SIAM J. OPTIM.
Vol. 25, No. 4, pp. 2120–2142

A GENERAL INERTIAL PROXIMAL POINT ALGORITHM FOR MIXED VARIATIONAL INEQUALITY PROBLEM*

CAIHUA CHEN†, SHIQIAN MA‡, AND JUNFENG YANG§

Abstract. In this paper, we first propose a general inertial *proximal point algorithm* (PPA) for the mixed *variational inequality* (VI) problem. Based on our knowledge, without stronger assumptions, a convergence rate result is not known in the literature for inertial type PPAs. Under certain conditions, we are able to establish the global convergence and nonasymptotic $O(1/k)$ convergence rate result (under a certain measure) of the proposed general inertial PPA. We then show that both the linearized *augmented Lagrangian method* (ALM) and the linearized *alternating direction method of multipliers* (ADMM) for structured convex optimization are applications of a general PPA, provided that the algorithmic parameters are properly chosen. Consequently, global convergence and convergence rate results of the linearized ALM and ADMM follow directly from results existing in the literature. In particular, by applying the proposed inertial PPA for mixed VI to structured convex optimization, we obtain inertial versions of the linearized ALM and ADMM whose global convergence is guaranteed. We also demonstrate the effect of the inertial extrapolation step via experimental results on the compressive principal component pursuit problem.

Key words. inertial proximal point algorithm, mixed variational inequality, inertial linearized augmented Lagrangian method, inertial linearized alternating direction method of multipliers

AMS subject classifications. 65K05, 65K10, 65J22, 90C25

DOI. 10.1137/140980910

1. Introduction.

Let $T : \Re^n \rightrightarrows \Re^n$ be a set-valued maximal monotone operator. The maximal monotone operator inclusion problem is to find $w^* \in \Re^n$ such that

$$0 \in T(w^*). \tag{1.1}$$

Due to the mathematical generality of maximal monotone operators, the problem (1.1) is very inclusive and serves as a unified model for many problems of fundamental importance, for example, the fixed point problem, the variational inequality (VI) problem, minimization of closed proper convex functions, and their extensions. Therefore, it becomes extremely important in many cases to solve (1.1) in practical and efficient ways.

The classical *proximal point algorithm* (PPA), which converts the maximal monotone operator inclusion problem into a fixed point problem of a firmly nonexpansive mapping via resolvent operators, is one of the most influential approaches for solving (1.1) and has been studied extensively both in theory and in practice. The PPA

*Received by the editors August 4, 2014; accepted for publication (in revised form) August 24, 2015; published electronically October 20, 2015.
http://www.siam.org/journals/siopt/25-4/98091.html

†International Center of Management Science and Engineering, School of Management and Engineering, Nanjing University, China (chchen@nju.edu.cn). This author's research was supported by the Natural Science Foundation of Jiangsu Province (BK20130550) and the Natural Science Foundation of China (NSFC-11401300, 71271112).

‡Department of Systems Engineering and Engineering Management, Chinese University of Hong Kong, Hong Kong (sqma@se.cuhk.edu.hk). This author was supported by Hong Kong Research Grants Council General Research Fund Early Career Scheme (project CUHK 439513).

§Corresponding author. Department of Mathematics, Nanjing University, China (jfyang@nju.edu.cn). This author was supported by the Natural Science Foundation of China (NSFC-11371192, 91130007), the Fundamental Research Funds for the Central Universities (20620140574), and a grant from Jiangsu Key Laboratory for Numerical Simulation of Large Scale Complex Systems. The work was done while this author was visiting the Chinese University of Hong Kong.

was originally proposed by Martinet [32] based on the work of Moreau [33] and was popularized by Rockafellar [41]. It turns out that the PPA is a very powerful algorithmic tool and contains many well-known algorithms as special cases. In particular, it was shown that the classical *augmented Lagrangian method* (ALM) for constrained optimization [27, 39], the Douglas–Rachford operator splitting method [19], and the *alternating direction method of multipliers* (ADMM) [23, 22] are all applications of the PPA; see [40, 20]. Various inexact, relaxed, and accelerated variants of the PPA were also very well studied in the literature; see, e.g., [41, 20, 24].

The PPA for minimizing a differentiable function $f: \Re^n \to \Re$ can be interpreted as an implicit one-step discretization method for the ordinary differential equations

$$w' + \nabla f(w) = 0, \tag{1.2}$$

where $w: \Re \to \Re^n$ is differentiable, w' denotes its derivative, and ∇f is the gradient of f. Suppose that f is closed and convex and its minimum value is attained; then every solution trajectory $\{w(t): \ t \geq 0\}$ of the differential system (1.2) converges to a minimizer of f as t goes to infinity. Similar conclusion can be drawn for (1.1) by considering the evolution differential inclusion problem $0 \in w'(t) + T(w(t))$ almost everywhere on $\Re_+$, provided that the operator T satisfies certain conditions; see e.g., [14].

The PPA is a one-step iterative method, i.e., each new iterate point does not depend on any iterate points already generated other than the current one. To speed up convergence, multistep methods have been proposed in the literature by discretizing a second-order ordinary differential system of the form

$$w'' + \gamma w' + \nabla f(w) = 0, \tag{1.3}$$

where $\gamma > 0$. Studies in this direction can be traced back to at least [38], which examined the system (1.3) in the context of optimization. In the two-dimensional case, the system (1.3) characterizes roughly the motion of a heavy ball which rolls under its own inertia over the graph of f until friction stops it at a stationary point of f. The three terms in (1.3) denote, respectively, inertial force, friction force, and gravity force. Therefore, the system (1.3) is usually referred to as the *heavy-ball with friction* (HBF) system. It is easy to show that the energy function $E(t) = \frac{1}{2}\|w'(t)\|^2 + f(w(t))$ is always decreasing with time t unless w' vanishes, which implies that the HBF system is dissipative. It was proved in [2] that if f is convex and its minimum value is attained, then each solution trajectory $\{w(t): t \geq 0\}$ of (1.3) converges to a minimizer of f. In theory the convergence of the solution trajectories of the HBF system to a stationary point of f can be faster than those of the first-order system (1.2), while in practice the second-order inertial term w'' can be exploited to design faster algorithms [1, 5]. Motivated by the properties of (1.3), an implicit discretization method was proposed in [2]. Specifically, given w^{k-1} and w^k, the next point w^{k+1} is determined via

$$\frac{w^{k+1} - 2w^k + w^{k-1}}{h^2} + \gamma \frac{w^{k+1} - w^k}{h} + \nabla f(w^{k+1}) = 0,$$

which results to an iterative algorithm of the form

$$w^{k+1} = (I + \lambda \nabla f)^{-1}(w^k + \alpha(w^k - w^{k-1})), \tag{1.4}$$

where $\lambda = h^2/(1+\gamma h)$ and $\alpha = 1/(1+\gamma h)$. Note that (1.4) is nothing but a proximal point step applied to the extrapolated point $w^k + \alpha(w^k - w^{k-1})$, rather than w^k as in the classical PPA. Thus the resulting iterative scheme (1.4) is a two-step method and is usually referred to as an inertial PPA. Convergence properties of (1.4) were studied in [2] under some assumptions on the parameters α and λ. Subsequently, this inertial technique was extended to solve the inclusion problem (1.1) of maximal monotone operators in [4]. See also [34] for approximate inertial PPA and [3, 31, 30] for some inertial type hybrid proximal algorithms. Recently, there has been increasing interest in studying inertial type algorithms. Recent work includes inertial forward-backward splitting methods for certain separable nonconvex optimization problems [37] and for strongly convex problems [36, 6], inertial versions of the Douglas–Rachford operator splitting method and the ADMM for the maximal monotone operator inclusion problem [12, 9], and the inertial forward-backward-forward method [11] based on Tseng's approach [42]. See also [28, 10, 8].

1.1. Contributions. In this paper, we focus on the mixed VI problem and study inertial PPA under a more general setting. In particular, a weighting matrix G in the proximal term is introduced. In our setting the matrix G is allowed to be positive semidefinite, as long as it is positive definite in the null space of a certain matrix. We establish its global convergence and a nonasymptotic $O(1/k)$ convergence rate result under certain conditions. To the best of our knowledge, without stronger assumptions, the convergence rate result is not known in the literature for general inertial type PPAs. This general setting allows us to propose inertial versions of the linearized ALM and ADMM, which are practical variants of the well-known ALM and ADMM that have recently found numerous applications [13]. Indeed, this is realized by showing that both the linearized ALM and the linearized ADMM for structured convex optimization are applications of a general PPA to the primal-dual optimality conditions, as long as the parameters are properly chosen. Another aim of this paper is to study the effect of the inertial extrapolation step via numerical experiments. Finally, we connect inertial type algorithms with the popular accelerated methods pioneered by Nesterov [35] and give some concluding remarks.

The main reason that we restrict our analysis to the mixed VI problem rather than the apparently more general problem (1.1) is because it is very convenient to represent the optimality conditions of linearly constrained separable convex optimization as mixed VI. In fact, our analysis for Theorems 1 and 2 can be generalized to the maximal monotone operator inclusion problem (1.1) without any difficulty.

1.2. Notation. We use the following notation. The standard inner product and ℓ_2 norm are denoted by $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$, respectively. The sets of symmetric, symmetric positive semidefinite, and symmetric positive definite matrices of order n are, respectively, denoted by S^n, S^n_+, and S^n_{++}. For any matrix $A \in S^n_+$ and vectors $u, v \in \Re^n$, we let $\langle u,v\rangle_A := u^T Av$ and $\|u\|_A := \sqrt{\langle u,u\rangle_A}$. The Frobenius norm is denoted by $\|\cdot\|_F$. The spectral radius of a square matrix M is denoted by $\rho(M)$.

2. A general inertial PPA for mixed VI. Let $\mathcal{W} \subseteq \Re^n$ be a closed and convex set, $\theta : \Re^n \to \Re$ be a closed convex function, and $F : \Re^n \to \Re^n$ be a monotone mapping. In this paper, we consider the mixed VI problem: find $w^* \in \mathcal{W}$ such that

$$\theta(w) - \theta(w^*) + \langle w - w^*, F(w^*)\rangle \geq 0 \; \forall w \in \mathcal{W}. \tag{2.1}$$

Let $G \in S^n_+$ and two sequences of parameters $\{\alpha_k \geq 0 : k = 0, 1, 2, \ldots\}$ and $\{\lambda_k > 0 : k = 0, 1, 2, \ldots\}$ be given. We study a general inertial PPA of the following form: given any $w^0 = w^{-1} \in \Re^n$, for $k = 0, 1, 2, \ldots$, find $w^{k+1} \in \mathcal{W}$ such that

(2.2a) $\bar{w}^k := w^k + \alpha_k(w^k - w^{k-1}),$

(2.2b)

$$\theta(w) - \theta(w^{k+1}) + \langle w - w^{k+1}, F(w^{k+1}) + \lambda_k^{-1} G(w^{k+1} - \bar{w}^k)\rangle \geq 0 \;\forall\, w \in \mathcal{W}.$$

We make the following assumptions.

Assumption 1. The set of solutions of (2.1), denoted by $\mathcal{W}^*$, is nonempty.

Assumption 2. The mapping F is H-monotone in the sense that

$$\langle u - v, F(u) - F(v)\rangle \geq \|u - v\|_H^2 \quad \forall u, v \in \Re^n, \tag{2.3}$$

where $H \in S_+^n$. Note that $H = 0$ if F is monotone, and $H \in S_{++}^n$ if F is strongly monotone.

Assumption 3. The sum of G and H, denoted by M, is positive definite, i.e., $M := G + H \in S_{++}^n$.

Under Assumptions 2 and 3, it can be shown that w^{k+1} is uniquely determined in (2.2b). Therefore, the algorithm (2.2a)–(2.2b) is well defined. Clearly, the algorithm reduces to the classical PPA if $G \in S_{++}^n$ and $\alpha_k = 0$ for all k. It is called inertial PPA because α_k can be greater than 0. We will impose conditions on α_k to ensure global convergence of the general inertial PPA (2.2). Our convergence results are extensions of those in [4].

THEOREM 1. *Assume that Assumptions* 1, 2, *and* 3 *hold. Let* $\{w^k\}_{k=0}^\infty \subseteq \Re^n$ *conform to algorithm* (2.2a)–(2.2b). *The parameters* $\{\alpha_k, \lambda_k\}_{k=0}^\infty$ *satisfy, for all* k, $0 \leq \alpha_k \leq \alpha$ *for some* $\alpha \in [0, 1)$ *and* $\lambda_k \geq \lambda$ *for some* $\lambda > 0$. *If*

$$\sum_{k=1}^{\infty} \alpha_k \|w^k - w^{k-1}\|_G^2 < \infty, \tag{2.4}$$

then the sequence $\{w^k\}_{k=0}^\infty$ *converges to some point in* $\mathcal{W}^*$ *as* $k \to \infty$.

Proof. First, we show that for any $w^* \in \mathcal{W}^*$, $\lim_{k\to\infty} \|w^k - w^*\|_M$ exists. As a result, $\{w^k\}_{k=0}^\infty$ is bounded and must have a limit point. Then, we show that any limit point of $\{w^k\}_{k=0}^\infty$ must lie in $\mathcal{W}^*$. Finally, we establish the convergence of $\{w^k\}_{k=0}^\infty$ to a point in $\mathcal{W}^*$ as $k \to \infty$.

Let $w^* \in \mathcal{W}^*$ be arbitrarily chosen and $k \geq 0$. It follows from setting $w = w^* \in \mathcal{W}^*$ in (2.2b) and the H-monotonicity (2.3) of F that

$$\begin{aligned}
\lambda_k^{-1}\langle w^{k+1} - w^*, w^{k+1} - \bar{w}^k\rangle_G &\leq \theta(w^*) - \theta(w^{k+1}) - \langle w^{k+1} - w^*, F(w^{k+1})\rangle \\
&\leq \theta(w^*) - \theta(w^{k+1}) - \langle w^{k+1} - w^*, F(w^*)\rangle \\
&\quad - \|w^{k+1} - w^*\|_H^2 \\
&\leq -\|w^{k+1} - w^*\|_H^2.
\end{aligned} \tag{2.5}$$

Define $\varphi_k := \|w^k - w^*\|_G^2$ and recall that $\bar{w}^k = w^k + \alpha_k(w^k - w^{k-1})$. Plugging the identities

$$\begin{aligned}
2\langle w^{k+1} - w^*,\, w^{k+1} - w^k\rangle_G &= \varphi_{k+1} - \varphi_k + \|w^{k+1} - w^k\|_G^2, \\
2\langle w^{k+1} - w^*,\, w^k - w^{k-1}\rangle_G &= \varphi_k - \varphi_{k-1} + \|w^k - w^{k-1}\|_G^2 \\
&\quad + 2\langle w^{k+1} - w^k,\, w^k - w^{k-1}\rangle_G
\end{aligned}$$

into (2.5) and reorganizing, we obtain

$$\begin{aligned} \psi_k &:= \varphi_{k+1} - \varphi_k - \alpha_k \left(\varphi_k - \varphi_{k-1}\right) \\ &\le -\|w^{k+1} - w^k\|_G^2 + 2\alpha_k \langle w^{k+1} - w^k,\, w^k - w^{k-1}\rangle_G + \alpha_k \|w^k - w^{k-1}\|_G^2 \\ &\quad -2\lambda_k \|w^{k+1} - w^*\|_H^2 \\ &= -\|w^{k+1} - \bar{w}^k\|_G^2 + (\alpha_k^2 + \alpha_k)\|w^k - w^{k-1}\|_G^2 - 2\lambda_k\|w^{k+1} - w^*\|_H^2 \\ &\le -\|w^{k+1} - \bar{w}^k\|_G^2 + 2\alpha_k\|w^k - w^{k-1}\|_G^2 - 2\lambda_k\|w^{k+1} - w^*\|_H^2 \\ &\le -\|w^{k+1} - \bar{w}^k\|_G^2 + 2\alpha_k\|w^k - w^{k-1}\|_G^2, \end{aligned} \tag{2.6}$$

where the first inequality is due to (2.5) and the second follows from $0 \le \alpha_k < 1$. Define

$$\nu_k := \varphi_k - \varphi_{k-1} \quad \text{and} \quad \delta_k := 2\alpha_k\|w^k - w^{k-1}\|_G^2.$$

Then, the inequality (2.6) implies that $\nu_{k+1} \le \alpha_k \nu_k + \delta_k \le \alpha[\nu_k]_+ + \delta_k$, where $[t]_+ := \max\{t, 0\}$ for $t \in \Re$. Therefore, we have

$$[\nu_{k+1}]_+ \le \alpha[\nu_k]_+ + \delta_k \le \alpha^{k+1}[\nu_0]_+ + \sum_{j=0}^{k} \alpha^j \delta_{k-j}. \tag{2.7}$$

Note that by our assumption $w^0 = w^{-1}$. This implies that $\nu_0 = [\nu_0]_+ = 0$ and $\delta_0 = 0$. Therefore, it follows from (2.7) that

$$\sum_{k=0}^{\infty} [\nu_k]_+ \le \frac{1}{1-\alpha} \sum_{k=0}^{\infty} \delta_k = \frac{1}{1-\alpha}\sum_{k=1}^{\infty} \delta_k < \infty. \tag{2.8}$$

Here the second inequality is due to the assumption (2.4). Let $\gamma_k := \varphi_k - \sum_{j=1}^{k} [\nu_j]_+$. From (2.8) and $\varphi_k \ge 0$, it follows that γ_k is bounded below. On the other hand,

$$\gamma_{k+1} = \varphi_{k+1} - [\nu_{k+1}]_+ - \sum_{j=1}^{k}[\nu_j]_+ \le \varphi_{k+1} - \nu_{k+1} - \sum_{j=1}^{k}[\nu_j]_+ = \varphi_k - \sum_{j=1}^{k}[\nu_j]_+ = \gamma_k,$$

i.e., γ_k is nonincreasing. As a result, $\{\gamma_k\}_{k=0}^{\infty}$ converges as $k \to \infty$, and the limit

$$\lim_{k\to\infty} \varphi_k = \lim_{k\to\infty} \left(\gamma_k + \sum_{j=1}^{k} [\nu_j]_+\right) = \lim_{k\to\infty} \gamma_k + \sum_{k=1}^{\infty}[\nu_k]_+$$

exists. That is, $\lim_{k\to\infty} \|w^k - w^*\|_G$ exists for any $w^* \in \mathcal{W}^*$. Furthermore, it follows from the second "$\le$" of (2.6) and the definition of ν_k and δ_k that

$$\begin{aligned} \|w^{k+1} - \bar{w}^k\|_G^2 + 2\lambda_k\|w^{k+1} - w^*\|_H^2 &\le \varphi_k - \varphi_{k+1} + \alpha_k\left(\varphi_k - \varphi_{k-1}\right) + \delta_k \\ &\le \varphi_k - \varphi_{k+1} + \alpha[\nu_k]_+ + \delta_k. \end{aligned} \tag{2.9}$$

By taking sum over k and noting that $\varphi_k \ge 0$, we obtain

$$\sum_{k=1}^{\infty} \left(\|w^{k+1} - \bar{w}^k\|_G^2 + 2\lambda_k \|w^{k+1} - w^*\|_H^2\right) \le \varphi_1 + \sum_{k=1}^{\infty}\left(\alpha[\nu_k]_+ + \delta_k\right) < \infty, \tag{2.10}$$

where the second inequality follows from (2.8) and assumption (2.4). Since $\lambda_k \geq \lambda > 0$ for all k, it follows from (2.10) that

$$\lim_{k\to\infty} \|w^k - w^*\|_H = 0. \tag{2.11}$$

Recall that $M = G + H$. Thus, $\lim_{k\to\infty} \|w^k - w^*\|_M$ exists. Since M is positive definite, it follows that $\{w^k\}_{k=0}^{\infty}$ is bounded and must have at least one limit point.

Again from (2.10) we have

$$\lim_{k\to\infty} \|w^{k+1} - \bar{w}^k\|_G = 0.$$

Thus, the positive semidefiniteness of G implies that $\lim_{k\to\infty} G(w^{k+1} - \bar{w}^k) = 0$. On the other hand, for any fixed $w \in \mathcal{W}$, it follows from (2.2b) that

$$\theta(w) - \theta(w^k) + \langle w - w^k, F(w^k)\rangle \geq \lambda_{k-1}^{-1}\langle w^k - w, G(w^k - \bar{w}^{k-1})\rangle. \tag{2.12}$$

Suppose that $w^\star$ is any limit point of $\{w^k\}_{k=0}^{\infty}$ and $w^{k_j} \to w^\star$ as $j \to \infty$. Since $\mathcal{W}$ is closed, $w^\star \in \mathcal{W}$. Furthermore, by taking the limit over $k = k_j \to \infty$ in (2.12) and noting that $G(w^k - \bar{w}^{k-1}) \to 0$ and $\lambda_{k-1} \geq \lambda > 0$, we obtain

$$\theta(w) - \theta(w^\star) + \langle w - w^\star, F(w^\star)\rangle \geq 0.$$

Since w can vary arbitrarily in $\mathcal{W}$, we conclude that $w^\star \in \mathcal{W}^*$. That is, any limit point of $\{w^k\}_{k=0}^{\infty}$ must also lie in $\mathcal{W}^*$.

Finally, we establish the uniqueness of limit points of $\{w^k\}_{k=0}^{\infty}$. Suppose that w_1^* and w_2^* are two limit points of $\{w^k\}_{k=0}^{\infty}$ and $\lim_{j\to\infty} w^{i_j} = w_1^*$, $\lim_{j\to\infty} w^{k_j} = w_2^*$. Assume that $\lim_{k\to\infty} \|w^k - w_i^*\|_M = v_i$ for $i = 1, 2$. By taking the limit over $k = i_j \to \infty$ and $k = k_j \to \infty$ in the equality

$$\|w^k - w_1^*\|_M^2 - \|w^k - w_2^*\|_M^2 = \|w_1^* - w_2^*\|_M^2 + 2\langle w_1^* - w_2^*, w_2^* - w^k\rangle_M,$$

we obtain $v_1 - v_2 = -\|w_1^* - w_2^*\|_M^2 = \|w_1^* - w_2^*\|_M^2$. Thus, $\|w_1^* - w_2^*\|_M = 0$. Since M is positive definite, this implies that $w_1^* = w_2^*$. Therefore, $\{w^k\}_{k=0}^{\infty}$ converges to some point in $\mathcal{W}^*$ and the proof of the theorem is completed. ∎

We have the following remarks on the assumptions and results of Theorem 1.

Remark 1. In practice, it is not hard to select α_k dynamically based on historical iterative information such that the condition (2.4) is satisfied.

Remark 2. If $\alpha_k = 0$ for all k, then the condition (2.4) is obviously satisfied. In this case, we reestablished the convergence of the classical PPA under the weaker condition that $G \in S_+^n$, provided that $\lambda_k \geq \lambda > 0$ and $H + G \in S_{++}^n$, e.g., when F is strongly monotone, i.e., $H \in S_{++}^n$.

Remark 3. Suppose that $H = 0$ and $G \in S_+^n$, but $G \notin S_{++}^n$. Then, the sequence $\{w^k\}_{k=0}^{\infty}$ may not be well defined since (2.2b) does not necessarily have a solution in general. In the case that $\{w^k\}_{k=0}^{\infty}$ is indeed well defined (which is possible), the conclusion that $\lim_{k\to\infty} \|w^k - w^*\|_G$ exists for any $w^* \in \mathcal{W}^*$ still holds under condition (2.4). However, since G is only positive semidefinite, the boundedness of $\{w^k\}_{k=0}^{\infty}$ cannot be guaranteed. If a limit point $w^\star$ of $\{w^k\}_{k=0}^{\infty}$ does exist, then the conclusion $w^\star \in \mathcal{W}^*$ still holds. Moreover, suppose that $w_1^\star$ and $w_2^\star$ are any two limit points of $\{w^k\}_{k=0}^{\infty}$; then it holds that $Gw_1^\star = Gw_2^\star$.

In the following theorem, we remove the assumption (2.4) by assuming that the sequence $\{\alpha_k\}_{k=0}^{\infty}$ satisfies some additional easily implementable conditions. Moreover, we establish a nonasymptotic $O(1/k)$ convergence rate result for the general inertial PPA (2.2). To the best of our knowledge, there is no convergence rate result known in the literature without stronger assumptions for inertial type PPAs.

THEOREM 2. *Assume that Assumptions* 1, 2, *and* 3 *hold. Suppose that the parameters* $\{\alpha_k, \lambda_k\}_{k=0}^\infty$ *satisfy, for all* k, $0 \le \alpha_k \le \alpha_{k+1} \le \alpha < \frac{1}{3}$ *and* $\lambda_k \ge \lambda$ *for some* $\lambda > 0$. *Let* $\{w^k\}_{k=0}^\infty$ *be the sequence generated by algorithm* (2.2a)–(2.2b). *Then, we have the following results:*

1. $\{w^k\}_{k=0}^\infty$ *converges to some point in* $\mathcal{W}^*$ *as* $k \to \infty$.
2. *For any* $w^* \in \mathcal{W}^*$ *and positive integer* k, *it holds that*

$$\min_{0\le i\le k-1} \|w^{i+1} - \bar{w}^i\|_G^2 \le \frac{\left(1 + \frac{2}{1-3\alpha}\right) \|w^0 - w^*\|_G^2}{k}. \tag{2.13}$$

Proof. Let $w^* \in \mathcal{W}^*$ be arbitrarily fixed and, for all $k \ge 0$, retain the notation $\varphi_k = \|w^k - w^*\|_G^2$,

$$\psi_k = \varphi_{k+1} - \varphi_k - \alpha_k (\varphi_k - \varphi_{k-1}) \quad \text{and} \quad \nu_k = \varphi_k - \varphi_{k-1}.$$

It follows from the first "$\le$" in (2.6) and $\lambda_k \ge 0$ that

$$\begin{aligned}\psi_k &\le -\|w^{k+1} - w^k\|_G^2 + 2\alpha_k \langle w^{k+1} - w^k, w^k - w^{k-1}\rangle_G + \alpha_k \|w^k - w^{k-1}\|_G^2 \\ &\le -\|w^{k+1} - w^k\|_G^2 + \alpha_k \left(\|w^{k+1} - w^k\|_G^2 + \|w^k - w^{k-1}\|_G^2\right) + \alpha_k \|w^k - w^{k-1}\|_G^2 \\ &= -(1-\alpha_k)\|w^{k+1} - w^k\|_G^2 + 2\alpha_k \|w^k - w^{k-1}\|_G^2,\end{aligned} \tag{2.14}$$

where the second "$\le$" follows from the Cauchy–Schwartz inequality. Define

$$\mu_k := \varphi_k - \alpha_k \varphi_{k-1} + 2\alpha_k \|w^k - w^{k-1}\|_G^2.$$

From $0 \le \alpha_k \le \alpha_{k+1} \le \alpha < \frac{1}{3}$, the fact that $\varphi_k \ge 0$, and (2.14), we have

$$\begin{aligned}\mu_{k+1} - \mu_k &= \varphi_{k+1} - \alpha_{k+1}\varphi_k + 2\alpha_{k+1}\|w^{k+1} - w^k\|_G^2 \\ &\quad - \left(\varphi_k - \alpha_k \varphi_{k-1} + 2\alpha_k \|w^k - w^{k-1}\|_G^2\right) \\ &\le \psi_k + 2\alpha_{k+1}\|w^{k+1} - w^k\|_G^2 - 2\alpha_k \|w^k - w^{k-1}\|_G^2 \\ &\le -(1-\alpha_k)\|w^{k+1} - w^k\|_G^2 + 2\alpha_{k+1}\|w^{k+1} - w^k\|_G^2 \\ &\le -(1-3\alpha)\|w^{k+1} - w^k\|_G^2 \\ &\le 0.\end{aligned} \tag{2.15}$$

Thus, $\mu_{k+1} \le \mu_k$ for all $k \ge 0$. Note that $w^0 = w^{-1}$ by our assumption. It follows from the definitions of μ_k and φ_k that $\mu_0 = (1-\alpha_0)\varphi_0 \le \varphi_0 = \|w^0 - w^*\|_G^2$. Therefore, we have

$$-\alpha\varphi_{k-1} \le \varphi_k - \alpha\varphi_{k-1} \le \varphi_k - \alpha_k\varphi_{k-1} \le \mu_k \le \mu_0 \le \varphi_0. \tag{2.16}$$

Further taking into account (2.15), we obtain

$$\varphi_k \le \alpha\varphi_{k-1} + \varphi_0 \le \alpha^k \varphi_0 + \varphi_0 \sum_{j=0}^{k-1} \alpha^j \le \alpha^k \varphi_0 + \frac{\varphi_0}{1-\alpha}. \tag{2.17}$$

The second-to-last "$\le$" in (2.15) implies that $(1-3\alpha)\|w^{k+1}-w^k\|_G^2 \le \mu_k - \mu_{k+1}$ for $k \ge 0$. Together with (2.16) and (2.17), this implies

$$(1-3\alpha)\sum_{j=0}^{k}\|w^{j+1}-w^j\|_G^2 \le \mu_0-\mu_{k+1} \le \varphi_0+\alpha\varphi_k \le \alpha^{k+1}\varphi_0+\frac{\varphi_0}{1-\alpha} \le 2\varphi_0, \tag{2.18}$$

where the second inequality is due to $\mu_0 \le \varphi_0$ and $-\alpha\varphi_k \le \mu_{k+1}$, the next one follows from (2.17), and the last one is due to $\alpha < 1/3$. By taking the limit $k \to \infty$, we obtain

$$\frac{1}{2}\sum_{k=1}^{\infty}\delta_k = \sum_{k=1}^{\infty}\alpha_k\|w^k-w^{k-1}\|_G^2 \le \alpha\sum_{k=1}^{\infty}\|w^k-w^{k-1}\|_G^2 \le \frac{2\varphi_0\alpha}{1-3\alpha} := C_1 < \infty. \tag{2.19}$$

The convergence of $\{w^k\}_{k=0}^{\infty}$ to a solution point in $\mathcal{W}^*$ follows from the proof of Theorem 1.

It follows from (2.9) that, for $i \ge 0$, $\|w^{i+1}-\bar{w}^i\|_G^2 \le \varphi_i - \varphi_{i+1} + \alpha[\nu_i]_+ + \delta_i$, from which we obtain

$$\sum_{i=0}^{k-1}\|w^{i+1}-\bar{w}^i\|_G^2 \le \varphi_0-\varphi_k+\alpha\sum_{i=1}^{k-1}[\nu_i]_+ + \sum_{i=1}^{k-1}\delta_i \le \varphi_0+\alpha C_2+2C_1, \tag{2.20}$$

where C_1 is defined in (2.19) and C_2 is defined as

$$C_2 := \frac{2C_1}{1-\alpha} \ge \frac{1}{1-\alpha}\sum_{i=1}^{\infty}\delta_i \ge \sum_{i=1}^{\infty}[\nu_i]_+.$$

Here the first "$\ge$" follows from the definition of C_1 in (2.19) and the second one follows from (2.8). Direct calculation shows that

$$\varphi_0+\alpha C_2+2C_1 = \left[1+\left(\frac{2\alpha}{1-\alpha}+2\right)\frac{2\alpha}{1-3\alpha}\right]\varphi_0 \le \left(1+\frac{2}{1-3\alpha}\right)\varphi_0, \tag{2.21}$$

where the "$\le$" follows from $\alpha < 1/3$. The estimate (2.13) follows immediately from (2.20) and (2.21). ∎

Remark 4. Note that w^{k+1} is obtained via a proximal point step from $\bar{w}^k$. Thus, the equality $w^{k+1} = \bar{w}^k$ implies that w^{k+1} is already a solution of (2.1) (even if G is only positive semidefinite; see (2.2b)). In this sense, the error estimate given in (2.13) can be viewed as a convergence rate result of the general inertial PPA (2.2). In particular, (2.13) implies that, to obtain an ε-optimal solution in the sense that $\|w^{k+1}-\bar{w}^k\|_G^2 \le \varepsilon$, the upper bound of iterations required by (2.2) is $(1+2/(1-3\alpha))\,\|w^0-w^*\|_G^2/\varepsilon$.

The convergence rate result (2.13) of the inertial PPA is in general weaker than that of the classical PPA, where it can be proved that $\|w^{k+1}-w^k\|^2 = O(1/k)$; see, e.g., [17, 18]. Thus one may ask whether the $O(1/k)$ convergence rate (measured by the residual $\|w^{k+1}-\bar{w}^k\|_G^2$) is still valid for the general inertial PPA (2.2). We answer this question affirmatively in the following theorem.

THEOREM 3. *Assume that Assumptions* 1, 2, *and* 3 *hold. Suppose that the parameters* $\{\alpha_k,\lambda_k\}_{k=0}^{\infty}$ *satisfy, for all* k, $0 \le \alpha_k \equiv \alpha < \frac{1}{3}$ *and* $\lambda_k \equiv \lambda > 0$. *Let* $\{w^k\}_{k=0}^{\infty}$ *be the sequence generated by algorithm* (2.2a)–(2.2b). *Then it holds, for any* $k \ge 2$, *that*

$$\|w^{k+1}-\bar{w}^k\|_G^2 \le \frac{4(1+7\alpha)(1+\alpha^2)}{(1-2\alpha)(1-3\alpha)(k-1)}\|w^0-w^*\|_G^2. \tag{2.22}$$

Proof. We prove this theorem by the following three steps.

Step 1. We show that the sequence $\{\|w^{k+1}-w^k\|_G^2+2\alpha\|w^{k+1}-2w^k+w^{k-1}\|_G^2-\alpha\|w^k-w^{k-1}\|_G^2\}_{k=0}^\infty$ is nonincreasing. In fact, it follows from (2.2b) and $w^k\in\mathcal{W}$ that

$$\theta(w^k)-\theta(w^{k+1})+\langle w^k-w^{k+1},F(w^{k+1})+\lambda^{-1}G(w^{k+1}-\bar{w}^k)\rangle\geq 0. \tag{2.23}$$

Note that (2.2b) is also true for the $(k-1)$th iteration and $w^{k+1}\in\mathcal{W}$. Then, we can deduce that

$$\theta(w^{k+1})-\theta(w^k)+\langle w^{k+1}-w^k,F(w^k)+\lambda^{-1}G(w^k-\bar{w}^{k-1})\rangle\geq 0,$$

which together with (2.23) yields that

$$\begin{aligned}\langle w^k-w^{k+1},G(w^{k+1}-\bar{w}^k)\rangle+\langle w^{k+1}-w^k,G(w^k-\bar{w}^{k-1})\\ \geq\lambda\langle w^k-w^{k+1},F(w^k)-F(w^{k+1})\rangle\geq 0,\end{aligned} \tag{2.24}$$

where the last "$\geq$" follows from Assumption 2. By using the definitions of $\bar{w}^{k-1}$ and $\bar{w}^k$, we know from (2.24) that

$$\begin{aligned}&\|w^k-w^{k+1}\|_G^2\\ &\leq(1+\alpha)\langle w^{k+1}-w^k,G(w^k-w^{k-1})\rangle-\alpha\langle w^{k+1}-w^k,G(w^{k-1}-w^{k-2})\rangle\\ &=\frac{1+\alpha}{2}\big(\|w^k-w^{k+1}\|_G^2+\|w^{k-1}-w^k\|_G^2-\|w^{k+1}-2w^k+w^{k-1}\|_G^2\big)\\ &\quad-\frac{\alpha}{2}\big(\|w^k-w^{k+1}\|_G^2+\|w^{k-2}-w^{k-1}\|_G^2-\|w^{k+1}-w^k-w^{k-1}+w^{k-2}\|_G^2\big)\\ &=\frac{1}{2}\|w^k-w^{k+1}\|_G^2+\frac{1}{2}\|w^{k-1}-w^k\|_G^2+\frac{\alpha}{2}\|w^{k-1}-w^k\|_G^2-\frac{\alpha}{2}\|w^{k-2}-w^{k-1}\|_G^2\\ &\quad-\frac{1+\alpha}{2}\|w^{k+1}-2w^k+w^{k-1}\|_G^2+\frac{\alpha}{2}\|w^{k+1}-w^k-w^{k-1}+w^{k-2}\|_G^2.\end{aligned} \tag{2.25}$$

Note that

$$\|w^{k+1}-w^k-w^{k-1}+w^{k-2}\|_G^2\leq 2\|w^{k+1}-2w^k+w^{k-1}\|_G^2+2\|w^k-2w^{k-1}+w^{k-2}\|_G^2.$$

By substituting the above inequality into (2.25), we have that

$$\begin{aligned}\|w^k-w^{k+1}\|_G^2\leq\|w^{k-1}-w^k\|_G^2+\alpha\|w^{k-1}-w^k\|_G^2-\alpha\|w^{k-2}-w^{k-1}\|_G^2\\ -(1-\alpha)\|w^{k+1}-2w^k+w^{k-1}\|_G^2+2\alpha\|w^k-2w^{k-1}+w^{k-2}\|_G^2,\end{aligned}$$

which together with $1-\alpha>2\alpha$ proves that $\{\|w^{k+1}-w^k\|_G^2+2\alpha\|w^{k+1}-2w^k+w^{k-1}\|_G^2-\alpha\|w^k-w^{k-1}\|_G^2\}_{k=0}^\infty$ is nonincreasing.

Step 2. We give an upper bound of the quantity $\|w^{k+1}-w^k\|_G^2$. It follows from Step 1 that

$$\begin{aligned}&k\big(\|w^{k+1}-w^k\|_G^2+2\alpha\|w^{k+1}-2w^k+w^{k-1}\|_G^2-\alpha\|w^k-w^{k-1}\|_G^2\big)\\ &\leq\sum_{i=1}^k\big(\|w^{i+1}-w^i\|_G^2+2\alpha\|w^{i+1}-2w^i+w^{i-1}\|_G^2-\alpha\|w^i-w^{i-1}\|_G^2\big)\\ &\leq\sum_{i=1}^k[(1+4\alpha)\|w^{i+1}-w^i\|_G^2+3\alpha\|w^i-w^{i-1}\|_G^2]\\ &\leq(1+7\alpha)\sum_{i=1}^{k+1}\|w^i-w^{i-1}\|_G^2\leq\frac{2(1+7\alpha)}{1-3\alpha}\|w^0-w^*\|_G^2,\end{aligned}$$

where the second "$\le$" follows from the elementary inequality $\|a+b\|_G^2 \le 2(\|a\|_G^2 + \|b\|_G^2)$, and the last inequality is due to (2.18). As a result, we have that, for any $k \ge 1$,

$$\|w^{k+1}-w^k\|_G^2+2\alpha\|w^{k+1}-2w^k+w^{k-1}\|_G^2-\alpha\|w^k-w^{k-1}\|_G^2 \le \frac{2(1+7\alpha)}{(1-3\alpha)k}\|w^0-w^*\|_G^2.$$

Moreover, it is easy to see that

$$\|w^k-w^{k-1}\|_G^2 \le 2\|w^{k+1}-w^k\|_G^2+2\|w^{k+1}-2w^k+w^{k-1}\|_G^2$$

and hence

$$\|w^{k+1}-w^k\|_G^2 \le \frac{2(1+7\alpha)}{(1-2\alpha)(1-3\alpha)k}\|w^0-w^*\|_G^2.$$

Step 3. We establish the $O(1/k)$ convergence rate of the inertial PPA measured by the residual $\|w^{k+1}-\bar{w}^k\|_G^2$. In fact, it follows from the definition of $\bar{w}^{k-1}$ that

$$\begin{aligned}&\|w^{k+1}-\bar{w}^k\|_G^2 = \|w^{k+1}-w^k-\alpha(w^k-w^{k-1})\|_G^2\\ &\le 2\|w^{k+1}-w^k\|_G^2+2\alpha^2\|w^k-w^{k-1}\|_G^2 \le \frac{4(1+7\alpha)(1+\alpha^2)}{(1-2\alpha)(1-3\alpha)(k-1)}\|w^0-w^*\|_G^2.\end{aligned}$$

This completes the proof of this theorem. ∎

3. Inertial linearized ALM and ADMM. In this section, we prove that under suitable conditions both the linearized ALM and the linearized ADMM are applications of PPA with weighting matrix $G \in S_{++}^n$. As byproducts, global iterate convergence and ergodic and nonergodic convergence rate results (measured by certain residues) for linearized ALM and ADMM follow directly from existing results for the PPA. Furthermore, inertial versions of the linearized ALM and ADMM are proposed, whose convergence is guaranteed by Theorems 1, 2, and 3. In the following, we first treat the linearized ALM in section 3.1 and then the linearized ADMM in section 3.2.

3.1. Inertial linearized ALM. Let $f : \Re^n \to \Re$ be a given convex function, $A \in \Re^{m\times n}$ and $b \in \Re^m$ be the given matrix and vector, respectively, and $\mathcal{X}$ be a closed convex subset of $\Re^n$. Consider convex optimization of the form

$$\min\{f(x) :\ \text{s.t. } Ax=b, x\in\mathcal{X}\}. \tag{3.1}$$

Assume that the set of KKT points of (3.1) is nonempty. We define $\mathcal{W}$, w, θ, and F, respectively, by $\mathcal{W} := \mathcal{X}\times\Re^m$,

$$w := \begin{pmatrix} x \\ p\end{pmatrix},\quad \theta(w) := f(x),\quad F(w) := \begin{pmatrix} 0 & -A^T \\ A & 0\end{pmatrix}\begin{pmatrix} x \\ p\end{pmatrix} - \begin{pmatrix} 0 \\ b\end{pmatrix}. \tag{3.2}$$

Then, (3.1) can be equivalently represented as the mixed VI problem: find $w^* \in \mathcal{W}$ such that

$$\theta(w)-\theta(w^*)+\langle w-w^*, F(w^*)\rangle \ge 0\ \forall w\in\mathcal{W}. \tag{3.3}$$

Given (x^k, p^k), the linearized ALM iterates as

$$g_k = A^T(Ax^k-b), \tag{3.4a}$$

$$x^{k+1} = \arg\min_{x\in\mathcal{X}} f(x) - \langle p^k, Ax-b\rangle + \frac{\beta}{2\tau}\|x-(x^k-\tau g_k)\|^2, \tag{3.4b}$$

$$p^{k+1} = p^k - \beta(Ax^{k+1}-b). \tag{3.4c}$$

Here $\tau > 0$ is a parameter. It is elementary to verify that the linearized ALM (3.4) is equivalent to finding w^{k+1}, with $w^k = (x^k, p^k)$ given, via

$$\theta(w) - \theta(w^{k+1}) + \langle w - w^{k+1}, F(w^{k+1}) + G(w^{k+1} - w^k)\rangle \geq 0 \ \forall w \in \mathcal{W}, \tag{3.5}$$

where G is defined by

$$G := \begin{pmatrix} \beta(\frac{1}{\tau}I - A^TA) & 0 \\ 0 & \frac{1}{\beta}I \end{pmatrix}. \tag{3.6}$$

Clearly, G defined in (3.6) is positive definite if $0 < \tau < 1/\rho(A^TA)$, and thus the linearized ALM is an application of PPA with a weighting matrix $G \in S^n_{++}$ to the primal-dual system (3.3).

The explanation of linearized ALM as a general PPA given in (3.5) enables us to propose an inertial linearized ALM as follows:

$$(\bar{x}^k, \bar{p}^k) = (x^k, p^k) + \alpha_k(x^k - x^{k-1}, p^k - p^{k-1}), \tag{3.7a}$$

$$g_k = A^T(A\bar{x}^k - b), \tag{3.7b}$$

$$x^{k+1} = \arg\min_{x \in \mathcal{X}} f(x) - \langle \bar{p}^k, Ax - b\rangle + \frac{\beta}{2\tau}\|x - (\bar{x}^k - \tau g_k)\|^2, \tag{3.7c}$$

$$p^{k+1} = \bar{p}^k - \beta(Ax^{k+1} - b). \tag{3.7d}$$

We note that in practice the penalty parameter β in both the linearized ALM and its inertial variant can vary adaptively to accelerate convergence, though in our present framework of analysis it must be fixed in order to freeze the weighting matrix G. The convergence of (3.7) is guaranteed by the following theorem, which, by considering (3.5)–(3.6), is a corollary of Theorems 2 and 3.

THEOREM 4. *Let G be defined in* (3.6) *and $\{(x^k, p^k)\}_{k=0}^\infty \subseteq \Re^n$ be generated by* (3.7) *from any starting point $(x^0, p^0) = (x^{-1}, p^{-1}) \in \mathcal{W}$. Suppose that $0 < \tau < 1/\rho(A^TA)$ and $\{\alpha_k\}_{k=0}^\infty$ satisfies, for all k, $0 \leq \alpha_k \leq \alpha_{k+1} \leq \alpha < \frac{1}{3}$. Then, $\{(x^k, p^k)\}_{k=0}^\infty$ converges to some point in $\mathcal{W}^*$, the set of solutions of* (3.3)*, as $k \to \infty$. Moreover, there exists a constant $C_1 > 0$ such that $\min_{0\leq i\leq k-1} \|(x^{i+1}, p^{i+1}) - (\bar{x}^i, \bar{p}^i)\|_G^2 \leq C_1/k$ for all $k \geq 1$. If we further assume that $\alpha_k \equiv \alpha$ for all k and $0 < \alpha < \frac{1}{3}$, then there exists a constant $C_2 > 0$ such that $\|(x^{k+1}, p^{k+1}) - (\bar{x}^k, \bar{p}^k)\|_G^2 \leq C_2/(k-1)$ for all $k \geq 2$.*

3.2. Inertial linearized ADMM. In this section, we prove results similar to those presented in section 3.1. The results are given in detail with proofs and remarks. Furthermore, an inertial version of the linearized ADMM is proposed, whose convergence is guaranteed by Theorems 1, 2, and 3.

Let $f : \Re^{n_1} \to \Re$ and $g : \Re^{n_2} \to \Re$ be convex functions and $\mathcal{X} \subseteq \Re^{n_1}$ and $\mathcal{Y} \subseteq \Re^{n_2}$ be closed convex sets. Consider a linearly constrained separable convex optimization problem of the form

$$\min_{x,y} \{f(x) + g(y) : \text{ s.t. } Ax + By = b, x \in \mathcal{X}, y \in \mathcal{Y}\}, \tag{3.8}$$

where $A \in \Re^{m\times n_1}$, $B \in \Re^{m\times n_2}$, and $b \in \Re^m$ are given. We assume that the set of KKT points of (3.8) is nonempty. Then (3.8) is equivalent to the mixed VI problem (2.1) with $\mathcal{W}$, w, θ, and F given, respectively, by $\mathcal{W} := \mathcal{X} \times \mathcal{Y} \times \Re^m$,

(3.9)

$$w := \begin{pmatrix} x \\ y \\ p \end{pmatrix}, \quad \theta(w) := f(x) + g(y), \quad F(w) := \begin{pmatrix} 0 & 0 & -A^T \\ 0 & 0 & -B^T \\ A & B & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ p \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ b \end{pmatrix}.$$

Since the coefficient matrix defining F is skew-symmetric, F is monotone, and thus Assumption 2 is satisfied with $H = 0$. Let $\beta > 0$ and define the Lagrangian and the augmented Lagrangian functions, respectively, as

$$\mathcal{L}(x, y, p) := f(x) + g(y) - \langle p, Ax + By - b\rangle, \tag{3.10a}$$

$$\bar{\mathcal{L}}(x, y, p) := \mathcal{L}(x, y, p) + \frac{\beta}{2}\|Ax + By - b\|^2. \tag{3.10b}$$

Given (y^k, p^k), the classical ADMM in "$x - p - y$" order iterates as follows:

$$x^{k+1} = \arg\min_{x\in\mathcal{X}} \bar{\mathcal{L}}(x, y^k, p^k), \tag{3.11a}$$

$$p^{k+1} = p^k - \beta(Ax^{k+1} + By^k - b), \tag{3.11b}$$

$$y^{k+1} = \arg\min_{y\in\mathcal{Y}} \bar{\mathcal{L}}(x^{k+1}, y, p^{k+1}). \tag{3.11c}$$

Note that here we still use the latest value of each variable in each step of the alternating computation. Therefore, it is equivalent to the commonly seen ADMM in "$y - x - p$" order in a cyclic sense. We use the order "$x - p - y$" because the resulting algorithm can be easily explained as a proximal-like algorithm applied to the primal-dual optimality conditions; see [15].

Given (x^k, y^k, p^k) and two parameters $\tau, \eta > 0$, the iteration of linearized ADMM in "$x - p - y$" order appears as

$$u^k = A^T(Ax^k + By^k - b), \tag{3.12a}$$

$$x^{k+1} = \arg\min_{x\in\mathcal{X}} f(x) - \langle p^k, Ax\rangle + \frac{\beta}{2\tau}\|x - (x^k - \tau u^k)\|^2, \tag{3.12b}$$

$$p^{k+1} = p^k - \beta(Ax^{k+1} + By^k - b), \tag{3.12c}$$

$$v^k = B^T(Ax^{k+1} + By^k - b), \tag{3.12d}$$

$$y^{k+1} = \arg\min_{y\in\mathcal{Y}} g(y) - \langle p^{k+1}, By\rangle + \frac{\beta}{2\eta}\|y - (y^k - \eta v^k)\|^2. \tag{3.12e}$$

In the following, we prove that under suitable assumptions $(x^{k+1}, y^{k+1}, p^{k+1})$ generated by (3.12) conforms to the classical PPA with an appropriate symmetric and positive definite weighting matrix G.

THEOREM 5. *Given $w^k = (x^k, y^k, p^k) \in \mathcal{W}$, then $w^{k+1} = (x^{k+1}, y^{k+1}, p^{k+1})$ generated by the linearized ADMM framework* (3.12) *satisfies*

$$w^{k+1} \in \mathcal{W},\ \theta(w) - \theta(w^{k+1}) + \langle w - w^{k+1}, F(w^{k+1}) + G(w^{k+1} - w^k)\rangle \geq 0\ \forall w \in \mathcal{W}, \tag{3.13}$$

where

$$G = \begin{pmatrix} \beta\left(\frac{1}{\tau}I - A^TA\right) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{\beta}{\eta}I & -B^T \\ \mathbf{0} & -B & \frac{1}{\beta}I \end{pmatrix}. \tag{3.14}$$

Here I denotes identity matrix of appropriate size.

Proof. The optimality conditions of (3.12b) and (3.12e) imply that

$$f(x) - f(x^{k+1}) + (x - x^{k+1})^T \left\{ -A^T p^k + \frac{\beta}{\tau}(x^{k+1} - x^k) + \beta A^T(Ax^k + By^k - b) \right\} \geq 0 \ \forall x \in \mathcal{X},$$

$$g(y) - g(y^{k+1}) + (y - y^{k+1})^T \left\{ -B^T p^{k+1} + \frac{\beta}{\eta}(y^{k+1} - y^k) + \beta B^T(Ax^{k+1} + By^k - b) \right\} \geq 0 \ \forall y \in \mathcal{Y}.$$

By noting (3.12c), the above relations can be rewritten as

$$f(x) - f(x^{k+1}) + (x - x^{k+1})^T \left\{ -A^T p^{k+1} + \beta \left(\frac{1}{\tau} I - A^T A \right) (x^{k+1} - x^k) \right\} \geq 0 \ \forall x \in \mathcal{X}, \tag{3.15a}$$

$$g(y) - g(y^{k+1}) + (y - y^{k+1})^T \left\{ -B^T p^{k+1} + \frac{\beta}{\eta}(y^{k+1} - y^k) - B^T(p^{k+1} - p^k) \right\} \geq 0 \ \forall y \in \mathcal{Y}. \tag{3.15b}$$

Note that (3.12c) can be equivalently represented as

$$(p - p^{k+1})^T \left\{ (Ax^{k+1} + By^{k+1} - b) - B(y^{k+1} - y^k) + \frac{1}{\beta}(p^{k+1} - p^k) \right\} \geq 0 \ \forall p \in \Re^m. \tag{3.16}$$

By the notation defined in (3.9), we see that the addition of (3.15a), (3.15b), and (3.16) yields (3.13), with G defined in (3.14). ∎

Remark 5. Clearly, the matrix G defined in (3.14) is symmetric and positive definite provided that the parameters τ and η are reasonably small. In particular, G is positive definite if $\tau < 1/\rho(A^T A)$ and $\eta < 1/\rho(B^T B)$. Using similar analysis, it is easy to verify that $w^{k+1} = (x^{k+1}, y^{k+1}, p^{k+1})$ generated by the ADMM framework (3.11) conforms to (3.13) with G defined by

$$G = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \beta B^T B & -B^T \\ \mathbf{0} & -B & \frac{1}{\beta} I \end{pmatrix}, \tag{3.17}$$

which is clearly never positive definite. See [15] for details.

Convergence results for the linearized ADMM (3.12) are summarized below in Theorem 6. We note that the results in Theorem 6 are not new. Specifically, the nonergodic convergence rate result (part (b) of the theorem) was first established in [25] for ADMM and Douglas–Rachford operator splitting, while the global iterate convergence (part (a) of the theorem) of PPA goes back to [41] for the monotone operator inclusion problem. See also [26] for an ergodic $O(1/k)$ convergence rate result for the ADMM, which is extendable to the linearized ADMM (3.12) too. Since this ergodic convergence result is less relevant to our results for the inertial PPA, we omit it here. It is also worth noting that similar convergence results and remarks are also valid for the linearized ALM (3.4).

THEOREM 6. *Assume that* $0 < \tau < 1/\rho(A^T A)$ *and* $0 < \eta < 1/\rho(B^T B)$. *Let* $\{w^k = (x^k, y^k, p^k)\}_{k=0}^{\infty}$ *be generated by the linearized ADMM framework* (3.12) *from any starting point* $w^0 = (x^0, y^0, p^0) \in \mathcal{W}$. *The following results hold:*

(a) *The sequence* $\{w^k = (x^k, y^k, p^k)\}_{k=0}^\infty$ *converges to a solution of* (2.1)*, i.e., there exists* $w^\star = (x^\star, y^\star, p^\star) \in \mathcal{W}^*$ *such that* $\lim_{k\to\infty} w^k = w^\star$*. Moreover,* $(x^\star, y^\star)$ *is a solution of* (3.8).

(b) *After* $k > 0$ *iterations, we have*

$$\|w^k - w^{k-1}\|_G^2 \le \frac{\|w^0 - w^*\|_G^2}{k}. \tag{3.18}$$

Remark 6. It is easy to see from (3.13) that w^{k+1} must be a solution if $w^{k+1} = w^k$. As such, the difference of two consecutive iterations can be viewed in some sense as a measure of how close the current point is to the solution set. Therefore, the result (3.18) estimates the convergence rate of w^k to the solution set using the measure $\|w^k - w^{k-1}\|_G^2$.

Remark 7. We note that all the results given in Theorem 6 remain valid if the conditions on τ and η are relaxed to $0 < \tau \le 1/\rho(A^TA)$ and $0 < \eta \le 1/\rho(B^TB)$, respectively. The proof is a little bit complicated and we refer interested readers to [21, 16].

Now we state the inertial version of the linearized ADMM, which is new to the best of our knowledge. Given $\beta, \tau, \eta > 0$, a sequence $\{\alpha_k \ge 0\}_{k=0}^\infty$, (x^k, y^k, p^k), and $(x^{k-1}, y^{k-1}, p^{k-1})$, the inertial linearized ADMM iterates as follows:

$$(\bar{x}^k, \bar{y}^k, \bar{p}^k) = (x^k, y^k, p^k) + \alpha_k(x^k - x^{k-1}, y^k - y^{k-1}, p^k - y^{k-1}), \tag{3.19a}$$

$$u^k = A^T(A\bar{x}^k + B\bar{y}^k - b), \tag{3.19b}$$

$$x^{k+1} = \arg\min_{x\in\mathcal{X}} f(x) - \langle \bar{p}^k, Ax\rangle + \frac{\beta}{2\tau}\|x - (\bar{x}^k - \tau u^k)\|^2, \tag{3.19c}$$

$$p^{k+1} = \bar{p}^k - \beta(Ax^{k+1} + B\bar{y}^k - b), \tag{3.19d}$$

$$v^k = B^T(Ax^{k+1} + B\bar{y}^k - b), \tag{3.19e}$$

$$y^{k+1} = \arg\min_{y\in\mathcal{Y}} g(y) - \langle p^{k+1}, By\rangle + \frac{\beta}{2\eta}\|y - (\bar{y}^k - \eta v^k)\|^2. \tag{3.19f}$$

The following convergence result is a consequence of Theorems 2, 3, and 5.

THEOREM 7. *Let* G *be defined in* (3.14) *and* $\{(x^k, y^k, p^k)\}_{k=0}^\infty \subseteq \Re^n$ *be generated by* (3.19) *from any starting point* $(x^0, y^0, p^0) = (x^{-1}, y^{-1}, p^{-1}) \in \mathcal{W}$*. Suppose that* $0 < \tau < 1/\rho(A^TA)$, $0 < \eta < 1/\rho(B^TB)$*, and* $\{\alpha_k\}_{k=0}^\infty$ *satisfies, for all* k, $0 \le \alpha_k \le \alpha_{k+1} \le \alpha < \frac{1}{3}$*. Then,* $\{(x^k, y^k, p^k)\}_{k=0}^\infty$ *converges to some point in* $\mathcal{W}^*$*, the set of solutions of* (2.1)*, as* $k \to \infty$*. Moreover, there exists a constant* $C_1 > 0$ *such that, for all* $k \ge 1$, $\min_{0\le i\le k-1} \|(x^{i+1}, y^{i+1}, p^{i+1}) - (\bar{x}^i, \bar{y}^i, \bar{p}^i)\|_G^2 \le C_1/k$*. If we further assume that* $\alpha_k \equiv \alpha$ *for all* k *and* $0 < \alpha < \frac{1}{3}$*, then there exists a constant* $C_2 > 0$ *such that for all* $k \ge 2$*,*

$$\|(x^{k+1}, y^{k+1}, p^{k+1}) - (\bar{x}^k, \bar{y}^k, \bar{p}^k)\|_G^2 \le C_2/(k-1).$$

4. Numerical results.

In this section, we present numerical results to compare the performance of the linearized ADMM (LADMM) (3.12) and the proposed inertial linearized ADMM (iLADMM) (3.19). Both algorithms were implemented in MATLAB. All the experiments were performed with Microsoft Windows 8 and MATLAB v7.13 (R2011b), running on a 64-bit Lenovo laptop with an Intel Core i7-3667U CPU at 2.00 GHz and 8 GB of memory.

4.1. Compressive principal component pursuit.

In our experiments, we focused on the compressive principal component pursuit problem [43], which aims to recover low-rank and sparse components from compressive or incomplete measurements. Let $\mathcal{A} : \Re^{m\times n} \to \Re^q$ be a linear operator and L_0 and S_0 be, respectively, low-rank and sparse matrices of size $m \times n$. The incomplete measurements are given by $b = \mathcal{A}(L_0 + S_0)$. Under certain technical conditions, such as L_0 is μ-incoherent, the support of S_0 is randomly distributed with nonzero probability ρ, and the signs of S_0 conform to Bernoulli distribution, it was proved in [43] that the low-rank and the sparse components L_0 and S_0 can be exactly recovered with high probability via solving the convex optimization problem

$$\min_{L,S} \{\|L\|_* + \lambda\|S\|_1 : \text{ s.t. } \mathcal{A}(L+S) = b\}, \tag{4.1}$$

as long as the range space of the adjoint operator $\mathcal{A}^*$ is randomly distributed according to the Haar measure and its dimension q is in the order $O\big((\rho mn + mr)\log^2 m\big)$. Here $\lambda = 1/\sqrt{m}$ is a constant, and $\|L\|_*$ and $\|S\|_1$ denote the nuclear norm of L (sum of all singular values) and the ℓ_1 norm of S (sum of absolute values of all components), respectively. Note that to determine a rank r matrix, it is sufficient to specify $(m + n - r)r$ elements. Let the number of nonzeros of S_0 be denoted by $\text{nnz}(S_0)$. Without considering the distribution of the support of S_0, we define the *degree of freedom* of the pair (L_0, S_0) by

$$\text{dof} := (m + n - r)r + \text{nnz}(S_0). \tag{4.2}$$

The augmented Lagrangian function of (4.1) is given by

$$\bar{\mathcal{L}}(L, S, p) := \|L\|_* + \lambda\|S\|_1 - \langle p, \mathcal{A}(L+S) - b\rangle + \frac{\beta}{2}\|\mathcal{A}(L+S) - b\|^2.$$

One can see that the minimization of $\bar{\mathcal{L}}$ with respect to either L or S, with the other two variables being fixed, does not have a closed form solution. To avoid inner loops for iteratively solving ADMM-subproblems, the linearized ADMM framework (3.12) and its inertial version (3.19) can obviously be applied. Note that it is necessary to linearize both ADMM-subproblems in order to avoid inner loops. Though the iterative formulas of LADMM and inertial LADMM for solving (4.1) can be derived very easily based on (3.12) and (3.19), we elaborate them below for clearness and subsequent references. Let (L^k, S^k, p^k) be given. The LADMM framework (3.12) for solving (4.1) appears as

$$U^k = \mathcal{A}^*(\mathcal{A}(L^k + S^k) - b), \tag{4.3a}$$
$$L^{k+1} = \arg\min_L \|L\|_* - \langle p^k, \mathcal{A}(L)\rangle + \frac{\beta}{2\tau}\|L - (L^k - \tau U^k)\|_F^2, \tag{4.3b}$$
$$p^{k+1} = p^k - \beta(\mathcal{A}(L^{k+1} + S^k) - b), \tag{4.3c}$$
$$V^k = \mathcal{A}^*(\mathcal{A}(L^{k+1} + S^k) - b), \tag{4.3d}$$
$$S^{k+1} = \arg\min_S \lambda\|S\|_1 - \langle p^{k+1}, \mathcal{A}(S)\rangle + \frac{\beta}{2\eta}\|S - (S^k - \eta V^k)\|_F^2. \tag{4.3e}$$

The inertial LADMM framework (3.19) for solving (4.1) appears as

$$(\bar{L}^k, \bar{S}^k, \bar{p}^k) = (L^k, S^k, p^k) + \alpha_k(L^k - L^{k-1}, S^k - S^{k-1}, p^k - p^{k-1}), \tag{4.4a}$$
$$U^k = \mathcal{A}^*(\mathcal{A}(\bar{L}^k + \bar{S}^k) - b), \tag{4.4b}$$
$$L^{k+1} = \arg\min_L \|L\|_* - \langle \bar{p}^k, \mathcal{A}(L)\rangle + \frac{\beta}{2\tau}\|L - (\bar{L}^k - \tau U^k)\|_F^2, \tag{4.4c}$$
$$p^{k+1} = \bar{p}^k - \beta(\mathcal{A}(L^{k+1} + \bar{S}^k) - b), \tag{4.4d}$$
$$V^k = \mathcal{A}^*(\mathcal{A}(L^{k+1} + \bar{S}^k) - b), \tag{4.4e}$$
$$S^{k+1} = \arg\min_S \lambda\|S\|_1 - \langle p^{k+1}, \mathcal{A}(S)\rangle + \frac{\beta}{2\eta}\|S - (\bar{S}^k - \eta V^k)\|_F^2. \tag{4.4f}$$

Note that the subproblems (4.3b) (or (4.4c)) and (4.3e) (or (4.4f)) have closed form solutions given, respectively, by the shrinkage operators of matrix nuclear norm and vector ℓ_1 norm; see, e.g., [29, 45]. The main computational cost per iteration of both algorithms is one singular value decomposition (SVD) required in solving the L-subproblem.

4.2. Generating experimental data. In our experiments, we set $m = n$ and tested different ranks of L_0 (denoted by r), sparsity levels of S_0 (i.e., $\mathrm{nnz}(S_0)/(mn)$), and sample ratios (i.e., $q/(mn)$). The low-rank matrix L_0 was generated by $\mathrm{randn}(m, r) * \mathrm{randn}(r, n)$ in MATLAB. The support of S_0 is randomly determined by uniform distribution, while the values of its nonzeros are uniformly distributed in $[-10, 10]$. Such types of synthetic data are roughly those tested in [43]. As for the linear operator $\mathcal{A}$, we tested three types of linear operators, i.e., two-dimensional partial discrete cosine transform (DCT), fast Fourier transform (FFT), and Walsh–Hadamard transform (WHT). The rows of these transforms are selected uniformly at random.

4.3. Parameters, stopping criterion, and initialization. The model parameter λ was set to $1/\sqrt{m}$ in our experiments, which is determined based on the exact recoverability theory in [43]. As for the other parameters (β, τ and η) common to LADMM and iLADMM, we used the same set of values and adaptive rules in all the tests. Now we elaborate how the parameters are chosen. Since $\mathcal{A}$ contains rows of orthonormal transforms, it holds that $\mathcal{A}\mathcal{A}^* = \mathcal{I}$, the identity operator. Therefore, it holds that $\rho(\mathcal{A}^*\mathcal{A}) = 1$. We set $\tau = \eta = 0.99$, which satisfies the convergence requirement specified in Theorems 6 and 7. The penalty parameter β was initialized at $0.1q/\|b\|_1$ and was tuned at the beginning stage of the algorithm. Specifically, we tuned β within the first 30 iterations according to the following rule:

$$\beta_{k+1} = \begin{cases} \max(0.5\beta_k, 10^{-3}) & \text{if } r_k < 0.1; \\ \min(2\beta_k, 10^2) & \text{if } r_k > 5; \\ \beta_k & \text{otherwise}, \end{cases} \qquad \text{where} \quad r_k := \frac{\beta_k \|\mathcal{A}(L^k + S^k) - b\|^2}{2s_k(\|L^k\|_* + \lambda\|S^k\|_1)}.$$

Here s_k is a parameter attached to the objective function $\|L\|_* + \lambda\|S\|_1$ and was chosen adaptively so that the quadratic term $\frac{\beta}{2}\|\mathcal{A}(L + S) - b\|^2$ and the objective term $\|L\|_* + \lambda\|S\|_1$ remain roughly in the same order. Note that the choice of β does not have much theory and is usually determined via numerical experiments; see, e.g., [44] for the influence of different β's in linearized ADMM for matrix completion problem. The extrapolation parameter α_k for iLADMM was set to be 0.28 and held constant in all our experiments. Note that this value of α_k is determined based on experiments and may be far from optimal. How to select α_k adaptively to achieve stable and faster convergence remains a research issue. Here our main goal is to illustrate the effect of the extrapolation steps. We also present some numerical results to compare the performance of iLADMM with different constant strategies for α_k.

It is easy to see from (3.13) that if two consecutive iterates generated by PPA are identical, then a solution is already obtained. Since LADMM is an application of a general PPA, we terminated it by the following rule:

$$\frac{\|(L^{k+1}, S^{k+1}, p^{k+1}) - (L^k, S^k, p^k)\|}{1 + \|(L^k, S^k, p^k)\|} < \varepsilon, \tag{4.5}$$

where $\varepsilon > 0$ is a tolerance parameter. Here $\|(L, S, p)\| := \sqrt{\|L\|_F^2 + \|S\|_F^2 + \|p\|^2}$. Since iLADMM generates the new point $(L^{k+1}, S^{k+1}, p^{k+1})$ by applying PPA to

$(\bar{L}^k, \bar{S}^k, \bar{p}^k)$, we used the same stopping rule as (4.5) except that (L^k, S^k, p^k) is replaced by $(\bar{L}^k, \bar{S}^k, \bar{p}^k)$. That is,

$$\frac{\|(L^{k+1}, S^{k+1}, p^{k+1}) - (\bar{L}^k, \bar{S}^k, \bar{p}^k)\|}{1 + \|(\bar{L}^k, \bar{S}^k, \bar{p}^k)\|} < \varepsilon. \tag{4.6}$$

In our experiments, we initialized all variables L, S, and p at zeros.

4.4. Experimental results. Recall that the matrix size is $m \times n$, the number of measurements is q, the rank of L_0 is r, and the degree of freedom of the pair (L_0, S_0) is defined in (4.2). In our experiments, we tested $m = n = 1024$. Let k be the number of nonzeros of S_0. We tested four different ranks for L_0, three levels of sparsity for S_0, and four levels of sample ratios. Specifically, in our experiments we tested $r \in \{5, 10, 15, 20\}$, $k/m^2 \in \{0.01, 0.05, 0.10\}$, and $q/m^2 \in \{0.4, 0.6, 0.8\}$.

Let (L, S) be the recovered solution. For each setting, we report the relative errors of L and S to the true low-rank and sparse matrices L_0 and S_0, i.e., $\|L - L_0\|_F/\|L_0\|_F$ and $\|S-S_0\|_F/\|S_0\|_F$, and the number of iterations to meet the condition (4.5) or (4.6), which are denoted by iter1 and iter2 for LADMM and iLADMM, respectively. We terminated both algorithms if the number of iterations reached 1000 but the stopping rule (4.5) or (4.6) still did not hold. For each problem scenario, we run 10 random trials for both algorithms and report the averaged results. Detailed experimental results for $\varepsilon = 10^{-5}$ and $r = 5, 10, 15$, and 20 are given in Tables 1–4, respectively. In each table, a dash represents that the maximum iteration number was reached.

It can be seen from Tables 1–4 that iLADMM is generally faster than LADMM to obtain solutions satisfying the aforementioned conditions. Specifically, within our setting the numbers of iterations consumed by iLADMM range, roughly, from 60% to 80% of those consumed by LADMM. If we take into account all the tests (except those cases where either LADMM or iLADMM failed to terminate within 1000 iterations, e.g., $(r, k/m^2, q/m^2) = (5, 0.1, 40\%)$, and $\mathcal{A}$ is partial DCT), the overall average number of iterations used by iLADMM is about 74% of that used by LADMM. Note that in some cases iLADMM obtained satisfactory results within the number of allowed iterations (1000 in our setting), while LADMM did not. For example, $(r, k/m^2, q/m^2) = (5, 0.1, 40\%)$ and $\mathcal{A}$ is partial DCT or partial WHT. In most cases, the recovered matrices L and S are close to the true low-rank and sparse components L_0 and S_0, respectively. The relative errors are usually on the order 10^{-5}–10^{-6}. For some cases, the recovered solutions are not of high quality (relative errors are large), which is mainly because the number of samples is small relative to the degree of freedom of (L_0, S_0). This can be seen from the values of q/dof listed in the tables. Roughly speaking, the recovered solutions are satisfactory (say, relative errors are less than 10^{-3}) provided that q/dof is no less than 3.5.

We note that the per-iteration cost of both LADMM and iLADMM for the compressive principal pursuit model (4.1) is dominated by one SVD and thus is roughly identical. The extra cost of the extrapolation inertial step in (4.4a) is negligible compared to the computational load of SVD. This is the main reason that we only reported the number of iterations but not CPU time consumed by both algorithms. The inertial technique actually accelerates the original algorithm to a large extent but without increasing the total computational cost.

To better understand the behavior of iLADMM relative to LADMM, we also tested different matrix sizes ($m = n = 256, 512$, and 1024) with different levels of stopping tolerance ($\varepsilon = 10^{-3}, 10^{-4}$, and 10^{-5} in (4.5)). For each case, we tested

TABLE 1
Results of rank$(L_0) = 5$*:* $\varepsilon = 10^{-5}$*, average results of* 10 *random trials.*

$m = n = 1024$				LADMM			iLADMM			
r	k/m^2	$(q/m^2, q/\text{dof})$	$\mathcal{A}$	$\frac{\Vert L-L_0\Vert_F}{\Vert L_0\Vert_F}$	$\frac{\Vert S-S_0\Vert_F}{\Vert S_0\Vert_F}$	iter1	$\frac{\Vert L-L_0\Vert_F}{\Vert L_0\Vert_F}$	$\frac{\Vert S-S_0\Vert_F}{\Vert S_0\Vert_F}$	iter2	$\frac{\text{iter2}}{\text{iter1}}$
5	1%	(40%, 20.26)	pdct	2.24e-5	4.45e-5	260.6	1.46e-5	3.97e-5	199.4	0.77
			pfft	1.68e-5	3.78e-5	223.8	1.35e-5	4.00e-5	173.3	0.77
			pwht	2.92e-5	5.24e-5	271.8	1.73e-5	3.99e-5	206.1	0.76
		(60%, 30.39)	pdct	1.11e-5	2.24e-5	170.3	1.24e-5	2.65e-5	123.1	0.72
			pfft	1.14e-5	2.24e-5	134.6	1.13e-5	2.67e-5	98.4	0.73
			pwht	1.23e-5	2.32e-5	160.9	1.12e-5	2.37e-5	119.3	0.74
		(80%, 40.52)	pdct	1.20e-5	1.28e-5	89.7	8.51e-6	1.52e-5	61.5	0.69
			pfft	3.68e-6	1.00e-5	67.2	7.06e-6	1.44e-5	42.2	0.63
			pwht	8.10e-6	1.09e-5	88.3	6.80e-6	1.22e-5	60.4	0.68
	5%	(40%, 6.70)	pdct	1.03e-5	4.07e-5	409.7	1.02e-5	3.72e-5	321.5	0.78
			pfft	1.32e-5	4.82e-5	349.5	1.10e-5	3.80e-5	278.4	0.80
			pwht	1.04e-5	4.14e-5	408.8	1.04e-5	3.74e-5	320.4	0.78
		(60%, 10.04)	pdct	8.54e-6	2.68e-5	198.5	5.35e-6	1.99e-5	149.5	0.75
			pfft	7.92e-6	3.13e-5	188.4	5.03e-6	1.84e-5	136.9	0.73
			pwht	8.12e-6	2.75e-5	204.0	5.19e-6	1.86e-5	150.1	0.74
		(80%, 13.39)	pdct	9.39e-6	1.93e-5	107.9	4.32e-6	1.39e-5	77.9	0.72
			pfft	5.94e-6	1.78e-5	99.7	3.54e-6	1.40e-5	70.1	0.70
			pwht	5.62e-6	1.50e-5	107.6	4.10e-6	1.43e-5	77.4	0.72
	10%	(40%, 3.64)	pdct	2.58e-2	7.19e-2	—	1.20e-5	4.08e-5	879.0	—
			pfft	1.53e-5	5.21e-5	757.1	1.21e-5	4.02e-5	563.3	0.74
			pwht	2.50e-2	6.93e-2	—	1.21e-5	4.02e-5	872.7	—
		(60%, 5.47)	pdct	8.58e-6	3.03e-5	344.9	7.64e-6	2.50e-5	243.9	0.71
			pfft	9.19e-6	3.23e-5	309.5	7.27e-6	2.42e-5	239.8	0.77
			pwht	8.49e-6	3.08e-5	345.5	7.77e-6	2.50e-5	243.8	0.71
		(80%, 7.29)	pdct	8.02e-6	2.13e-5	170.7	6.45e-6	1.71e-5	123.8	0.73
			pfft	7.82e-6	1.79e-5	160.7	5.98e-6	1.39e-5	116.0	0.72
			pwht	8.12e-6	2.00e-5	171.0	6.54e-6	1.83e-5	124.0	0.73

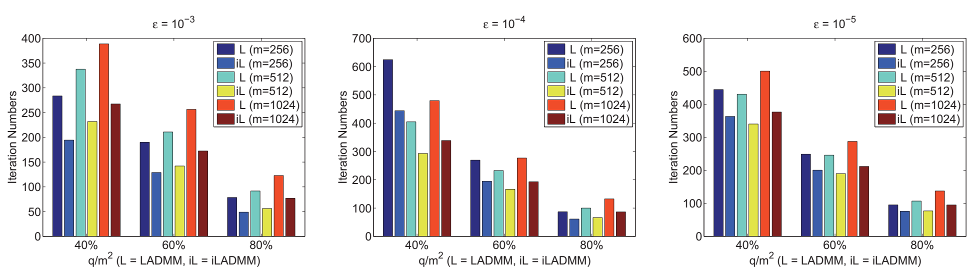

FIG. 1. *Comparison results on different matrix sizes and stopping tolerance: Average results of* 10 *random trials* ($m = n = 256, 512, 1024$*, and from left to right* $\varepsilon = 10^{-3}, 10^{-4}, 10^{-5}$*, respectively).*

$r \in \{5, 10, 15, 20\}$ and $k/m^2 \in \{0.01, 0.05, 0.10\}$ for a fixed q such that $q/m^2 \in \{0.4, 0.6, 0.8\}$. For each q, we accumulated the iteration numbers for different (r, k) and the three types of linear operators and took an average finally. The results are summarized in Figure 1. Again, these results are the average of 10 random trials for each case. From the results we can see that iLADMM is faster and terminates earlier than LADMM with different levels of stopping tolerance. Roughly speaking, iLADMM reduced the cost of LADMM by about 30%.

We also run iLADMM with various constant strategies for α_k. In particular, we set $m = n = 512$ and tested different values of q such that $q/\text{dof} \in \{5, 10, 15\}$. For each

TABLE 2
Results of rank$(L_0) = 10$*:* $\varepsilon = 10^{-5}$*, average results of* 10 *random trials.*

$m = n = 1024$				LADMM			iLADMM			
r	k/m^2	$(q/m^2, q/\text{dof})$	$\mathcal{A}$	$\frac{\|L-L_0\|_F}{\|L_0\|_F}$	$\frac{\|S-S_0\|_F}{\|S_0\|_F}$	iter1	$\frac{\|L-L_0\|_F}{\|L_0\|_F}$	$\frac{\|S-S_0\|_F}{\|S_0\|_F}$	iter2	$\frac{\text{iter2}}{\text{iter1}}$
10	1%	(40%, 13.59)	pdct	2.20e-5	5.42e-5	340.5	1.74e-5	5.33e-5	253.0	0.74
			pfft	1.87e-5	5.09e-5	300.1	1.65e-5	5.46e-5	226.8	0.76
			pwht	2.22e-5	5.50e-5	350.5	1.91e-5	5.54e-5	257.4	0.73
		(60%, 20.38)	pdct	2.12e-5	3.85e-5	198.1	1.20e-5	3.16e-5	141.6	0.71
			pfft	1.03e-5	2.75e-5	163.1	9.56e-6	3.07e-5	117.0	0.72
			pwht	1.84e-5	3.61e-5	191.0	1.32e-5	3.38e-5	135.9	0.71
		(80%, 27.18)	pdct	7.82e-6	1.55e-5	101.1	7.16e-6	1.60e-5	66.6	0.66
			pfft	8.63e-6	1.67e-5	79.3	6.36e-6	1.70e-5	47.0	0.59
			pwht	7.43e-6	1.40e-5	97.8	6.74e-6	1.56e-5	64.5	0.66
	5%	(40%, 5.76)	pdct	9.22e-6	3.91e-5	536.2	1.41e-5	4.95e-5	406.5	0.76
			pfft	1.30e-5	5.29e-5	448.9	1.15e-5	4.31e-5	346.6	0.77
			pwht	9.16e-6	3.93e-5	537.4	1.42e-5	5.08e-5	407.7	0.76
		(60%, 8.64)	pdct	7.86e-6	3.18e-5	241.9	9.93e-6	3.28e-5	183.4	0.76
			pfft	8.45e-6	3.55e-5	223.3	9.87e-6	3.24e-5	170.1	0.76
			pwht	1.42e-5	3.71e-5	244.6	9.32e-6	3.21e-5	184.4	0.75
		(80%, 11.52)	pdct	8.78e-6	2.17e-5	120.4	4.36e-6	1.78e-5	84.5	0.70
			pfft	6.74e-6	2.82e-5	109.8	4.22e-6	1.76e-5	76.3	0.69
			pwht	8.55e-6	2.01e-5	121.2	4.61e-6	1.68e-5	85.1	0.70
	10%	(40%, 3.35)	pdct	6.84e-2	2.24e-1	—	2.54e-2	8.67e-2	—	—
			pfft	1.09e-5	4.23e-5	965.2	1.26e-5	4.39e-5	702.5	0.73
			pwht	6.79e-2	2.21e-1	—	2.48e-2	8.38e-2	—	—
		(60%, 5.02)	pdct	8.02e-6	3.27e-5	398.4	7.75e-6	2.76e-5	299.1	0.75
			pfft	9.53e-6	3.58e-5	352.3	7.75e-6	2.75e-5	266.9	0.76
			pwht	8.13e-6	3.27e-5	397.6	7.70e-6	2.75e-5	298.4	0.75
		(80%, 6.70)	pdct	8.49e-6	2.57e-5	185.9	7.08e-6	2.08e-5	132.6	0.71
			pfft	8.62e-6	2.59e-5	172.8	6.85e-6	1.84e-5	121.2	0.70
			pwht	8.49e-6	2.43e-5	185.9	7.06e-6	2.21e-5	132.4	0.71

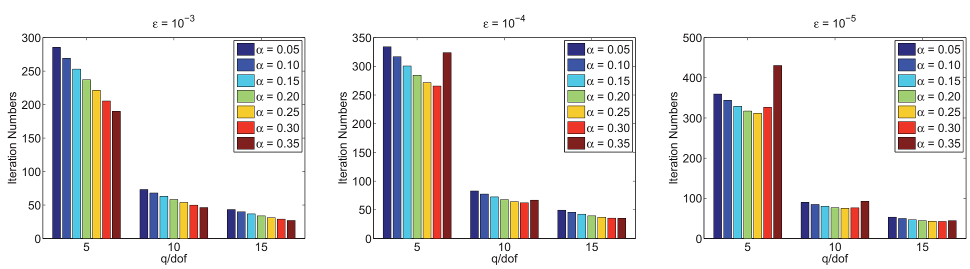

FIG. 2. *Comparison results on different* $\alpha_k \equiv \alpha$ *and stopping tolerance: Average results of* 10 *random trials (*$m = n = 512$*,* α *ranges from* 0.05 *to* 0.35*, and from left to right* $\varepsilon = 10^{-3}, 10^{-4}, 10^{-5}$*, respectively).*

case, we varied $r \in \{5, 10, 15, 20\}$ and $k/m^2 \in \{0.01, 0.05, 0.10\}$ for the three types of aforementioned measurement matrices. We accumulated the number of iterations and took an average finally. The detailed average results of 10 random trials for $\alpha_k \equiv \alpha$ from 0.05 to 0.35 are given in Figure 2.

From the results in Figure 2 we see that, for the tested seven values of α, iLADMM is slightly faster if α is larger, provided that α does not exceed 0.3. We have also observed that for $\alpha > 0.3$ iLADMM either slows down or performs not very stably, especially when q/dof is small. This is the main reason that we set α_k a constant value that is near 0.3 but not larger.

Table 3
Results of rank$(L_0) = 15$*:* $\varepsilon = 10^{-5}$*, average results of* 10 *random trials.*

$m = n = 1024$				LADMM			iLADMM			
r	k/m^2	$(q/m^2, q/\text{dof})$	$\mathcal{A}$	$\frac{\Vert L-L_0\Vert_F}{\Vert L_0\Vert_F}$	$\frac{\Vert S-S_0\Vert_F}{\Vert S_0\Vert_F}$	iter1	$\frac{\Vert L-L_0\Vert_F}{\Vert L_0\Vert_F}$	$\frac{\Vert S-S_0\Vert_F}{\Vert S_0\Vert_F}$	iter2	$\frac{\text{iter2}}{\text{iter1}}$
15	1%	(40%, 10.23)	pdct	1.87e-5	5.29e-5	418.6	1.78e-5	5.87e-5	310.5	0.74
			pfft	1.78e-5	5.23e-5	367.5	1.57e-5	5.68e-5	276.1	0.75
			pwht	1.59e-5	4.86e-5	415.5	1.59e-5	5.49e-5	311.1	0.75
		(60%, 15.35)	pdct	2.01e-5	4.31e-5	223.7	1.26e-5	3.59e-5	158.2	0.71
			pfft	1.53e-5	3.73e-5	193.9	1.16e-5	3.77e-5	137.0	0.71
			pwht	2.20e-5	4.55e-5	225.1	1.20e-5	3.47e-5	159.0	0.71
		(80%, 20.47)	pdct	8.30e-6	1.72e-5	112.7	6.86e-6	1.67e-5	72.7	0.64
			pfft	1.06e-5	1.98e-5	90.5	7.07e-6	2.08e-5	51.2	0.57
			pwht	1.17e-5	2.28e-5	117.4	8.37e-6	1.95e-5	76.5	0.65
	5%	(40%, 5.06)	pdct	2.08e-5	7.79e-5	674.1	1.33e-5	5.12e-5	506.5	0.75
			pfft	1.28e-5	5.46e-5	548.2	1.60e-5	5.84e-5	414.0	0.76
			pwht	2.07e-5	7.40e-5	673.6	1.34e-5	5.20e-5	505.0	0.75
		(60%, 7.59)	pdct	7.46e-6	3.45e-5	282.4	1.02e-5	3.79e-5	209.0	0.74
			pfft	9.17e-6	3.94e-5	256.7	1.12e-5	3.91e-5	190.8	0.74
			pwht	6.90e-6	3.41e-5	282.6	1.02e-5	3.80e-5	210.3	0.74
		(80%, 10.12)	pdct	6.76e-6	2.29e-5	134.1	5.00e-6	2.02e-5	93.3	0.70
			pfft	7.24e-6	2.66e-5	121.3	4.96e-6	2.01e-5	82.5	0.68
			pwht	7.42e-6	2.06e-5	136.5	5.58e-6	2.02e-5	93.3	0.68
	10%	(40%, 3.10)	pdct	1.01e-1	3.50e-1	—	6.15e-2	2.25e-1	—	—
			pfft	2.61e-2	8.19e-2	—	1.40e-5	4.98e-5	888.4	—
			pwht	1.02e-1	3.51e-1	—	6.22e-2	2.27e-1	—	—
		(60%, 4.65)	pdct	7.98e-6	3.37e-5	455.3	7.99e-6	2.93e-5	336.5	0.74
			pfft	9.79e-6	3.87e-5	394.3	8.41e-6	3.07e-5	294.9	0.75
			pwht	7.78e-6	3.35e-5	454.9	7.90e-6	2.86e-5	336.6	0.74
		(80%, 6.20)	pdct	8.96e-6	2.81e-5	201.9	7.75e-6	2.16e-5	143.2	0.71
			pfft	9.38e-6	3.12e-5	186.0	7.63e-6	2.53e-5	129.0	0.69
			pwht	8.93e-6	2.75e-5	202.6	7.86e-6	2.48e-5	142.5	0.70

5. Concluding remarks. In this paper, we proposed and analyzed a general inertial PPA within the setting of a mixed VI problem (2.1). The proposed method adopts a weighting matrix and allows more flexibility. Our convergence results require weaker conditions in the sense that the weighting matrix G is not necessarily positive definite, as long as the function F is H-monotone and G is positive definite in the null space of H. We also established a nonasymptotic $O(1/k)$ convergence rate result of the proposed method, which was previously not known. The convergence analysis can be easily adapted to the monotone inclusion problem (1.1). We also showed that both the linearized ALM and the LADMM for the structured convex optimization problem are applications of the PPA to the primal-dual optimality conditions, as long as the parameters are reasonably small. As a result, the global convergence and convergence rate results of the linearized ALM and LADMM follow directly from results existing in the literature. This proximal reformulation also allows us to propose inertial versions of the linearized ALM and LADMM, whose convergence is guaranteed under suitable conditions. Our preliminary implementation of the algorithms and extensive experimental results on the compressive principal component pursuit problem have shown that the inertial LADMM is generally faster than the original LADMM. Though in a sense the acceleration is not very significant, we note that the inertial LADMM does not require any additional and unnegligible computational cost either.

Throughout our experiments the extrapolation steplength α_k held constant. How to select α_k adaptively based on the current information such that the overall

TABLE 4
Results of rank$(L_0) = 20$*:* $\varepsilon = 10^{-5}$*, average results of* 10 *random trials.*

$m = n = 1024$				LADMM			iLADMM			
r	k/m^2	$(q/m^2, q/\text{dof})$	$\mathcal{A}$	$\frac{\|L-L_0\|_F}{\|L_0\|_F}$	$\frac{\|S-S_0\|_F}{\|S_0\|_F}$	iter1	$\frac{\|L-L_0\|_F}{\|L_0\|_F}$	$\frac{\|S-S_0\|_F}{\|S_0\|_F}$	iter2	$\frac{\text{iter2}}{\text{iter1}}$
20	1%	(40%, 8.22)	pdct	2.21e-5	6.28e-5	506.4	1.83e-5	6.41e-5	374.9	0.74
			pfft	1.83e-5	6.62e-5	434.6	1.81e-5	6.93e-5	327.7	0.75
			pwht	1.55e-5	5.48e-5	500.9	1.62e-5	6.12e-5	375.2	0.75
		(60%, 12.33)	pdct	1.19e-5	3.38e-5	249.8	1.15e-5	3.83e-5	176.1	0.70
			pfft	1.12e-5	3.32e-5	218.2	1.00e-5	3.88e-5	155.9	0.71
			pwht	2.11e-5	4.71e-5	253.6	1.57e-5	4.68e-5	176.4	0.70
		(80%, 16.43)	pdct	1.29e-5	2.84e-5	123.8	7.86e-6	2.31e-5	78.5	0.63
			pfft	1.03e-5	2.47e-5	100.4	5.63e-6	2.08e-5	56.6	0.56
			pwht	1.20e-5	2.50e-5	128.2	7.84e-6	2.51e-5	81.2	0.63
	5%	(40%, 4.51)	pdct	1.89e-5	7.60e-5	837.6	1.72e-5	6.57e-5	616.0	0.74
			pfft	1.22e-5	5.35e-5	655.8	1.73e-5	6.64e-5	488.6	0.75
			pwht	1.89e-5	7.65e-5	832.9	1.73e-5	6.52e-5	613.1	0.74
		(60%, 6.77)	pdct	7.29e-6	3.62e-5	325.5	1.04e-5	4.31e-5	237.8	0.73
			pfft	9.55e-6	4.30e-5	292.2	1.17e-5	4.06e-5	214.4	0.73
			pwht	7.11e-6	3.66e-5	323.8	1.09e-5	4.24e-5	236.9	0.73
		(80%, 9.02)	pdct	1.10e-5	2.85e-5	149.2	5.43e-6	2.28e-5	101.9	0.68
			pfft	7.68e-6	3.13e-5	133.7	5.26e-6	2.47e-5	88.6	0.66
			pwht	9.83e-6	2.67e-5	151.4	6.33e-6	2.30e-5	102.8	0.68
	10%	(40%, 2.88)	pdct	1.32e-1	4.65e-1	—	9.35e-2	3.52e-1	—	—
			pfft	6.03e-2	1.98e-1	—	1.25e-2	4.32e-2	—	—
			pwht	1.32e-1	4.62e-1	—	9.35e-2	3.49e-1	—	—
		(60%, 4.33)	pdct	7.23e-6	3.27e-5	517.0	1.22e-5	4.09e-5	375.3	0.73
			pfft	9.63e-6	4.08e-5	441.0	8.89e-6	3.32e-5	325.2	0.74
			pwht	7.41e-6	3.33e-5	516.5	1.24e-5	4.26e-5	374.9	0.73
		(80%, 5.77)	pdct	9.20e-6	3.05e-5	219.3	8.41e-6	2.43e-5	154.1	0.70
			pfft	9.81e-6	3.23e-5	200.1	8.14e-6	2.79e-5	137.3	0.69
			pwht	9.23e-6	3.16e-5	219.7	8.46e-6	2.49e-5	154.0	0.70

algorithm performs more efficiently and stable is a practically very important question and deserves further investigation. Another theoretical issue is to investigate worst-case complexity analysis for general inertial type algorithms. In fact, complexity results of inertial type algorithms for minimizing closed proper convex functions already exist in the literature. The pioneering work in this direction is due to Nesterov [35], where the algorithm can also be viewed in the perspective of inertial algorithms. Refined analyses for more general problems can be found in [7, 24]. Let $f : \Re^n \to \Re \cup \{+\infty\}$ be an extended real-valued proper convex function and be bounded below. Based on [35, 7, 24], the following algorithm can be studied. Let $w^0 \in \Re^n$ be given. Set $w^0 = w^{-1}$, $t_0 = 1$, and $k = 0$. For $k \geq 0$, the algorithm iterates as

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}, \tag{5.1a}$$

$$\bar{w}^k = w^k + \frac{t_k - 1}{t_{k+1}}(w^k - w^{k-1}), \tag{5.1b}$$

$$w^{k+1} = \arg\min_w f(w) + \frac{1}{2\lambda_k}\|w - \bar{w}^k\|^2. \tag{5.1c}$$

Using analyses similar to those in [35, 24, 7], one can show that the sequence $\{w^k\}_{k=0}^{\infty}$ satisfies

$$f(w^k) - \min_{w \in \Re^n} f(w) = O(1/k^2).$$

Algorithm (5.1) is nothing but an inertial PPA with steplength $\alpha_k = \frac{t_k - 1}{t_{k+1}}$. It is interesting to note that α_k is monotonically increasing as $k \to \infty$ and converges to 1, which is much larger than the upper bound condition $\alpha < 1/3$ required in Theorem 2. Also note that the convergence for (5.1) is measured by the objective residue. Without further assumptions on f, it seems difficult to establish convergence of the sequence $\{w^k\}_{k=0}^{\infty}$; see, e.g., [24]. In comparison, our results impose a smaller upper bound on α_k but guarantee the convergence of the sequence of iterates $\{w^k\}_{k=0}^{\infty}$. Even though we have derived some nonasymptotic $O(1/k)$ convergence rate results, there seems to be a certain gap between the classical results [35, 7, 24] for minimizing closed proper convex functions and the results presented in the present paper. Further research in this direction is interesting.

REFERENCES

[1] F. Aluffi-Pentini, V. Parisi, and F. Zirilli, *Algorithm* 617. *DAFNE: A differential-equations algorithm for nonlinear equations*, ACM Trans. Math. Software, 10 (1984), pp. 317–324.

[2] F. Alvarez, *On the minimizing property of a second order dissipative system in Hilbert spaces*, SIAM J. Control Optim., 38 (2000), pp. 1102–1119.

[3] F. Alvarez, *Weak convergence of a relaxed and inertial hybrid projection-proximal point algorithm for maximal monotone operators in Hilbert space*, SIAM J. Optim., 14 (2004), pp. 773–782.

[4] F. Alvarez and H. Attouch, *An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping*, Set-Valued Anal., 9 (2001), pp. 3–11.

[5] A. S. Antipin, *Minimization of convex functions on convex sets by means of differential equations*, Differential Equations, 30 (1994), pp. 1475–1486, 1652.

[6] H. Attouch, J. Peypouquet, and P. Redont, *A dynamical approach to an inertial forward-backward algorithm for convex minimization*, SIAM J. Optim., 24 (2014), pp. 232–256.

[7] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202.

[8] R. I. Bot and E. R. Csetnek, *A hybrid proximal-extragradient algorithm with inertial effects*, preprint, arXiv:1407.0214, 2014.

[9] R. I. Bot and E. R. Csetnek, *An inertial alternating direction method of multipliers*, preprint, arXiv:1404.4582, 2014.

[10] R. I. Bot and E. R. Csetnek, *An inertial forward-backward-forward primal-dual splitting algorithm for solving monotone inclusion problems*, preprint, arXiv:1402.5291, 2014.

[11] R. I. Bot and E. R. Csetnek, *An inertial Tseng's type proximal algorithm for nonsmooth and nonconvex optimization problems*, preprint, arXiv:1406.0724, 2014.

[12] R. I. Bot, E. R. Csetnek, and C. Hendrich, *Inertial Douglas-Rachford splitting for monotone inclusion problems*, preprint, arXiv:1403.3330, 2014.

[13] 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 Machine Learning, 3 (2011), pp. 1–122.

[14] R. E. Bruck, Jr., *Asymptotic convergence of nonlinear contraction semigroups in Hilbert space*, J. Funct. Anal., 18 (1975), pp. 15–26.

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

[16] C. Chen, *Numerical Algorithms for a Class of Matrix Norm Approximation Problems*, Ph.D. thesis, Nanjing University, 2012.

[17] E. Corman and X. Yuan, *A generalized proximal point algorithm and its convergence rate*, SIAM J. Optim., 24 (2014), pp. 1614–1638.

[18] D. Davis and W. Yin, *Convergence rate analysis of several splitting schemes*, preprint, arXiv:1406.4834, 2014.

[19] J. Douglas, Jr., and H. H. Rachford, Jr., *On the numerical solution of heat conduction problems in two and three space variables*, Trans. Amer. Math. Soc., 82 (1956), pp. 421–439.

[20] 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.

[21] M. Fazel, T. K. Pong, D. F. Sun, and P. Tseng, *Hankel matrix rank minimization with applications to system identification and realization*, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 946–977.

[22] D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite element approximation*, Comput. Math. Appl., 2 (1976), pp. 17–40.

[23] 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*, ESAIM Math. Model. Numer. Anal., 9 (1975), pp. 41–76.

[24] O. Güler, *New proximal point algorithms for convex minimization*, SIAM J. Optim., 2 (1992), pp. 649–664.

[25] B. S. He and X. M. Yuan, *On non-ergodic convergence rate of Douglas-Rachford alternating directions method of multipliers*, Numer. Math., 130 (2015), pp. 567–577.

[26] B. He and X. Yuan, *On the $O(1/n)$ convergence rate of the Douglas–Rachford alternating direction method*, SIAM J. Numer. Anal., 50 (2012), pp. 700–709.

[27] M. R. Hestenes, *Multiplier and gradient methods*, J. Optim. Theory Appl., 4 (1969), pp. 303–320.

[28] D. A. Lorenz and T. Pock, *An inertial forward-backward algorithm for monotone inclusions*, J. Math. Imaging Vis., 51 (2015), pp. 311–325.

[29] S. Ma, D. Goldfarb, and L. Chen, *Fixed point and Bregman iterative methods for matrix rank minimization*, Math. Program., 128 (2011), pp. 321–353.

[30] P.-E. Maingé and N. Merabet, *A new inertial-type hybrid projection-proximal algorithm for monotone inclusions*, Appl. Math. Comput., 215 (2010), pp. 3149–3162.

[31] P.-E. Maingé and A. Moudafi, *A proximal method for maximal monotone operators via discretization of a first order dissipative dynamical system*, J. Convex Anal., 14 (2007), pp. 869–878.

[32] B. Martinet, *Régularisation d'inéquations variationnelles par approximations successives*, Rev. Française Informat. Recherche Opérationnelle, 4 (1970), pp. 154–158.

[33] J.-J. Moreau, *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. France, 93 (1965), pp. 273–299.

[34] A. Moudafi and E. Elissabeth, *Approximate inertial proximal methods using the enlargement of maximal monotone operators*, Int. J. Pure Appl. Math., 5 (2003), pp. 283–299.

[35] Y. E. Nesterov, *A method for solving the convex programming problem with convergence rate $O(1/k^2)$*, Dokl. Akad. Nauk SSSR, 269 (1983), pp. 543–547.

[36] P. Ochs, T. Brox, and T. Pock, *iPiasco: Inertial proximal algorithm for strongly convex optimization*, manuscript, 2014.

[37] P. Ochs, Y. Chen, T. Brox, and T. Pock, *iPiano: Inertial proximal algorithm for non-convex optimization*, SIAM J. Imaging Sci., to appear.

[38] B. T. Poljak, *Some methods of speeding up the convergence of iterative methods*, Ž. Vyčisl. Mat. i Mat. Fiz., 4 (1964), pp. 791–803.

[39] M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, in Optimization, Academic Press, London, 1969, pp. 283–298.

[40] R. T. Rockafellar, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Math. Oper. Res., 1 (1976), pp. 97–116.

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

[42] P. Tseng, *A modified forward-backward splitting method for maximal monotone mappings*, SIAM J. Control Optim., 38 (2000), pp. 431–446.

[43] J. Wright, A. Ganesh, K. Min, and Y. Ma, *Compressive principal component pursuit*, Inform. and Inference, 2 (2013), pp. 32–68.

[44] J. Yang and X. Yuan, *Linearized augmented lagrangian and alternating direction methods for nuclear norm minimization*, Math. Comp., 82 (2013), pp. 301–329.

[45] J. Yang and Y. Zhang, *Alternating direction algorithms for ℓ_1-problems in compressive sensing*, SIAM J. Sci. Comput., 33 (2011), pp. 250–278.