Journal of Inequalities and Applications
a SpringerOpen Journal

RESEARCH

Open Access

CrossMark

An accelerated proximal augmented Lagrangian method and its application in compressive sensing

Min Sun[1,2*] and Jing Liu[3]

*Correspondence:
ziyouxiaodou@163.com
[1] School of Management, Qufu
Normal University, Shandong,
276826, P.R. China
[2] School of Mathematics and
Statistics, Zaozhuang University,
Shandong, 277160, P.R. China
Full list of author information is
available at the end of the article

Abstract

As a first-order method, the augmented Lagrangian method (ALM) is a benchmark solver for linearly constrained convex programming, and in practice some semi-definite proximal terms are often added to its primal variable's subproblem to make it more implementable. In this paper, we propose an accelerated PALM with indefinite proximal regularization (PALM-IPR) for convex programming with linear constraints, which generalizes the proximal terms from semi-definite to indefinite. Under mild assumptions, we establish the worst-case $\mathcal{O}(1/t^2)$ convergence rate of PALM-IPR in a non-ergodic sense. Finally, numerical results show that our new method is feasible and efficient for solving compressive sensing.

MSC: 90C25; 90C30

Keywords: proximal augmented Lagrangian multiplier method; convex programming; global convergence; compressive sensing

1 Introduction

Let $\mathcal{R}$ denote the set of all real numbers, $\mathcal{R}^n$ the Euclidean space of all real vectors with n coordinates. In this paper, we are going to solve the following linearly constrained convex programming:

$$\min\{f(x)|Ax = b, x \in \mathcal{R}^n\}, \tag{1}$$

where $f(x): \mathcal{R}^n \to \mathcal{R}$ is a closed proper convex function, $A \in \mathcal{R}^{m \times n}$, $b \in \mathcal{R}^m$. Throughout, we assume that the solution set of Problem (1) is nonempty. By choosing different objective function $f(x)$, a variety of problems encountered in compressive processing, machine learning and statistics can be cast into Problem (1) (see [1–8] and the references therein). The following are two concrete examples of Problem (1):

- The compressive sensing (CS):

$$\min_x f(x) = \|x\|_1 + \frac{1}{2\mu}\|x\|_2^2, \tag{2}$$

s.t. $Ax = b$,

Springer

where $\mu > 0, A \in \mathcal{R}^{m \times n} (m \ll n)$ is the sensing matrix; $b \in \mathcal{R}^m$ is the observed signal, and the ℓ_1-norm and ℓ_2-norm of the vector x are defined by $\|x\|_1 = \sum_{i=1}^{n} |x_i|$ and $\|x\|_2 = (\sum_{i=1}^{n} x_i^2)^{1/2}$, respectively.

- The wavelet-based image processing problem:

$$\min_x f(x) = \|x\|_1,$$
$$\text{s.t. } BWx = b,$$

(3)

where $B \in \mathcal{R}^{m \times l}$ is a diagonal matrix whose elements are either 0 (missing pixels) or 1 (known pixels), and $W \in \mathcal{R}^{l \times n}$ is a wavelet dictionary.

Problem (1) can be converted into the following strongly convex programming:

$$\min \left\{ f(x) + \frac{\beta}{2} \|Ax - b\|^2 \mid Ax = b, x \in \mathcal{R}^n \right\},$$

(4)

where the constant $\beta > 0$ is a penalty parameter. Introducing the Lagrange multiplier $\lambda \in \mathcal{R}^m$ to the linear constraints $Ax = b$, we get the Lagrangian function associated with Problem (4):

$$\mathcal{L}(x, \lambda) = f(x) + \lambda^\top (Ax - b) + \frac{\beta}{2} \|Ax - b\|^2,$$

which is also the augmented Lagrangian function associated with Problem (1). Then the dual function is denoted by

$$G(\lambda) = \inf_x \mathcal{L}(x, \lambda),$$

and the dual problem of (4) is

$$\max_{\lambda \in \mathcal{R}^m} G(\lambda).$$

Due to the strong convexity of the objective function of Problem (4), $G(\lambda)$ is continuously differentiable at any $\lambda \in \mathcal{R}^m$, and $\nabla G(\lambda) = -(Ax(\lambda) - b)$, where $x(\lambda) = \arg\min_x \mathcal{L}(x, \lambda)$ (see, e.g., Theorem 1 of [9]). Solving the above dual problem by the gradient ascent method, we get a benchmark solver for Problem (1): the augmented Lagrangian method (ALM) [10, 11], which first minimizes the Lagrangian function of Problem (4) with respect to x by fixing $\lambda = \lambda^k$ to get $x(\lambda^k)$, and set $x^{k+1} = x(\lambda^k)$; then it updates the Lagrange multiplier λ. Specifically, for given λ^k, the kth iteration of PALM for Problem (1) reads

$$\begin{cases} x^{k+1} = \arg\min_{x \in \mathcal{R}^n} \{\mathcal{L}(x, \lambda^k)\}, \\ \lambda^{k+1} = \lambda^k + \gamma\beta(Ax^{k+1} - b), \end{cases}$$

(5)

where $\gamma \in (0, 2)$ is a relaxation factor. Though ALM plays a fundamental role in the algorithmic development of Problem (1), the cost of solving its first subproblem is often high for general $f(\cdot)$ and A. To address this issue, many proximal ALMs [3, 12–14] are developed by adding the proximal term $\frac{1}{2}\|x - x^k\|_G^2$ to the x-related subproblem, where $G \in \mathcal{R}^{n \times n}$ is a

semi-definite matrix. By setting $G = \tau I_n - \beta A^\top A$ with $\tau > \beta \|A^\top A\|$, the x-related subproblem reduces to the following form:

$$x^{k+1} = \arg\min_{x \in \mathcal{R}^n} \left\{ f(x) + \frac{\tau}{2} \|x - u^k\|^2 \right\},$$

where $u^k = \frac{1}{\tau}(Gx^k - A^\top \lambda^k + \beta A^\top b)$. The above subproblem is often simple enough to have a closed-form solution or can be easily solved up to a high precision. The proximal ALM is so instructive, and along this philosophy, a lot of efficient proximal ALM-type methods [3, 15–18] have been proposed. However, a new difficult problem has arisen for the proximal ALM-type methods, which is how to determine the optimal value of the proximal parameter τ. Numerical results indicate that smaller values of τ can often speed up the corresponding proximal ALM-type method [19].

In [20], Fazel et al. pointed out that the proximal matrix G should be as small as possible, while the subproblem related to the primal variable x is still relatively easy to tackle. Furthermore, though any $\tau > \beta \|A^\top A\|$ can ensure the global convergence of proximal ALM-type methods, the computation of the norm $\|A^\top A\|$ is often high for some problems in practice, especially for large n. Therefore, it is meaningful to relax the feasible region of the proximal parameter τ. Quite recently some efficient methods with relaxed proximal matrix [21–24] have been developed for Problem (1). Specifically, He et al. [23] proposed a positive-indefinite proximal augmented Lagrangian method, in which the proximal matrix is $G = G_0 - (1 - \alpha)\beta A^\top A$, where G_0 is an arbitrarily positive definite matrix in $\mathcal{R}^n$, and $\alpha \in (\frac{2+\gamma}{4}, 1)$. If we set $G_0 = \tau I_n - \alpha\beta A^\top A$ with $\tau > \alpha\beta \|A^\top A\|$, then $G = \tau I_n - \beta A^\top A$, which maybe indefinite.

Many research results on the convergence speed of the ALM-type methods have been presented recently. For the classical ALM, He et al. [25] firstly established the worst-case $\mathcal{O}(1/t)$ convergence rate in the non-ergodic sense, and further developed a new ALM-type method, which has $\mathcal{O}(1/t^2)$ convergence rate. However, for inequality (3.13) in [25], the authors only proved $L(x^*, \lambda^*) - L(\tilde{x}^k, \tilde{\lambda}^k) \leq C/(k+1)^2$ with $C > 0$, and we cannot ensure $L(x^*, \lambda^*) - L(\tilde{x}^k, \tilde{\lambda}^k) \geq 0$. Similar problem occurs in Theorem 2.2 of [26]. In the following, we shall prove that the inequality $L(x^*, \lambda^*) - L(\tilde{x}^k, \tilde{\lambda}^k) \geq 0$ holds for the method in [26]. Quite recently, by introducing an ingeniously designed sequence $\{\theta_k\}$, Lu et al. [27] proposed a fast PALM-type method without proximal term, which has $\mathcal{O}(1/t^2)$ convergence rate.

In this paper, based on the study of [27], we are going to further study the augmented Lagrangian method and develop a new fast proximal ALM-type method with indefinite proximal regularization, whose worst-case convergence rate is $\mathcal{O}(1/t^2)$ in a non-ergodic sense. Furthermore, a relaxation factor $\gamma \in (0, 2)$ is attached to the updated formula of our new method, which is often beneficial to speed up convergence in practice.

The rest of this paper is organized as follows. In Section 2, we list some necessary notations. We then give the proximal ALM with indefinite proximal regularization (PALM-IPR) and show its worst-case $\mathcal{O}(1/t^2)$ convergence rate in Section 3. In Section 4, numerical experiments are conducted to illustrate the efficiency of PALM-IPR. Finally, some conclusions are drawn in Section 5.

2 Preliminaries

In this section, we give some notations used in the following analysis and present two criteria to measure the worst-case $\mathcal{O}(1/t)$ convergence rate of PALM type methods. At the end of this section, we prove that the inequality $L(x^*, \lambda^*) - L(\tilde{x}^k, \tilde{\lambda}^k) \geq 0$ holds for the method in [26].

Throughout, we use the following standard notations. For any two vectors $x, y \in \mathcal{R}^n$, $\langle x, y \rangle$ or $x^\top y$ denote their inner products. The symbols $\| \cdot \|_1$ and $\| \cdot \|$ represent the ℓ_1-norm and ℓ_2-norm for vector variables, respectively. I_n denotes the n-dimensional identity matrix. If the matrix $G \in \mathcal{R}^{n \times n}$ is symmetric, we use the symbol $\|x\|_G^2$ to denote $x^\top G x$ even if G is indefinite; $G \succ 0$ (resp., $G \succeq 0$) denotes that the matrix G is positive definite (resp., semi-definite).

The following identity will be used in the following analysis:

$$\langle x - y, x - z \rangle = \frac{1}{2}\left(\|x - y\|^2 + \|x - z\|^2 - \|y - z\|^2\right), \quad x, y, z \in \mathcal{R}^n. \tag{6}$$

Definition 2.1 A point (x^*, λ^*) is a Karush-Kuhn-Tucker (KKT) point of Problem (1) if the following conditions are satisfied:

$$\begin{cases} 0 \in \partial f(x^*) + A^\top \lambda^*, \\ Ax^* = b. \end{cases} \tag{7}$$

Note that the two conditions of (7) correspond to the dual feasibility and the primal feasibility of Problem (1), respectively. The solution set of KKT system (7), denoted by $\mathcal{W}^*$, is nonempty under the nonempty solution set of (1). By (7) and the property of the convex function $f(\cdot)$, for any $(x^*, \lambda^*) \in \mathcal{W}^*$, we have

$$f(x) - f(x^*) + (x - x^*)^\top (A^\top \lambda^*) \geq 0, \quad \forall x \in \mathcal{R}^n. \tag{8}$$

Based on (8) and $Ax^* = b$, we have the following proposition.

Proposition 2.1 ([28]) *Vector $\tilde{x} \in \mathcal{R}^n$ is an optimal solution to Problem* (1) *if and only if there exists $r > 0$ such that*

$$f(\tilde{x}) - f(x^*) + (\tilde{x} - x^*)^\top (A^\top \lambda^*) + \frac{r}{2}\|A\tilde{x} - b\|^2 = 0, \tag{9}$$

where $(x^*, \lambda^*) \in \mathcal{W}^*$.

Now let us review two different criteria to measure the worst-case $\mathcal{O}(1/t^2)$ convergence rate in [28, 29].

(1) In [29], Xu presented the following criterion:

$$\begin{cases} |f(x^t) - f(x^*)| \leq \frac{C}{(t+1)^2}, \\ \|Ax^t - b\| \leq \frac{C}{\|\lambda^*\|(t+1)^2}, \end{cases} \tag{10}$$

where $C > 0$. The second inequality of (10) implies that there must exist at least one $(x^*, \lambda^*) \in \mathcal{W}^*$ with $\lambda^* \neq 0$.

(2) In [28], Lin et al. proposed the following criterion:

$$f(x^t) - f(x^*) + (x^t - x^*)^\top (A^\top \lambda^*) + \frac{c}{2} \|Ax^t - b\|^2 \leq \frac{C}{(t+1)^2}, \tag{11}$$

where $c, C > 0$. Obviously, inequality (11) is motivated by equality (9). Compared with (10), the criterion (11) is more reasonable. Therefore, we shall use (11) to measure the $\mathcal{O}(1/t^2)$ convergence rate of our new method.

Now, we prove that the inequality $L(x^*, \lambda^*) - L(\tilde{x}^k, \tilde{\lambda}^k) \geq 0$ holds for the iteration method proposed in [26].

Theorem 2.1 *For the KKT point (x^*, λ^*) of Problem (1), the tuple $(x^{k+1}, \tilde{\lambda}^{k+1})$ generated by the method in* [26] *satisfies*

$$\mathcal{L}(x^{k+1}, \tilde{\lambda}^{k+1}) \leq \mathcal{L}(x^*, \lambda^*).$$

Proof Since x^{k+1} is generated by the following subproblem:

$$x^{k+1} = \arg\min_{x \in \mathcal{R}^n} \left\{ f(x) + \frac{\beta}{2} \left\| Ax - b + \frac{\lambda^k}{\beta} \right\|^2 \right\},$$

then it holds that

$$0 \in \partial f(x^{k+1}) + A^\top (\lambda^k + \beta(Ax^{k+1} - b)),$$

i.e.,

$$-A^\top \tilde{\lambda}^{k+1} \in \partial f(x^{k+1}).$$

This and the convexity of the function $f(\cdot)$ yield

$$f(x^*) - f(x^{k+1}) \geq \langle -A^\top \tilde{\lambda}^{k+1}, x^* - x^{k+1} \rangle.$$

So, it holds that

$$\begin{aligned}
\mathcal{L}(x^*, \lambda^*) &= f(x^*) \\
&\geq f(x^{k+1}) + \langle -A^\top \tilde{\lambda}^{k+1}, x^* - x^{k+1} \rangle \\
&\geq f(x^{k+1}) + \langle -\tilde{\lambda}^{k+1}, Ax^* - Ax^{k+1} \rangle \\
&\geq f(x^{k+1}) + \langle \tilde{\lambda}^{k+1}, Ax^{k+1} - b \rangle \\
&= \mathcal{L}(x^{k+1}, \tilde{\lambda}^{k+1}).
\end{aligned}$$

This completes the proof. □

3 PALM-IPR and its convergence rate

In this section, we first present the proximal ALM with indefinite proximal regularization (PALM-IPR) for Problem (1) and then prove its convergence rate step by step.

To present our new method, let us define a sequence $\{\theta_k\}$ which satisfies $\theta_0 = 1$, $\theta_k > 0$ and

$$\frac{1 - \theta_{k+1}}{\theta_{k+1}^2} = \frac{1}{\theta_k^2}. \tag{12}$$

Then the sequence $\{\theta_k\}$ has the following properties [27]:

$$\sum_{k=0}^{t} \frac{1}{\theta_k} = \frac{1}{\theta_t^2}, \tag{13}$$

and

$$\theta_k \leq \frac{2}{k+2}, \quad \forall k \geq 0. \tag{14}$$

Algorithm 3.1 (PALM-IPR for Problem (1))

Step 1. Input the parameters $\beta_0 = \theta_0 = 1, \gamma \in (0, 2)$, the tolerance $\varepsilon > 0$. Initialize $x^0 \in \mathcal{R}^n, z^0 \in \mathcal{R}^n, \lambda^0 \in \mathcal{R}^m$ and $G_0 \in \mathcal{R}^{n \times n}$. Set $k := 0$.

Step 2. Compute

$$z^{k+1} = \arg\min_{x \in \mathcal{R}^n} \left\{ f(x) + \frac{\beta_k}{2} \left\| Ax - b + \frac{\lambda^k}{\beta_k} \right\|^2 + \frac{\beta_k}{2} \left\| x - z^k \right\|_{G_k}^2 \right\}; \tag{15}$$

$$x^{k+1} = (1 - \theta_k)x^k + \theta_k z^{k+1}; \tag{16}$$

$$\lambda^{k+1} = \lambda^k + \gamma \beta_k (Az^{k+1} - b); \tag{17}$$

$$\theta_{k+1} = \frac{-\theta_k^2 + \sqrt{\theta_k^4 + 4\theta_k^2}}{2}; \tag{18}$$

$$\beta_{k+1} = \frac{1}{\theta_{k+1}}. \tag{19}$$

Step 3. If $\|x^k - x^{k+1}\| + \|\lambda^k - \lambda^{k+1}\| \leq \varepsilon$, then stop; otherwise, choose $G_{k+1} \in \mathcal{R}^{n \times n}$, and set $k := k + 1$. Go to Step 2.

To prove the global convergence of PALM-IPR, we need to impose some restrictions on the proximal matrix G_k, which is stated as follows.

Assumption 3.1 The proximal matrix G_k is set as $G_k = \theta_k^2(D_k - (1 - \alpha)A^\top A)$, where $\alpha \in (\frac{2+\gamma}{4}, 1)$ and the matrix $D_k \succeq 0, \forall k \geq 1$.

Remark 3.1 The proximal matrix G_k maybe indefinite. For example, if we set $D_k = 0$, then $G_k = -(1 - \alpha)\theta_k^2 A^\top A$, which is indefinite when A is full-column rank.

Using the first-order optimality condition of the subproblem of PALM-IPR, we can deduce the following one-iteration result.

Lemma 3.1 *Let* $\{(x^k, y^k, z^k, \lambda^k)\}_{k \geq 0}$ *be the sequence generated by PALM-IPR. For any* $x \in \mathcal{R}^n$, *it holds that*

$$
\frac{1 - \theta_{k+1}}{\theta_{k+1}^2} \left(f(x^{k+1}) - f(x) \right) - \frac{1}{\gamma \theta_k} \langle \lambda^{k+1} + (\gamma - 1)\lambda^k, Ax - Az^{k+1} \rangle
$$
$$
\leq \frac{1 - \theta_k}{\theta_k^2} \left(f(x^k) - f(x) \right) + \frac{1}{2\theta_k^2} \left(\| z^k - x \|_{G_k}^2 - \| z^{k+1} - x \|_{G_k}^2 - \| z^{k+1} - z^k \|_{G_k}^2 \right). \tag{20}
$$

Proof From the first-order optimality condition for z-related subproblem (15), we have

$$
0 \in \partial f(z^{k+1}) + \beta_k A^\top \left(Az^{k+1} - b + \frac{1}{\beta_k} \lambda^k \right) + \beta_k G_k(z^{k+1} - z^k)
$$
$$
= \partial f(z^{k+1}) + A^\top \tilde{\lambda}^k + \beta_k G_k(z^{k+1} - z^k), \tag{21}
$$

where $\tilde{\lambda}^k = \lambda^k + \beta_k(Az^{k+1} - b)$. Then, from the convexity of $f(\cdot)$ and (21), we have

$$
f(x) - f(z^{k+1}) \geq -\langle A^\top \tilde{\lambda}^k + \beta_k G_k(z^{k+1} - z^k), x - z^{k+1} \rangle. \tag{22}
$$

From (16) and the convexity of $f(\cdot)$ again, we thus get

$$
f(x^{k+1}) \leq (1 - \theta_k) f(x^k) + \theta_k f(z^{k+1})
$$
$$
\leq (1 - \theta_k) f(x^k) + \theta_k \left(f(x) + \langle A^\top \tilde{\lambda}^k + \beta_k G_k(z^{k+1} - z^k), x - z^{k+1} \rangle \right),
$$

where the second inequality follows from (22). Then, by rearranging terms of the above inequality, we arrive at

$$
f(x^{k+1}) - f(x) - \theta_k \langle A^\top \tilde{\lambda}^k, x - z^{k+1} \rangle
$$
$$
\leq (1 - \theta_k)\left(f(x^k) - f(x) \right) + \langle G_k(z^{k+1} - z^k), x - z^{k+1} \rangle
$$
$$
\leq (1 - \theta_k)\left(f(x^k) - f(x) \right) + \frac{1}{2} \left(\| z^k - x \|_{G_k}^2 - \| z^{k+1} - x \|_{G_k}^2 - \| z^{k+1} - z^k \|_{G_k}^2 \right),
$$

where the second inequality uses (19), and the third inequality comes from identity (6). Dividing both sides of the above inequality by θ_k^2, we get

$$
\frac{1}{\theta_k^2} f(x^{k+1}) - f(x) - \frac{1}{\theta_k} \langle A^\top \tilde{\lambda}^k, x - z^{k+1} \rangle
$$
$$
\leq \frac{1 - \theta_k}{\theta_k^2} \left(f(x^k) - f(x) \right) + \frac{1}{2\theta_k^2} \left(\| z^k - x \|_{G_k}^2 - \| z^{k+1} - x \|_{G_k}^2 - \| z^{k+1} - z^k \|_{G_k}^2 \right).
$$

From (17), we have

$$
\tilde{\lambda}^k = \lambda^k + \frac{\lambda^{k+1} - \lambda^k}{\gamma} = \frac{\lambda^{k+1} + (\gamma - 1)\lambda^k}{\gamma}.
$$

Substituting this into the above inequality and using (12) lead to (20). This completes the proof. $\qquad\square$

Lemma 3.2 *Let* $\{(x^k, y^k, z^k, \lambda^k)\}_{k \geq 0}$ *be the sequence generated by PALM-IPR. For any* $(x, \lambda) \in \mathcal{R}^{m+n}$ *with* $Ax = b$, *it holds that*

$$
\frac{1 - \theta_{k+1}}{\theta_{k+1}^2} \left(f(x^{k+1}) - f(x) \right) - \frac{1 - \theta_k}{\theta_k^2} \left(f(x^k) - f(x) \right) + \frac{1}{\theta_k} \langle \lambda, Az^{k+1} - b \rangle
$$
$$
\leq \frac{1}{2\theta_k^2} \left(\|z^k - x\|_{G_k}^2 - \|z^{k+1} - x\|_{G_k}^2 - \|z^{k+1} - z^k\|_{G_k}^2 \right)
$$
$$
+ \frac{1}{2\gamma^2} \left[\gamma \left(\|\lambda^k - \lambda\|^2 - \|\lambda^{k+1} - \lambda\|^2 \right) - (2 - \gamma) \|\lambda^{k+1} - \lambda^k\|^2 \right]. \tag{23}
$$

Proof Adding the term $\frac{1}{\theta_k} \langle \lambda, Az^{k+1} - b \rangle$ to both sides of (20), we get

$$
\frac{1 - \theta_{k+1}}{\theta_{k+1}^2} \left(f(x^{k+1}) - f(x) \right) - \frac{1 - \theta_k}{\theta_k^2} \left(f(x^k) - f(x) \right) + \frac{1}{\theta_k} \langle \lambda, Az^{k+1} - b \rangle
$$
$$
\leq \frac{1}{2\theta_k^2} \left(\|z^k - x\|_{G_k}^2 - \|z^{k+1} - x\|_{G_k}^2 - \|z^{k+1} - z^k\|_{G_k}^2 \right)
$$
$$
+ \frac{1}{\theta_k} \left\langle \lambda - \frac{\lambda^{k+1} + (\gamma - 1)\lambda^k}{\gamma}, Az^{k+1} - b \right\rangle. \tag{24}
$$

Now, let us deal with the term $\langle \lambda - \frac{\lambda^{k+1} + (\gamma-1)\lambda^k}{\gamma}, Az^{k+1} - b \rangle$. By (17), we have

$$
\left\langle \lambda - \frac{\lambda^{k+1} + (\gamma - 1)\lambda^k}{\gamma}, Az^{k+1} - b \right\rangle
$$
$$
= \frac{1}{\gamma^2 \beta_k} \langle \gamma\lambda - \lambda^{k+1} - (\gamma - 1)\lambda^k, \lambda^{k+1} - \lambda^k \rangle
$$
$$
= \frac{1}{\gamma^2 \beta_k} \left(\gamma \langle \lambda - \lambda^k, \lambda^{k+1} - \lambda^k \rangle - \|\lambda^{k+1} - \lambda^k\|^2 \right)
$$
$$
= \frac{1}{2\gamma^2 \beta_k} \left[\gamma \left(\|\lambda^k - \lambda\|^2 - \|\lambda^{k+1} - \lambda\|^2 \right) - (2 - \gamma) \|\lambda^{k+1} - \lambda^k\|^2 \right].
$$

Substituting the above equality into (24) and using (19), we get (23) immediately. This completes the proof. □

Let us further deal with the term $\|z^k - x\|_{G_k}^2 - \|z^{k+1} - x\|_{G_k}^2 - \|z^{k+1} - z^k\|_{G_k}^2$ on the right-hand side of (24).

Lemma 3.3 *Let* $\{(x^k, y^k, z^k, \lambda^k)\}_{k \geq 0}$ *be the sequence generated by PALM-IPR. For any* $(x, \lambda) \in \mathcal{R}^{m+n}$ *with* $Ax = b$, *it holds that*

$$
\frac{1}{\theta_k^2} \left(\|z^k - x\|_{G_k}^2 - \|z^{k+1} - x\|_{G_k}^2 - \|z^{k+1} - z^k\|_{G_k}^2 \right)
$$
$$
\leq \|z^k - x\|_{D_k}^2 - \|z^{k+1} - x\|_{D_k}^2 - \|z^{k+1} - z^k\|_{D_k}^2
$$
$$
+ (1 - \alpha) \left(\|Az^k - b\|^2 - \|Az^{k+1} - b\|^2 \right) + 4(1 - \alpha) \|Az^{k+1} - b\|^2. \tag{25}
$$

Proof By Assumption 3.1, we have

$$
\frac{1}{\theta_k^2} \left(\|z^k - x\|_{G_k}^2 - \|z^{k+1} - x\|_{G_k}^2 - \|z^{k+1} - z^k\|_{G_k}^2 \right)
$$

$$
\begin{aligned}
&= \left\| z^k - x \right\|_{D_k}^2 - \left\| z^{k+1} - x \right\|_{D_k}^2 - \left\| z^{k+1} - z^k \right\|_{D_k}^2 \\
&\quad + (1-\alpha)\left(-\left\| Az^k - Ax \right\|^2 + \left\| Az^{k+1} - Ax \right\|^2 + \left\| Az^k - Az^{k+1} \right\|^2 \right) \\
&= \left\| z^k - x \right\|_{D_k}^2 - \left\| z^{k+1} - x \right\|_{D_k}^2 - \left\| z^{k+1} - z^k \right\|_{D_k}^2 \\
&\quad + (1-\alpha)\left(-\left\| Az^k - b \right\|^2 + \left\| Az^{k+1} - b \right\|^2 + \left\| \left(Az^k - b \right) - \left(Az^{k+1} - b \right) \right\|^2 \right).
\end{aligned}
$$

Using the inequality $\|\xi - \eta\|^2 \le 2\|\xi\|^2 + 2\|\eta\|^2$ with $\xi = Az^k - b$ and $\eta = Az^{k+1} - b$, we get

$$
\left\| \left(Az^k - b \right) - \left(Az^{k+1} - b \right) \right\|^2 \le 2\left\| Az^k - b \right\|^2 + 2\left\| Az^{k+1} - b \right\|^2.
$$

Substituting this inequality into the right-hand side of the above equality, we obtain assertion (25) immediately. This completes the proof. □

Then, from (23) and (25), we have

$$
\begin{aligned}
&\frac{1-\theta_{k+1}}{\theta_{k+1}^2}\left(f\left(x^{k+1}\right) - f(x) \right) - \frac{1-\theta_k}{\theta_k^2}\left(f\left(x^k\right) - f(x) \right) + \frac{1}{\theta_k}\langle \lambda, Az^{k+1} - b \rangle \\
&\le \frac{1}{2}\Big[\left\| z^k - x \right\|_{D_k}^2 - \left\| z^{k+1} - x \right\|_{D_k}^2 - \left\| z^{k+1} - z^k \right\|_{D_k}^2 + (1-\alpha)\left(\left\| Az^k - b \right\|^2 - \left\| Az^{k+1} - b \right\|^2 \right) \Big] \\
&\quad + 4(1-\alpha)\left\| Az^{k+1} - b \right\|^2 \\
&\quad + \frac{1}{2\gamma^2}\Big[\gamma\left(\left\| \lambda^k - \lambda \right\|^2 - \left\| \lambda^{k+1} - \lambda \right\|^2 \right) - (2-\gamma)\left\| \lambda^{k+1} - \lambda^k \right\|^2 \Big] \\
&\le \frac{1}{2\theta_k^2}\Big[\left\| z^k - x \right\|_{D_k}^2 - \left\| z^{k+1} - x \right\|_{D_k}^2 - \left\| z^{k+1} - z^k \right\|_{D_k}^2 \\
&\quad + (1-\alpha)\left(\left\| Az^k - b \right\|^2 - \left\| Az^{k+1} - b \right\|^2 \right) \Big] \\
&\quad + \frac{1}{2\gamma}\left(\left\| \lambda^k - \lambda \right\|^2 - \left\| \lambda^{k+1} - \lambda \right\|^2 \right) + \frac{2+\gamma - 4\alpha}{2\gamma^2}\left\| \lambda^{k+1} - \lambda^k \right\|^2,
\end{aligned}
\tag{26}
$$

where the second inequality comes from (17) and the fact $\beta_k \ge 1$.

Based on (26), we are going to prove the worst-case $\mathcal{O}(1/t^2)$ convergence rate of PALM-IPR in an ergodic sense.

Theorem 3.1 *Let $\{(x^k, y^k, z^k, \lambda^k)\}_{k \ge 0}$ be the sequence generated by PALM-IPR. Then*

$$
\begin{aligned}
&f\left(x^{t+1}\right) - f\left(x^*\right) + \left\langle A^\top \lambda^*, x^{t+1} - x^* \right\rangle + \frac{4\alpha - 2 - \gamma}{2}\left\| Ax^{t+1} - b \right\|^2 \\
&\le \frac{2}{(t+2)^2}\left(\left\| z^0 - x^* \right\|_{D_0}^2 + (1-\alpha)\left\| Az^0 - b \right\|^2 \right) + \frac{2}{\gamma(t+2)^2}\left\| \lambda^0 - \lambda^* \right\|^2.
\end{aligned}
$$

Proof Setting $x = x^*$ and $\lambda = \lambda^*$ in (26), we get

$$
\begin{aligned}
&\frac{1-\theta_{k+1}}{\theta_{k+1}^2}\left(f\left(x^{k+1}\right) - f\left(x^*\right) \right) - \frac{1-\theta_k}{\theta_k^2}\left(f\left(x^k\right) - f\left(x^*\right) \right) + \frac{1}{\theta_k}\langle \lambda^*, Az^{k+1} - b \rangle \\
&\le \frac{1}{2}\Big[\left\| z^k - x^* \right\|_{D_k}^2 - \left\| z^{k+1} - x^* \right\|_{D_k}^2 - \left\| z^{k+1} - z^k \right\|_{D_k}^2 \\
&\quad + (1-\alpha)\left(\left\| Az^k - b \right\|^2 - \left\| Az^{k+1} - b \right\|^2 \right) \Big]
\end{aligned}
$$

$$+ \frac{1}{2\gamma} \left(\left\| \lambda^k - \lambda^* \right\|^2 - \left\| \lambda^{k+1} - \lambda^* \right\|^2 \right) + \frac{2 + \gamma - 4\alpha}{2\gamma^2} \left\| \lambda^{k+1} - \lambda^k \right\|^2.$$

Summing the above inequality over $k = 1, 2, \ldots, t$ and by (17), we have

$$\frac{1 - \theta_{t+1}}{\theta_{t+1}^2} \left(f(x^{t+1}) - f(x^*) \right) + \sum_{k=0}^{t} \frac{1}{\theta_k} \langle \lambda^*, Az^{k+1} - b \rangle$$

$$\leq \frac{1}{2} \left(\left\| z^0 - x^* \right\|_{D_0}^2 + (1 - \alpha) \left\| Az^0 - b \right\|^2 \right) + \frac{1}{2\gamma} \left\| \lambda^0 - \lambda^* \right\|^2$$

$$+ \frac{2 + \gamma - 4\alpha}{2\theta_k^2} \sum_{k=0}^{t} \left\| Az^{k+1} - b \right\|^2$$

$$\leq \frac{1}{2} \left(\left\| z^0 - x^* \right\|_{D_0}^2 + (1 - \alpha) \left\| Az^0 - b \right\|^2 \right) + \frac{1}{2\gamma} \left\| \lambda^0 - \lambda^* \right\|^2$$

$$+ \frac{2 + \gamma - 4\alpha}{2\theta_k} \sum_{k=0}^{t} \left\| Az^{k+1} - b \right\|^2, \tag{27}$$

where the second inequality follows from $\theta_k \leq 1$ and $\alpha \in (\frac{2+\gamma}{4}, 1)$. By equations (65), (68) of [27], we have

$$\sum_{k=0}^{t} \frac{z^{k+1}}{\theta_k} = \frac{1}{\theta_t^2} x^{t+1},$$

and

$$\sum_{k=0}^{t} \frac{1}{\theta_k} \left\| Az^{k+1} - b \right\|^2 \geq \frac{1}{\theta_t^2} \left\| Ax^{t+1} - b \right\|^2.$$

Substituting the above two relationships into (27) and using (13), we get

$$\frac{1 - \theta_{t+1}}{\theta_{t+1}^2} \left(f(x^{t+1}) - f(x^*) \right) + \frac{1}{\theta_t^2} \langle \lambda^*, Ax^{t+1} - b \rangle + \frac{4\alpha - 2 - \gamma}{2\theta_t^2} \left\| Ax^{t+1} - b \right\|^2$$

$$\leq \frac{1}{2} \left(\left\| z^0 - x^* \right\|_{D_0}^2 + (1 - \alpha) \left\| Az^0 - b \right\|^2 \right) + \frac{1}{2\gamma} \left\| \lambda^0 - \lambda^* \right\|^2.$$

Then multiplying two sides of the above inequality by θ_t^2 and using (12) lead to

$$f(x^{t+1}) - f(x^*) + \langle \lambda^*, Ax^{t+1} - b \rangle + \frac{4\alpha - 2 - \gamma}{2} \left\| Ax^{t+1} - b \right\|^2$$

$$\leq \frac{\theta_t^2}{2} \left(\left\| z^0 - x^* \right\|_{D_0}^2 + (1 - \alpha) \left\| Az^0 - b \right\|^2 \right) + \frac{\theta_t^2}{2\gamma} \left\| \lambda^0 - \lambda^* \right\|^2$$

$$\leq \frac{2}{(t+2)^2} \left(\left\| z^0 - x^* \right\|_{D_0}^2 + (1 - \alpha) \left\| Az^0 - b \right\|^2 \right) + \frac{2}{\gamma(t+2)^2} \left\| \lambda^0 - \lambda^* \right\|^2,$$

where the second inequality follows from (14). The above inequality with $Ax^* = b$ results in the conclusion of the theorem. The proof is completed. □

4 Numerical results

In this section, we apply PALM-IPR to some practical applications and report the numerical results. All the codes were written by Matlab R2010a and conducted on ThinkPad notebook with 2GB of memory.

Problem 4.1 (Quadratic programming) Firstly, let us test PALM-IPR on equality constrained quadratic programming (ECQP) [29] to validate its stability:

$$\min F(x) = \frac{1}{2} x^\top Q x + c^\top x, \quad \text{s.t. } Ax = b.$$

We set the problem size to $m = 20, n = 100$ and generate $A \in \mathcal{R}^{m \times n}, b, c$ and $Q \in \mathcal{R}^{n \times n}$ according to the standard Gaussian distribution. We compare PALM-IPR with the classical ALM with $\beta = 1$. For PALM-IPR, we set $G_k = 0$ for simplicity. We have tested the experiment sixty times, and the numerical results are listed in Table 1, in which 'NC' means the number of convergence; 'ND' means the number of divergence and 'Ratio' means the ratio of succession. From Table 1, we can see that PALM-IPR performs much more stably than the classical ALM.

Problem 4.2 (Compressive sensing: the linearly constrained $\ell_1 - \ell_2$ minimization problem (2)) Now, let us test PALM-IPR on the compressive sensing to validate its acceleration. For this problem, we firstly elaborate on how to solve subproblem (15) resulting from PALM-IPR. Due to the existence of the augmented term $\|Ax - b\|^2$, we cannot get the closed-form solution if we do not attach the proximal term $\frac{1}{2}\|x - z^k\|_{G_k}^2$. In this case, we have to solve the subproblem inexactly, which is often time-consuming. Therefore, we choose the proximal matrix G_k in (15) as $G_k = \theta_k^2(D_k - (1 - \alpha)A^\top A)$, in which $D_k = \frac{1}{\beta_k \theta_k^2} P_k + (1 - \alpha)A^\top A$, $P_k = \tau_k I_n - \beta_k A^\top A$ and $\tau_k > \beta_k \|A^\top A\|$, and subproblem (15) can be written as

$$z^{k+1} = \arg\min_{x \in \mathcal{R}^n} \left\{ f(x) + \frac{\beta_k}{2} \left\| Ax - b + \frac{\lambda^k}{\beta_k} \right\|^2 + \frac{1}{2} \left\| x - z^k \right\|_{P_k}^2 \right\},$$

which is equivalent to

$$z^{k+1} = \arg\min_{x \in \mathcal{R}^n} \left\{ \|x\|_1 + \frac{1 + \mu \tau_k}{2\mu} \left\| x - \frac{\mu}{1 + \mu \tau_k} \left(P_k z^k - A^\top \lambda^k + \beta_k A^\top b \right) \right\|^2 \right\},$$

and has a closed-form solution as follows:

$$z^{k+1} = \text{shrink}_{1,2}\left(a^k, \mu/(1 + \mu \tau_k)\right) \doteq \text{sign}\left(a^k\right) \cdot \max\left\{0, \left|a^k\right| - \mu/(1 + \mu \tau_k)\right\},$$

in which $a^k = \frac{\mu}{1+\mu\tau_k}(P_k z^k - A^\top \lambda^k + \beta_k A^\top b)$. Note that all computations are component-wise.

Table 1 Comparison of PALM-IPR with ALM

Method	NC	ND	Ratio
PALM-IPR	59	1	98.33%
ALM	36	24	60.00%

Table 2 Comparison of PALM-IPR with PALM-SDPR and PALM-PIPR

n	a	b	IPR		SDPR		PIPR	
			Iter	RelErr	Iter	RelErr	Iter	RelErr
500	0.2	0.2	62.3	0.0488	93.2	0.0473	98.6	0.0491
	0.2	0.1	20.5	0.0482	73.6	0.0440	69.2	0.0443
1,000	0.2	0.2	91.9	0.0492	80.0	0.0476	84.3	0.0484
	0.2	0.1	20.6	0.0462	83.3	0.0469	79.9	0.0450
1,500	0.2	0.2	66.4	0.0493	77.5	0.0482	87.5	0.0485
	0.2	0.1	20.3	0.0471	76.6	0.0459	75.9	0.0453
2,000	0.2	0.2	44.7	0.0492	58.3	0.0491	54.0	0.0488
	0.2	0.1	19.1	0.0474	85.2	0.0481	86.5	0.0459
3,000	0.2	0.2	46.3	0.0488	86.8	0.0480	93.2	0.0477
	0.2	0.1	20.2	0.0474	69.3	0.0472	64.6	0.0496

In this experiment, we set $m = \text{floor}(a \times n)$ and $k = \text{floor}(b \times m)$ with $n \in \{500, 1{,}000, 1{,}500, 2{,}000, 3{,}000\}$, where k is the number of random nonzero elements contained in the original signal. The sensing matrix A is generated by the following Matlab scripts: $\bar{A} = \text{randn}(m, n)$, $[Q, R] = \text{qr}(\bar{A}', 0); A = Q'$, and the nonzero entries of the true signal x^*, whose values are sampled from the standard Gaussian: $x^* = \text{zeros}(n, 1)$; $p = \text{randperm}(n); x^*(p(1:k)) = \text{randn}(k, 1)$, are selected at random. The observed signal b is generated by $b = R' \setminus Ax^*$. In addition, we set $\mu = 500$, $\gamma = 1.3$. In this experiment, we set the proximal parameter $\tau_k = 4\beta_k \|A^\top A\|$. Furthermore, the stopping criterion is

$$\text{RelErr} = \frac{\|x^k - x^*\|}{\|x^*\|} \leq 5\%,$$

or the number of iterations exceeds 10^4, where x^k is the iterate generated by PALM-IPR. Furthermore, all initial points are set as $x^0 = A^\top b, \lambda^0 = 0$. For comparison, we also give the numerical results of PALM-SDPR [14] and the proximal PALM with positive-indefinite proximal regularization (PALM-PIPR) [21], and the proximal matrix is set to be $G = \tau I_n - \beta A^\top A$ in PALM-SDPR and $G = 0.9\tau I_n - \beta A^\top A$ in PALM-PIPR, and $\tau = 1.1\beta \|A^\top A\|$, $\beta = 2\text{mean}(\text{abs}(b))$. Furthermore, we set $\gamma = 1$ in PALM-PIPR. Since the computational load of all three methods are almost the same at each iteration, we only list the number of iterations ('Iter'), the relative error ('RelErr') when three methods achieve the stopping criterion. The numerical results are listed in Table 1, and in the view of statistics, all the results are the average of 10 runs for each pair (n, a, b).

Numerical results in Table 2 indicate that: (1) All methods have succeeded in solving Problem (2) for all the scenarios; (2) The new method PALM-IPR outperforms PALM-SDPR and PALM-PIPR by taking a fewer number of iterations to converge except $(n, a, b) = (1{,}000, 0.2, 0.2)$.

5 Conclusions

In this paper, an accelerated augmented proximal Lagrangian method with indefinite proximal regularization (PALM-IPR) for linearly constrained convex programming is proposed. Under mild conditions, we have established the worst-case $\mathcal{O}(1/t^2)$ convergence rate in a non-ergodic sense of PALM-IPR. Two sets of numerical results, which illustrate that PALM-IPR performs better than some state-of-the-art solvers, are given.

Similar to our proposed method, the methods in [26, 30] also have the worst-case $\mathcal{O}(1/t^2)$ convergence rate in a non-ergodic sense, but they often have to solve a difficult subproblem at each iteration, and some inner iteration has to be executed. A prominent characteristic of the methods in [26, 30] is that the parameter β can be any positive constant, but the parameter β in PALM-IPR changes with respect to the iteration counter k, and it can actually go to infinity as $k \rightarrow \infty$. In practice, we often observe that larger β usually induces to slower convergence. Therefore, the method with proximal term, faster convergence rate and constant parameter β deserves further research.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
All the authors conceived the study, participated in its design and read and approved the final manuscript.

Author details
[1]School of Management, Qufu Normal University, Shandong, 276826, P.R. China. [2]School of Mathematics and Statistics, Zaozhuang University, Shandong, 277160, P.R. China. [3]School of Data Sciences, Zhejiang University of Finance and Economics, Zhejiang, 310018, P.R. China.

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

References
1. Boyd, S, Parikh, N, Chu, E, Peleato, B, Eckstein, J: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. **3**, 1-122 (2011)
2. He, BS, Yuan, XM, Zhang, WX: A customized proximal point algorithm for convex minimization with linear constraints. Comput. Optim. Appl. **56**, 559-572 (2013)
3. Ma, F, Ni, MF: A class of customized proximal point algorithms for linearly constrained convex optimization. Comput. Appl. Math. (2016). doi:10.1007/s40314-016-0371-3
4. Sun, HC, Sun, M, Wang, YJ: Proximal ADMM with larger step size for two-block separable convex programming and its application to the correlation matrices calibrating problems. J. Nonlinear Sci. Appl. **10**(9), 5038-5051 (2017). doi:10.22436/jnsa.010.09.40
5. Liu, J, Duan, YR, Sun, M: A symmetric version of the generalized alternating direction method of multipliers for two-block separable convex programming. J. Inequal. Appl. **2017**, Article ID 129 (2017)
6. Sun, HC, Tian, MY, Sun, M: The symmetric ADMM with indefinite proximal regularization and its application. J. Inequal. Appl. **2017**, Article ID 172 (2017)
7. Sun, HC, Liu, J, Sun, M: A proximal fully parallel splitting method for stable principal component pursuit. Math. Probl. Eng. (2017, in press)
8. Sun, M, Sun, HC, Wang, YJ: Two proximal splitting methods for multi-block separable programming with applications to stable principal component pursuit. J. Appl. Math. Comput. (2017). doi:10.1007/s12190-017-1080-9
9. Nesterov, YE: Smooth minimization of non-smooth functions. Math. Program., Ser. A **103**(1), 127-152 (2005). doi:10.1007/s10107-004-0552-5
10. Hestenes, MR: Multiplier and gradient methods. J. Optim. Theory Appl. **4**, 303-320 (1969)
11. Powell, MJ: A method for non-linear constraints in minimization problems. In: Fletcher, R (ed.) Optimization, pp. 283-298. Academic Press, San Diego (1969)
12. He, BS, Liao, LZ, Han, DR, Yang, H: A new inexact alternating directions method for monotone variational inequalities. Math. Program. **92**(1), 103-118 (2002)
13. He, BS, Yuan, XM: On the $\mathcal{O}(1/n)$ convergence rate of the Douglas-Rachford alternating direction method. SIAM J. Numer. Anal. **50**(2), 700-709 (2012)
14. Yang, JF, Yuan, XM: Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization. Math. Comput. **82**(281), 301-329 (2013)
15. Wang, YJ, Zhou, GL, Caccetta, L, Liu, WQ: An alternative Lagrange-dual based algorithm for sparse signal reconstruction. IEEE Trans. Signal Process. **59**, 1895-1901 (2011)

16. Qiu, HN, Chen, XM, Liu, WQ, Zhou, GL, Wang, YJ, Lai, JH: A fast ℓ_1-solver and its applications to robust face recognition. J. Ind. Manag. Optim. **8**, 163-178 (2012)
17. Wang, YJ, Liu, WQ, Caccetta, L, Zhou, G: Parameter selection for nonnegative ℓ_1 matrix/tensor sparse decomposition. Oper. Res. Lett. **43**, 423-426 (2015)
18. Sun, M, Wang, YJ, Liu, J: Generalized Peaceman-Rachford splitting method for multiple-block separable convex programming with applications to robust PCA. Calcolo **54**(1), 77-94 (2017)
19. Li, M, Sun, DF, Toh, KC: A majorized ADMM with indefinite proximal terms for linearly constrained convex composite optimization. SIAM J. Optim. **26**(2), 922-950 (2016)
20. Fazel, M, Pong, TK, Sun, DF, Tseng, P: Hankel matrix rank minimization with applications to system identification and realization. SIAM J. Matrix Anal. Appl. **34**, 946-977 (2013)
21. He, BS, Ma, F, Yuan, XM: Linearized alternating direction method of multipliers via positive-indefinite proximal regularization for convex programming. Unpublished manuscript (2016)
22. He, BS, Yuan, XM: Improving an ADMM-like splitting method via positive-indefinite proximal regularization for three-block separable convex minimization. Unpublished manuscript (2016)
23. He, BS, Ma, F, Yuan, XM: Positive-indefinite proximal augmented Lagrangian method and its application to full Jacobian splitting for multi-block separable convex minimization problems. Unpublished manuscript (2016)
24. Sun, M, Liu, J: The convergence rate of the proximal alternating direction method of multipliers with indefinite proximal regularization. J. Inequal. Appl. **2017**, Article ID 19 (2017)
25. He, BS, Yuan, XM: On the acceleration of augmented Lagrangian method for linearly constrained optimization. Unpublished manuscript (2010)
26. Ke, YF, Ma, CF: An accelerated augmented Lagrangian method for linearly constrained convex programming with the rate of convergence $\mathcal{O}(1/k^2)$. Appl. Math. J. Chin. Univ. **32**(1), 117-126 (2017). doi:10.1007/s11766-017-3381-z
27. Lu, CY, Li, H, Lin, ZC, Yan, SC: Fast proximal linearized alternating direction method of multiplier with parallel splitting. In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI'16), pp. 739-745. AAAI Press, Menlo Park (2016)
28. Lin, ZC, Liu, RS, Li, H: Linearized alternating direction method with parallel splitting and adaptive penalty for separable convex programs in machine learning. Mach. Learn. **99**(2), 287-325 (2015)
29. Xu, YY: Accelerated first-order primal-dual proximal methods for linearly constrained composite convex programming. Unpublished manuscript (2016)
30. Kang, M, Yun, S, Woo, H, Kang, M: Accelerated Bregman method for linearly constrained $\ell_1 - \ell_2$ minimization. J. Sci. Comput. **56**, 515-534 (2013)