AMO - Advanced Modeling and Optimization, Volume 19, Number 2, 2017

The proximal augmented Lagrangian method with indefinite proximal regularization and its application in image restoration

Shengmin Zhou[1]
Zhejiang Hanma Optoelectronic Equipment Co., Ltd
30# of RenMin Road, LinPu Town, Xiaoshan District, Hangzhou, Zhejiang, 311251, China
E-mail: ccdinfo@126.com

Abstract. In this paper, we generalize the proximal matrix in the proximal augmented Lagrangian method (PALM) from semi-definite to indefinite, and propose a proximal ALM with indefinite proximal regularization (PALM-IPR) for convex programming with linear constraints, which inherits the easily implementable property of the proximal ALM, and often has better numerical performance than the latter. Under mild assumptions, the global convergence of PALM-IPR is proved. Furthermore, we prove its worst-case $\mathcal{O}(1/t)$ convergence rate in an ergodic sense under a more reasonable criterion than those adopted in [5, 17]. Finally, numerical results show that PALM-IPR is feasible and efficient for solving some problems in image reconstruction.

Keywords. Proximal augmented Lagrangian multiplier method; convex programming; global convergence; image reconstruction.

AMS(2000) Subject Classification: 90C25, 90C30

1 Introduction

In this paper, we aim to solve the following linearly constrained convex programming:

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

where $A \in \mathcal{R}^{m\times n}$, $b \in \mathcal{R}^m$. Throughout, we make the following assumptions:

Assumption 1.1. The function $\theta(x)$ is convex, lower semi-continuous and it is simple in the sense that its proximal operator $(I_n + \frac{1}{r}\partial\theta)^{-1}(a)$, defined by

$$\left(I_n + \frac{1}{r}\partial\theta\right)^{-1}(a) := \operatorname{argmin}\left\{\theta(x) + \frac{r}{2}\|x - a\|^2\right\},$$

[1] ***AMO-Advanced Modeling and optimization. ISSN: 1841-4311**

can be easily solved, where $r > 0, a \in \mathcal{R}^n$ and $\partial\theta(\cdot)$ denotes the subdifferential of the convex function $\theta(\cdot)$, which is a set-valued function defined by

$$\partial\theta(x) = \{\xi \in \mathcal{R}^n : \theta(\bar{x}) \geq \theta(x) + \langle \xi, \bar{x} - x \rangle, \text{ for all } \bar{x} \in \mathcal{R}^n\}.$$

Assumption 1.2. The solution set of problem (1) is nonempty.

By choosing different objective function $\theta(x)$, a variety of problems encountered in compressive processing, machine learning, statistics can be cast into problem (1) (see [1, 3, 11] and reference therein). For example, it includes the famous $\ell_1 - \ell_2$ basis pursuit (BP) model of compressive sensing as a special case.

Let $\lambda \in \mathcal{R}^m$ be the Lagrange multiplier to the linear constraints $Ax = b$, and we get the augmented Lagrangian function associated with problem (1):

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

The famous augmented Lagrangian method (ALM) [9, 13] for (1) is as follows: for given λ^k, the $k-$th iteration of ALM for problem (1) is

$$\left\{ \begin{array}{l} x^{k+1} = \text{argmin}_{x \in \mathcal{R}^n} \left\{ \mathcal{L}(x, \lambda^k) \right\}, \\ \lambda^{k+1} = \lambda^k - \gamma\beta(Ax^{k+1} - b), \end{array} \right. \tag{2}$$

where $\gamma \in (0, 2)$ is a relaxation factor. In practice, many proximal ALMs [4, 5, 18, 11] 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 special proximal matrix G, the above subproblem is often easy to compute under Assumption 1.1.

It has been pointed out by Fazel et al. [2] that the proximal matrix G should be as small as possible under the condition that the corresponding subproblem is still easy to tackle. Then, in recent years, some ALM type methods [6, 7, 8, 14] with indefinite proximal regularization are developed for some special cases of problem (1), including

$$\min \left\{ \theta_1(x_1) + \theta(x) | A_1 x_1 + Ax = b \right\} \quad \text{in} \quad [6,\ 14],$$

and

$$\min \left\{ \theta_1(x_1) + \theta(x) + \theta_3(x_3) | A_1 x_1 + Ax + A_3 x_3 = b \right\} \quad \text{in} \quad [7],$$

and

$$\min \left\{ \sum_{i=1}^{m} \theta_i(x_i) | \sum_{i=1}^{m} A_i x_i = b \right\} \quad \text{in} \quad [8].$$

Global convergence of these ALM type methods are ensured under the convexity of $\theta_i(\cdot)(\forall i)$, and their worst-case $\mathcal{O}(1/t)$ convergence rate in an ergodic sense are also established.

To the best of our knowledge, there is only one paper discussing the ALM with indefinite proximal regularization for problem (1). That is [6], in which He et al. have proved the convergence results of the linearized ALM with indefinite proximal regularization. Note that the relaxation factor γ is set to 1 in the linearized ALM in [6], and numerical results indicate that the over relaxation, i.e., $\gamma \in (1, 2)$, can often speed up the convergence speed of the corresponding ALM. Therefore it is meaningful to extend the feasible set of γ from $\{1\}$ to the interval $(0, 2)$. Because problem (1) can also be viewed as a special case of the above three problems by setting $\theta_1(x_1) = \theta(x), A_1 = A$ and $\theta_i(x_i) = A_i = 0 (i \neq 1)$. By doing so, the feasible set of γ still cannot be extended to the interval $(0, 2)$. For example, the feasible set of γ in [14] is $(0, \frac{1+\sqrt{5}}{2})$, which is a proper subset of $(0, 2)$. In this paper, we are going to study the proximal ALM with $\gamma \in (0, 2)$, and prove its global convergence and establish its worst-case $\mathcal{O}(1/t)$ convergence rate under a more reasonable criterion than those used in [5, 17, 6, 7, 8, 14].

2 Preliminaries

Given any $x, y, z \in \mathcal{R}^n$, we have the identity

$$\langle x - y, x - z \rangle = \frac{1}{2}(\|x - y\|^2 + \|x - z\|^2 - \|y - z\|^2). \tag{3}$$

Definition 2.1. Some point (x^*, λ^*) is said to be a Karush-Kuhn-Tucker (KKT) point of problem (1) if the following two relationships hold:

$$\begin{cases} 0 \in \partial\theta(x^*) - A^\top \lambda^*, \\ Ax^* = b. \end{cases} \tag{4}$$

The solution set of KKT system (4), denoted by $\mathcal{W}^*$, is nonempty under the Assumption 1.2. By (4), we can characterize problem (1) as a mixed variational inequality problem: Find a vector $w^* = (x^*, \lambda^*) \in \mathcal{R}^{m+n}$ such that

$$\theta(x) - \theta(x^*) + (w - w^*)^\top F(w^*) \geq 0, \quad \forall w \in \mathcal{R}^{m+n}, \tag{5}$$

where

$$F(w) = \begin{pmatrix} -A^\top \lambda \\ Ax - b \end{pmatrix} = \begin{pmatrix} 0 & -A^\top \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} - \begin{pmatrix} 0 \\ b \end{pmatrix}.$$

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

Proposition 2.1. [10] The vector $\tilde{x} \in \mathcal{R}^n$ is an optimal solution to problem (1) if and only if there exists $r > 0$, such that

$$\theta(\tilde{x}) - \theta(x^*) + (\tilde{x} - x^*)^\top (-A^\top \lambda^*) + \frac{r}{2}\|A\tilde{x} - b\|^2 = 0, \tag{6}$$

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

Based on Proposition 2.1, in [10], Lin et al. proposed the following criterion to measure the convergence rate of their ALM

$$\theta(x_t) - \theta(x^*) + (x_t - x^*)^\top(-A^\top\lambda^*) + \frac{c}{2}\|Ax_t - b\|^2 \le \frac{C}{t+1}, \tag{7}$$

where $c > 0$. Obviously, the inequality (7) is motivated by the equality (6). Then, we shall use (7) to measure the convergence rate of our new method.

3 PALM-IPR and its global convergence

We now present the proximal ALM with indefinite proximal regularization (PALM-IPR) for problem (1).

Algorithm 1 The PALM-IPR for problem (1)

Input $\beta > 0, \gamma \in (0, 2)$, one symmetric matrix $G \in \mathcal{R}^{n\times n}$. Initialize $(x, \lambda) := (x^0, \lambda^0)$, $k := 0$.

while "not converged", **do**

(1) Compute (x^{k+1}, λ^{k+1}) in the alternating order by the following PALM-IPR procedure.

$$\begin{cases} x^{k+1} \in \text{argmin}_{x\in\mathcal{R}^n}\{\theta(x) + \frac{\beta}{2}\|Ax - b - \frac{\lambda^k}{\beta}\|^2 + \frac{1}{2}\|x - x^k\|_G^2\}, \\ \lambda^{k+1} = \lambda^k - \gamma\beta(Ax^{k+1} - b). \end{cases} \tag{8}$$

(2) Set $k := k + 1$.

end while

Output x^{k+1}.

Assumption 3.1. The proximal matrix G is set as $G = \alpha\tau I_n - \beta A^\top A$ with $\tau > \beta\|A^\top A\|$, $\max\left\{1 - c, \dfrac{1 + 4c - \varrho_3}{1 + 4c}\right\} < \alpha < 1$, where the constant c is set as $c = 1$ if $0 < \gamma < \frac{1+\sqrt{5}}{2}$, and $c \in \left(0, \dfrac{2-\gamma}{(\gamma-1)^2}\right)$ if $\frac{1+\sqrt{5}}{2} \le \gamma < 2$; and ϱ_3 is a constant defined in the following Lemma 3.3.

Let us further define two matrices to simplify our notation. Set

$$\bar{G} = \tau I_n - \beta A^\top A, \quad M = G + \beta c A^\top A. \tag{9}$$

Remark 3.1. From the definitions of G, $\bar{G}$ and M, we have

$$G = \alpha\bar{G} - (1 - \alpha)\beta A^\top A, \tag{10}$$

and

$$\bar{G} \succ 0, \quad M = \alpha\tau I_n + \beta(c - 1)A^\top A \succ 0.$$

Using the first-order optimization condition of the subproblem of PALM-IPR, we can deduce the following one-iteration result, and its proof is motivated by Lemma 3.1 of [14].

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

$$\begin{array}{rl} & \theta(x^*)-\theta(x^{k+1}) \\ \geq & \frac{1}{2}(\|x^{k+1}-x^*\|_G^2-\|x^k-x^*\|_G^2+\|x^{k+1}-x^k\|_G^2)+\langle Ax^{k+1}-Ax^*,-\lambda^*\rangle+\frac{2-\gamma}{2\beta\gamma^2}\|\lambda^{k+1}-\lambda^k\|^2 \\ & +\frac{1}{2\beta\gamma}(\|\lambda^{k+1}-\lambda^*\|^2-\|\lambda^k-\lambda^*\|^2)+\beta c\langle Ax^{k+1}-b,Ax^k-Ax^{k+1}\rangle \\ & +\frac{\beta c}{2}(\|Ax^{k+1}-Ax^*\|^2-\|Ax^k-Ax^*\|^2+\|Ax^{k+1}-Ax^k\|^2). \end{array} \tag{11}$$

Proof. From the first-order optimality condition for $x-$related subproblem of (8), we have

$$\begin{aligned} 0 &\in \partial\theta(x^{k+1})+\beta A^\top(Ax^{k+1}-b-\tfrac{1}{\beta}\lambda^k)+G(x^{k+1}-x^k) \\ &= \partial\theta(x^{k+1})-A^\top\tilde{\lambda}^k+G(x^{k+1}-x^k), \end{aligned} \tag{12}$$

where $\tilde{\lambda}^k=\lambda^k-\beta(Ax^{k+1}-b)$. Then, from the convexity of $\theta(\cdot)$ and (12), we have

$$\begin{array}{rl} & \theta(x^*)-\theta(x^{k+1}) \\ \geq & \langle A^\top\tilde{\lambda}^k-G(x^{k+1}-x^k),x^*-x^{k+1}\rangle \\ = & \langle x^{k+1}-x^*,G(x^{k+1}-x^k)\rangle+\langle A(x^{k+1}-x^*),-\tilde{\lambda}^k\rangle \\ = & \langle x^{k+1}-x^*,G(x^{k+1}-x^k)\rangle+\langle Ax^{k+1}-Ax^*,-\tilde{\lambda}^k\rangle \\ & +\beta c\langle Ax^{k+1}-b,Ax^k-Ax^{k+1}\rangle+\beta c\langle Ax^*-Ax^{k+1},Ax^k-Ax^{k+1}\rangle \\ = & \frac{1}{2}(\|x^{k+1}-x^*\|_G^2-\|x^k-x^*\|_G^2+\|x^{k+1}-x^k\|_G^2)+\langle Ax^{k+1}-Ax^*,-\tilde{\lambda}^k\rangle \\ & +\beta c\langle Ax^{k+1}-b,Ax^k-Ax^{k+1}\rangle \\ & +\frac{\beta c}{2}(\|Ax^{k+1}-Ax^*\|^2-\|Ax^k-Ax^*\|^2+\|Ax^{k+1}-Ax^k\|^2), \end{array} \tag{13}$$

where the second equality comes from $Ax^*=b$, and the third equality follows from identity (3).

Now, let us deal with the crossing term $\langle Ax^{k+1}-Ax^*,-\tilde{\lambda}^k\rangle$ on the right side of (13). From the updating formula for λ in (8), we have

$$\begin{array}{rl} & \langle Ax^{k+1}-Ax^*,-\tilde{\lambda}^k\rangle \\ = & \langle Ax^{k+1}-Ax^*,-\lambda^*\rangle+\frac{1}{\beta}\langle\lambda^k-\tilde{\lambda}^k,\lambda^*-\tilde{\lambda}^k\rangle \\ = & \langle Ax^{k+1}-Ax^*,-\lambda^*\rangle+\frac{1}{\beta}\langle\lambda^k-\frac{1}{\gamma}\lambda^{k+1}+\frac{1-\gamma}{\gamma}\lambda^k,\lambda^*-\frac{1}{\gamma}\lambda^{k+1}+\frac{1-\gamma}{\gamma}\lambda^k\rangle \\ = & \langle Ax^{k+1}-Ax^*,-\lambda^*\rangle+\frac{1}{\beta}\langle\frac{1}{\gamma}\lambda^k-\frac{1}{\gamma}\lambda^{k+1},\lambda^*-\lambda^k+\frac{1}{\gamma}\lambda^k-\frac{1}{\gamma}\lambda^{k+1}\rangle \\ = & \langle Ax^{k+1}-Ax^*,-\lambda^*\rangle+\frac{1}{\beta\gamma}\langle\lambda^k-\lambda^{k+1},\lambda^*-\lambda^k\rangle+\frac{1}{\beta\gamma^2}\|\lambda^{k+1}-\lambda^k\|^2 \\ = & \langle Ax^{k+1}-Ax^*,-\lambda^*\rangle+\frac{1}{2\beta\gamma}(\|\lambda^{k+1}-\lambda^*\|^2-\|\lambda^k-\lambda^*\|^2)+\frac{2-\gamma}{2\beta\gamma^2}\|\lambda^{k+1}-\lambda^k\|^2, \end{array} \tag{14}$$

where the second equality comes from $\lambda^{k+1}=\lambda^k-\gamma(\lambda^k-\tilde{\lambda}^k)$. Substituting (14) into (13), we obtain (11) immediately. This completes the proof.

The following lemma aims to further deal with the crossing term $\beta\langle Ax^{k+1}-b, Ax^k - Ax^{k+1}\rangle$ on the right side of (11), which is reduced to an expression with respect to x^{k-1}, x^k, x^{k+1}, and this expression is obtained by two-iteration results. Note that, the proof is motivated by Lemma 3.3 of [15].

Lemma 3.2. Let $\{(x^k,\lambda^k)\}_{k\geq 0}$ be the sequence generated by PALM-IPR. Then, we have

$$\begin{array}{rl} & \beta\langle Ax^{k+1}-b, Ax^k - Ax^{k+1}\rangle \\ \geq & (1-\gamma)\beta\langle Ax^k - Ax^{k+1}, Ax^k - b\rangle + \|x^{k+1}-x^k\|_G^2 \\ & +\langle x^k - x^{k+1}, G(x^k - x^{k-1})\rangle. \end{array} \tag{15}$$

Proof. From (12) and the convexity of $\theta(x)$, it holds that

$$\theta(x) - \theta(x^{k+1}) \geq \langle A^\top\tilde{\lambda}^k - G(x^{k+1}-x^k), x - x^{k+1}\rangle. \tag{16}$$

By setting $x = x^k$ in (16), we get

$$\theta(x^k) - \theta(x^{k+1}) - \langle A(x^{k+1}-x^k), -\tilde{\lambda}^k\rangle \geq \|x^{k+1}-x^k\|_G^2. \tag{17}$$

Due to the same reason in the $(k-1)$th iteration, it follows that

$$\theta(x^{k+1}) - \theta(x^k) - \langle A(x^k - x^{k+1}), -\tilde{\lambda}^{k-1}\rangle \geq \langle x^k - x^{k+1}, G(x^k - x^{k-1})\rangle. \tag{18}$$

From (17) and (18), we get

$$\langle A(x^k - x^{k+1}), \tilde{\lambda}^{k-1} - \tilde{\lambda}^k\rangle \geq \|x^{k+1}-x^k\|_G^2 + \langle x^k - x^{k+1}, G(x^k - x^{k-1})\rangle. \tag{19}$$

On the other hand, from the definition of $\tilde{\lambda}^k$, we have

$$\begin{array}{rl} & \tilde{\lambda}^{k-1} - \tilde{\lambda}^k \\ = & \tilde{\lambda}^{k-1} - [\lambda^k - \beta(Ax^{k+1}-b)] \\ = & \tilde{\lambda}^{k-1} - [\lambda^{k-1} - \gamma\beta(Ax^k - b) - \beta(Ax^{k+1}-b)] \\ = & -(1-\gamma)\beta(Ax^k - b) + \beta(Ax^{k+1}-b). \end{array}$$

Then, substituting the above equality into the left side of (19) and by some simple manipulations, we get (15). The proof is completed.

Now, substituting (15) into the right side of (11), we have an inequality with respective to x^{k-1}, x^k, x^{k+1}.

$$\begin{array}{rl} & \theta(x^*) - \theta(x^{k+1}) \\ \geq & \frac{1}{2}(\|x^{k+1}-x^*\|_G^2 - \|x^k - x^*\|_G^2 + \|x^{k+1}-x^k\|_G^2) + \langle Ax^{k+1}-Ax^*, -\lambda^*\rangle + \frac{2-\gamma}{2\beta\gamma^2}\|\lambda^{k+1}-\lambda^k\|^2 \\ & +\frac{1}{2\beta\gamma}(\|\lambda^{k+1}-\lambda^*\|^2 - \|\lambda^k - \lambda^*\|^2) + (1-\gamma)\beta c\langle Ax^k - Ax^{k+1}, Ax^k - b\rangle + c\|x^{k+1}-x^k\|_G^2 \\ & +c\langle x^k - x^{k+1}, G(x^k - x^{k-1})\rangle + \frac{\beta c}{2}(\|Ax^{k+1}-Ax^*\|^2 - \|Ax^k - Ax^*\|^2 + \|Ax^{k+1}-Ax^k\|^2). \end{array} \tag{20}$$

Now, there are three crossing terms on the right side of (20). The first one $\langle Ax^{k+1}-b,-\lambda^*\rangle$ is remaind temporarily. Let us deal with the second crossing term $(1-\gamma)\beta c\langle Ax^k - Ax^{k+1}, Ax^k - b\rangle$.

To make the following proof more concision, let us define several constants in advance:

(1). For $\gamma \in [\frac{1+\sqrt{5}}{2}, 2)$, we set $a_1 = (\gamma-1)c/(\gamma(2-\gamma))$ and $a_2 = 1/(\gamma(\gamma-1))$. Then by the definition of the constant c (see Assumption 3.1), we have $0 < a_1 < a_2$. Then, we further set $a = a_1 + (a_2 - a_1)/10$, which obviously belongs to the interval (a_1, a_2). Furthermore, we set

$$a_3 = \frac{2-\gamma}{2\beta\gamma^2} - \frac{(\gamma-1)c}{2a\beta\gamma^3}, \quad \text{and} \quad a_4 = \frac{\beta c}{2} - \frac{(\gamma-1)\gamma\beta a c}{2}. \tag{21}$$

(2). For $\gamma \in (0, 2)$, we define

$$\varrho_1 = \begin{cases} \frac{1}{2\beta\gamma^2}\max\{1-\gamma, 1-\gamma^{-1}\}, & \text{if } (0, \frac{1+\sqrt{5}}{2}), \\ \frac{(\gamma-1)c}{2a\beta\gamma^3}, & \text{if } [\frac{1+\sqrt{5}}{2}, 2), \end{cases}$$

and

$$\varrho_2 = \begin{cases} \frac{\min\{\gamma, 1+\gamma-\gamma^2\}}{2\beta\gamma^2}, & \text{if } (0, \frac{1+\sqrt{5}}{2}), \\ a_3, & \text{if } [\frac{1+\sqrt{5}}{2}, 2), \end{cases} \qquad \varrho_3 = \begin{cases} \min\{\gamma, 1+\gamma-\gamma^2\}, & \text{if } (0, \frac{1+\sqrt{5}}{2}), \\ \frac{2a_4}{\beta}, & \text{if } [\frac{1+\sqrt{5}}{2}, 2). \end{cases}$$

Lemma 3.3. Let $\{(x^k, \lambda^k)\}_{k\ge 0}$ be the sequence generated by PALM-IPR. Then, we have

$$\begin{aligned} &\frac{2-\gamma}{2\beta\gamma^2}\|\lambda^{k+1}-\lambda^k\|^2 + (1-\gamma)\beta c\langle A(x^k - x^{k+1}), Ax^k - b\rangle + \frac{\beta c}{2}\|Ax^{k+1} - Ax^k\|^2 \\ \ge\ & \varrho_1(\|\lambda^{k+1}-\lambda^k\|^2 - \|\lambda^k - \lambda^{k-1}\|^2) + \varrho_2\|\lambda^{k+1}-\lambda^k\|^2 + \frac{\varrho_3\beta}{2}\|A(x^k - x^{k+1})\|^2 . \end{aligned} \tag{22}$$

Proof. From the updating formula for λ in (8), we get

$$\langle A(x^k - x^{k+1}), Ax^k - b\rangle = \frac{1}{\beta\gamma}\langle A(x^k - x^{k+1}), \lambda^{k-1} - \lambda^k\rangle. \tag{23}$$

Then, by the Cauchy-Schwartz inequality and the definition of the constant c, we obtain

$$\begin{cases} (1-\gamma)\beta c\langle A(x^k - x^{k+1}), (\lambda^{k-1}-\lambda^k)/(\beta\gamma)\rangle \ge -\frac{(1-\gamma)\beta}{2}(\|A(x^k - x^{k+1})\|^2 + \frac{1}{\gamma^2\beta^2}\|\lambda^{k-1}-\lambda^k\|^2), \\ \hfill \text{if } \gamma \in (0,1] \\ (1-\gamma)\beta c\langle A(x^k - x^{k+1}), (\lambda^{k-1}-\lambda^k)/(\beta\gamma)\rangle \ge -\frac{(\gamma-1)\beta}{2}(\gamma\|A(x^k - x^{k+1})\|^2 + \frac{1}{\gamma^3\beta^2}\|\lambda^{k-1}-\lambda^k\|^2), \\ \hfill \text{if } \gamma \in (1, \frac{1+\sqrt{5}}{2}) \\ (1-\gamma)\beta c\langle A(x^k - x^{k+1}), (\lambda^{k-1}-\lambda^k)/(\beta\gamma)\rangle \ge -\frac{(\gamma-1)\beta c}{2}(\gamma a\|A(x^k - x^{k+1})\|^2 + \frac{1}{a\gamma^3\beta^2}\|\lambda^{k-1}-\lambda^k\|^2), \\ \hfill \text{if } \gamma \in [\frac{1+\sqrt{5}}{2}, 2). \end{cases}$$

(1). If $\gamma \in (0, \frac{1+\sqrt{5}}{2})$, substituting the first two inequalities into the left side of (22), we get

$$\begin{aligned} &\frac{2-\gamma}{2\beta\gamma^2}\|\lambda^{k+1}-\lambda^k\|^2 + (1-\gamma)\beta c\langle A(x^k - x^{k+1}), Ax^k - b\rangle + \frac{\beta c}{2}\|Ax^{k+1} - Ax^k\|^2 \\ \ge\ & \frac{1}{2\beta\gamma^2}\max\{1-\gamma, 1-\gamma^{-1}\}(\|\lambda^{k+1}-\lambda^k\|^2 - \|\lambda^k - \lambda^{k-1}\|^2) \\ &+\frac{\beta}{2}\min\{\gamma, 1+\gamma-\gamma^2\}(\frac{1}{\gamma^3\beta^2}\|\lambda^{k+1}-\lambda^k\|^2 + \|A(x^k - x^{k+1})\|^2). \end{aligned}$$

(2). If $\gamma \in [\frac{1+\sqrt{5}}{2}, 2)$, substituting the third inequality into the left side of (22), we have

$$\begin{aligned}
&\frac{2-\gamma}{2\beta\gamma^2}\|\lambda^{k+1}-\lambda^k\|^2 + (1-\gamma)\beta c\langle A(x^k - x^{k+1}), Ax^k - b\rangle + \frac{\beta c}{2}\|Ax^{k+1} - Ax^k\|^2 \\
\geq\ & \frac{(\gamma-1)c}{2a\beta\gamma^3}(\|\lambda^{k+1}-\lambda^k\|^2 - \|\lambda^k - \lambda^{k-1}\|^2) \\
&+\left(\frac{2-\gamma}{2\beta\gamma^2} - \frac{(\gamma-1)c}{2a\beta\gamma^3}\right)\|\lambda^{k+1}-\lambda^k\|^2 + \left(\frac{\beta c}{2} - \frac{(\gamma-1)\gamma\beta a c}{2}\right)\|A(x^k - x^{k+1})\|^2 \\
=\ & \frac{(\gamma-1)c}{2a\beta\gamma^3}(\|\lambda^{k+1}-\lambda^k\|^2 - \|\lambda^k - \lambda^{k-1}\|^2) + a_3\|\lambda^{k+1}-\lambda^k\|^2 + a_4\|A(x^k - x^{k+1})\|^2,
\end{aligned}$$

where the constants a_3, a_4 are defined by (21). Combining the above two cases, we get the assertion (22), and the lemma is proved.

Remark 3.2. The proximal matrix G defined in Assumption 3.1 maybe indefinite. In fact, when $A = I_n, \gamma = 1.7, \beta = 1, \tau = 1.01$, then we set $c = 0.99 \times \frac{2-\gamma}{(\gamma-1)^2} \cong 0.3031$, and so $a_1 \cong 0.4160, a_2 \cong 0.8403$, $a \cong 0.4584$, $a_4 \cong 0.1377$, and $\varrho_3 \cong 0.1377$. Therefore, the feasible set of α is the interval $(0.9377, 1)$ by Assumption 3.1. If we set $\alpha = 0.94$, then $G = -0.0506 I_n$, which is obviously indefinite.

Now, substituting (22) into the right side of (20), we get

$$\begin{aligned}
&\theta(x^*) - \theta(x^{k+1}) \\
\geq\ & \frac{1}{2}(\|x^{k+1}-x^*\|_G^2 - \|x^k - x^*\|_G^2 + \|x^{k+1}-x^k\|_G^2) + \langle Ax^{k+1} - Ax^*, -\lambda^*\rangle \\
&+\frac{1}{2\beta\gamma}(\|\lambda^{k+1}-\lambda^*\|^2 - \|\lambda^k - \lambda^*\|^2) + c\|x^{k+1}-x^k\|_G^2 \\
&+c\langle x^k - x^{k+1}, G(x^k - x^{k-1})\rangle + \frac{\beta c}{2}(\|Ax^{k+1} - Ax^*\|^2 - \|Ax^k - Ax^*\|^2) \\
&+ \varrho_1(\|\lambda^{k+1}-\lambda^k\|^2 - \|\lambda^k - \lambda^{k-1}\|^2) + \varrho_2\|\lambda^{k+1}-\lambda^k\|^2 + \frac{\varrho_3\beta}{2}\|A(x^k - x^{k+1})\|^2
\end{aligned} \tag{24}$$

Lemma 3.4. Let $\{(x^k, \lambda^k)\}_{k\geq 0}$ be the sequence generated by PALM-IPR. Then, we have

$$\begin{aligned}
&\frac{1}{2}(\|x^{k+1}-x^*\|_G^2 - \|x^k - x^*\|_G^2 + \|x^{k+1}-x^k\|_G^2) + c\|x^{k+1}-x^k\|_G^2 \\
&+c\langle x^k - x^{k+1}, G(x^k - x^{k-1})\rangle + \frac{\beta c}{2}(\|Ax^{k+1} - Ax^*\|^2 - \|Ax^k - Ax^*\|^2) \\
&+\frac{\varrho_3\beta}{2}\|A(x^k - x^{k+1})\|^2 \\
\geq\ & \frac{1}{2}(\|x^{k+1}-x^*\|_M^2 - \|x^k - x^*\|_M^2) + \frac{\alpha}{2}\|x^{k+1}-x^k\|_{\bar{G}}^2 + \frac{\alpha c}{2}(\|x^{k+1}-x^k\|_{\bar{G}}^2 - \|x^k - x^{k-1}\|_{\bar{G}}^2) \\
&+\frac{(1-\alpha)\beta c}{2}(\|A(x^k - x^{k+1})\|^2 - \|A(x^k - x^{k-1})\|^2) + \frac{\beta(\varrho_3 - (1-\alpha)(1+4c))}{2}\|A(x^k - x^{k+1})\|^2.
\end{aligned} \tag{25}$$

Proof. By the definitions of the matrices G, M and (10), the left side of (25) can be written as

$$
\begin{aligned}
& \frac{1}{2}(\|x^{k+1}-x^*\|^2_{G+\beta cA^\top A}-\|x^k-x^*\|^2_{G+\beta cA^\top A})+\frac{1}{2}\|x^{k+1}-x^k\|^2_{G+\beta\varrho_3 A^\top A} \\
& +c\|x^{k+1}-x^k\|^2_G+c\langle x^k-x^{k+1}, G(x^k-x^{k-1})\rangle \\
= \; & \frac{1}{2}(\|x^{k+1}-x^*\|^2_M-\|x^k-x^*\|^2_M)+\frac{1}{2}\|x^{k+1}-x^k\|^2_{(1+2c)\alpha\bar{G}-((1+2c)(1-\alpha)-\varrho_3)\beta A^\top A} \\
& +\alpha c(x^k-x^{k+1})^\top\bar{G}(x^k-x^{k-1})-(1-\alpha)\beta c(x^k-x^{k+1})^\top(A^\top A)(x^k-x^{k-1}) \\
\geq \; & \frac{1}{2}(\|x^{k+1}-x^*\|^2_M-\|x^k-x^*\|^2_M)+\frac{1}{2}\|x^{k+1}-x^k\|^2_{(1+2c)\alpha\bar{G}-((1+2c)(1-\alpha)-\varrho_3)\beta A^\top A} \\
& -\frac{\alpha c}{2}(\|x^k-x^{k+1}\|^2_{\bar{G}}+\|x^k-x^{k-1}\|^2_{\bar{G}})-\frac{(1-\alpha)\beta c}{2}(\|A(x^k-x^{k+1})\|^2+\|A(x^k-x^{k-1})\|^2) \\
= \; & \frac{1}{2}(\|x^{k+1}-x^*\|^2_M-\|x^k-x^*\|^2_M)+\frac{\alpha}{2}\|x^{k+1}-x^k\|^2_{\bar{G}}+\frac{\alpha c}{2}(\|x^{k+1}-x^k\|^2_{\bar{G}}-\|x^k-x^{k-1}\|^2_{\bar{G}}) \\
& +\frac{(1-\alpha)\beta c}{2}(\|A(x^k-x^{k+1})\|^2-\|A(x^k-x^{k-1})\|^2)+\frac{\beta(\varrho_3-(1-\alpha)(1+4c))}{2}\|A(x^k-x^{k+1})\|^2,
\end{aligned}
$$

where the inequality follows from the Cauchy-Schwartz inequality. The proof is completed.

Substituting (25) into (24), we have

$$
\begin{aligned}
& \theta(x^*)-\theta(x^{k+1})+\langle x^{k+1}-x^*, A^\top\lambda^*\rangle \\
\geq \; & \frac{1}{2\beta\gamma}(\|\lambda^{k+1}-\lambda^*\|^2-\|\lambda^k-\lambda^*\|^2)+\varrho_1(\|\lambda^{k+1}-\lambda^k\|^2-\|\lambda^k-\lambda^{k-1}\|^2)+\varrho_2\|\lambda^{k+1}-\lambda^k\|^2 \\
& +\frac{1}{2}(\|x^{k+1}-x^*\|^2_M-\|x^k-x^*\|^2_M)+\frac{\alpha}{2}\|x^{k+1}-x^k\|^2_{\bar{G}}+\frac{\alpha c}{2}(\|x^{k+1}-x^k\|^2_{\bar{G}}-\|x^k-x^{k-1}\|^2_{\bar{G}}) \\
& +\frac{(1-\alpha)\beta c}{2}(\|A(x^k-x^{k+1})\|^2-\|A(x^k-x^{k-1})\|^2)+\frac{\beta(\varrho_3-(1-\alpha)(1+4c))}{2}\|A(x^k-x^{k+1})\|^2 .
\end{aligned}
\tag{26}
$$

Based on (26), now we can prove the global convergence of PALM-IPR.

Theorem 3.1. Let $\{(x^k,\lambda^k)\}_{k\geq 0}$ be the sequence generated by PALM-IPR. Then, $\{(x^k,\lambda^k)\}_{k\geq 0}$ is bounded and converges to a point $(x^\infty,\lambda^\infty)$ that satisfies the KKT condition in (4).

Proof. From (26), we get

$$
\begin{aligned}
& \theta(x^{k+1})-\theta(x^*)+\langle x^{k+1}-x^*, -A^\top\lambda^*\rangle+\varrho_2\|\lambda^{k+1}-\lambda^k\|^2+\frac{\alpha}{2}\|x^{k+1}-x^k\|^2_{\bar{G}} \\
& +\frac{\beta(\varrho_3-(1-\alpha)(1+4c))}{2}\|A(x^k-x^{k+1})\|^2 \\
\leq \; & \frac{1}{2\beta\gamma}(\|\lambda^k-\lambda^*\|^2-\|\lambda^{k+1}-\lambda^*\|^2)+\varrho_1(\|\lambda^k-\lambda^{k-1}\|^2-\|\lambda^{k+1}-\lambda^k\|^2) \\
& +\frac{1}{2}(\|x^k-x^*\|^2_M-\|x^{k+1}-x^*\|^2_M)+\frac{\alpha c}{2}(\|x^k-x^{k-1}\|^2_{\bar{G}}-\|x^{k+1}-x^k\|^2_{\bar{G}}) \\
& +\frac{(1-\alpha)\beta c}{2}(\|A(x^k-x^{k-1})\|^2-\|A(x^k-x^{k+1})\|^2) .
\end{aligned}
$$

This implies that

$$\begin{aligned} & \frac{1}{2\beta\gamma}\|\lambda^{k+1}-\lambda^*\|^2+\varrho_1\|\lambda^{k+1}-\lambda^k\|^2+\frac{1}{2}\|x^{k+1}-x^*\|_M^2+\frac{\alpha c}{2}\|x^{k+1}-x^k\|_{\bar{G}}^2 \\ & +\frac{(1-\alpha)\beta c}{2}\|A(x^k-x^{k+1})\|^2 \\ \leq \; & \frac{1}{2\beta\gamma}\|\lambda^k-\lambda^*\|^2+\varrho_1\|\lambda^k-\lambda^{k-1}\|^2+\|x^k-x^*\|_M^2+\frac{\alpha c}{2}\|x^k-x^{k-1}\|_{\bar{G}}^2 \\ & +\frac{(1-\alpha)\beta c}{2}\|A(x^k-x^{k-1})\|^2 \\ & -\left(\varrho_2\|\lambda^{k+1}-\lambda^k\|^2+\frac{\alpha}{2}\|x^{k+1}-x^k\|_{\bar{G}}^2+\frac{\beta(\varrho_3-(1-\alpha)(1+4c))}{2}\|A(x^k-x^{k+1})\|^2\right). \end{aligned}$$

Since $\dfrac{1+4c-\varrho_3}{1+4c}<\alpha<1$, the above inequality indicates that the sequences $\{\|x^{k+1}-x^*\|_M^2\}$, $\{\|\lambda^{k+1}-\lambda^*\|\}$ are both bounded, and

$$\lim_{k\to\infty}\|x^k-x^{k+1}\|_{\bar{G}}=\lim_{k\to\infty}\|\lambda^k-\lambda^{k+1}\|=0, \tag{27}$$

Since $\bar{G}\succ 0$, we get

$$\lim_{k\to\infty}\|x^k-x^{k+1}\|=0. \tag{28}$$

On the other hand, since the sequence $\{\|x^{k+1}-x^*\|_M\}$ is bounded and $M\succ 0$, we obtain the sequence $\{\|x^{k+1}-x^*\|\}$ is bounded. Therefore, the sequence $\{(x^k,\lambda^k)\}$ generated by PALM-IPR is bounded. Then, it has at least one cluster point, saying $(x^\infty,\lambda^\infty)$, and suppose that the subsequence $\{(x^{k_i},\lambda^{k_i})\}$ converges to $(x^\infty,\lambda^\infty)$. From (12), we have

$$0\in\partial\theta(x^{k+1})+\beta A^\top(Ax^{k+1}-b-\frac{1}{\beta}\lambda^k)+G(x^{k+1}-x^k).$$

That is,

$$0\in\partial\theta(x^{k+1})+A^\top[(\lambda^k-\lambda^{k+1})/\gamma-\lambda^k]+G(x^{k+1}-x^k).$$

Then, taking limits on both sides of the above equality along the subsequence $\{(x^{k_i},\lambda^{k_i})\}$ and using (27)-(28), we have

$$A^\top\lambda^\infty\in\partial\theta(x^\infty).$$

Furthermore, taking limits on both sides of the updating formula for λ in (8) along the subsequence $\{(x^{k_i},\lambda^{k_i})\}$ and using (28), we have

$$Ax^\infty=b.$$

Therefore, $(x^\infty, \lambda^\infty)$ satisfies the KKT condition (4). Hence, we have

$$\begin{aligned}
& \frac{1}{2\beta\gamma}\|\lambda^{k+1}-\lambda^\infty\|^2+\varrho_1\|\lambda^{k+1}-\lambda^k\|^2+\frac{1}{2}\|x^{k+1}-x^\infty\|_M^2+\frac{\alpha c}{2}\|x^{k+1}-x^k\|_{\bar{G}}^2 \\
& +\frac{(1-\alpha)\beta c}{2}\|A(x^k-x^{k+1})\|^2 \\
\leq\ & \frac{1}{2\beta\gamma}\|\lambda^k-\lambda^\infty\|^2+\varrho_1\|\lambda^k-\lambda^{k-1}\|^2+\|x^k-x^\infty\|_M^2+\frac{\alpha c}{2}\|x^k-x^{k-1}\|_{\bar{G}}^2 \\
& +\frac{(1-\alpha)\beta c}{2}\|A(x^k-x^{k-1})\|^2 .
\end{aligned}$$

From (27)-(28), we have that for any given $\varepsilon > 0$, there exists $l_0 > 0$, such that

$$\varrho_1\|\lambda^k-\lambda^{k-1}\|^2+\frac{\alpha c}{2}\|x^k-x^{k-1}\|_{\bar{G}}^2+\frac{(1-\alpha)\beta c}{2}\|A(x^k-x^{k-1})\|^2<\frac{\varepsilon}{2}, \quad \forall k \geq l_0.$$

Since $(x^{k_i}, \lambda^{k_i}) \to (x^\infty, \lambda^\infty)$ for $i \to \infty$, there exists $k_l > l_0$, such that

$$\frac{1}{2\beta\gamma}\|\lambda^{k_l}-\lambda^\infty\|^2+\|x^{k_l}-x^\infty\|_M^2<\frac{\varepsilon}{2}.$$

Then, the above three inequalities lead to that, for any $k > k_l$, we have

$$\begin{aligned}
& \frac{1}{2\beta\gamma}\|\lambda^k-\lambda^\infty\|^2+\frac{1}{2}\|x^k-x^\infty\|_M^2 \\
\leq\ & \frac{1}{2\beta\gamma}\|\lambda^{k_l}-\lambda^\infty\|^2+\|x^{k_l}-x^\infty\|_M^2+\varrho_1\|\lambda^{k_l}-\lambda^{k_l-1}\|^2 \\
& +\frac{\alpha c}{2}\|x^{k_l}-x^{k_l-1}\|_{\bar{G}}^2+\frac{(1-\alpha)\beta c}{2}\|A(x^{k_l}-x^{k_l-1})\|^2 \\
<\ & \varepsilon.
\end{aligned}$$

Therefore the whole sequence $\{(x^k, \lambda^k)\}_{k\geq 0}$ converges to the $(x^\infty, \lambda^\infty)$. The proof is completed.

4 Convergence rate

In this section, we are going to prove the worst-case $\mathcal{O}(1/t)$ convergence rate in an ergodic sense of PALM-IPR.

Theorem 4.1. Let $\{(x^k, \lambda^k)\}_{k\geq 0}$ be the sequence generated by PALM-IPR and let

$$x_t=\frac{1}{t}\sum_{k=1}^{t}x^{k+1} ,$$

where t is a positive integer. Then, we have

$$\theta(x_t)-\theta(x^*)+\langle x_t-x^*,-A^\top\lambda^*\rangle+\varrho_2\gamma^2\beta^2\|Ax_t-b\|^2\leq\frac{D}{t}. \tag{29}$$

where (x^*, λ^*) is a vector satisfying the KKT system (4), and D is a constant defined by

$$\begin{aligned} D = & \frac{1}{2\beta\gamma}\|\lambda^1 - \lambda^*\|^2 + \varrho_1\|\lambda^1 - \lambda^0\|^2 + \frac{1}{2}\|x^1 - x^*\|_M^2 \\ & + \frac{\alpha c}{2}\|x^1 - x^0\|_{\bar{G}}^2 + \frac{(1-\alpha)\beta c}{2}\|A(x^1 - x^0)\|^2. \end{aligned} \tag{30}$$

Proof. Summing (26) over $k = 1, 2, \ldots, t$, we have

$$\begin{aligned} & \sum_{k=1}^{t}\Big[\theta(x^{k+1}) - \theta(x^*) + \langle x^{k+1} - x^*, -A^\top\lambda^*\rangle + \varrho_2\|\lambda^{k+1} - \lambda^k\|^2\Big] \\ \le & \frac{1}{2\beta\gamma}\|\lambda^1 - \lambda^*\|^2 + \varrho_1\|\lambda^1 - \lambda^0\|^2 + \frac{1}{2}\|x^1 - x^*\|_M^2 \\ & + \frac{\alpha c}{2}\|x^1 - x^0\|_{\bar{G}}^2 + \frac{(1-\alpha)\beta c}{2}\|A(x^1 - x^0)\|^2. \end{aligned} \tag{31}$$

Therefore, dividing (31) by t and using the convexity of $\theta(\cdot)$ lead to

$$\theta(x_t) - \theta(x^*) + \langle x_t - x^*, -A^\top\lambda^*\rangle + \frac{\varrho_2}{t}\sum_{k=1}^{t}\|\lambda^{k+1} - \lambda^k\|^2 \le \frac{D}{t}, \tag{32}$$

where the constant D is defined by (30).

Compared (32) with (7), we have to deal with the term $\frac{\varrho_2}{t}\sum_{k=1}^{t}\|\lambda^{k+1} - \lambda^k\|^2$ of (32). In fact, from the updating formula for λ in (8), we get

$$\begin{aligned} & \frac{\varrho_2}{t}\sum_{k=1}^{t}\|\lambda^{k+1} - \lambda^k\|^2 \\ = & \frac{\varrho_2}{t}\sum_{k=1}^{t}\|\gamma\beta(Ax^{k+1} - b)\|^2 \\ = & \varrho_2\gamma^2\beta^2\sum_{k=1}^{t}\frac{1}{t}\|Ax^{k+1} - b\|^2 \\ \ge & \varrho_2\gamma^2\beta^2\left\|A\frac{\sum_{k=1}^{t}x^{k+1}}{t} - b\right\|^2 \\ = & \varrho_2\gamma^2\beta^2\|Ax_t - b\|^2, \end{aligned}$$

where the inequality follows from the convexity of $\|\cdot\|^2$. Then, substituting the above inequality into (32), we get the desired result (29). This completes the proof.

5 Numerical results

In this section, we apply PALM-IPR to some practical applications of the studied problem, and report the numerical results. All the codes were written by Matlab R2010a.

Problem Image reconstruction

Let the matrix $W \in \mathcal{R}^{l\times n}$ be a wavelet dictionary with $l = l_1 \times l_2$, the vector $\mathbf{x} \in \mathcal{R}^l$ a digital image. Set $\mathbf{x} = Wx$, and x is a sparse vector. The image reconstruction problem is to recover the clean image $x \in \mathcal{R}^n$ based on some observation b, which can modelled as:

$$\min\{\|x\|_1 | BWx = b\},$$

where $B \in \mathcal{R}^{m\times l}$ is a diagonal matrix whose elements are either 0 (missing pixels) or 1 (known pixels), and $B = SH$, where $S \in \mathcal{R}^{m\times l}$ is a downsampling matrix generated by the following subroutines

```
dtx = 64; dty = 64; S1 = rand(dtx,dty)>0.6;
```

and

```
S2 = ones(n1/dtx,n2/dty); S = kron(S1,S2).
```

Furthermore, the blurry matrix $H =$ `fspecial('disk',7)`, which can be diagonalized by the discrete consine transform.

In order to demonstrate its efficiency, we compare PALM-IPR with the primal-dual hybrid gradient method in [3] (namely PDHG-HY). In the numerical experiment, for PALM-IPR, we set $\gamma = 1.9, \beta = 0.25 * \texttt{mean}(\texttt{mean}(\texttt{abs}(A^\top b)))$ and other parameters are the same as problem 1; For the parameters of PDHG-HY, we use the default values in [3]. Furthermore, the maximize iteration number is set 500. The initial point is $(x^0, \lambda^0) = (W^\top(b), 0)$. We test the 256×256 image of Cameraman.tif for the image reconstruction problem. We use the signal to noise ratio (SNR) defined as

$$\mathrm{SNR} = 20\log_{10}\frac{\|\mathbf{x}\|}{\|\tilde{\mathbf{x}} - \mathbf{x}\|}.$$

to assess the restoration performance qualitatively, in which $\mathbf{x}$ is the true image, and $\tilde{\mathbf{x}}$ is the restored image.

The original image, degraded image and restored images are shown in Figure 1, which indicate that both methods have recovered the degraded image with high quality. Here we also plot the evolutions of SNR with respect to the number of iteration counter k in Figure 2, which indicate that the SNR of PALM-IPR is smaller than that of PDHG-HY at the first stage, and after $k = 57$, the former becomes larger than the latter. In fact, at $k = 500$, the SNRs of PALM-IPR and PDHG-HY are 20.6829 and 19.4507, respectively.

6 Conclusions

In this paper, we have proposed a proximal augmented Lagrangian method with indefinite proximal regularization for linearly constrained convex programming. By introducing a constant c, we have proved

Original Degraded PALM−IPR PDHG−HY

Figure 1: Original image, degraded image and restored images

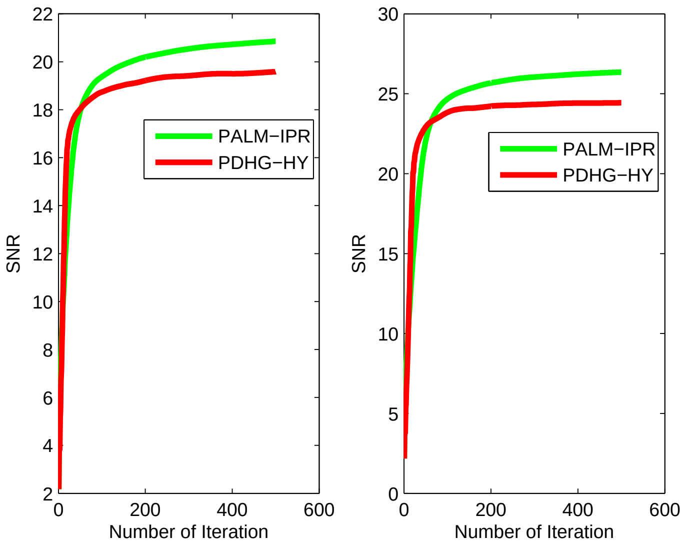

Figure 2: Evolution of SNR with respective to the number of iteration counter

the global convergence of PALM-IPR with the relaxation factor $\gamma \in (0,2)$. Furthermore, under a more reasonable criterion than those in [5, 17], we have established the worst-case $\mathcal{O}(1/t)$ convergence rate in an ergodic sense of PALM-IPR. Some numerical results are given, which illustrate that PALM-IPR performs better than some state-of-the-art solvers.

References

[1] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Foundations and Trends in Machine Learning, 3, 1-122, 2011.

[2] M. Fazel, T.K. Pong, D.F. Sun, P. Tseng, Hankel matrix rank minimization with applications to system identification and realization, SIAM J. Matrix Anal. Appl., 34: 946-977, 2013.

[3] B.S. He, X.M. Yuan, X, W.X. Zhang, A customized proximal point algorithm for convex minimization with linear constraints, Comput. Optim. Appl., 56, 559-572, 2013.

[4] B.S. He, L.Z. Liao, D.R. Han, H Yang, A new inexact alternating directions method for monotone variational inequalities, Mathematical Programming, 92(1), 103-118, 2002.

[5] B.S. He, X.M. Yuan, On the $\mathcal{O}(1/n)$ convergence rate of the Douglas-Rachford alternating direction method, SIAM J. Numer. Anal., 50(2), 700-709, 2012.

[6] B.S. He, F. Ma, X.M. Yuan, Linearized alternating direction method of multipliers via positive-indefinite proximal regularization for convex programming, 2016, manuscript.

[7] B.S. He, X.M. Yuan, Improving an ADMM-like splitting method via positive-indefinite proximal regularization for three-block separable convex minimization, 2016, manuscript.

[8] B.S. He, F. Ma, X.M. Yuan, Positive-indefinite proximal augmented Lagrangian method and its application to full Jacobian splitting for multi-block separable convex minimization problems, 2016, manuscript.

[9] M.R. Hestenes, Multiplier and gradient methods, J. Optim. Theory. Appl., 4, 303-320, 1969.

[10] Z.C. Lin, R.S. Liu, H. Li, Linearized alternating direction method with parallel splitting and adaptive penalty for separable convex programs in machine learning, Machine Learning (Special Issue for ACML2013), 99(2), 287-325, 2015.

[11] F. Ma, M.F. Ni, A class of customized proximal point algorithms for linearly constrained convex optimization, Comp. Appl. Math., 2016 (Accepted).

[12] Y.E. Nesterov, Smooth minimization of non-smooth functions, Math. Program., 103, no. 1, Ser. A, 127-152, 2005.

[13] M.J. Powell, A method for non-linear constraints in minimization problems. In: Fletcher R (ed) Optimization. Academic Press, Waltham Massachusetts, 283-298, 1969.

[14] M. Sun, J. Liu, The convergence rate of the proximal alternating direction method of multipliers with indefinite proximal regularization, Journal of Inequalities and Applications (2017) 2017: 19.

[15] M.H. Xu, Proximal alternating directions method for structured variational inequalities, J. Optim. Theory. Appl., 134, 107-117, 2007.

[16] M.H. Xu, T. Wu, A class of linearized proximal alternating direction methods, J. Optim. Theory Appl., 155, 321-337, 2011.

[17] Y.Y. Xu, Accelerated first-order primal-dual proximal methods for linearly constrained composite convex programming, 2016, manuscript.

[18] J.F. Yang, X.M. Yuan, Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization, Mathematics of Computation, 82(281), 301-329, 2013.

[19] M. Sun, Y.J. Wang, J. Liu, Generalized Peaceman-Rachford splitting method for multiple-block separable convex programming with applications to robust PCA, Calcolo, 54(1), 77-94, 2017.

[20] H.C. Sun, M. Sun, Y.J. Wang, Proximal ADMM with larger step size for two-block separable convex programming and its application to the correlation matrices calibrating problems, Journal of Nonlinear Sciences and Applications, 2017 (Accepted)

[21] J. Liu, Y.R. Duan, M. Sun, A symmetric version of the generalized alternating direction method of multipliers for two-block separable convex programming, Journal of Inequalities and Applications, (2017) 2017: 129.

[22] H.C. Sun, M.Y. Tian, M. Sun, The symmetric ADMM with indefinite proximal regularization and its application, Journal of Inequalities and Applications, (2017) 2017: 172.

[23] H.C. Sun, J. Liu, M. Sun, A proximal fully parallel splitting method for stable principal component pursuit, Mathematical Problems in Engineering, 2017 (Accepted)

[24] M. Sun, H.C. Sun, Y.J. Wang, Two proximal splitting methods for multi-block separable programming with applications to stable principal component pursuit, Journal of Applied Mathematics and Computing, 2017 (Accepted)