On Relaxation of Some Customized Proximal Point Algorithms for Convex Minimization: From Variational Inequality Perspective

Feng Ma[1]

Abstract. The proximal point algorithm (PPA) is a fundamental method for convex programming. When PPA applied to solve linearly constrained convex problems, we may prefer to choose an appropriate metric matrix to define the proximal regularization, so that the computational burden of the resulted PPA can be reduced, and in most cases, even admit closed form or efficient solutions. This idea results in the so called customized PPA (also known as preconditioned PPA), and it covers the linearized ALM, the primal-dual hybrid gradient algorithm (PDHG), ADMM as special cases. To accelerate the convergence of the customized PPA, a simple scheme is to use a relaxation parameter $\gamma \in (0, 2)$ to over-relax the essential variables (the variables that are really involved in the iterations), and has worked well in practice. However, when the primal variables are included in the essential, this relaxation strategy may destroy the inherent structure we may preferred in the primal variables, e.g., sparsity, low rank, non-negative, etc. In this paper we treat some customized PPA algorithms uniformly by a mixed variational inequality approach, and propose an alternative relaxation strategy to speedup the convergence. Our analysis is based on correcting the dual variables individually. From variational inequality perspective, we establish the global convergence and a worst-case $\mathcal{O}(1/t)$ convergence rate for this series of relaxed algorithms. Finally, we demonstrate the performance improvements by some numerical results.

Keywords. Convex minimization, Proximal point algorithm, Relaxation, Augmented Lagrangian method.

1 Introduction

The problem concerned in this paper is the following convex minimization problems

$$\min\{\theta(x) \mid Ax = b,\, x \in \mathcal{X}\}, \tag{1.1}$$

where $A \in \Re^{m\times n}$, $b \in \Re^m$, $\mathcal{X} \subset \Re^n$ is a closed convex set, $\theta : \Re^n \to \Re$ is a generic convex but not necessarily smooth function. Throughout, the solution set of (1.1) is assumed to be nonempty.

We first reformulate (1.1) as a variational inequality form, so we can take a brief look at the algorithms to solve this problem. The Lagrangian function of the optimization problem (1.1) is

$$L(x, \lambda) = \theta(x) - \lambda^T(Ax - b),$$

which is defined on $\mathcal{X} \times \Re^m$. Let (x^*, λ^*) be a saddle point of the Lagrange function. Then we have

$$L_{\lambda\in\Re^m}(x^*, \lambda) \le L(x^*, \lambda^*) \le L_{x\in\mathcal{X}}(x, \lambda^*).$$

[1]High-Tech Institute of Xi'an, Xi'an, 710025, Shaanxi, China. Email: mafengnju@gmail.com.

Rearrange the above saddle point inequalities, we have the following optimal conditions:

$$\left\{\begin{array}{ll} x^* \in \mathcal{X}, & \theta(x) - \theta(x^*) + (x - x^*)^T(-A^T\lambda^*) \geq 0, \quad \forall\, x \in \mathcal{X}, \\ \lambda^* \in \Re^m, & (\lambda - \lambda^*)^T(Ax^* - b) \geq 0, \quad \forall\, \lambda \in \Re^m. \end{array}\right. \tag{1.2}$$

More compactly, the inequalities (1.2) can be characterized as a monotone variational inequality (VI):

$$\text{VI}(\Omega, F, \theta) \quad w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega, \tag{1.3a}$$

where

$$w = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad u = x, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ Ax - b \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{X} \times \Re^m. \tag{1.3b}$$

It is easily verified that $(w_1 - w_2)^T\big(F(w_1) - F(w_2)\big) \geq 0$ for any $w_1, w_2 \in \Omega$, thus the operator F defined in (1.3b) is monotone. Since the solution set of (1.1) is assumed to be nonempty, the solution set of VI (1.3), denoted by Ω^*, is also nonempty.

As we have mentioned, our analysis will be conducted in the variational inequality context. So, the VI reformulation (1.3) serves as the starting point for further analysis. The algorithms used to solve the model (1.1) amounts to finding a solution point of VI (1.3). We now briefly revisit some relevant algorithms in VI form.

The classical proximal point algorithm (PPA) [21, 24] is a fundamental method to solve the VI. Starting from an initial point w^0, the PPA iterates as follows

$$w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T\{F(w^{k+1}) + r(w^{k+1} - w^k)\} \geq 0, \quad \forall\, w \in \Omega. \tag{1.4}$$

where $r > 0$ is a regularization parameter. PPA plays a fundamental role both theoretically and algorithmically in the area of the optimization. Indeed, the augmented Lagrangian method (ALM) [19, 23] can be interpreted as applying the PPA to the dual of (1.1) [24]. We now verify the interpretation by revisiting the ALM (1.5) in the VI context. For (1.1), the ALM reads as

$$\left\{\begin{array}{ll} x^{k+1} = \arg\min\{\theta(x) - (\lambda^k)^T(Ax - b) + \frac{\beta}{2}\|Ax - b\|^2 \,|\, x \in \mathcal{X}\}, & (1.5\text{a}) \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} - b), & (1.5\text{b}) \end{array}\right.$$

where λ is the Lagrange multiplier and $\beta > 0$ is a penalty parameter for the linear constraints. Note that the first-order optimality condition of the x-subproblem (1.5a) is

$$\theta(x) - \theta(x^{k+1}) + (x - x^{k+1})^T\{-A^T\lambda^k + \beta A^T(Ax^{k+1} - b)\} \geq 0, \ \forall x \in \mathcal{X}.$$

Recall (1.5b), it can be rewritten as

$$\theta(x) - \theta(x^{k+1}) + (x - x^{k+1})^T\{-A^T\lambda^{k+1}\} \geq 0, \ \forall x \in \mathcal{X}. \tag{1.6}$$

Next, (1.5b) can be written as

$$(\lambda - \lambda^{k+1})^T\{Ax^{k+1} - b + \frac{1}{\beta}(\lambda^{k+1} - \lambda^k)\} \geq 0, \ \forall \lambda \in \Re^m. \tag{1.7}$$

Combining (1.6) and (1.7), we have

$$\theta(x) - \theta(x^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\lambda^{k+1} \\ Ax^{k+1} - b \end{pmatrix} + \begin{pmatrix} 0 \\ \frac{1}{\beta}(\lambda^{k+1} - \lambda^k) \end{pmatrix} \right\} \geq 0, \quad \forall (x, \lambda) \in \Omega, \tag{1.8}$$

By using the notations in (1.3), the compact form is

$$w^{k+1} \in \Omega, \;\; \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T \{F(w^{k+1}) + Q(w^{k+1} - w^k)\} \geq 0, \;\; \forall w \in \Omega, \tag{1.9a}$$

with

$$Q = \begin{pmatrix} 0 & \\ & \frac{1}{\beta} I_m \end{pmatrix}. \tag{1.9b}$$

Compared with the PPA (1.4), it is easy to see that only the dual proximal term is regularized by the coefficient $\frac{1}{\beta}$ and the primal is not required at all. Thus, the primal variable x in ALM can be viewed as an intermediate variable and the dual variable λ plays an essential role in the iterations.

ALM is a benchmark method that has been successfully used in many important application areas (e.g., linear inverse problems in imaging processing [1, 20]). At each iteration, the computational effort of ALM is dominated by evaluating the operator $(A^T A + \frac{1}{\beta}\partial\theta)^{-1}(\cdot)$. When $A^T A$ is not be an identity matrix, or even large and dense, such evaluation usually does not have easy or closed form solutions. On the other hand, we have encountered some concrete applications arising in sparse or low-rank optimization models abstracted by (1.1), the convex functions in most of them have some special structures to utilize. An important structure we often encountered in practice is that the proximal mapping related to the objective $\theta(\cdot)$ defined by

$$\mathrm{Prox}_\theta(a, r) := \operatorname*{argmin}_{r>0} \{\theta(x) + \frac{r}{2}\|x - a\|^2 \mid x \in \mathcal{X}\}, \tag{1.10}$$

can be computed easily. Here "easy" means a closed-form solution exits or efficient solvers are available. Examples of such structure exists in linear and quadratic programming, basis pursuit, nuclear norm minimization, and model fitting problems, etc. Let us briefly review two typical algorithms developed for employing the structure.

The first is the primal-dual hybrid gradient (PDHG) algorithm, which is used for TV image restoration problems in [31], and subsequently modified by Chambolle and Pock in [3]. For solving (1.1), the PDHG iterates as

$$\text{(PDHG)} \quad \begin{cases} x^{k+1} = \arg\min\{L(x, \lambda^k) + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}, & (1.11a) \\ \bar{x}^k = x^{k+1} + (x^{k+1} - x^k), & (1.11b) \\ \lambda^{k+1} = \arg\max\big\{L\big(\bar{x}^k, \lambda\big) - \frac{s}{2}\big\|\lambda - \lambda^k\big\|^2\big\}. & (1.11c) \end{cases}$$

where r and s are required to satisfy $r > 0, s > 0, rs > \|A^T A\|$.

Similar as the derivation of ALM, by using the notations in (1.3), the VI form of PDHG is

$$w^{k+1} \in \Omega, \;\; \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T \{F(w^{k+1}) + Q(w^{k+1} - w^k)\} \geq 0, \;\; \forall w \in \Omega, \tag{1.12a}$$

where

$$Q = \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix}. \tag{1.12b}$$

Since $rs > \|A^T A\|$, Q is symmetric positive definite, and can be used to define Q-norm by $\|w\|_Q := \sqrt{w^T Q w}$. With the VI characterization (1.12), we now derive the contraction property of PDHG.

Setting $w = w^*$ in (1.12), we have

$$(w^{k+1} - w^*)^T Q(w^k - w^{k+1}) \geq \theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^{k+1}).$$

Note that (see (1.3b))

$$\theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^{k+1}) \geq \theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^*) \geq 0,$$

we obtain

$$(w^{k+1} - w^*)^T Q(w^k - w^{k+1}) \geq 0.$$

Using the last inequality, we obtain

$$\|w^k - w^*\|_Q^2 = \|(w^{k+1} - w^*) + (w^k - w^{k+1})\|_Q^2 \geq \|w^{k+1} - w^*\|_Q^2 + \|w^k - w^{k+1}\|_Q^2,$$

and thus the sequence generated by PDHG (1.12) obeys

$$\|w^{k+1} - w^*\|_Q^2 \leq \|w^k - w^*\|_Q^2 - \|w^k - w^{k+1}\|_Q^2. \tag{1.13}$$

The above inequality implies the sequence $\{w^k\}$ is strictly contractive to the solution set in Q-norm. Recall that in the original PPA, the proximal parameter is a scalar r, means that the proximal regularization makes no difference on different coordinates of w, thus all the coordinates of w are proximally regularized with equivalent weights. While for the VI form of PDHG (1.12), the proximal parameter is a positive definite matrix Q, and the coordinates of w are proximally regularized with different weights. Using this judicious choice of Q, simple subproblems can be yields under the assumption (1.10). The PDHG's convergence is guaranteed by noting the monotonic decrease of the distance under the Q-norm (see (1.13)). Further, we observe that it owes similar algorithmic structure with the PPA (1.12). In this sense, by viewing the PDHG in the VI form (1.12), it is essentially a customized application of PPA. The VI revisit (1.12) and PPA interpretation of PDHG is first shown in [17], and used in [4] to simplify the convergence analysis.

In (1.12), when we choose Q as

$$Q = \begin{pmatrix} rI_n & -A^T \\ -A & sI_m \end{pmatrix},$$

where $rs > \|A^T A\|$, the reduced subproblems amounts to the following iterate schemes

$$\text{(C-PPA)} \quad \begin{cases} \lambda^{k+1} = \arg\max\big\{L\big(x^k, \lambda\big) - \dfrac{s}{2}\big\|\lambda - \lambda^k\big\|^2\big\}, & (1.14\text{a}) \\ \bar{\lambda}^k = \lambda^{k+1} + (\lambda^{k+1} - \lambda^k), & (1.14\text{b}) \\ x^{k+1} = \arg\min\{L(x, \bar{\lambda}^k) + \dfrac{r}{2}\|x - x^k\|^2 \,|\, x \in \mathcal{X}\}, & (1.14\text{c}) \end{cases}$$

Obviously, the specified algorithm (1.14) is also the customized PPA (C-PPA) with dual primal iteration order, see [16] for details.

Since the PDHG (1.11) and C-PPA (1.14) take an algorithmic structure analogous to that of the generic PPA, the existing relaxation strategies for the PPA in the literature can be naturally used for these two customized PPA. One popular relaxation scheme is to combine a computationally trivial relaxation step with the original PPA [11]. Specially, let $\tilde{w}^k = (\tilde{x}^k, \tilde{\lambda}^k)$ be the output of the customized PPA, the new iterate v^{k+1} combined with the relaxation step in [11] is given by

VI framework of relaxed PPA for (1.3)

1. PPA step: generate $\tilde{w}^k$ via solving

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \quad \forall w \in \Omega, \tag{1.15a}$$

where Q is a symmetric positive definite matrix.

2. Relaxation step: generate the new iterate v^{k+1} via

$$v^{k+1} = v^k - \gamma(v^k - \tilde{v}^k), \tag{1.15b}$$

where the relaxation factor $\gamma \in (0, 2)$

To make our analysis uniform, in the above VI expression of relaxed PPA (1.15), we use u to denote the primal variable, and introduce a new variable v to denote the essential variables, which is an appropriate sub-vector of w that is really involved in iterations, e.g., for ALM, $u = x, v = \lambda, w = (x, \lambda)$; for PDHG (1.12), $u = x, v = w = (x, \lambda)$. These notations will be used in all VI expressions throughout the papers.

As demonstrated in many cases, see, e.g. [17, 16, 14], the relaxation step (1.15b) with $\gamma \in (1, 2)$ can accelerate the customized PPA. However, it has several drawbacks. One drawback is that, the predicted variables $\tilde{w}^k$ lies in the constrained set Ω, when we relax the essential variables $\tilde{v}^k$ along the direction $v^k - \tilde{v}^k$, the final corrected point x^{k+1} can't be guaranteed to lie in the set $\mathcal{X}$ at each iteration; another drawback is, when we use the relaxed PPA to solve some optimization problems, $\tilde{x}^k$ is directly obtained by the minimization subproblem, and it usually preserves the structure we preferred, while x^{k+1} is the linear combinations of the two previous iterates of $\tilde{x}^k$ and x^k, it may lose this structure. A feasible way to overcome these drawbacks is that, we use the relaxation term (1.15b) to accelerate the customized PPA in the iteration loops, while when the specified termination criterion is reached, we use the customized PPA (1.15a) without the relaxation step to make one final iteration to get the desired solution.

The reader may give rise to a natural question: can we have any other simple relaxation strategies for the customized PPA, e.g., PDHG (1.11), C-PPA (1.14), while still preserve the structure of the primal variable? We here give an affirmative answer. Indeed, we can do this by only relaxing the dual variables. Now, we specify our relaxed customized PPA as follows

The new VI framework of relaxed C-PPA for (1.3)

1. PPA step: generate $\tilde{u}^k$ via solving

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \quad \forall w \in \Omega, \tag{1.16a}$$

where Q is a symmetric positive definite matrix.

2. Relaxation step: Set $u^{k+1} = \tilde{u}^k$, and generate the dual iterate λ^{k+1} via

$$\lambda^{k+1} = \lambda^k - \mathcal{A}(\gamma)(v^k - \tilde{v}^k), \tag{1.16b}$$

where $\mathcal{A}(\gamma)$ is a linear operator regarding of the relaxation factor γ.

The paper is organized as follows: In Section 2, we introduce our relaxation steps for some customized PPA algorithms. In Section 3, we discuss the relaxation strategy for the ADMM. We then analyse the convergence of these methods in Section 4 and further establish their convergence rate in both the non-ergodic and ergodic sense in Section 5. After that, we present some numerical results to demonstrate the improved performance of the relaxation strategy in Section 6. Finally, some conclusions are made in Section 7.

2 Relaxed customized PPA

In this section, we show how to relax the customized PPA via only correcting the dual variable. To make our illustration more clear, the output variable w^{k+1} of the customized PPA is relabeled as $\tilde{w}^k$ below. Since the notation v is an appropriate sub-vector of w, the intended meaning of the notations $v^k, \tilde{v}^k, v^*$ are also well denoted under the routine of w, and we use $\mathcal{V}^*$ to collect v^* for all subvectors of w^* in Ω^*.

Note that the relaxation falls into the category of prediction-correction methods. We use the prototype algorithm framework proposed in [18] to unify our analysis.

Prediction-correction method for the VI problem (1.3)

[Prediction Step.] With given v^k, find a vector $\tilde{w}^k \in \Omega$ and a matrix Q satisfying

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \quad \forall\, w \in \Omega, \tag{2.1a}$$

where the matrix Q has the property: $Q^T + Q$ is positive definite.

[Correction Step.] Determine a nonsingular matrix M; and generate the new iterate v^{k+1} via

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k). \tag{2.1b}$$

Then, the convergence of the prototype algorithm (2.1) can be guaranteed if the following conditions is fulfilled.

Convergence conditions

For the matrices Q and M in (2.1), The matrices

$$H := QM^{-1} \succ 0 \tag{2.2a}$$

and

$$G := Q^T + Q - M^T HM \succ 0. \tag{2.2b}$$

In the following, we show that each PPA with the correction step falls into the prototype algorithm (2.1). Then we demonstrate that the conditions specified in (2.2) are fulfilled.

2.1 Relaxed PDHG

Recall that the PDHG in variational form is (1.16a) with

$$Q = \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix}.$$

Then we judiciously correct the dual variable via

$$\lambda^{k+1} = \lambda^k - \frac{\gamma - 1}{s} A(x^k - \tilde{x}^k) - \gamma(\lambda^k - \tilde{\lambda}^k), \tag{2.3}$$

where the relaxation factor $\gamma \in (0,2)$. Since we always have $x^{k+1} = \tilde{x}^k$, using the notations in (1.3), we get the correction step as

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k), \tag{2.4a}$$

where

$$M = \begin{pmatrix} I_n & 0 \\ \frac{\gamma-1}{s}A & \gamma I_m \end{pmatrix}. \tag{2.4b}$$

Now, we check the positive definiteness of H and G to verify whether the condition (2.2) is fulfilled by the relaxed PDHG. First, we need to deduce the expression of H and G. Recall M (see (2.4b)), we have

$$M^{-1} = \begin{pmatrix} I_n & 0 \\ -\frac{\gamma-1}{s\gamma}A & \frac{1}{\gamma}I_m \end{pmatrix}.$$

Then

$$\begin{aligned} H = QM^{-1} &= \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix} \begin{pmatrix} I_n & 0 \\ -\frac{\gamma-1}{s\gamma}A & \frac{1}{\gamma}I_m \end{pmatrix} \\ &= \begin{pmatrix} rI_n - \frac{\gamma-1}{s\gamma}A^TA & \frac{1}{\gamma}A^T \\ \frac{1}{\gamma}A & \frac{s}{\gamma}I_m \end{pmatrix}. \end{aligned} \tag{2.5}$$

Since $rs > \|A^TA\|$, according to the Schur complement, H is positive definite.

The matrix G is symmetric with $Q^T = Q$, we next deduce the expression of G.

$$\begin{aligned} G &= Q^T + Q - M^THM = 2Q - M^TQ \\ &= 2\cdot \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix} - \begin{pmatrix} I_n & \frac{\gamma-1}{s}A^T \\ 0 & \gamma I_m \end{pmatrix} \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix} \\ &= \begin{pmatrix} 2rI_n & 2A^T \\ 2A & 2sI_m \end{pmatrix} - \begin{pmatrix} rI_n + \frac{\gamma-1}{s}A^TA & \gamma A^T \\ \gamma A & s\gamma I_m \end{pmatrix} \\ &= \begin{pmatrix} rI_n - \frac{\gamma-1}{s}A^TA & (2-\gamma)A^T \\ (2-\gamma)A & (2-\gamma)sI_m \end{pmatrix}. \end{aligned} \tag{2.6}$$

Since $rI_n \succ \frac{1}{s}A^TA$ and $0 < \gamma < 2$, according to the Schur complement, G is positive definite. The relaxed PDHG scheme is a special case of the prototype algorithm (2.2).

The relaxation step (2.3) looks somehow complicated. However, combining PDHG and (2.3), the update steps of the relaxed PDHG can be simple. By straightforward manipulations, the PDHG can be rewritten as a prediction step, which gives

$$\left\{ \begin{aligned} &\tilde{x}^k = \operatorname{argmin}\{\theta(x) + \frac{r}{2}\|x - x^k - \frac{1}{r}A^T\lambda^k\|^2 \mid x \in \mathcal{X}\}, && (2.7a) \\ &\tilde{\lambda}^k = \lambda^k - \frac{1}{s}(A(2\tilde{x}^k - x^k) - b). && (2.7b) \end{aligned} \right.$$

Together with (2.3) and noting $x^{k+1} = \tilde{x}^k$, we obtain

Relaxed PDHG for (1.1)

$$\left\{ \begin{aligned} &x^{k+1} = \operatorname{argmin}\{\theta(x) + \frac{r}{2}\|x - x^k - \frac{1}{r}A^T\lambda^k\|^2 \mid x \in \mathcal{X}\}, && (2.8a) \\ &\lambda^{k+1} = \lambda^k - \frac{1}{s}\{A\big((1+\gamma)x^{k+1} - x^k\big) - \gamma b\}, && (2.8b) \end{aligned} \right.$$

where $r > 0, s > 0$, $rs > \|A^TA\|$, and $\gamma \in (0,2)$.

2.2 Relaxed C-PPA

The C-PPA (1.14) in VI form is (1.16a) with

$$Q = \begin{pmatrix} rI_n & -A^T \\ -A & sI_m \end{pmatrix}.$$

We then judiciously correct the dual variable by

$$\lambda^{k+1} = \lambda^k + \frac{\gamma - 1}{s} A(x^k - \tilde{x}^k) - \gamma(\lambda^k - \tilde{\lambda}^k), \tag{2.9}$$

where the relaxation factor $\gamma \in (0, 2)$.

Since we set $x^{k+1} = \tilde{x}^k$, using the notations in (1.3), we have the correction step as

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k), \tag{2.10a}$$

where

$$M = \begin{pmatrix} I_n & 0 \\ -\frac{\gamma-1}{s}A & \gamma I_m \end{pmatrix}. \tag{2.10b}$$

Then, we check the positive definiteness of H and G to verify whether the condition (2.2) is fulfilled by the relaxed C-PPA scheme. First, recall M (see (2.10b)), we have

$$M^{-1} = \begin{pmatrix} I_n & 0 \\ \frac{\gamma-1}{s\gamma}A & \frac{1}{\gamma}I_m \end{pmatrix}.$$

Then

$$\begin{aligned} H = QM^{-1} &= \begin{pmatrix} rI_n & -A^T \\ -A & sI_m \end{pmatrix} \begin{pmatrix} I_n & 0 \\ \frac{\gamma-1}{s\gamma}A & \frac{1}{\gamma}I_m \end{pmatrix} \\ &= \begin{pmatrix} rI_n - \frac{\gamma-1}{s\gamma}A^TA & -\frac{1}{\gamma}A^T \\ -\frac{1}{\gamma}A & \frac{s}{\gamma}I_m \end{pmatrix}. \end{aligned} \tag{2.11}$$

Since $rI_n \succ \frac{1}{s}A^TA$, according to the Schur complement, H is positive definite.

Next, we check the positive definiteness of G. We first deduce the expression of G.

$$\begin{aligned} G &= 2Q - M^THM = 2Q - M^TQ \\ &= 2 \cdot \begin{pmatrix} rI_n & -A^T \\ -A & sI_m \end{pmatrix} - \begin{pmatrix} I_n & -\frac{\gamma-1}{s}A^T \\ 0 & \gamma I_m \end{pmatrix} \begin{pmatrix} rI_n & -A^T \\ -A & sI_m \end{pmatrix} \\ &= \begin{pmatrix} 2rI_n & -2A^T \\ -2A & 2sI_m \end{pmatrix} - \begin{pmatrix} rI_n + \frac{\gamma-1}{s}A^TA & -\gamma A^T \\ -\gamma A & s\gamma I_m \end{pmatrix} \\ &= \begin{pmatrix} rI_n - \frac{\gamma-1}{s}A^TA & -(2-\gamma)A^T \\ -(2-\gamma)A & (2-\gamma)sI_m \end{pmatrix}. \end{aligned} \tag{2.12}$$

Since $rI_n \succ \frac{1}{s}A^TA$ and $0 < \gamma < 2$, according to the Schur complement, G is positive definite. The relaxed C-PPA scheme is therefore a special case of the prototype algorithm (2.2).

We can also write the relaxed C-PPA scheme in a compact form. First, with simple manipulation, the C-PPA (1.14) is written as

$$\begin{cases} \tilde{\lambda}^k = \lambda^k - \frac{1}{s}(Ax^k - b), & (2.13a) \\ \tilde{x}^k = \operatorname{argmin}\{\theta(x) + \frac{r}{2}\|x - x^k - \frac{1}{r}A^T(2\tilde{\lambda}^k - \lambda^k)\|^2 \mid x \in \mathcal{X}\}, & (2.13b) \end{cases}$$

Inserting (2.13a) in (2.13b), we get

$$\tilde{x}^k = \operatorname{argmin}\{\theta(x) + \frac{r}{2}\|x - x^k - \frac{1}{r}A^T[2(\lambda^k - \frac{1}{s}(Ax^k - b)] - \lambda^k)\|^2 \mid x \in \mathcal{X}\},$$

and thus we have

$$\tilde{x}^k = \operatorname{argmin}\{\theta(x) + \frac{r}{2}\|x - x^k - \frac{1}{r}A^T(\lambda^k - \frac{2}{s}(Ax^k - b))\|^2 \mid x \in \mathcal{X}\}.$$

Inserting (2.13a) in (2.9), we get

$$\begin{aligned} \lambda^{k+1} &= \lambda^k + \frac{\gamma - 1}{s}A(x^k - \tilde{x}^k) - \frac{\gamma}{s}(Ax^k - b) \\ &= \lambda^k - \frac{\gamma}{s}(A\tilde{x}^k - b) - \frac{1}{s}A(x^k - \tilde{x}^k) \\ &= \lambda^k - \frac{1}{s}(A[x^k + (\gamma - 1)\tilde{x}^k] - \gamma b). \end{aligned}$$

Noting $x^{k+1} = \tilde{x}^k$, the resulting relaxed C-PPA can be directly written as

Relaxed C-PPA for (1.1)

$$\begin{cases} x^{k+1} = \operatorname{argmin}\{\theta(x) + \frac{r}{2}\|x - x^k - \frac{1}{r}A^T(\lambda^k - \frac{2}{s}(Ax^k - b))\|^2 \mid x \in \mathcal{X}\}, & (2.14a) \\ \lambda^{k+1} = \lambda^k - \frac{1}{s}(A[x^k + (\gamma - 1)x^{k+1}] - \gamma b), & (2.14b) \end{cases}$$

where $r > 0, s > 0$, $rs > \|A^TA\|$, and $\gamma \in (0, 2)$.

2.3 Relaxed linearized ALM

In this section, we show that the linearized ALM (L-ALM) [27, 30] is also a special case of customized PPA, and it can be relaxed by (1.16b). The main idea of L-ALM is to linearize the augmented Lagrangian function (1.5a) in the ALM via adding a quadratic term $\frac{1}{2}\|x - x^k\|^2_{rI - \beta A^TA}$. Ignoring some constant terms in the x-subproblem, the iterate scheme of L-ALM is

$$\begin{cases} x^{k+1} = \operatorname{argmin}\{\theta(x) + \frac{r}{2}\|x - x^k - \frac{1}{r}A^T(\lambda^k - \beta(Ax^k - b))\|^2 \mid x \in \mathcal{X}\}, & (2.15a) \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} - b), & (2.15b) \end{cases}$$

Note that the quadratic term should be positive, we have $r > \beta\|A^TA\|$. Now, we revisit L-ALM (2.15) in VI form.

The first-order optimality condition of the x-subproblem of (2.15a) is

$$\theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T\{-A^T\lambda^k + \beta A^T(\tilde{x}^k - b) + (rI - \beta A^TA)(\tilde{x}^k - x^k)\} \geq 0, \ \forall x \in \mathcal{X},$$

and it can be written as

$$\theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T\{-A^T\tilde{\lambda}^k + (rI - \beta A^T A)(\tilde{x}^k - x^k)\} \geq 0, \ \forall x \in \mathcal{X}. \tag{2.16}$$

Next, we rewrite $\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k - b)$ (see (2.15b)) as

$$(\lambda - \tilde{\lambda}^k)^T\{A\tilde{x}^k - b + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \ \forall \lambda \in \Re^m. \tag{2.17}$$

Combining (2.16) and (2.17), we have

$$\begin{array}{l} \forall w \in \Omega, \quad \theta(x) - \theta(\tilde{x}^k) + \\ \begin{pmatrix} x - \tilde{x}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\tilde{\lambda}^k \\ A\tilde{x}^k - b \end{pmatrix} + \begin{pmatrix} (rI_n - \beta A^T A)(\tilde{x}^k - x^k) \\ \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) \end{pmatrix} \right\} \geq 0. \end{array} \tag{2.18}$$

Recall the notations in (1.3), the L-ALM (2.15) can thus be interpreted as a customized PPA (1.16a) with Q given by

$$Q = \begin{pmatrix} rI_n - \beta A^T A & 0 \\ 0 & \frac{1}{\beta}I_m \end{pmatrix}. \tag{2.19}$$

To accelerate the L-ALM, we suggest to relax the dual variable by

$$\lambda^{k+1} = \lambda^k - \gamma(\lambda^k - \tilde{\lambda}^k). \tag{2.20}$$

where the relaxation factor $\gamma \in (0, 2)$. Note that x^{k+1} always amounts to $\tilde{x}^k$, we have

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k), \tag{2.21a}$$

where

$$M = \begin{pmatrix} I_n & 0 \\ 0 & \gamma I_m \end{pmatrix}. \tag{2.21b}$$

Next, we show that the condition (2.2) is fulfilled by the relaxed L-ALM scheme. We have

$$\begin{aligned} H = QM^{-1} &= \begin{pmatrix} rI_n - \beta A^T A & 0 \\ 0 & \frac{1}{\beta}I_m \end{pmatrix} \begin{pmatrix} I_n & 0 \\ 0 & \frac{1}{\gamma}I_m \end{pmatrix} \\ &= \begin{pmatrix} rI_n - \beta A^T A & 0 \\ 0 & \frac{1}{\gamma\beta}I_m \end{pmatrix}, \end{aligned} \tag{2.22}$$

and

$$\begin{aligned} G &= Q^T + Q - M^T H M = 2Q - M^T Q \\ &= \begin{pmatrix} 2(rI_n - \beta A^T A) & 0 \\ 0 & \frac{2}{\beta}I_m \end{pmatrix} - \begin{pmatrix} I_n & 0 \\ 0 & \gamma I_m \end{pmatrix} \begin{pmatrix} rI_n - \beta A^T A & 0 \\ 0 & \frac{1}{\beta}I_m \end{pmatrix} \\ &= \begin{pmatrix} 2(rI_n - \beta A^T A) & 0 \\ 0 & \frac{2}{\beta}I_m \end{pmatrix} - \begin{pmatrix} rI_n - \beta A^T A & 0 \\ 0 & \frac{\gamma}{\beta}I_m \end{pmatrix} \\ &= \begin{pmatrix} rI_n - \beta A^T A & 0 \\ 0 & \frac{2-\gamma}{\beta}I_m \end{pmatrix}. \end{aligned} \tag{2.23}$$

Under the condition that $r > 0, \beta > 0, r > \beta\|A^TA\|$ and the relaxation factor $\gamma \in (0, 2)$, the matrix H and G are both symmetric and positive definite; and thus the relaxed L-ALM scheme is a special case of the prototype algorithm (2.2) and the condition (2.2) is fulfilled. We refer to [26] for the relaxed L-ALM with (1.15b).

Now, combining the relaxed term (2.9) with the L-ALM, the relaxed L-ALM scheme is

Relaxed L-ALM for (1.1)

$$\begin{cases} x^{k+1} = \text{argmin}\{\theta(x) + \frac{r}{2}\|x - x^k - \frac{1}{r}A^T(\lambda^k - \beta(Ax^k - b))\|^2 \mid x \in \mathcal{X}\}, & (2.24a) \\ \lambda^{k+1} = \lambda^k - \gamma\beta(Ax^{k+1} - b), & (2.24b) \end{cases}$$

where $r > 0, s > 0$, $rs > \|A^TA\|$, and $\gamma \in (0, 2)$.

3 Relaxed ADMM for separable case

Motivated by recent popular applications, we also discuss a separable case of (1.1), in which the objective function $\theta(x)$ can be decomposed as a sum of two individual convex functions without coupled variables:

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b,\ x \in \mathcal{X}, y \in \mathcal{Y}\}, \tag{3.1}$$

where $\theta_1 : \Re^{n_1} \to \Re$, $\theta_2 : \Re^{n_2} \to \Re$ are both closed convex (not necessarily smooth) functions; $A \in \Re^{m\times n_1}$, $B \in \Re^{m\times n_2}$, $\mathcal{X} \subset \Re^{n_1}$, $\mathcal{Y} \subset \Re^{n_2}$ are closed convex sets.

The optimality condition of the problem (3.1) can be analogously written as the following variational inequality:

$$\text{VI}(\Omega, F, \theta) \quad w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega, \tag{3.2a}$$

where

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \qquad F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix}, \tag{3.2b}$$

$$\theta(u) = \theta_1(x) + \theta_2(y), \qquad \text{and} \qquad \Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m.$$

Note that the mapping $F(w)$ defined in (3.2b) also has $(w_1 - w_2)^T\big(F(w_1) - F(w_2)\big) \ge 0$ for any $w_1, w_2 \in \Omega$, thus it is monotone.

The classical alternating directions method of multipliers (ADMM) originates in [5, 10] is suitable for solving (3.1). Starting with an initial iterate $(y^0, \lambda^0) \in \mathcal{Y}\times\Re^m$, the ADMM iterates via the scheme

$$\begin{cases} x^{k+1} = \arg\min\big\{\theta_1(x) - (\lambda^k)^T Ax + \frac{\beta}{2}\|Ax + By^k - b\|^2 \bigm| x \in \mathcal{X}\big\}, & (3.3a) \\ y^{k+1} = \arg\min\big\{\theta_2(y) - (\lambda^k)^T By + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \bigm| y \in \mathcal{Y}\big\}, & (3.3b) \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). & (3.3c) \end{cases}$$

In [8], it was shown that ADMM is actually equivalent to the Douglas-Rachford splitting method (DRSM) applied to the dual of (3.1). We also refer to [13] for the primal application of DRSM. In this section, to make our analysis more general, we consider the proximal ADMM where a proximal term is added on the second subproblem. Note that the original ADMM produces the new iterate in the forward order $x \to y \to \lambda$. In cyclical sense, it can also be written as the $x \to \lambda \to y$ order, which gives

$$
\begin{cases}
x^{k+1} = \arg\min\{\theta_1(x) - (\lambda^k)^T Ax + \frac{\beta}{2}\|Ax + By^k - b\|^2 \mid x \in \mathcal{X}\}, & \text{(3.4a)}\\
\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^k - b), & \text{(3.4b)}\\
y^{k+1} = \arg\min\{\theta_2(y) - (\lambda^{k+1})^T By + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 + \|y - y^k\|_D^2 \mid y \in \mathcal{Y}\}, & \text{(3.4c)}
\end{cases}
$$

where D is a semidefinite matrix. When $D = 0$, the scheme (3.4) reduces to ADMM. When $D = rI - \beta B^T B$ ($r > \beta\|B^T B\|$), the so-called split inexact Uzawa method [29, 30] is recovered; and the y-subproblem is in form of (1.10), which is simple under the assumption (1.10).

Note that the variable x in (3.4) (also in (3.3)) is not required to generate the new $(k+1)$-th iteration. Thus, as referred in [2], x is an intermediate variable, $v = (y, \lambda)$ are essential variables. The output of (3.4) serves as a predictor for our relaxation, we here relabel it as $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda})$ as before. Now, we follow [6] to revisit the proximal ADMM (3.4) in VI form.

Lemma 3.1. *For given $v^k = (y^k, \lambda^k)$, let $\tilde{w}^k$ be generated by (3.4). Then, we have*

$$
\tilde{w}^k \in \Omega, \;\; \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \;\; \forall w \in \Omega, \tag{3.5a}
$$

where

$$
Q = \begin{pmatrix} D + \beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}. \tag{3.5b}
$$

Proof. The optimality condition of the x-subproblem in (3.4) is

$$
\tilde{x}^k \in \mathcal{X}_1, \;\; \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{-A^T\lambda^k + \beta A^T(A\tilde{x}^k + By^k - b)\} \geq 0, \;\; \forall\, x \in \mathcal{X},
$$

and it can be written as (by using $\tilde{\lambda}^k$ generated by (3.4b))

$$
\tilde{x}^k \in \mathcal{X}, \;\; \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T(-A^T\tilde{\lambda}^k) \geq 0, \;\; \forall\, x \in \mathcal{X}. \tag{3.6a}
$$

The optimality condition of the y-subproblem of (3.4) is

$$
\tilde{y}^k \in \mathcal{Y}, \; \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\{-B^T\tilde{\lambda}^k + \beta B^T(A\tilde{x}^k + B\tilde{y}^k - b) + D(\tilde{y}^k - y^k)\} \geq 0, \; \forall y \in \mathcal{Y}.
$$

Now, we treat the $\{\cdot\}$ term in the last inequality. Using $\beta(A\tilde{x}^k + By^k - b) = -(\tilde{\lambda}^k - \lambda^k)$ (see (3.4b)), we obtain

$$
\begin{aligned}
&-B^T\tilde{\lambda}^k + \beta B^T(A\tilde{x}^k + B\tilde{y}^k - b) + D(\tilde{y}^k - y^k) \\
&\quad = \;\; -B^T\tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k) - B^T(\tilde{\lambda}^k - \lambda^k) + D(\tilde{y}^k - y^k) \\
&\quad = \;\; -B^T\tilde{\lambda}^k + (D + \beta B^T B)(\tilde{y}^k - y^k) - B^T(\tilde{\lambda}^k - \lambda^k).
\end{aligned}
$$

Then the optimality condition of the y-subproblem can be written as

$$
\tilde{y}^k \in \mathcal{Y}, \;\; \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\{-B^T\tilde{\lambda}^k + (D + \beta B^T B)(\tilde{y}^k - y^k) - B^T(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \;\; \forall y \in \mathcal{Y}. \tag{3.6b}
$$

From (3.4), we have

$$
(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k) = 0,
$$

and it can be written as

$$\tilde{\lambda}^k \in \Re^m, \;\; (\lambda - \tilde{\lambda}^k)^T\big\{(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k)\big\} \geq 0, \;\; \forall\, \lambda \in \Re^m. \tag{3.6c}$$

Combining (3.6a), (3.6b) and (3.6c), and using the notations of (3.2b), the assertion of this lemma is proved. □

Remark 3.2. The matrix Q (3.5b) is symmetric positive definite. Thus, the proximal ADMM (3.4) is a customized PPA.

To accelerate the ADMM (3.4), we suggest to relax the dual variable by

$$\lambda^{k+1} = \lambda^k + (\gamma - 1)\beta B(y^k - \tilde{y}^k) - \gamma(\lambda^k - \tilde{\lambda}^k), \tag{3.7}$$

where the relaxation factor $\gamma \in (0, 2)$. Note that y^{k+1} always amounts to $\tilde{y}^k$, we have

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k), \tag{3.8a}$$

where

$$M = \begin{pmatrix} I_{n_2} & 0 \\ -(\gamma-1)\beta B & \gamma I_m \end{pmatrix}. \tag{3.8b}$$

Then, we check the positive definiteness of H and G to verify whether the condition (2.2) is fulfilled for the relaxed ADMM scheme. First, recall M (see (3.8b)), we have

$$M^{-1} = \begin{pmatrix} I_{n_2} & 0 \\ \frac{\gamma-1}{\gamma}\beta B & \frac{1}{\gamma} I_m \end{pmatrix},$$

and

$$\begin{aligned} H = QM^{-1} &= \begin{pmatrix} D + \beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} I_{n_2} & 0 \\ \frac{\gamma-1}{\gamma}\beta B & \frac{1}{\gamma} I_m \end{pmatrix} \\ &= \begin{pmatrix} D + \frac{1}{\gamma}\beta B^T B & -\frac{1}{\gamma} B^T \\ -\frac{1}{\gamma} B & \frac{1}{\gamma\beta} I_m \end{pmatrix}. \end{aligned} \tag{3.9}$$

thus H is positive definite. Next, we check the positive definiteness of G. We first deduce the expression of G.

$$\begin{aligned} G &= Q^T + Q - M^T H M = 2Q - M^T Q \\ &= \begin{pmatrix} 2(D + \beta B^T B) & -2B^T \\ -2B & \frac{2}{\beta} I_m \end{pmatrix} - \begin{pmatrix} I_n & -(\gamma-1)\beta B^T \\ 0 & \gamma I_m \end{pmatrix} \begin{pmatrix} D + \beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} \\ &= \begin{pmatrix} 2(D + \beta B^T B) & -2B^T \\ -2B & \frac{2}{\beta} I_m \end{pmatrix} - \begin{pmatrix} D + r\beta B^T B & -rB^T \\ -rB & \frac{\gamma}{\beta} I_m \end{pmatrix} \\ &= \begin{pmatrix} D + (2-r)\beta B^T B & -(2-r)B^T \\ -(2-r)B & \frac{2-\gamma}{\beta} I_m \end{pmatrix}. \end{aligned} \tag{3.10}$$

G is positive definite with $0 < \gamma < 2$. The relaxed ADMM scheme is thus a special case of the prototype algorithm (2.2). Note that

$$\begin{aligned} \lambda^{k+1} &= \tilde{\lambda}^k + (\gamma-1)\beta B(y^k - \tilde{y}^k) - (\gamma-1)(\lambda^k - \tilde{\lambda}^k) \\ &= \tilde{\lambda}^k + (\gamma-1)\beta B(y^k - \tilde{y}^k) + (\gamma-1)\beta(A\tilde{x}^k + By^k - b) \\ &= \tilde{\lambda}^k - (\gamma-1)\beta(A\tilde{x}^k + B\tilde{y}^k - b). \end{aligned}$$

Write (3.4) and (3.7) together in a compact form and recall $u^{k+1} = \tilde{u}^k$, the relaxed ADMM is

Relaxed ADMM for (3.1)

$$\begin{cases} x^{k+1} = \arg\min\{\theta_1(x) - (\lambda^k)^T Ax + \frac{\beta}{2}\|Ax + By^k - b\|^2 \mid x \in \mathcal{X}\}, & \text{(3.11a)} \\ \tilde{\lambda}^k = \lambda^k - \beta(Ax^{k+1} + By^k - b), & \text{(3.11b)} \\ y^{k+1} = \arg\min\{\theta_2(y) - (\tilde{\lambda}^k)^T By + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 + \|y - y^k\|_D^2 \mid y \in \mathcal{Y}\}, & \text{(3.11c)} \\ \lambda^{k+1} = \tilde{\lambda}^k - (\gamma - 1)\beta(Ax^{k+1} + By^{k+1} - b), & \text{(3.11d)} \end{cases}$$

where $\gamma \in (0, 2)$.

4 Global convergence

We have characterized the relaxed PDHG, relaxed C-PPA, relaxed L-ALM, relaxed ADMM as special cases of the Prototype Framework (2.1). Using the Prototype Framework (2.1), we can easily analyze their convergence property in a unified way.

Under the condition that H and G are positive definite, we now characterize the right hand side of (2.1a) in terms of $\|v - v^k\|_H$ and $\|v - v^{k+1}\|_H$. This is given in the following lemma.

Lemma 4.1. *Let $\tilde{w}^k$ be generated by the step* (2.1a) *and $\{v^k\}$ be generated by* (2.1b)*. We have*

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq \frac{1}{2}(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2) + \frac{1}{2}\|v^k - \tilde{v}^k\|_G^2, \ \forall w \in \Omega. \quad (4.1)$$

where H and G are defined respectively in (2.2a) *and* (2.2b)*.*

Proof. Observe that $M(v^k - \tilde{v}^k) = (v^k - v^{k+1})$ (see (2.1b)) and $Q = HM$, the right hand side of (2.1a) can be written as

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T H(v^k - v^{k+1}), \ \forall w \in \Omega. \quad (4.2)$$

Applying the identity

$$(a - b)^T H(c - d) = \frac{1}{2}(\|a - d\|_H^2 - \|a - c\|_H^2) + \frac{1}{2}(\|c - b\|_H^2 - \|d - b\|_H^2),$$

with $a = v, b = \tilde{v}^k, c = v^k, d = v^{k+1}$ to the right hand side of (4.2), we have

$$(v - \tilde{v}^k)^T H(v^k - v^{k+1}) = \frac{1}{2}(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2) + \frac{1}{2}(\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2). \quad (4.3)$$

For the last term of (4.3), we have

$$\begin{aligned} &\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \\ &= \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - (v^k - v^{k+1})\|_H^2 \\ &\overset{(2.1b)}{=} \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - M(v^k - \tilde{v}^k)\|_H^2 \\ &= 2(v^k - \tilde{v}^k)^T HM(v^k - \tilde{v}^k) - (v^k - \tilde{v}^k)^T M^T HM(v^k - \tilde{v}^k) \\ &\overset{(2.2)}{=} (v^k - \tilde{v}^k)^T(Q^T + Q - M^T HM)(v^k - \tilde{v}^k) \\ &\overset{(2.2b)}{=} (v^k - \tilde{v}^k)^T G(v^k - \tilde{v}^k). \end{aligned} \quad (4.4)$$

Combing (4.2), (4.3) and (4.4), the lemma is thus proved. □

Theorem 4.2. *Let $\tilde{w}^k$ be generated by the step* (2.1a) *and* $\{v^k\}$ *be generated by* (2.1b)*. Then we have*

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \ \forall v^* \in \mathcal{V}^*. \tag{4.5}$$

Proof. Setting $v = v^*$ in (4.1), we get

$$\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \ge \|v^k - \tilde{v}^k\|_G^2 + 2\{\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k)\}.$$

By using the optimality of w^* and the monotonicity of $F(w)$, we have

$$\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k) \ge \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(w^*) \ge 0,$$

and thus

$$\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \ge \|v^k - \tilde{v}^k\|_G^2.$$

The assertion (4.5) follows directly. □

Since H and G are positive definite, Theorem 4.2 implies that the sequence $\{v^k\}$ generated by (2.1) is strictly contractive with respect to the solution set $\mathcal{V}^*$. Thus the global convergence of these relaxed PPA can be easily established. This is given by the following theorem.

Theorem 4.3. *Let* $\{v^k\}$ *be a sequence generated by* (2.1)*. Then it globally converges to a point* v^∞ *which belongs to* $\mathcal{V}^*$*.*

Proof. From (4.5), we can see that $\{v^k\}$ is bounded, and

$$\lim_{k\to\infty} \|v^k - \tilde{v}^k\| = 0. \tag{4.6}$$

Thus, the sequence $\{\tilde{v}^k\}$ is also bounded. Let v^∞ be a cluster point of the sequence $\{\tilde{v}^k\}$. Since the sequence $\{\tilde{v}^k\}$ is bounded, it has a subsequence $\{\tilde{v}_j^k\}$ that converges to v^∞. From (2.1a), we have

$$\tilde{w}_j^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}_j^k) + (w - \tilde{w}_j^k)^T F(\tilde{w}_j^k) \ge (v - \tilde{v}_j^k)^T Q(v_j^k - \tilde{v}_j^k), \quad \forall w \in \Omega.$$

Since Q is not singular, from the continuity of $\theta(u)$ and $F(w)$, we have

$$w^\infty \in \Omega, \quad \theta(u) - \theta(x^\infty) + (w - w^\infty)^T F(w^\infty) \ge 0, \quad \forall w \in \Omega.$$

The above variational inequality implies that w^∞ is a solution point of VI$(\Omega; F; \theta)$. By using (4.6) and $\lim_{k\to\infty} \tilde{v}_j^k = v^\infty$, we get $\lim_{k\to\infty} v_j^k = v^\infty$. Recall (4.5), we have

$$\|v^{k+1} - v^\infty\|_H \le \|v^k - v^\infty\|_H.$$

Thus the sequence $\{v^k\}$ converges to v^∞. The proof is completed. □

5 Convergence rate

In this section, we analyze the iteration complexity in both the non-ergodic and ergodic sense for the sequence generated by these relaxed customized PPA. Note that by letting the matrix $M = I$, the analysis is also applicable for the original customized PPA.

5.1 Worse-case convergence rate in the non-ergodic sense

In this subsection, we establish the $\mathcal{O}(1/t)$ non-ergodic convergence rate of these relaxed PPA.

Lemma 5.1. *Let $\tilde{w}^k$ be generated by the step* (2.1a) *and $\{v^k\}$ be generated by* (2.1b)*. Then we have*

$$(\tilde{v}^k - \tilde{v}^{k+1})^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \geq 0. \tag{5.1}$$

Proof. Setting $w = \tilde{w}^{k+1}$ in (2.1a), we get

$$\theta(\tilde{u}^{k+1}) - \theta(\tilde{u}^k) + (\tilde{w}^{k+1} - \tilde{w}^k)^T F(\tilde{w}^k) \geq (\tilde{v}^{k+1} - \tilde{v}^k)^T Q(v^k - \tilde{v}^k). \tag{5.2}$$

Notice that (2.1a) is also suitable for $k := k+1$, thus

$$\theta(u) - \theta(\tilde{u}^{k+1}) + (w - \tilde{w}^{k+1})^T F(\tilde{w}^{k+1}) \geq (v - \tilde{v}^{k+1})^T Q(v^{k+1} - \tilde{v}^{k+1}), \quad \forall w \in \Omega. \tag{5.3}$$

Let $w = \tilde{w}^k$ in the above inequality, we get

$$\theta(\tilde{u}^k) - \theta(\tilde{u}^{k+1}) + (\tilde{w}^k - \tilde{w}^{k+1})^T F(\tilde{w}^{k+1}) \geq (\tilde{v}^k - \tilde{v}^{k+1})^T Q(v^{k+1} - \tilde{v}^{k+1}). \tag{5.4}$$

Summing (5.2) and (5.4) together and using the monotonicity of F, the assertion (5.1) is proved. □

Lemma 5.2. *Let $\tilde{w}^k$ be generated by the step* (2.1a) *and $\{v^k\}$ be generated by* (2.1b)*. Then we have*

$$(v^k - \tilde{v}^k)^T M^T H M\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \geq \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_Q^2. \tag{5.5}$$

Proof. Note that Q is symmetric positive definite. By adding the equation

$$\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\}^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} = \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_Q^2$$

to the both sides of (5.1), we get

$$(v^k - v^{k+1})^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \geq \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_Q^2. \tag{5.6}$$

Substituting the terms

$$M(v^k - \tilde{v}^k) = v^k - v^{k+1} \text{ and } \; Q = HM$$

into (5.6), we obtain

$$(v^k - \tilde{v}^k)^T M^T H M\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \geq \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_Q^2.$$

This completes the proof. □

Theorem 5.3. *Let $\{v^k\}$ be a sequence generated by* (2.1)*. Then we have*

$$\|v^{k+1} - v^{k+2}\|_H^2 \leq \|v^k - v^{k+1}\|_H^2, \tag{5.7}$$

where H is defined in (2.2a)*.*

Proof. Let $a = M(v^k - \tilde{v}^k)$ and $b = M(v^{k+1} - \tilde{v}^{k+1})$ be in the following identity

$$\|a\|_H^2 - \|b\|_H^2 = 2a^T H(a - b) - \|a - b\|_H^2,$$

we get

$$
\begin{aligned}
&\|M(v^k - \tilde{v}^k)\|^2 - \|M(v^{k+1} - \tilde{v}^{k+1})\|^2 \\
&= 2(v^k - \tilde{v}^k)^T M^T HM\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} - \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|^2_{M^T HM}.
\end{aligned}
\tag{5.8}
$$

Substitute (5.5) into the right hand side of above equality, we get

$$
\|M(v^k - \tilde{v}^k)\|^2 - \|M(v^{k+1} - \tilde{v}^{k+1})\|^2 \geq \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|^2_{Q^T + Q - M^T HM}.
$$

Note that

$$
M(v^k - \tilde{v}^k) = v^k - v^{k+1} \text{ and } M(v^{k+1} - \tilde{v}^{k+1}) = v^{k+1} - v^{k+2}.
$$

Combining with $G = Q^T + Q - M^T HM \succ 0$ (see (2.2b)), the assertion is proved. □

Theorem 5.3 implies that the sequences $\{\|v^k - v^{k+1}\|^2_H\}$ is monotonically nonincreasing, which is crucial for our proof. Next, we need to establish the connection between $\|v^k - \tilde{v}^k\|^2_G$ and $\|v^k - v^{k+1}\|^2_H$.

Lemma 5.4. *For each relaxed customized PPA, there exists a constant c, such that*

$$
c\|v^k - v^{k+1}\|^2_H \prec \|v^k - \tilde{v}^k\|^2_G. \tag{5.9}
$$

Proof. From $v^k - v^{k+1} = M(v^k - \tilde{v}^k)$ (see (2.1b)), we have

$$
\|v^k - v^{k+1}\|^2_H = \|v^k - \tilde{v}^k\|^2_{M^T HM}.
$$

The remaining task is to find a constant c, such that

$$
G \succ cM^T HM,
$$

for each G, H pair emerged in these algorithms. Note that from (2.2b), we have $M^T HM \prec 2Q$. From (**??**) and (**??**), we have $G \succcurlyeq (2 - \gamma)Q$ for the relaxed PDHG and relaxed C-PPA, respectively. From (2.23) and (3.10), $G \succcurlyeq \frac{2-\gamma}{2}Q$ for relaxed L-ALM and relaxed ADMM, respectively. The assertion is thus proved. □

Theorem 5.5. *Let $\{v^k\}$ be a sequence generated by* (2.1)*. Then we have*

$$
\|v^k - v^{k+1}\|^2_H \leq \frac{c}{k+1}\|v^0 - v^*\|^2_H, \ \forall v^* \in \mathcal{V}^*, \tag{5.10}
$$

where the matrix H is defined as (2.2a)*.*

Proof. Inserting (5.9) into (4.5), we get

$$
\|v^{k+1} - v^*\|^2_H \leq \|v^k - v^*\|^2_H - c\|v^k - v^{k+1}\|^2_H, \ \forall v^* \in \mathcal{V}^*.
$$

Summing this inequality over $t = 0, 1, \cdots, \infty$, we have

$$
\frac{1}{c}\sum_{t=0}^{\infty} \|v^t - v^{t+1}\|^2_H \leq \|v^0 - v^*\|^2_H, \ \forall v^* \in \mathcal{V}^*. \tag{5.11}
$$

Now observe that the sequences $\{\|v^k - v^{k+1}\|^2_H\}$ is monotonically nonincreasing in Theorem 5.3, we have

$$
(k+1)\|v^k - v^{k+1}\|^2_H \leq \sum_{i=0}^{k} \|v^i - v^{i+1}\|^2_H. \tag{5.12}
$$

Using the above two inequalities (5.11) and (5.12), we obtain (5.10). □

Recall that if $\|v^k - v^{k+1}\|_H^2 = 0$, then w^k is the solution point of VI$(\Omega; F; \theta)$. This allows us to use $\|v^k - v^{k+1}\|_H^2$ as an error measurement in terms of the distance to the solution set of VI$(\Omega; F; \theta)$ for the t-th iteration of the algorithm. Let $d := \inf\{\|v^0 - v^*\|_H \mid v^* \in \mathcal{V}^*\}$. Then, for any given $\varepsilon > 0$, Theorem 5.5 implies that these relaxed PPA needs at most $\lceil \frac{cd^2}{\varepsilon} \rceil$ iterations to ensure that $\|v^k - v^{k+1}\|_H^2 \leq \varepsilon$. Therefore, a worse-case $\mathcal{O}(1/t)$ convergence rate in the non-ergodic sense for these relaxed PPA is established.

5.2 Worse-case convergence rate in ergodic sense

In this section, we establish $\mathcal{O}(1/t)$ worse-case convergence rate in the ergodic sense for these relaxed C-PPA. First we introduce the following lemma, which is convenient for analyzing the ergodic iteration complexity of the algorithm. Its proof can be found in [17] or Theorem 2.3.5 in [9].

Theorem 5.6. *The solution set of VI$(\Omega; F; \theta)$ ((1.3), (3.2)) is closed and convex, and can be characterized via*

$$\Omega^* := \bigcap_{w\in\Omega} \{\tilde{w} \in \Omega : \ \theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(w) \geq 0\}.$$

Computing exactly a primal-dual solution w^* is impractical, but we can find an approximate solution $\tilde{w} \in \Omega$. With the Theorem 5.6, we define an ε-approximation solution of VI$(\Omega; F; \theta)$ as follows: Given an accuracy $\varepsilon > 0$, $\tilde{w}$ is said to be an ϵ-approximation solution of VI$(\Omega; F; \theta)$, if it satisfies

$$\sup_{u\in\mathcal{D}}\{\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w)\} \leq \varepsilon, \ \forall w \in \Omega,$$

where $\mathcal{D} \in \Omega$ is a defined compact set, i.e., $\mathcal{D} = \{w \mid \|w - \tilde{w}\| \leq 1\}$. Next, our purpose is to prove that we can find an ε-solution after t iterations.

Theorem 5.7. *Let $\tilde{w}^k$ be generated by the step* (2.1a) *and $\{v^k\}$ be generated by* (2.1b). *For any integer number $t > 0$, define $\tilde{w}_t$*

$$\tilde{w}_t = \frac{1}{t+1}\sum_{k=0}^{t} \tilde{w}^k, \tag{5.13}$$

then we have $\tilde{w}_t \in \Omega$ and

$$\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w) \leq \frac{1}{2(t+1)}\|v - v_0\|_H^2, \ \forall w \in \Omega, \tag{5.14}$$

where H is defined in (2.2a).

Proof. First, because of $\tilde{w}^k \in \Omega$, it holds that $\tilde{w}_t \in \Omega$ for all integer t.

Next, by using the definitions of F (see (1.3b) and (3.2b)), we obtain,

$$(w - \tilde{w}^k)^T F(w) \geq (w - \tilde{w}^k)^T F(\tilde{w}^k).$$

Also, using the positive definiteness of G and (4.1), we get

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) + \frac{1}{2}\|v - v^k\|_H^2 \geq \frac{1}{2}\|v - v^{k+1}\|_H^2, \ \forall w \in \Omega. \tag{5.15}$$

Summing inequality (5.15) from $k = 0$ to $k = t$, we derive

$$(t+1)\theta(u) - \sum_{k=1}^{t}\theta(\tilde{u}^k) + \big((t+1)w - \sum_{k=1}^{t}\tilde{w}^k\big)^T F(w) + \frac{1}{2}\|v - v^0\|_H^2 \geq 0, \ \forall w \in \Omega.$$

From the definition of $\tilde{w}_t$ (5.13), we can write above inequality as

$$\frac{1}{t+1}\sum_{k=1}^{t}\theta(\tilde{u}^k) - \theta(u) + (\tilde{w}_t - w)^T F(w) \le \frac{1}{2(t+1)}\|v - v^0\|_H^2, \ \forall w \in \Omega. \tag{5.16}$$

Noticing the fact that $\theta(u)$ is convex, for $\tilde{w}_t$, we have

$$\theta(\tilde{u}_t) \le \frac{1}{t+1}\sum_{k=1}^{t}\theta(\tilde{u}^k). \tag{5.17}$$

Substituting (5.17) into (5.16), we obtain (5.14). □

Theorem 5.7 implies that, after t iterations of (2.1), $\tilde{w}_t$ defined in (5.13) satisfies

$$\sup_{w\in\mathcal{D}}\{\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w)\} \le \frac{1}{2(t+1)}\|v - v_0\|_H^2, \ \forall w \in \Omega.$$

Thus, we obtained computational complexity estimates for the average of $\tilde{w}^k$, which is of order $\mathcal{O}(1/t)$ to get an $1/t$-approximation solution for VI$(\Omega; F; \theta)$.

6 Numerical results

In this section, we report some numerical results of the proposed algorithms. Note that the numerical comparison with related methods for the customized PPA have already been conducted in the literatures, see, e.g., [3, 16, 30, 27]. Thus it is omitted in our tests and we focus on reporting some numerical results of the new relaxed methods by comparing them to the original customized PPA, and verify the accelerated assertions in the computational sense. All codes were written in Matlab 2015b and all experiments were conducted on a laptop with Intel Core (TM) CPU 2.40GHz and 8G memory.

6.1 Matrix completion problem

The matrix completion problem consists of reconstructing an unknown low rank matrix from a given subset of observed entries. Specifically, let $M \in \Re^{m\times n}$ be a low rank matrix, Ω be a subset of the indices of entries $\{1, \cdots, l\} \times \{1, \cdots, n\}$, the convex model is

$$\min\{\|X\|_* \mid X_{ij} = M_{ij}, \ \forall\{ij\} \in \Omega\}, \tag{6.1}$$

where $\|\cdot\|_*$ denotes the nuclear norm of a matrix. Note that the constraints can also be viewed as a projection equation $P_\Omega(X) = P_\Omega(M)$, thus we have $\|A^T A\| = 1$.

In our implementations, for the SVD decomposition, we resort to the widely recommended PROPACK package [22]. we generate the rank-r_a matrix M as a product $M_L M_R^T$, where M_L and M_R are independent $n \times r_a$ matrices who have i.i.d. Gaussian entries. The set of observations Ω is sampled uniformly at random among all sets of cardinality $|\Omega|$. The degrees of freedom of the matrix is $d_{r_a} = r_a(2m - r_a)$, then the oversampling factor $|\Omega|/d_{r_a}$ is the ratio of the number of samples to the degrees of freedom.

In addition, the penalty parameters is set to be $r = 0.004$, $s = 1.01/r$ and the relaxation factor is $\gamma = 1.99$. The primal variable X^0 and the dual variable Λ^0 were all initialized as zeros(n), The stopping criteria is set to be

$$\frac{\|X_\Omega^k - M_\Omega\|_F}{\|M_\Omega\|_F} \le 10^{-4}. \tag{6.2}$$

Tables 1–3 shows the results of the PPA and their relaxed versions on problem (6.1). For various scenarios of r_a, $|\Omega|/d_{r_a}$ and $|\Omega|/n^2$, we report the number of iterations ("No. It"), computation time in seconds ("CPU."), the relative error ("RelErr") defined in (6.2). From the table, we see that the proposed relaxed PPA are significantly faster. Figure 1,2,3 shows the convergence behavior of three algorithms with $n = 300, rank = 12, |\Omega|/d_{r_a} = 5$, respectively. We can see that these relaxed PPA quickly reach to objective value at early iterations and the relative error (6.2) descends much faster.

Table 1: Performance comparison of PDHG, Relaxed PDHG.

Unknown $n \times n$ matrix M				PDHG		Relaxed PDHG	
n	$rank(r_a)$	$\|\Omega\|/d_{r_a}$	$\|\Omega\|/n^2$	No. It	RelErr	No. It	RelErr
500	5	5	0.10	179	9.77e-05	91	9.58e-05
1000	10	6	0.12	91	9.51e-05	68	9.01e-05
1500	10	8	0.11	104	9.06e-05	80	9.07e-05

Table 2: Performance comparison of C-PPA, Relaxed C-PPA.

Unknown $n \times n$ matrix M				C-PPA		Relaxed C-PPA	
n	$rank(r_a)$	$\|\Omega\|/d_{r_a}$	$\|\Omega\|/n^2$	No. It	RelErr	No. It	RelErr
500	5	5	0.10	177	9.77e-05	91	8.58e-05
1000	10	6	0.12	89	9.65e-05	66	9.61e-05
1500	10	8	0.11	102	9.04e-05	78	9.06e-05

Table 3: Performance comparison of L-ALM, Relaxed L-ALM.

Unknown $n \times n$ matrix M				L-ALM		Relaxed L-ALM	
n	$rank(r_a)$	$\|\Omega\|/d_{r_a}$	$\|\Omega\|/n^2$	No. It	RelErr	No. It	RelErr
500	5	5	0.10	178	9.77e-05	91	8.61e-05
1000	10	6	0.12	90	9.51e-05	67	9.12e-05
1500	10	8	0.11	103	9.06e-05	79	9.36e-05

6.2 Robust principal component analysis

In this subsection, we test the robust principal component analysis (RPCA) which has the form

$$\begin{array}{ll} \min_{L,S} & \|L\|_* + \tau \|S\|_1, \\ \text{s.t.} & L + S = M. \end{array} \tag{6.3}$$

where M is a known matrix. It is noted that the problem (6.3) is a standard model of (3.1). We now numerically verify that the iterates generated by ADMM and its relaxation for solving (6.3). In the experiments, we generated the data similar with [28]. More specially, we generate the rank-r_a matrix L as a product $L_L L_R^T$, where L_L and L_R are independent $n \times r_a$ matrices who have i.i.d. Gaussian entries. For a sparsity ratio sr, we generate $S = zeros(n,n); p = randperm(n*n); K = round(sr*n*n); S(p(1:K)) = randn(K,1);$ Then we use $L+S$ to generate the matrix M. The model parameter τ is set to $1/\sqrt{n}$. We simply set $\beta = 10 * \tau$, and the stopping criteria is set to be

$$\frac{\|M - L - S\|_F}{\|M\|_F} \leq 10^{-6}. \tag{6.4}$$

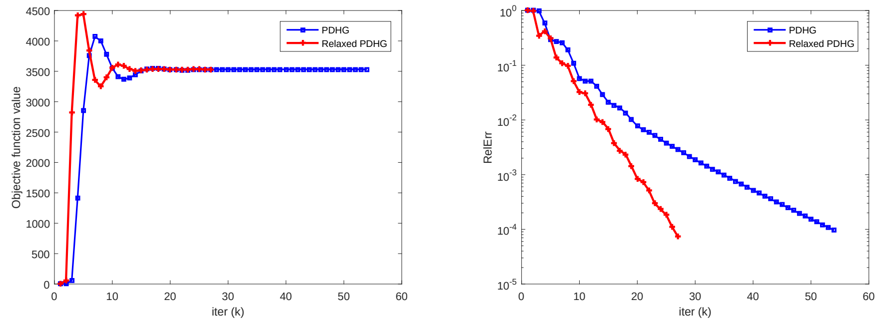

Figure 1: Evolution of objective value, residual for PDHG and its relaxation.

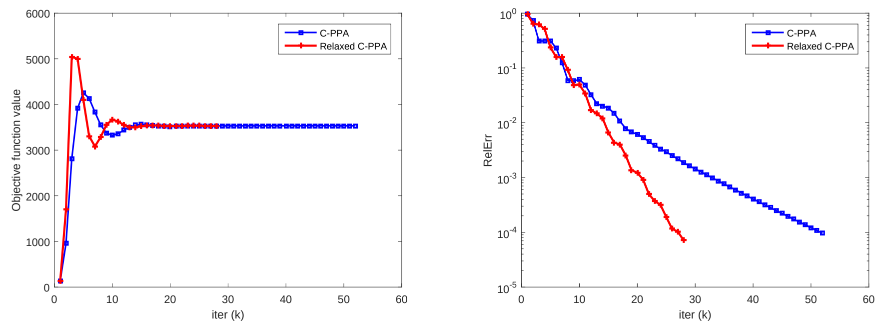

Figure 2: Evolution of the objective value, residual for C-PPA and its relaxation.

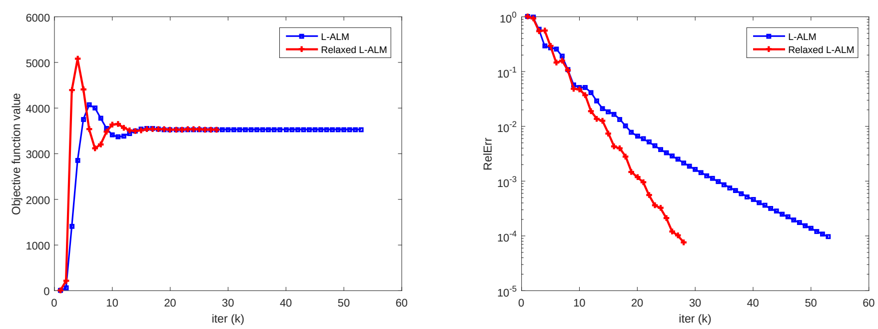

Figure 3: Evolution of objective value, residual for L-ALM and its relaxation.

We choose the initial point to be zero. The relaxation factor for relaxed ADMM is set to be $\gamma = 1.7$.

The results and performance of these algorithms are reported in Table 4. As we can see through this example that relaxed ADMM requires less iterations than ADMM while produces the same accurate solutions. To show the convergence behavior of the algorithms more clearly, we plot the results in Figure 4 with $n = 500, rank = 5, \mathrm{sr} = 0.1$.

7 Conclusions

In the literature, it is well known that the PPA can be accelerated by using the relaxation strategy proposed in [11]. For many cases, the relaxed PPA with $\gamma \in [1.5, 1.8]$ can obtain a guaranteed improvement of the convergence. In this paper, different from most previous works, we suggest to

Table 4: Performance comparison of ADMM, Relaxed ADMM.

$n \times n$ matrix			ADMM		Relaxed ADMM	
n	$rank(r_a)$	sr	No. It	RelErr	No. It	RelErr
500	5	0.1	118	3.02e-07	89	9.67e-07
1000	10	0.1	144	3.74e-07	103	8.22e-07
1500	15	0.1	180	4.31e-07	124	7.71e-07

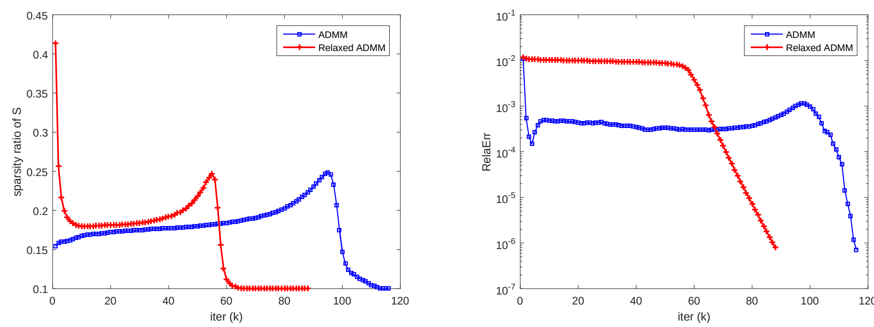

Figure 4: Sparsity ratio (S), and residual for ADMM and its relaxation.

relax the PPA by correcting the dual variables individually. For the PDHG, the PPA proposed in [16, 12], the linearized augmented Lagrangian method (L-ALM) and the alternating direction method of multipliers (ADMM), which are special cases of the customized PPA, we studied their relaxed schemes and observed the corresponding speedup. With the guidance of the variational inequality, we easily established the global convergence of these relaxed algorithms and a worst-case $\mathcal{O}(1/t)$ convergence rate in both ergodic and nonergodic sense. In conclusion, our results theoretically provide an alternative way of relaxing PPA.

Finally, we mention that this paper only discussed the customized PPA for the convex problems less than two block. In fact, our relaxation strategy can be further extended to the customized PPA for multi-block case, such as the Jacobian ADMM [25, 7] and its linearized version, the partial parallel ADMM [15], etc. We leave the details to the interested reader.

References

[1] Beck, Amir, and Marc Teboulle. *A fast iterative shrinkage-thresholding algorithm for linear inverse problems.* SIAM J. Imag. Sci., 2(2009), pp. 183-202.

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn., 3 (2010), pp. 1-122.

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

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

[5] T. F. Chan and R. Glowinski, *Finite Element Approximation and Iterative Solution of a Class of Mildly Nonlinear Elliptic Equations*, Technical Report STAN-CS-78-674, Computer Science Department, Stanford University, 1978.

[6] X. J. Cai, G. Y. Gu, B. S. He, and X. M. Yuan. *A proximal point algorithm revisit on the alternating direction method of multipliers*. Sci. China Math., 56(2013), pp. 2179-2186.

[7] W. Deng, M. J. Lai, Z. M. Peng and W. T. Yin. *Parallel multi-block ADMM with o(1/k) convergence*. , J. Sci. Comput., (2016), pp. 1-25.

[8] J. Eckstein and D.P. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Program., 55(1992), pp. 293-318.

[9] F. Facchinei and J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity problems, Volume I*, Springer Series in Operations Research, Springer-Verlag, 2003.

[10] R. Glowinski, and A. Marrocco, *Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires*, Revue Fr. Autom. Inform. Rech. Opér., Anal. Numér. 2(1975), pp. 41–76.

[11] E. G. Gol'shtein, N. V. Tret'yakov. *Modified Lagrangians in convex programming and their generalizations*. Math. Program. Stud., 10(1979), pp. 86-97.

[12] G. Y. Gu, B. S. He, and X. M. Yuan. *Customized proximal point algorithms for linearly constrained convex minimization and saddle-point problems: a unified approach*. Comput. Optim. Appl., 59(2014), pp. 135-161.

[13] D. R. Han, H. J. He, H. Yang, and X. M. Yuan. *A customized Douglas–Rachford splitting algorithm for separable convex minimization with linear constraints*. Numer. Math., 127(2014), pp. 167-200.

[14] B. S. He. *PPA-Like contraction methods for convex optimization: a framework using variational inequality approach*. J. Oper. Res. Soc. China, 3(2015), pp. 391-420.

[15] B. S. He and X. M. Yuan. *Block-wise alternating direction method of multipliers for multiple-block convex programming and beyond*, preprint, 2014.

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

[17] B. S. He and X. M. Yuan, *Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective*, SIAM J. Imaging Sci., 5(2012), pp. 119C149.

[18] B. S. He and X. M. Yuan, *A class of ADMM-based algorithms for multi-block separable convex programming*, preprint, 2015.

[19] M. R. Hestenes, *Multiplier and gradient methods*, J. Optim. Theory Appli., 4(1969), pp. 303-320.

[20] I. Daubechies, M. Defrise, and C. D. Mol. *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Commun. Pur. Appl. Math., 57(2004), pp. 1413-1457.

[21] Martinet, B. *Regularisation, d'inéquations variationelles par approximations succesives*, Rev. Francaise d'Inform. Recherche Oper. 4(1970), pp. 154-159.

[22] R. M. Larsen, *PROPACK Software for large and sparse svd calculations*, Standford University, 1969, url=http://sun.stanford.edu/ rmunk/PROPACK/.

[23] M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, In Optimization edited by R. Fletcher, pp. 283-298, Academic Press, New York, 1969.

[24] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim., 14(1976), pp. 877–898.

[25] X. F. Wang, M. Y. Hong, S. Q. Ma and Z. Q. Luo. *Solving Multiple-Block Separable Convex Minimization Problems Using Two-Block Alternating Direction Method of Multipliers*, Pacific J. Optim., 11(2015), pp. 645-667.

[26] Y. Shen and H. Y. Wang, *New augmented Lagrangian-based proximal point algorithm for convex optimization with equality constraints*, J. Optim. Theory Appli., 171(2016), pp. 251-261.

[27] J. F. Yang and X. M. Yuan, *Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization*, Math. Comput., 82(2013), pp. 301–329.

[28] X. M. Yuan and J. F. Yang, *Sparse and low-rank matrix decomposition via alternating direction methods*, Pacific J. Optim., 9(2013), pp. 167–180.

[29] X. Q. Zhang, M. Burger, X. Bresson and S. Osher, *Bregmanized nonlocal regularization for deconvolution and sparse reconstruction*, SIAM J. Imag. Sci., 3(2010), pp. 253-276.

[30] X. Q. Zhang, M. Burger, and S. Osher, *A unified primal-dual algorithm framework based on Bregman iteration*, J. Sci. Comput., 46(2010), pp. 20–46.

[31] M. Zhu and T.F. Chan, *An efficient primal-dual hybrid gradient algorithm for total variation image restoration*, CAM Report 08-34, UCLA, USA (2008).