obvious. $\square$

Lemma 3.2. *Let Assumptions 2.1 and 3.1 hold. For all $n \geq 1$, we have*

$$\begin{aligned} L(x_{n+1}, y_{n+1}) \leq L(x_n, y_n) &- \left(\frac{1}{\tau} - \frac{L_f + L_g}{2} \right) \|x_{n+1} - x_n\|^2 \\ &+ \big\langle y_{n+1} - y_n, \tau A\big(\nabla f(x_{n-1}) - \nabla f(x_n)\big) \big\rangle \\ &- \theta \langle y_{n+1} - y_n, M(y_{n-1} - y_{n-2}) \rangle, \end{aligned}$$

where $M = \tau A A^\top$.

Proof. Combining (2.1) and (2.2) yields

$$\begin{aligned} &f(x_{n+1}) + g(x_{n+1}) \\ &\leq f(x_n) + g(x_n) + \langle \nabla f(x_n) + \nabla g(x_n), x_{n+1} - x_n \rangle + \frac{L_f + L_g}{2} \|x_{n+1} - x_n\|^2 \\ &= f(x_n) + g(x_n) + \frac{1}{\tau} \langle x_n - x_{n+1}, x_{n+1} - x_n \rangle - \langle y_n, A(x_{n+1} - x_n) \rangle + \frac{L_f + L_g}{2} \|x_{n+1} - x_n\|^2 \\ &= f(x_n) + g(x_n) - \langle y_n, A(x_{n+1} - x_n) \rangle - \left(\frac{1}{\tau} - \frac{L_f + L_g}{2} \right) \|x_{n+1} - x_n\|^2. \end{aligned}$$

By (3.5) and the convexity of h^*,

$$-h^*(y_{n+1}) \leq -h^*(y_n) + \big\langle y_n - y_{n+1}, A\big(2x_n - x_{n-1} + \tau \nabla g(x_{n-1}) - \tau \nabla g(x_n)\big) - M(y_n - z_{n-1}) \big\rangle.$$

Adding the term $\langle y_{n+1}, Ax_{n+1}\rangle$ on both sides of the above inequality, we get

$$\begin{aligned} L(x_{n+1}, y_{n+1}) \leq L(x_n, y_n) &- \left(\frac{1}{\tau} - \frac{L_f + L_g}{2}\right) \|x_{n+1} - x_n\|^2 \\ &+ \langle y_{n+1} - y_n, A(x_{n+1} - x_n + x_{n-1} - x_n)\rangle \\ &+ \big\langle y_{n+1} - y_n, \tau A\big(\nabla g(x_n) - \nabla g(x_{n-1})\big) \\ &\quad + M\big(y_n - y_{n-1} - \theta(y_{n-1} - y_{n-2})\big)\big\rangle. \end{aligned} \tag{3.10}$$

Note that $x_{n+1} = x_n - \tau(A^\top y_n + \nabla f(x_n) + \nabla g(x_n))$ and $M = \tau AA^\top$, we have

$$\begin{aligned} &\langle y_{n+1} - y_n, A(x_{n+1} - x_n + x_{n-1} - x_n)\rangle \\ &\quad + \big\langle y_{n+1} - y_n, \tau A\big(\nabla g(x_n) - \nabla g(x_{n-1})\big) + M\big(y_n - y_{n-1} - \theta(y_{n-1} - y_{n-2})\big)\big\rangle \\ &= \big\langle y_{n+1} - y_n, (\tau AA^\top - M)(y_{n-1} - y_n)\big\rangle + \big\langle y_{n+1} - y_n, \tau A\big(\nabla f(x_{n-1}) - \nabla f(x_n)\big)\big\rangle \\ &\quad - \theta\langle y_{n+1} - y_n, M(y_{n-1} - y_{n-2})\rangle. \end{aligned}$$

This together with (3.10) yields the conclusion. This completes the proof. □

Lemma 3.3. *Let Assumptions* 2.1 *and* 3.1 *hold, and* $\tau \leq 1/(5(L_f + L_g))$*. Then, for all* $n \geq 1$,

$$\begin{aligned} &\mathcal{L}(x_{n+1}, y_{n+1}, x_{n+2}, x_n, x_{n-1}) \\ &\quad + d\big(\|x_{n+1} - x_n\|^2 + \|x_n - x_{n-1}\|^2\big) + e\|x_{n+1} - x_{n-1}\|^2 \\ &\leq \mathcal{L}(x_n, y_n, x_{n+1}, x_{n-1}, x_{n-2}), \end{aligned} \tag{3.11}$$

where d *and* e *are defined in* (3.8).

Proof. From Lemma 3.2 and the Cauchy-Schwarz inequality, we have

$$\begin{aligned} L(x_{n+1}, y_{n+1}) &\leq L(x_n, y_n) - \left(\frac{1}{\tau} - \frac{L_f + L_g}{2}\right) \|x_{n+1} - x_n\|^2 \\ &\quad - \theta\langle y_{n+1} - y_n, M(y_{n-1} - y_{n-2})\rangle + \big\langle y_{n+1} - y_n, \tau A\big(\nabla f(x_{n-1}) - \nabla f(x_n)\big)\big\rangle \\ &\leq L(x_n, y_n) - \frac{9}{10\tau}\|x_{n+1} - x_n\|^2 - \left(\frac{1}{10\tau} - \frac{L_f + L_g}{2}\right) \|x_{n+1} - x_n\|^2 \\ &\quad + \frac{\tau\theta}{2}\big\|A^\top(y_{n+1} - y_n)\big\|^2 + \frac{\tau L_f^2}{2\theta}\|x_n - x_{n-1}\|^2 \\ &\quad - \theta\langle y_{n+1} - y_n, M(y_{n-1} - y_n)\rangle - \theta\langle y_{n+1} - y_n, M(y_n - y_{n-2})\rangle \\ &\leq L(x_n, y_n) - \frac{9}{10\tau}\|x_{n+1} - x_n\|^2 - \left(\frac{2}{25\tau} - \frac{2(L_f + L_g)}{5}\right) \|x_{n+1} - x_{n-1}\|^2 \\ &\quad + \left(\frac{2}{5\tau} - 2(L_f + L_g)\right) \|x_n - x_{n-1}\|^2 + \frac{\tau\theta}{2}\big\|A^\top(y_{n+1} - y_n)\big\|^2 \\ &\quad + \frac{\tau L_f^2}{2\theta}\|x_n - x_{n-1}\|^2 + \tau\theta\big\|A^\top(y_{n+1} - y_n)\big\|^2 \\ &\quad + \frac{\tau\theta}{2}\big\|A^\top(y_{n-1} - y_n)\big\|^2 + \frac{\tau\theta}{2}\big\|A^\top(y_{n-2} - y_n)\big\|^2, \end{aligned} \tag{3.12}$$

where the second inequality follows from the Lipschitz continuity of ∇f and the fact

$$\langle x, y\rangle \leq \frac{\theta}{2}\|x\|^2 + \frac{1}{2\theta}\|y\|^2,$$

the third inequality follows from the

$$-\alpha\|x+y\|^2 \le -\frac{4}{5}\alpha\|x\|^2 + 4\alpha\|y\|^2 \quad (\alpha > 0).$$

Combining (3.6) and (3.12) yields

$$\begin{aligned} L(x_{n+1}, y_{n+1}) \le L(x_n, y_n) &- \left(\frac{9}{10\tau} - \frac{3\tau\theta}{2} \times 2\left(\frac{1}{\tau} + L_f + L_g\right)^2 - \frac{\tau\theta}{2} \times \frac{2}{\tau^2}\right)\|x_{n+1} - x_n\|^2 \\ &+ \left(\frac{2}{5\tau} - 2(L_f + L_g) + \frac{\tau L_f^2}{2\theta} + \frac{\tau\theta}{2} \times \frac{2}{\tau^2}\right)\|x_n - x_{n-1}\|^2 \\ &+ \left(\frac{3\tau\theta}{2} \times \frac{2}{\tau^2}\right)\|x_{n+2} - x_{n+1}\|^2 \\ &- \left(\frac{2}{25\tau} - \frac{2(L_f + L_g)}{5} - \frac{\tau\theta}{2} \times \frac{2}{\tau^2}\right)\|x_{n+1} - x_{n-1}\|^2 \\ &+ \left(\frac{\tau\theta}{2} \times 2\left(\frac{1}{\tau} + L_f + L_g\right)^2\right)\|x_n - x_{n-2}\|^2. \end{aligned}$$

Using (3.8) and (3.9), we obtain

$$\begin{aligned} L(x_{n+1}, y_{n+1}) \le L(x_n, y_n) &- (a+b+d)\|x_{n+1} - x_n\|^2 + (b-d)\|x_n - x_{n-1}\|^2 \\ &+ a\|x_{n+2} - x_{n+1}\|^2 - (c+e)\|x_{n+1} - x_{n-1}\|^2 + c\|x_n - x_{n-2}\|^2, \end{aligned}$$

or equivalently,

$$\begin{aligned} &L(x_{n+1}, y_{n+1}) - a\|x_{n+2} - x_{n+1}\|^2 + b\|x_{n+1} - x_n\|^2 + c\|x_{n+1} - x_{n-1}\|^2 \\ &\quad + d\big(\|x_{n+1} - x_n\|^2 + \|x_n - x_{n-1}\|^2\big) + e\|x_{n+1} - x_{n-1}\|^2 \\ &\le L(x_n, y_n) - a\|x_{n+1} - x_n\|^2 + b\|x_n - x_{n-1}\|^2 + c\|x_n - x_{n-2}\|^2. \end{aligned}$$

It follows that (3.11) holds. This completes the proof. □

Denote

$$r_n := (x_n, y_n, x_{n+1}, x_{n-1}, x_{n-2}).$$

By Lemma 3.3, the sequence $\{\mathcal{L}(r_n)\}$ is nonincreasing. Let

$$d_n := \big(\nabla_x \mathcal{L}(r_n), Ax_n - s_n, \nabla_u \mathcal{L}(r_n), \nabla_v \mathcal{L}(r_n), \nabla_w \mathcal{L}(r_n)\big),$$

where

$$s_n = -M(y_n - z_{n-1}) + A\big(2x_n - x_{n-1} + \tau\nabla g(x_{n-1}) - \tau\nabla g(x_n)\big).$$

It follows from (3.5) that $s_n \in \partial h^*(y_n)$. And so $d_n \in \partial\mathcal{L}(r_n)$ by the definition of $\mathcal{L}$. In next lemma, we give the bound of d_n.

Lemma 3.4. *Let Assumptions* 2.1 *and* 3.1 *hold. Then*

$$\|d_n\| \le t_1\|x_n - x_{n-1}\| + t_2\|x_{n+1} - x_n\| + t_3\|x_n - x_{n-2}\|,$$

where

$$\begin{aligned} t_1 &:= 2(L_f + L_g) + 4b + \frac{2}{\tau} + \big((1+\theta)\big(2 + \tau(L_f + L_g)\big) + \tau L_g\big)\|A\|, \\ t_2 &:= 4a + \frac{1}{\tau} + (1+2\theta)\|A\|, \quad t_3 := \theta\big(1 + \tau(L_f + L_g)\big)\|A\| + 4c. \end{aligned}$$

Proof. From the definition of $\mathcal{L}$, we have

$$\begin{aligned}
\|\nabla_x \mathcal{L}(r_n)\| &= \big\|\nabla f(x_n) + \nabla g(x_n) + A^\top y_n - 2a(x_n - x_{n+1}) + 2b(x_n - x_{n-1}) - 2c(x_n - x_{n-2})\big\| \\
&= \big\|\nabla f(x_n) - \nabla f(x_{n-1}) + \nabla g(x_n) - \nabla g(x_{n-1}) + A^\top (y_n - y_{n-1}) + \nabla f(x_{n-1}) \\
&\quad + \nabla g(x_{n-1}) + A^\top y_{n-1} - 2a(x_n - x_{n+1}) + 2b(x_n - x_{n-1}) - 2c(x_n - x_{n-2})\big\| \\
&\le (L_f + L_g + 2b)\|x_n - x_{n-1}\| + 2a\|x_{n+1} - x_n\| + \big\|A^\top (y_n - y_{n-1})\big\| \\
&\quad + \big\|\nabla f(x_{n-1}) + \nabla g(x_{n-1}) + A^\top y_{n-1}\big\| + 2c\|x_n - x_{n-2}\|.
\end{aligned}$$

This together with (3.6) gives

$$\begin{aligned}
\|\nabla_x \mathcal{L}(r_n)\| \le &\left(2a + \frac{1}{\tau}\right)\|x_{n+1} - x_n\| + 2\left(L_f + L_g + b + \frac{1}{\tau}\right)\|x_n - x_{n-1}\| \\
&+ 2c\|x_n - x_{n-2}\|.
\end{aligned} \tag{3.13}$$

On the other hand, from (3.5), we obtain

$$\begin{aligned}
\|Ax_n - s_n\| &= \big\|M(y_n - z_{n-1}) - A(x_n - x_{n-1}) + \tau A\big(\nabla g(x_n) - \nabla g(x_{n-1})\big)\big\| \\
&= \big\|M\big(y_n - y_{n-1} - \theta(y_{n-1} - y_{n-2})\big) - A(x_n - x_{n-1}) + \tau A\big(\nabla g(x_n) - \nabla g(x_{n-1})\big)\big\| \\
&\le \tau\|A\|\big\|A^\top (y_n - y_{n-1})\big\| + \tau\theta\|A\|\big\|A^\top (y_{n-1} - y_n)\big\| + \tau\theta\|A\|\big\|A^\top (y_n - y_{n-2})\big\| \\
&\quad + \|A\|\|x_n - x_{n-1}\| + \tau\|A\|L_g\|x_n - x_{n-1}\|.
\end{aligned}$$

This fact together with (3.6) yields that

$$\begin{aligned}
\|Ax_n - s_n\| &= (1+\theta)\tau\|A\|\big\|A^\top (y_{n-1} - y_n)\big\| + \tau\theta\|A\|\big\|A^\top (y_n - y_{n-2})\big\| \\
&\quad + \|A\|\|x_n - x_{n-1}\| + \tau\|A\|L_g\|x_n - x_{n-1}\| \\
&\le (1+\theta)\tau\|A\|\left(\frac{1}{\tau}\|x_n - x_{n+1}\| + \left(\frac{1}{\tau} + L_f + L_g\right)\|x_n - x_{n-1}\|\right) \\
&\quad + \tau\theta\|A\|\left(\frac{1}{\tau}\|x_{n+1} - x_{n-1}\| + \left(\frac{1}{\tau} + L_f + L_g\right)\|x_n - x_{n-2}\|\right) \\
&\quad + \|A\|\|x_n - x_{n-1}\| + \tau\|A\|L_g\|x_n - x_{n-1}\| \\
&= \left((1+\theta)\tau\|A\|\left(\frac{1}{\tau} + L_f + L_g\right) + \|A\| + \tau\|A\|L_g\right)\|x_n - x_{n-1}\| \\
&\quad + \left((1+\theta)\tau\|A\| \times \frac{1}{\tau}\right)\|x_{n+1} - x_n\| \\
&\quad + \tau\theta\|A\|\left(\frac{1}{\tau} + L_f + L_g\right)\|x_n - x_{n-2}\| \\
&\quad + \left(\tau\theta\|A\| \times \frac{1}{\tau}\right)\|x_{n+1} - x_{n-1}\| \\
&= \big((1+\theta)\big(1 + \tau(L_f + L_g)\big) + 1 + \tau L_g\big)\|A\|\|x_n - x_{n-1}\| \\
&\quad + (1+\theta)\|A\|\|x_{n+1} - x_n\| \\
&\quad + \theta\big(1 + \tau(L_f + L_g)\big)\|A\|\|x_n - x_{n-2}\| \\
&\quad + \theta\|A\|\|x_{n+1} - x_n + x_n - x_{n-1}\| \\
&\le \big((1+\theta)\big(1 + \tau(L_f + L_g)\big) + 1 + \tau L_g + \theta\big)\|A\|\|x_n - x_{n-1}\|
\end{aligned}$$

$$
\begin{aligned}
&+ (1+\theta+\theta)\|A\|\|x_{n+1}-x_n\| \\
&+ \theta\big(1+\tau(L_f+L_g)\big)\|A\|\|x_n - x_{n-2}\| \\
=& \big((1+\theta)\big(2+\tau(L_f+L_g)\big)+\tau L_g\big)\|A\|\|x_n - x_{n-1}\| \\
&+ (1+2\theta)\|A\|\|x_{n+1}-x_n\| + \theta\big(1+\tau(L_f+L_g)\big)\|A\|\|x_n-x_{n-2}\|. \qquad (3.14)
\end{aligned}
$$

It is easy to see that

$$
\begin{aligned}
\|\nabla_u \mathcal{L}(r_n)\| &= 2a\|x_{n+1}-x_n\|, \\
\|\nabla_v \mathcal{L}(r_n)\| &= 2b\|x_n-x_{n-1}\|, \\
\|\nabla_w \mathcal{L}(r_n)\| &= 2c\|x_n-x_{n-2}\|.
\end{aligned}
$$

This together with (3.13) and (3.14) yields

$$
\begin{aligned}
\|d_n\| \le& \left(2(L_f+L_g)+4b+\frac{2}{\tau}+\big((1+\theta)\big(2+\tau(L_f+L_g)\big)+\tau L_g\big)\|A\|\right)\|x_n-x_{n-1}\| \\
&+\left(4a+\frac{1}{\tau}+(1+2\theta)\|A\|\right)\|x_{n+1}-x_n\| + \big(\theta\big(1+\tau(L_f+L_g)\big)\|A\|+4c\big)\|x_n-x_{n-2}\|.
\end{aligned}
$$

This completes the proof. □

By Lemmas 3.1-3.3, we get the following results.

Theorem 3.1. *Let Assumptions* 2.1 *and* 3.1 *hold, and the sequence* $\{(x_n, y_n)\}$ *be bounded. Then,*

(i) $\sum_{n=1}^{\infty}\|x_{n+1}-x_n\|^2<\infty$ *and* $\sum_{n=1}^{\infty}\|y_{n+1}-y_n\|^2<\infty$.

(ii) S *is a nonempty compact set and* $\lim_{n\to\infty}\operatorname{dist}((x_n,y_n),S)=0$.

(iii) $S\subseteq \operatorname{crit}L$.

(iv) L *is finite and constant on* S.

Proof. (i) By Assumption 3.1(ii), $\inf_n L(x_n,y_n)>-\infty$. Note that $\{x_n\}$ is bounded. Then, $\inf_n \mathcal{L}(r_n)>-\infty$. It follows from Lemma 3.2 that there exists a constant $\bar{\mathcal{L}}$ such that

$$
\lim_{n\to\infty}\mathcal{L}(r_n)=\bar{\mathcal{L}}. \qquad (3.15)
$$

Summing (3.11) over $n=1,\dots,m$, we get

$$
d\sum_{n=1}^{m}\left(\|x_{n+1}-x_n\|^2+\|x_n-x_{n-1}\|^2\right)+e\sum_{n=1}^{m}\|x_{n+1}-x_{n-1}\|^2\le \mathcal{L}(r_1)-\mathcal{L}(r_{m+1}).
$$

Let $m\to\infty$ and by the convergence of $\{\mathcal{L}(r_n)\}$,

$$
\sum_{n=1}^{\infty}\|x_{n+1}-x_n\|^2<\infty.
$$

Combining (3.7) and (3.6) yields

$$
\sum_{n=1}^{\infty}\|y_{n+1}-y_n\|^2<\infty.
$$

(ii) From (i), we obtain

$$\lim_{n\to\infty}\|x_{n+1}-x_n\|=0,\quad \lim_{n\to\infty}\|y_{n+1}-y_n\|=0. \tag{3.16}$$

Similar to the proof of [7, Lemma 5(iii)], we can conclude the compactness of S. Note that the sequence $\{(x_n,y_n)\}$ is bounded. Then we obtain that S is nonempty and for any $(x_0,y_0)\in S$, there exists a subsequence $\{(x_{n_k},y_{n_k})\}$ of $\{(x_n,y_n)\}$ such that

$$\lim_{k\to\infty}\|x_{n_k}-x_0\|=0,\quad \lim_{k\to\infty}\|y_{n_k}-y_0\|=0. \tag{3.17}$$

From the definition of the distance function, we obtain

$$\begin{aligned}\operatorname{dist}\big((x_n,y_n),S\big) &\le \|x_n-x_0\|+\|y_n-y_0\|\\ &\le \|x_n-x_{n_k}\|+\|x_{n_k}-x_0\|+\|y_n-y_{n_k}\|+\|y_{n_k}-y_0\|.\end{aligned}$$

This together with (3.16) and (3.17) gives that $\operatorname{dist}((x_n,y_n),S)$ converges to 0.

(iii) For any $(x_0,y_0)\in S$, we need to prove $(x_0,y_0)\in \operatorname{crit}L$. Let $r_0:=(x_0,y_0,x_0,x_0,x_0)$. We observe that $r_{n_k}\to r_0, d_{n_k}\in\partial\mathcal{L}(r_{n_k})$, by Lemma 3.4, $d_{n_k}\to 0$. Since $\partial\mathcal{L}$ is outer semicontinuous, $0\in\partial\mathcal{L}(r_0)$. And so $(x_0,y_0,x_0,x_0,x_0)\in\operatorname{crit}\mathcal{L}$. Therefore, $(x_0,y_0)\in\operatorname{crit}L$ from Lemma 3.1.

(iv) From Remark 3.4(iii), we know that the conjugate function h^* is proper lower semicontinuous and convex. By [5, Theorem 2.22], h^* is continuous over its domain $\operatorname{dom}h^*$. It follows that L is continuous on $X\times\operatorname{dom}h^*$. Therefore,

$$\lim_{k\to\infty}L(x_{n_k},y_{n_k})=L(x_0,y_0).$$

It follows that

$$\begin{aligned}\lim_{k\to\infty}\mathcal{L}(r_{n_k}) &= \lim_{k\to\infty}\big(L(x_{n_k},y_{n_k})-a\|x_{n_k}-x_{n_k+1}\|^2+b\|x_{n_k}-x_{n_k-1}\|^2+c\|x_{n_k}-x_{n_k-2}\|^2\big)\\ &= L(x_0,y_0)=\mathcal{L}(r_0).\end{aligned}$$

This together with (3.15) yields

$$L(x_0,y_0)=\bar{\mathcal{L}}. \tag{3.18}$$

By the arbitrariness of (x_0,y_0) in S, we can get the conclusion. □

Remark 3.5. The condition that the sequence $\{(x_n,y_n)\}$ is bounded, which is a standard assumption in the global convergence analysis of nonconvex optimization algorithms, see, for instance, [7, 8, 10].

The following theorem is the main convergence result for Algorithm 3.1 which requires the objective function satisfying the KL property.

Theorem 3.2. *Let Assumptions* 2.1 *and* 3.1 *hold. Suppose that $\mathcal{L}$ is a KL function and the sequence $\{(x_n,y_n)\}$ generated by Algorithm* 3.1 *is bounded. Then, $\{(x_n,y_n)\}$ converges to a critical point of L and*

$$\sum_{n=1}^{\infty}\|x_{n+1}-x_n\|<\infty,\quad \sum_{n=1}^{\infty}\|y_{n+1}-y_n\|<\infty.$$

Proof. By Theorem 3.1(iv), $\lim_{n\to\infty}\mathcal{L}(r_n)=\bar{\mathcal{L}}$, where $\bar{\mathcal{L}}$ is the constant value of L over S.

If there exists a number $l_0>0$ such that $\mathcal{L}(r_{l_0})=\bar{\mathcal{L}}$, then by Lemma 3.3, $\mathcal{L}(r_n)=\bar{\mathcal{L}}$ and $x_{n+1}=x_n$ for all $n\geq l_0$. From (3.6), we have $y_{n+1}=y_n$ for all $n\geq l_0$. It follows that $(x_n,y_n)=(x_{n+1},y_{n+1})$ for all $n\geq l_0$. This shows the claim.

Otherwise, since $\{\mathcal{L}(r_n)\}$ is nonincreasing by Lemma 3.3, thus $\mathcal{L}(r_n)>\bar{\mathcal{L}}$ for any $n>0$. Note that $\lim_{n\to\infty}\mathcal{L}(r_n)=\bar{\mathcal{L}}$. It follows that for any $\eta>0$, there exists an integer $l_1>0$ such that

$$\mathcal{L}(r_n)<\bar{\mathcal{L}}+\eta,\quad \forall\, n\geq l_1.$$

Let $\mathcal{D}$ be the set of cluster points of $\{r_n\}$. In the similar way to Theorem 3.1(ii) and 3.1(iv), we can get that the function $\mathcal{L}$ is constant on the nonempty compact set $\mathcal{D}$ and $\operatorname{dist}(r_n,\mathcal{D})\to 0$ as $n\to\infty$. This implies that for any $\varepsilon>0$, there exists $l_2>0$ such that

$$\operatorname{dist}(r_n,\mathcal{D})<\varepsilon,\quad n\geq l_2.$$

Let $N_0:=\max\{l_1,l_2\}$. By the above discussion, we have

$$r_n\in\{r:\operatorname{dist}(r,\mathcal{D})<\varepsilon\}\cap[\bar{\mathcal{L}}<\mathcal{L}<\bar{\mathcal{L}}+\eta],\quad \forall\, n\geq N_0.$$

Since $\mathcal{L}$ is a KL function, by Lemma 2.1, there exists a continuous concave function φ such that for all $n\geq N_0$,

$$\varphi'\big(\mathcal{L}(r_n)-\bar{\mathcal{L}}\big)\cdot\operatorname{dist}\big(0,\partial\mathcal{L}(r_n)\big)\geq 1. \tag{3.19}$$

By the concavity of φ,

$$\varphi\big(\mathcal{L}(r_{n+1})-\bar{\mathcal{L}}\big)\leq\varphi\big(\mathcal{L}(r_n)-\bar{\mathcal{L}}\big)+\varphi'\big(\mathcal{L}(r_n)-\bar{\mathcal{L}}\big)\cdot\big(\mathcal{L}(r_{n+1})-\mathcal{L}(r_n)\big). \tag{3.20}$$

Using Lemma 3.4, we obtain

$$\begin{aligned}\operatorname{dist}\big(0,\partial\mathcal{L}(r_n)\big)\leq\, & t_1\|x_n-x_{n-1}\|+t_2\|x_{n+1}-x_n\|+t_3\|x_n-x_{n-2}\|\\ \leq & (t_1+t_3)\|x_n-x_{n-1}\|+t_2\|x_{n+1}-x_n\|+t_3\|x_{n-1}-x_{n-2}\|\\ \leq & t(\|x_{n+1}-x_n\|+\|x_n-x_{n-1}\|+\|x_{n-1}-x_{n-2}\|),\end{aligned} \tag{3.21}$$

where $t:=\max\{t_2,(t_1+t_3)\}$. From (3.19)-(3.21), we have

$$\mathcal{H}_{m,n}:=\varphi\big(\mathcal{L}(r_m)-\bar{\mathcal{L}}\big)-\varphi\big(\mathcal{L}(r_n)-\bar{\mathcal{L}}\big)$$

satisfies

$$\begin{aligned}\mathcal{H}_{n,n+1}&\geq\varphi'\big(\mathcal{L}(r_n)-\bar{\mathcal{L}}\big)\cdot\big(\mathcal{L}(r_n)-\mathcal{L}(r_{n+1})\big)\geq\frac{\mathcal{L}(r_n)-\mathcal{L}(r_{n+1})}{\operatorname{dist}\big(0,\partial\mathcal{L}(r_n)\big)}\\ &\geq\frac{d\big(\|x_{n+1}-x_n\|^2+\|x_n-x_{n-1}\|^2\big)+e\|x_n-x_{n-1}\|^2}{t(\|x_{n+1}-x_n\|+\|x_n-x_{n-1}\|+\|x_{n-1}-x_{n-2}\|)}\\ &\geq\frac{d\big(\|x_{n+1}-x_n\|^2+\|x_n-x_{n-1}\|^2\big)}{t(\|x_{n+1}-x_n\|+\|x_n-x_{n-1}\|+\|x_{n-1}-x_{n-2}\|)}.\end{aligned}$$

It follows that

$$\begin{aligned}&\|x_{n+1}-x_n\|^2+\|x_n-x_{n-1}\|^2\\ \leq&\frac{t}{d}\mathcal{H}_{n,n+1}(\|x_{n+1}-x_n\|+\|x_n-x_{n-1}\|+\|x_{n-1}-x_{n-2}\|).\end{aligned}$$

This implies that

$$\begin{aligned}\|x_{n+1}-x_n\| &\leq \sqrt{\frac{t}{d}\mathcal{H}_{n,n+1}(\|x_{n+1}-x_n\|+\|x_n-x_{n-1}\|+\|x_{n-1}-x_{n-2}\|)}\\ &\leq \frac{3t}{2d}\mathcal{H}_{n,n+1}+\frac{1}{6}(\|x_{n+1}-x_n\|+\|x_n-x_{n-1}\|+\|x_{n-1}-x_{n-2}\|),\end{aligned}$$

or equivalently,

$$\frac{1}{2}\|x_{n+1}-x_n\| \leq \frac{3t}{2d}\mathcal{H}_{n,n+1}+\frac{1}{6}(\|x_n-x_{n-1}\|+\|x_{n-1}-x_{n-2}\|-2\|x_{n+1}-x_n\|).$$

Summing up from $n=N_0$ to m with $m>N_0$, we obtain

$$\begin{aligned}&\frac{1}{2}\sum_{n=N_0}^{m}\|x_{n+1}-x_n\|\\ \leq\ &\frac{3t}{2d}\mathcal{H}_{N_0,m+1}+\frac{1}{6}(\|x_{N_0}-x_{N_0-1}\|-\|x_{m+1}-x_m\|)\\ &\quad+\frac{1}{6}(\|x_{N_0-1}-x_{N_0-2}\|+\|x_{N_0}-x_{N_0-1}\|-\|x_m-x_{m-1}\|-\|x_{m+1}-x_m\|)\\ \leq\ &\frac{3t}{2d}\mathcal{H}_{N_0,m+1}+\frac{1}{3}\|x_{N_0}-x_{N_0-1}\|-\frac{1}{3}\|x_{m+1}-x_m\|\\ &\quad+\frac{1}{6}\|x_{N_0-1}-x_{N_0-2}\|-\frac{1}{6}\|x_m-x_{m-1}\|\\ \leq\ &\frac{3t}{2d}\mathcal{H}_{N_0,m+1}+\frac{1}{3}\|x_{N_0}-x_{N_0-1}\|+\frac{1}{6}\|x_{N_0}-x_{N_0-1}\|,\end{aligned}\tag{3.22}$$

and so

$$\begin{aligned}\sum_{n=N_0}^{m}\|x_{n+1}-x_n\| &\leq \frac{3t}{d}\mathcal{H}_{N_0,m+1}+\frac{2}{3}\|x_{N_0}-x_{N_0-1}\|+\frac{1}{3}\|x_{N_0}-x_{N_0-1}\|\\ &\leq \frac{3t}{d}\varphi\big(\mathcal{L}(r_{N_0})-\bar{\mathcal{L}}\big)+\frac{2}{3}\|x_{N_0}-x_{N_0-1}\|+\frac{1}{3}\|x_{N_0}-x_{N_0-1}\|,\end{aligned}\tag{3.23}$$

where the second inequality follows from the fact that $\varphi>0$ on $(0,\eta)$. Similarly, we can derive

$$\begin{aligned}\sum_{n=N_0}^{m}\|x_n-x_{n-1}\| &= \sum_{n=N_0}^{m}\|x_{n+1}-x_n\|+\|x_{N_0}-x_{N_0-1}\|\\ &\leq \frac{3t}{d}\varphi\big(\mathcal{L}(r_{N_0})-\bar{\mathcal{L}}\big)+\frac{5}{3}\|x_{N_0}-x_{N_0-1}\|+\frac{1}{3}\|x_{N_0-1}-x_{N_0-2}\|,\end{aligned}\tag{3.24}$$

$$\begin{aligned}\sum_{n=N_0}^{m}\|x_{n-1}-x_{n-2}\| &\leq \sum_{n=N_0}^{m}\|x_n-x_{n-1}\|+\|x_{N_0-1}-x_{N_0-2}\|\\ &\leq \frac{3t}{d}\varphi\big(\mathcal{L}(r_{N_0})-\bar{\mathcal{L}}\big)+\frac{5}{3}\|x_{N_0}-x_{N_0-1}\|+\frac{4}{3}\|x_{N_0-1}-x_{N_0-2}\|,\end{aligned}\tag{3.25}$$

Let $m\to\infty$ in (3.23), using the first term of (3.16), we obtain

$$\sum_{n=N_0}^{\infty}\|x_{n+1}-x_n\|<\infty.$$

This together with (3.6) yields

$$\sum_{n=N_0}^{\infty}\|y_{n+1}-y_n\|<\infty.$$

It follows that $\{(x_n, y_n)\}$ is a Cauchy sequence by the same line of analysis as [5, Theorem 1(ii)]. Therefore, the sequence $\{(x_n, y_n)\}$ converges to a limit (x_0, y_0) that is a critical point of L by Theorem 3.1(iii). This completes the proof. $\square$

The following result illustrates the convergence rate of Algorithm 3.1.

Theorem 3.3. *Let Assumptions* 2.1 *and* 3.1 *hold. Assume that the sequence* $\{(x_n, y_n)\}$ *is bounded and* $\mathcal{L}$ *is a KL function with* $\varphi(s) = \sigma s^{1-\alpha}$ *for some* $\sigma > 0$ *and* $\theta \in [0, 1)$*. Let* (x_0, y_0) *be the limit of* $\{(x_n, y_n)\}$*, then the following assertions hold:*

(i) If $\alpha = 0$*, then the sequence* $\{(x_n, y_n)\}$ *converges in finite steps.*

(ii) If $\alpha \in (0, 1/2]$*, then there exist constants* $\mu > 0, 0 < \epsilon < 1$ *and a positive integer* N *such that*

$$\|x_n - x_0\| \le \mu\epsilon^{n-N}, \quad \|y_n - y_0\| \le \mu'\epsilon^{n-N}, \quad \forall\, n \ge N,$$

where $\mu' := \mu(1/\tau + L_f + L_g)/\hat{\lambda}$ *and* $\hat{\lambda}$ *is given in* (3.7)*.*

(iii) If $\alpha \in (1/2, 1)$*, then there exist a constant* $\nu > 0$ *and a positive integer* N' *such that*

$$\|x_n - x_0\| \le \nu n^{-\frac{1-\alpha}{2\alpha-1}}, \quad \|y_n - y_0\| \le \nu' n^{-\frac{1-\alpha}{2\alpha-1}}, \quad \forall\, n \ge N',$$

where $\nu' := \nu(1/\tau + L_f + L_g)/\hat{\lambda}$ *and* $\hat{\lambda}$ *is given* (3.7)*.*

Proof. (i) Let $\alpha = 0$ and $N_1 := \max\{n : x_{n+1} \ne x_n\}$. We claim that N_1 is a finite number. Suppose by contradiction that N_1 is sufficiently large such that (3.19) holds for all $n \ge N_1$. Since $\varphi(s) = \sigma s$, by (3.19) and (3.21),

$$t(\|x_{n+1} - x_n\| + \|x_n - x_{n-1}\| + \|x_{n-1} - x_{n-2}\|) \ge \operatorname{dist}\big(0, \partial\mathcal{L}(r_n)\big) \ge \frac{1}{\sigma}, \quad \forall\, n \ge N_1,$$

which

$$\begin{aligned}
\mathcal{L}(r_{n+1}) &\le \mathcal{L}(r_n) - d\big(\|x_{n+1} - x_n\|^2 + \|x_n - x_{n-1}\|^2\big) - e\|x_{n+1} - x_{n-1}\|^2, \\
\mathcal{L}(r_n) &\le \mathcal{L}(r_{n-1}) - d\big(\|x_n - x_{n-1}\|^2 + \|x_{n-1} - x_{n-2}\|^2\big) - e\|x_n - x_{n-2}\|^2.
\end{aligned}$$

Adding the above two equations, together with Lemma 3.3 and using the fact

$$a^2 + b^2 + c^2 \ge \frac{(a+b+c)^2}{3}$$

and Lemma 3.3, we have

$$\begin{aligned}
\mathcal{L}(r_{n+1}) &\le \mathcal{L}(r_{n-1}) - d\|x_{n+1} - x_n\|^2 - 2d\|x_n - x_{n-1}\|^2 - d\|x_{n-1} - x_{n-2}\|^2 \\
&\le \mathcal{L}(r_{n-1}) - d\big(\|x_{n+1} - x_n\|^2 + \|x_n - x_{n-1}\|^2 + \|x_{n-1} - x_{n-2}\|^2\big) \\
&\le \mathcal{L}(r_{n-1}) - \frac{d}{3t^2\sigma^2}.
\end{aligned}$$

Taking the limit $n \to \infty$ in the previous inequality and using (3.15) and (3.18), we obtain

$$L(x_0, y_0) \le L(x_0, y_0) - \frac{d}{3t^2\sigma^2},$$

which gives a contradiction. Thus, N_1 is a finite number and $\{x_n\}$ converges in finite steps. Combining (3.6) and (3.7) yields

$$\begin{aligned}\|y_{n+1}-y_n\| &\le \frac{1}{\tau\hat{\lambda}}\|x_{n+1}-x_{n+2}\| + \frac{1/\tau+L_f+L_g}{\hat{\lambda}}\|x_{n+1}-x_n\| \\ &\le \frac{1/\tau+L_f+L_g}{\hat{\lambda}}(\|x_{n+1}-x_{n+2}\|+\|x_{n+1}-x_n\|) \\ &\le \frac{1/\tau+L_f+L_g}{\hat{\lambda}}(\|x_{n+1}-x_{n+2}\|+\|x_{n+1}-x_n\|+\|x_n-x_{n-1}\|).\end{aligned}$$

This implies that $\{y_n\}$ also converges in finite steps. And so the conclusion holds.

(ii) Let

$$\triangle_n := \sum_{k=n}^{\infty}(\|x_{k+1}-x_k\|+\|x_k-x_{k-1}\|+\|x_{k-1}-x_{k-2}\|).$$

From (3.23)-(3.25), we have that $\triangle_n < \infty$ for any $n \ge 1$ and $\{(x_n, y_n)\}$ converges to (x_0, y_0), where (x_0, y_0) is a critical point of L. It follows that $\|x_n - x_0\| \le \triangle_n$ and

$$\|y_n-y_0\| \le \sum_{k=n}^{\infty}\|y_{k+1}-y_k\| \le \frac{1/\tau+L_f+L_g}{\hat{\lambda}}\triangle_n.$$

Now, we estimate the term of $\triangle_n$. If $\triangle_n = 0$ for some n, then $\|x_{k+1} - x_k\| = 0$ for $k \ge n$ and $\{(x_n, y_n)\}$ converges in finite steps. Hence, without loss of generality, we may assume that $\triangle_n > 0$ for any $n \ge 1$.

For $\alpha \in (0, 1)$, since $\varphi(s) = \sigma s^{(1-\alpha)}$, taking $m \to \infty$ in (3.23) and by (3.19), for any $n \ge N_0$,

$$\begin{aligned}\triangle_{n+1} \le \triangle_n &\le \frac{9t\sigma}{d}\big(\mathcal{L}(r_n)-L(x_0,y_0)\big)^{1-\alpha} + 4\|x_n-x_{n-1}\| + 2\|x_{n-1}-x_{n-2}\| \\ &\le \frac{9t\sigma^{\frac{1}{\alpha}}}{d}\big((1-\alpha)\mathrm{dist}\big(0,\partial\mathcal{L}(r_n)\big)\big)^{\frac{1-\alpha}{\alpha}} + 4\|x_n-x_{n-1}\| + 2\|x_n-x_{n-2}\|.\end{aligned}$$

By the definition of $\triangle_n$ and (3.21),

$$\begin{aligned}\triangle_{n+1} &\le \frac{9t\sigma^{1/\alpha}}{d}[t(1-\alpha)(\triangle_n-\triangle_{n+1})]^{\frac{1-\alpha}{\alpha}} + 4(\triangle_n-\triangle_{n+1}) \\ &= t'(\triangle_n-\triangle_{n+1})^{\frac{1-\alpha}{\alpha}} + 4(\triangle_n-\triangle_{n+1}),\end{aligned} \tag{3.26}$$

where

$$t' := \frac{9}{d}(t\alpha)^{\frac{1}{\alpha}}(1-\alpha)^{\frac{1-\alpha}{\alpha}}.$$

Let $\alpha \in (0, 1/2]$. Since $0 < \triangle_n - \triangle_{n+1} < 1$ when $n \ge N$ and $N \ge N_0$ is large enough, from (3.26) and $(1-\alpha)/\alpha \ge 1$, we have

$$\triangle_{n+1} \le (t'+4)(\triangle_n-\triangle_{n+1}).$$

Let $\epsilon := (t'+4)/(t'+5)$. It is easy to see that $\epsilon < 1$. The above inequality implies that

$$\triangle_{n+1} \le \epsilon\triangle_n.$$

It follows that for any $n \ge N$,

$$\triangle_n \le \mu\epsilon^{n-N},$$

where $\mu := \triangle_N$ is a finite number. This shows the claim.

(iii) Let $\alpha \in (1/2, 1)$. Let $N' \geq N_0$ be large enough such that $0 < \triangle_n - \triangle_{n+1} < 1$ for all $n \geq N'$. By the fact $0 < (1-\alpha)/\alpha < 1$ and (3.26), for all $n \geq N'$,

$$\triangle_{n+1} \leq (t'+4)(\triangle_n - \triangle_{n+1})^{\frac{1-\alpha}{\alpha}},$$

or equivalently,

$$\triangle_{n+1}^{\frac{\alpha}{1-\alpha}} \leq (t'+4)^{\frac{\alpha}{1-\alpha}}(\triangle_n - \triangle_{n+1}).$$

Define $H : (0, +\infty) \to \mathbb{R}$ by $H(s) = s^{-\alpha/(1-\alpha)}$ and let $Q \in (1, +\infty)$. Taking $n \geq N'$ such that $H(\triangle_{n+1}) \leq QH(\triangle_n)$. By rewriting the above equation as

$$1 \leq \frac{C_1(\triangle_n - \triangle_{n+1})}{\triangle_{n+1}^{\alpha/(1-\alpha)}},$$

where $C_1 = (t'+4)^{\alpha/(1-\alpha)}$. It follows that

$$\begin{aligned} 1 &\leq C_1(\triangle_n - \triangle_{n+1})H(\triangle_{n+1}) \leq QC_1(\triangle_n - \triangle_{n+1})H(\triangle_n) \\ &\leq QC_1 \int_{\triangle_{n+1}}^{\triangle_n} H(s)ds \leq QC_1 \frac{1-\alpha}{1-2\alpha}\Big[\triangle_n^{\frac{1-2\alpha}{1-\alpha}} - \triangle_{n+1}^{\frac{1-2\alpha}{1-\alpha}}\Big]. \end{aligned}$$

Let $\nu_1 = (2\alpha - 1)/((1-\alpha)QC_1)$ and $\rho := (1-2\alpha)/(1-\alpha)$. Obviously, $\nu_1 > 0$ and $\rho < 0$. The above inequality gives that

$$(\triangle_{n+1})^\rho - (\triangle_n)^\rho \geq \nu_1.$$

Summing up from $n \geq N'$ to m for any $m \geq N'$, we have

$$(\triangle_m)^\rho \geq (m - N')\nu_1 + (\triangle_{N'})^\rho.$$

From $\rho < 0$, it follows that there exists positive constant ν such that

$$\triangle_m \leq \big[(m-N')\nu_1 + (\triangle_{N'})^\rho\big]^{\frac{1}{\rho}} \leq \nu m^{\frac{1}{\rho}}.$$

This shows the claim. This completes the proof. □

4. Numerical Experiments

In this section, we give the applications of Algorithm 3.1 to image processing and sparse signal recovery. We focus on All numerical experiments are carried out using MATLAB R2018a on a PC Intel(R) Core(TM) i5-8250U CPU (1.60 GHz).

In all experiments, to apply Algorithm 3.1, we should calculate

$$y_{n+1} = \mathrm{Prox}_{h^*}^M\big(z_n + M^{-1}A\big(2x_{n+1} - x_n + \tau\nabla g(x_n) - \tau\nabla g(x_{n+1})\big)\big)$$

with $M = \tau AA^\top$ at each iteration. To avoid computing the inverse of M, in practice we calculate the following term as an approximation:

$$y_{n+1} = \mathrm{Prox}_{\beta h^*}\big(z_n + \beta A\big(2x_{n+1} - x_n + \tau\nabla g(x_n) - \tau\nabla g(x_{n+1})\big)\big),$$

where $\beta = 1/(\tau\|A\|^2)$.

The stopping condition of all algorithms is

$$\frac{\|x_{n+1} - x_n\|}{\max\{\|x_{n+1}\|, 1\}} < err \quad \text{or} \quad n \leq maxiter,$$

where *err* is a given error and *maxiter* is a given max iteration.

4.1. A nonconvex image processing problem

In this subsection, we apply Algorithm 3.1 to solve the following image processing problem:

$$\min_{x\in\Gamma} \frac{1}{2}\|x-b\|_F^2 + \lambda\|\nabla x\|_0, \tag{4.1}$$

where $b \in R^{q\times p}$ represents the noisy input image, x is the result after denoising, $\lambda > 0$ is a regularization parameter, and $\Gamma = \{x \in R^{q\times p} : c_1 \leq (\nabla x)_{i,j} \leq c_2\}$ for two given constants c_1, c_2. Problem (4.1) can be expressed in the form of (1.1) with $f(x) = \|x-b\|_F^2/2$, $h(z) = \lambda\|z\|_0 + \mathcal{I}_D(z)$ and $g = 0$, where $D = \{z : c_1 \leq z_{i,j} \leq c_2\}$, $\mathcal{I}_D(\cdot)$ is the indicator function of D, and A is the linear operator associated with ∇x such that $Ax = \nabla x$.

In this experiment, we illustrate the performance of Algorithm 3.1 by comparing with the algorithms PPDG [17] and ADMM [22]. The following signal-to-noise ratio (SNR) is used as a measure of the quality of the denoised image:

$$\mathrm{SNR} := 20\log_{10}\frac{\|x^*\|}{\|x_{n+1}-x^*\|},$$

where x^* is the original image without any noisy, x_{n+1} is the output image.

In the following, we test three images named "Boat", "Man" and "Lena". The test images are added the Gaussian white noise with a mean 0 and a standard deviation 0.01. Taking $c_1 = -1, c_2 = 1$, $\lambda = 0.1, err = 10^{-5}$ and $maxiter = 200$. The original clean images, the input noisy images and the denoising images recovered by Algorithm 3.1, PPDG and ADMM, respectively, are shown in Fig. 4.1.

Table 4.1 reports the number of iterations, CPU time and the SNR value of Algorithm 3.1, PPDG and ADMM for all test problems. From Table 4.1, we conclude that these algorithms can deal with problem (4.1) effectively. It is easy to see that Algorithm 3.1 has significant advantages over PPDG and ADMM in terms of taking less iterations and computing time in seconds (time for short) for all test problems. We can observe that, from Fig. 4.1, Algorithm 3.1 outperforms ADMM and PPDG in terms of denoising capability, and from Table 4.1, Algorithm 3.1 is superior to ADMM and PPDG in view of running time and SNR.

4.2. A compressive sensing problem

In this subsection, we will investigate the numerical performance of Algorithm 3.1 by solving a compressive sensing problem (1.2).

In our numerical experiments, for given (p, q, s), we generate $C_{p\times q}$ and b as below: first, each entry of C is drawn from a standard Gaussian, and then each column of C is normalized; second, let $b = Cx^* + 0.01\widehat{p}$, where $\widehat{p} \in R^p$ is a random vector with Gaussian entries and $x^* = \{(\widehat{x}, 0, \ldots, 0)\} \in R^q$ $(q > s)$ in which $\widehat{x} \in R^s$ is a standard Gaussian vector.

Table 4.1: Comparison between Algorithm 3.1, PPDG and ADMM for test problems.

	Image "Boat"			Image "Man"			Image "Lena"		
	Iter	time	SNR	Iter	time	SNR	Iter	time	SNR
Algorithm 3.1	93	32.5625	39.6524	95	33.1406	40.0870	94	32.7813	40.9717
PPDG	168	67.5781	39.6503	169	68.9531	40.0861	169	66.1094	40.9550
ADMM	200	102.4688	32.8312	200	102.7656	33.0176	200	103.2031	32.0099

(a) Original image

(b) Original image

(c) Original image

(d) Noisy image

(e) Noisy image

(f) Noisy image

(g) Algorithm 3.1

(h) Algorithm 3.1

(i) Algorithm 3.1

(j) PPDG

(k) PPDG

(l) PPDG

(m) ADMM

(n) ADMM

(o) ADMM

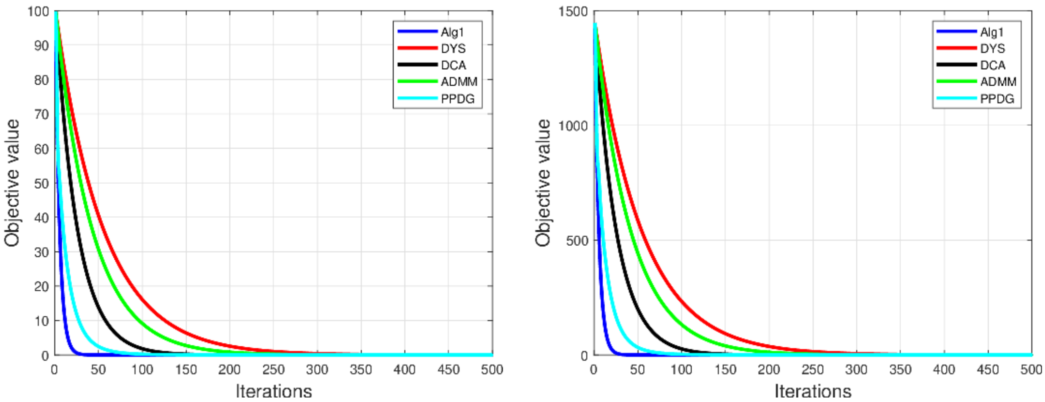
Fig. 4.1. The first row contains the original images; the second row represents the Gaussian noised images; the third, fourth and fifth rows are the denoised images by Algorithm 3.1, PPDG and ADMM, respectively.

Let $(p, q, s) = (720i, 2880i, 90i)$ for $i = 1/4, 1/2, 1, 2, 3, 4$. We generate 10 instances randomly as described above and report the number of iterations (Iter), CPU times in seconds (Time) and the relative error ($rerr$), where

$$rerr := \frac{\|\hat{x} - x^*\|}{\|x^*\|},$$

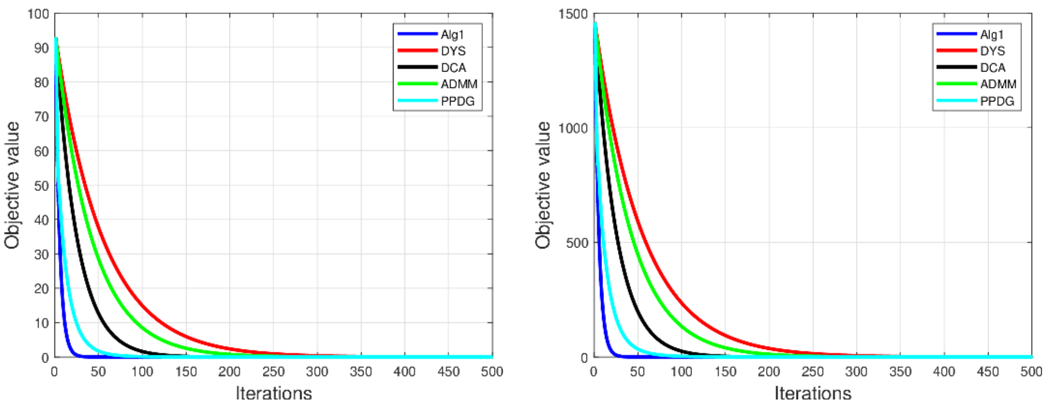
where $\hat{x}$ is the recovered sparse solution by algorithms. We compare Algorithm 3.1 with algorithms ADMM [23], DCA [25], DYS [6] and PPDG [17]. The numerical results are presented in Tables 4.2 and 4.3, and Figs. 4.2 and 4.3, where $\lambda_1 = 1 \times 10^{-6}$ and $\lambda_1 = 1 \times 10^{-5}$ respectively. All parameters of these compared algorithms follow the default settings as used in the original papers.

It can be seen from Tables 4.2 and 4.3 that the size has no influence on the number of iterations and the relative error of algorithms. It is easy to see that Algorithm 3.1 is superior to algorithms ADMM [23], DCA [25], DYS [6] and PPDG [17] in Iter, time and $rerr$, respectively. Moreover, Figs. 4.2 and 4.3 also show that by running a same number of iterations, Algorithm 3.1 outperforms all other algorithms.



(a) $p = 180, q = 720$ (b) $p = 2880, q = 11520$

Fig. 4.2. Objective value with $\lambda_1 = 10^{-6}$ under different dimentions.



(a) $p = 180, q = 720$ (b) $p = 2880, q = 11520$

Fig. 4.3. Objective value with $\lambda_1 = 10^{-5}$ under different dimentions.

Table 4.2: Comparison of four Algorithms with $\lambda_1 = 10^{-6}$.

p	q	s	Algorithm 3.1			DYS			DCA		
			Iter	Time	$rerr$	Iter	Time	$rerr$	Iter	Time	$rerr$
180	720	20	161	0.0356	1.9244e-04	427	0.0409	71.8169e-04	401	0.0449	0.0012
360	1440	40	163	0.0996	6.4215e-05	443	0.1256	4.6097e-04	416	0.1359	3.1493e-05
720	2880	80	162	0.5756	6.0756e-06	432	1.0367	5.7965e-05	405	0.9971	3.2699e-05
1440	5760	160	162	2.3314	3.4608e-06	431	4.3895	3.6912e-05	405	4.1318	2.3416e-05
2160	8640	240	166	5.7996	7.6193e-07	433	11.4936	1.4763e-06	406	10.7057	3.5314e-06
2880	11520	320	164	9.3141	2.0543e-07	434	19.8208	3.7273e-06	407	18.6239	1.4575e-06

p	q	s	ADMM			PPDG		
			Iter	Time	$rerr$	Iter	Time	$rerr$
180	720	20	348	0.0585	0.0014	169	0.0518	7.6869e-04
360	1440	40	361	0.1495	3.1874e-05	174	0.1030	4.5422e-04
720	2880	80	352	0.8803	3.3088e-05	171	0.6781	5.7072e-05
1440	5760	160	352	3.6357	2.3697e-05	170	3.1788	3.6407e-05
2160	8640	240	353	9.3718	3.5728e-06	171	7.6091	1.4510e-06
2880	11520	320	354	16.2507	1.1713e-06	171	15.3682	3.6692e-06

Table 4.3: Comparison of four Algorithms with $\lambda_1 = 10^{-5}$.

p	q	s	Algorithm 3.1			DYS			DCA		
			Iter	Time	$rerr$	Iter	Time	$rerr$	Iter	Time	$rerr$
180	720	20	107	0.0106	1.7677e-05	263	0.0170	8.9156e-04	254	0.0167	8.4106e-04
360	1440	40	119	0.0235	6.0148e-06	288	0.0706	3.5545e-05	278	0.0528	3.3962e-05
720	2880	80	119	0.4547	4.3010e-06	280	0.6962	4.2654e-05	271	0.5919	2.1097e-05
1440	5760	160	121	1.8941	2.1491e-06	281	2.7577	2.9787e-05	272	2.6437	1.0635e-05
2160	8640	240	120	3.9854	7.2764e-07	283	7.2395	1.4242e-06	274	7.0496	3.1340e-06
2880	11520	320	124	7.1472	8.4316e-07	281	12.5600	1.9829e-06	271	12.1165	1.9022e-06

p	q	s	ADMM			PPDG		
			Iter	Time	$rerr$	Iter	Time	$rerr$
180	720	20	223	0.0160	9.0620e-04	119	0.0111	8.7183e-04
360	1440	40	243	0.0519	3.6891e-05	127	0.0284	3.4787e-05
720	2880	80	237	0.5826	2.5836e-05	125	0.3810	4.1734e-05
1440	5760	160	237	2.3527	1.1518e-05	125	6.9893	2.9148e-05
2160	8640	240	239	6.3154	3.3931e-06	126	4.4888	1.3937e-06
2880	11520	320	237	10.5621	2.0607e-06	129	8.1427	5.8535e-06

Acknowledgements. This work was supported by the National Natural Science Foundation of China (Grant No. 12271067), by the NSF of Chongqing (Grant No. CSTB2024NSCQ-MSX1282), by the Education Committee Project Research Foundation of Chongqing (Grant No. KJZD-K202500805) and by the Team Building Project for Graduate Tutors in Chongqing (Grant No. yds223010).

References

[1] M. Ahookhosh, A. Themelis, and P. Patrinos, A Bregman forward-backward linesearch algorithm for nonconvex composite optimization: Superlinear convergence to nonisolated local minima, *SIAM J. Optim.*, **31**:1 (2021), 653–685.

[2] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality, *Math. Oper. Res.*, **35**:2 (2010), 438–457.

[3] J. Bai, L. Jia, and Z. Peng, A new insight on augmented Lagrangian method with applications in machine learning, *J. Sci. Comput.*, **99** (2024), 53.

[4] H.H. Bauschke and P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, 2017.

[5] A. Beck, *First-Order Methods in Optimization*, SIAM, 2017.

[6] F. Bian and X.Q. Zhang, A three-operator splitting algorithm for nonconvex sparsity regularization, *SIAM Sci. Comput.*, **43**:4 (2021), A2809–A2839.

[7] J. Bolte, H. Sabach, and M. Teboulle, Proximal alternating linearized minimization for nonconvex and nonsmooth problems, *Math. Program.*, **146** (2014), 459–494.

[8] J. Bolte, H. Sabach, and M. Teboulle, Nonconvex Lagrangian-based optimization: Monitoring schemes and global convergence, *Math. Oper. Res.*, **43** (2018), 1210–1232.

[9] J.F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer-Verlag, 2000.

[10] R.I. Bot, E.R. Csetnek, and D.K. Nguyen, A proximal minimization algorithm for structured nonconvex and nonsmooth problems, *SIAM J. Optim.*, **29**:2 (2019), 1300–1328.

[11] A. Chambolle and T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, *J. Math. Imaging Vision*, **40**:1 (2011), 120–145.

[12] A. Chambolle and T. Pock, On the ergodic convergence rates of a first-order primal-dual algorithm, *Math. Program.*, **159** (2016), 253–287.

[13] X.K. Chang, J.F. Yang, and H.C. Zhang, Golden ratio primal-dual algorithm with linesearch, *SIAM J. Optim.*, **32**:3 (2022), 1584–1613.

[14] L. Condat, A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms, *J. Optim. Theory Appl.*, **158** (2013), 460–479.

[15] D. Davis and Y.T. Yin, A three-operator splitting scheme and its optimization applications, *Set-Valued Var. Anal.*, **25** (2017), 829–858.

[16] E. Esser, Y. Lou, and J. Xin, A method for finding structured sparse solutions to non-negative least squares problems with applications, *SIAM J. Imaging Sci.*, **6** (2013), 2010–2046.

[17] J.H. Guo, X. Wang, and X.T. Xiao, Preconditioned primal-dual gradient methods for nonconvex composite and finite-sum optimization, *arXiv:2309.13416v2*, 2023.

[18] C.Y. Huang et al., Deep block proximal linearized minimization algorithm for nonconvex inverse problems, *SIAM J. Math. Data Sci.*, **7** (2025), 1729–1754.

[19] W.L. Huang and Y.C. Tang, Primal-dual fixed point algorithm based on adapted metric method for solving convex minimization problem with application, *Appl. Numer. Math.*, **157** (2020), 236–254.

[20] X. Jiang and L. Vandenberghe, Bregman three-operator splitting methods, *J. Optim. Theory Appl.*, **199** (2023), 936–972.

[21] S.C. Laszlo, A forward-backward algorithm with different inertial terms for structured non-convex minimization problems, *J. Optim. Theory Appl.*, **198** (2023), 387–427.

[22] G.Y. Li and T.K. Pong, Global convergence of splitting methods for nonconvex composite optimization, *SIAM J. Optim.*, **25** (2015), 2434–2460.

[23] Y. Lou and M. Yan, Fast L1-L2 minimization via a proximal operator, *J. Sci. Comput.*, **74** (2018), 767–785.

[24] Y. Lou, T. Zeng, S. Osher, and J. Xin, A weighted difference of anisotropic and isotropic total variation model for image processing, *SIAM J. Imaging Sci.*, **8** (2015), 1798–1823.

[25] T. Pham and E.B. Souad, Algorithms for solving a class of nonconvex optimization problems. Methods of subgradients, in: *Fermat Days 85: Mathematics for Optimization. NorthHolland Mathematics Studies*, Vol. 129, North-Holland, (1986), 249–271.

[26] J. Pock and A. Chambolle, Diagonal preconditioning for first order primal-dual algorithms in convex optimization, in: *2011 International Conference on Computer Vision*, IEEE, (2011), 1762–1769.

[27] Y. Qian and S. Pan, Convergence of a class of nonmonotone descent methods for Kurdyka-Łojasiewicz optimization problems, *SIAM J. Optim.*, **33** (2023), 638–651.

[28] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.

[29] R.T. Rockafellar and R.J.B. Wets, *Variational Analysis*, in: *Grundlehren der Mathematischen Wissenschaften*, Vol. 317, Springer-Verlag, 1998.

[30] A. Salim, L. Condat, K. Mishchenko, and P. Richtarik, Dualize, split, randomize: Toward fast nonsmooth optimization algorithms, *J. Optim. Theory Appl.*, **195** (2022), 102–130.

[31] A. Themelis, L. Stella, and P. Panagiotis, Forward-backward envelope for the sum of two nonconvex functions: Further properties and nonmonotone linesearch algorithms, *SIAM J. Optim.*, **28** (2018), 2274–2303.

[32] H. Uzawa, Iterative methods for concave programming, in: *Studies in Linear and Nonlinear Programming*, Stanford University Press, (1958), 154–165.

[33] B.C. Vu, A splitting algorithm for dual monotone inclusions involving cocoercive operators, *Adv. Comput. Math.*, **38** (2013), 667–681.

[34] T. Wang and H. Liu, A class of modified accelerated proximal gradient methods for nonsmooth and nonconvex minimization problems, *Numer. Algorithms*, **95** (2024), 207–241.

[35] Y.J. Wang, C. Ouyang, L.F. Lv, and G.L. Yuan, Analysis of a new BFGS algorithm and conjugate gradient algorithms and their applications in image restoration and machine learning, *Appl. Numer. Math.*, **210** (2025), 199–221.

[36] B. Wen, X. Chen, and T.K. Pong, Linear convergence of proximal gradient algorithm with extrapolation for a class of nonconvex nonsmooth minimization problems, *SIAM J. Optim.*, **27** (2017), 124–145.

[37] Z.M. Wu, C.Y. Huang, and T.Y. Zeng, Extrapolated plug-and-play three-operator splitting methods for nonconvex optimization with applications to image restoration, *SIAM J. Imaging Sci.*, **17**:20 (2024), 1145–1181.

[38] Z.M. Wu, C. Li, M. Li, and A. Lim, Inertial proximal gradient methods with Bregman regularization for a class of nonconvex optimization problems, *J. Glob. Optim.*, **79**:3 (2021), 617–644.

[39] M. Yan, A new primal-dual algorithm for minimizing the sum of three functions with a linear operator, *J. Sci. Comput.*, **76** (2018), 1698–1717.

[40] L. Yang, Proximal gradient method with extrapolation and line search for a class of non-convex and non-smooth problems, *J. Optim. Theory Appl.*, **200** (2024), 68–103.

[41] P. Yin, Y. Lou, Q. He, and J. Xin, Minimization of l_{1-2} for compressed sensing, *SIAM J. Sci. Comput.*, **37** (2015), A536–A567.

[42] H. Zou and T. Hastie, Regularization and variable selection via the elastic net, *J. R. Stat. Soc. B.*, **67** (2005), 301–320.