ON THE LINEAR CONVERGENCE RATE OF A GENERALIZED PROXIMAL POINT ALGORITHM

MIN TAO¹ AND XIAOMING YUAN²

February 17, 2015

Abstract. The proximal point algorithm (PPA) has been well studied in the literature. In particular, its linear convergence rate has been studied by Rockafellar in 1976 under certain condition. We consider a generalized PPA in the generic setting of finding a zero point of a maximal monotone operator, and show that the condition proposed by Rockafellar can also sufficiently ensure the linear convergence rate for this generalized PPA. Both the exact and inexact versions of this generalized PPA are discussed. The motivation to consider this generalized PPA is that it includes as special cases the relaxed versions of some splitting methods that are originated from PPA. Thus, linear convergence results of this generalized PPA can be used to better understand the convergence of some widely used algorithms in the literature. We focus on the particular convex minimization context and specify Rockafellar’s condition to see how to ensure the linear convergence rate for some efficient numerical schemes, including the classical augmented Lagrangian method proposed by Hensen and Powell in 1969 and its relaxed version, the original alternating direction method of multipliers (ADMM) by Glowinski and Marrocco in 1975 and its relaxed version (i.e., the generalized ADMM by Eckstein and Bertsekas in 1992). Some refined conditions weaker than existing ones are proposed in these particular contexts.

Key words. Convex programming, proximal point algorithm, augmented Lagrangian method, alternating direction method of multipliers, linear convergence rate

1. Introduction. Let \( H \) be a real Hilbert space with inner product \( \langle \cdot, \cdot \rangle \). A set-valued mapping \( T : H \to 2^H \) is said to be monotone if
\[
\langle z - z', w - w' \rangle \geq 0, \quad \forall z, z' \in H, w \in T(z), w' \in T(z').
\]
\( T \) is said to be maximal monotone if, in addition, its graph is not properly contained in the graph of any other monotone operator. A fundamental problem is finding a zero point, denoted by \( z^* \), of a maximal monotone set-valued mapping \( T \):
\[
0 \in T(z).
\]
Throughout, the set of \( T \)'s zero point, denoted by \( zer(T) \), is assumed to be nonempty.

The proximal point algorithm (PPA), which traces back to [24, 26], has been playing an important role both theoretically and algorithmically for (1.1). Starting from an arbitrary point \( z^0 \) in \( H \), the PPA iteratively generates its sequence \( \{z^k\} \) by the scheme
\[
0 \in c_k T(z^{k+1}) + z^{k+1} - z^k,
\]
where \( \{c_k\} \), called proximal parameter, is a sequence of positive real numbers. Indeed, as shown in [33], the convergence of PPA can be ensured when \( \{c_k\} \) is bounded away from zero. Moreover, an inexact version...
of PPA was proposed in [33], allowing the subproblem (1.2) to be solved approximately subject to some inexactness criteria. Conceptually, the inexact version of PPA can be written as

$$0 \approx c_k T(z^{k+1}) + z^{k+1} - z^k, \quad (1.3)$$

in which the accuracy should be judiciously chosen to guarantee its convergence. Let

$$J_{c_k T} := (I + c_k T)^{-1} \quad (1.4)$$

denote the resolvent operator of the maximal monotone set-valued mapping $T$ for a positive scalar $c_k$ (Note that $J_{c_k T}$ is single-valued, see, e.g., [12]). Then, the exact and inexact versions of the PPA can be written, respectively, as

$$z^{k+1} = J_{c_k T}(z^k) \quad (1.5)$$

and

$$z^{k+1} \approx J_{c_k T}(z^k). \quad (1.6)$$

Technically, (1.6) includes (1.5) as the special case where the tolerance of accuracy is zero. But we still discuss them individually because (1.3) is of particular interest and it may have stronger convergence, because it requires estimating the resolvent operator accurately.

Research results on the convergence of PPA can be found in earlier literature. For example, when $T$ is specified as the sum of a single-valued, monotone and hemicontinuous mapping and the normal cone to a bounded set, i.e., the problem (1.1) reduces to a variational inequality, then some convergence of the exact version of PPA (1.2) with $c_k \equiv c$ in the weak topology was investigated in [24, 25]. In [33], the convergence of both the exact and inexact versions of PPA was comprehensively studied; it is indeed the work [33] that popularized PPA in optimization community. More specifically, under the condition that $\{c_k\}$ is bounded away from zero, the convergence for “$\approx$” in (1.6) is specified into certain forms (see (A) and (B) of Section 1 in [33]). In fact, the exact version (1.5) was shown to find a solution point of (1.1) after finitely many iterations in [33]. Note that the strong monotonicity of $T$ is not required for the analysis in [33]. Moreover, if the inverse of $T$ (denoted by $T^{-1}$) is Lipschitz continuous at 0 (see Definition 2.4 in Section 2 or Section 3 in [33]) and $\{c_k\}$ is bounded away from zero, it was proved in [33] that the (1.6) (thus also (1.5)) with some relative error control in its accuracy is linearly convergent. There are many other articles studying the PPA from various perspectives. For example, the PPA application to nonconvex problems studied in [14], the PPA with variable metric in [6], a unified convergence rate analysis for some PPA-based decomposition methods in [35], accelerated PPA schemes with a worst-case $O(1/k^2)$ convergence rate proposed in [19], the logarithmic quadratic proximal extension considered in [12], and some other proximal-type algorithms in [36]. We particularly refer to [18, 27] for some insightful analysis on the iteration complexity of PPA, which can be regarded as a measure of its worst-case convergence rate. Algorithmically, the PPA is the basis of a large number of celebrated methods, e.g., the projected gradient method [30], the extragradient method [22], the extended extragradient and hyperplane schemes in [1], the forward-backward operator splitting method [23], and the accelerated projected gradient method [28].

Our main purpose is to extend the analysis in [33] to more general PPA schemes, and establish their linear convergence rates under the same assumption as [33]. More specifically, under the assumption:

[Assumption]: $T^{-1}$ is Lipschitz continuous at 0,
we show the linear convergence for the generalized PPA scheme

\[ z^{k+1} = z^k - \gamma (z^k - J_{c_k} T(z^k)), \]  

(1.7)

and its inexact version

\[ z^{k+1} \approx z^k - \gamma (z^k - J_{c_k} T(z^k)). \]  

(1.8)

In (1.7) and (1.8), the proximal parameter sequence \{c_k\} is also required to be bound away from zero (\(c_k \geq \kappa > 0\) for all \(k\)), and the relaxation factor \(\gamma \in (0, 2)\). Obviously, the schemes (1.7) and (1.8) include (1.5) and (1.6) as special cases with \(\gamma = 1\), respectively. Note that, as analyzed in [33], “the assumption of Lipschitz continuity of \(T^{-1}\) at 0 turns out to be very natural in applications to convex programming”. Indeed, we will consider several specific convex optimization contexts of the abstract model (1.1) and investigate how the assumption “\(T^{-1}\) is Lipschitz continuous at 0” can be specified in these special contexts to ensure the linear convergence rates for some benchmark algorithms in the literature.

One particular motivation for studying the generalized PPA schemes (1.7) and (1.8) is that they represent a unified consideration for accelerating a series of well known splitting algorithms in the convex optimization context. Indeed, it has been well studied in the literature that some popular iterative schemes such as the Douglas-Rachford splitting method (DRSM) in [11, 23], the Peaceman-Rachford splitting method in [29, 23], the augmented Lagrangian method (ALM) in [21, 31], and the alternating direction method of multipliers (ADMM) in [16] are all special cases of the PPA (1.5) with specific choices of \(T\). That is, the PPA scheme (1.5) indeed is a unified illustration of some different schemes for different models. A theoretical generalization of (1.5), the generalized PPA (1.7) indeed can numerically accelerate the original PPA (1.5) and thus it is of numerical interest. One example is that the generalized PPA scheme (1.7) with \(c_k \equiv c\) was suggested in [17] and its numerical efficiency has been verified in, e.g., [3, 7]. Moreover, as analyzed in [12] (see also Section 8 for details), the DRSM in [11, 23] is also a special case of the PPA (1.5). Thus, the DRSM can be immediately generalized by (1.7). Indeed, since it was proved in [13] that the ADMM in [16] is a special case of the DRSM and thus it is also a special case of (1.5), the ADMM can be accelerated by the scheme (1.7) and it results in the generalized ADMM in [12]. Recently, the acceleration of the generalized ADMM was demonstrated in [13] by some applications in statistical learning. In [8], the generalization of the PRSM via (1.7) was also studied (only the case \(c_k \equiv c\) was considered therein). In [8], the sublinear and linear convergence rates of the schemes (1.7) and (1.8) with \(c_k \equiv c\) was studied; and these results were especially specified for the DRSM and PRSM scenarios. The results in [8] were refined in [2] for the special DRSM and PRSM cases of (1.2). Indeed, the case of \(\gamma \geq 2\) in (1.7) and (1.8) was also discussed in [8]. We will show later in Section 5 that for the case \(\gamma \in (0, 2)\), some conditions weaker than those in [8] can be found based on the analysis in [33] to ensure the linear convergence of the schemes (1.7) and (1.8). Moreover, we will show that for a convex minimization context of the problem (1.1), the condition in [33] under our discussion is weaker than the conditions in [9].

The rest of this paper is organized as follows. In Section 2 some preliminaries useful for further analysis are summarized. Then, we discuss the convergence and the linear convergence rate of the exact version of the generalized PPA (1.7) in Section 3. In Section 4, the convergence and linear convergence rate of its inexact version (1.8), in which the inexactness criterion is also specified, is studied. Then, we revisit the assumption “\(T^{-1}\) is Lipschitz continuous at 0” in Section 5 and show that it can be further relaxed. In Section 6, we discuss the possibility of deriving the superlinear convergence for the schemes (1.7) and (1.8). In Section 7, we apply the scheme (1.7) to a canonical convex minimization model with linear constraints and discuss the linear convergence for the resulting generalized ALM scheme. In Section 8, we focus on the analysis for the
linear convergence of the ADMM and the generalized ADMM scheme, both are special cases of the scheme [15]. Finally, some conclusions are made in Section 9.

2. Preliminaries. In this section, we recall some definitions and known results for further discussions.

2.1. Some Definitions. We first recall some basic definitions to be used in our analysis.

Definition 2.1. Let $T : H \rightarrow 2^H$ be set-valued and maximal monotone. Then, $T$ is said to be nonexpansive if $\|w' - w\| \leq \|z' - z\|$, $\forall z, z' \in H$, $w \in T(z)$, $w' \in T(z')$.

Definition 2.2. Let $T : H \rightarrow 2^H$ be set-valued and maximal monotone. Then, $T$ is said to be firmly nonexpansive if $\|w' - w\|^2 \leq \langle z' - z, w' - w \rangle$, $\forall z, z' \in H$, $w \in T(z)$, $w' \in T(z')$.

Definition 2.3. Let $T : H \rightarrow 2^H$ be set-valued and maximal monotone. Then, $T$ is called $\alpha$-coercive if $\langle z - z', w - w' \rangle \geq \alpha \|z - z'\|^2$, $\forall z, z' \in H$, $w \in T(z)$, $w' \in T(z')$ for $\alpha > 0$.

Definition 2.4. Let $T$ be set-valued and be defined on $H$. Then, $T^{-1}$ is called Lipschitz continuous at 0 with modulus $\alpha \geq 0$ if there is a unique solution $z^*$ to $0 \in T(z)$ (i.e. $T^{-1}(0) = \{z^*\}$), and for some $\tau > 0$ we have $\|z - z'\| \leq \alpha \|w\|$ whenever $z \in T^{-1}(w)$ and $\|w\| \leq \tau$.

Definition 2.5 is quoted from [53]. Based on these definitions, we have some immediate conclusions. For examples, if $T$ is nonexpansive, then it is Lipschitz continuous. Note that Definition 2.3 corresponds to the $\alpha$-strong-monotonicity defined in [8]. Moreover, clearly, the problem (1.1) has a unique solution point when $T^{-1}$ is Lipschitz continuous at 0.

2.2. Some Known Results. Then, we summarize some known results that are relevant to our analysis. The following lemma summarizes some well-known properties of a firmly nonexpansive operator. The proof is straightforward and thus omitted, or see, e.g., [12].

Lemma 2.5. We have the following facts.

i) All firmly nonexpansive operators are nonexpansive.

ii) An operator $T$ is firmly nonexpansive if and only if $2T - I$ is nonexpansive.

iii) An operator is firmly nonexpansive if and only if it is of the form $\frac{1}{2}(C + I)$, where $C$ is nonexpansive.

iv) An operator $T$ is firmly nonexpansive if and only if $I - T$ is firmly nonexpansive.

In the following lemma, we show some simple conclusions for the resolvent operator of a maximal monotone operator.

Lemma 2.6. Let $T : H \rightarrow 2^H$ be set-valued and maximal monotone; $J_{cT}$ be defined in (1.4), and $c > 0$ be a scalar. Then, we have

i) $\langle J_{cT}(z) - J_{cT}(z'), (I - J_{cT})(z) - (I - J_{cT})(z') \rangle \geq 0$, $\forall z, z' \in H$.

ii) $\|z - z'\|^2 \geq \|J_{cT}(z) - J_{cT}(z')\|^2 + \|(I - J_{cT})(z) - (I - J_{cT})(z')\|^2$, $\forall z, z' \in H$.

Proof. Obviously, $J_{cT}$ defined in (1.4) is nonexpansive, and it implies the first property immediately. The second property is trivial because of Property (i).

Last, let us recall the representation lemma, see, e.g., [12].

Lemma 2.7. (The representation lemma) Let $c > 0$ and let $T$ be monotone on $H$. Then every element $z$ of $H$ can be written in at most one way as $x + cy$, where $y \in T(x)$. If $T$ is maximal, then every element $z$ of $H$ can be written in exactly one way as $x + cy$, where $y \in T(x)$.

3. Convergence of the Exact Version (1.7). In this section, we show that the generalized PPA also converges linearly to a zero point of $T$ under the assumption “$T^{-1}$ is Lipschitz continuous at 0 with positive modulus”*, the same one as that in [33]. For a lighter notation in analysis, we use the notation $z^k = J_{\alpha T}(z^k)$ in the following analysis.
3.1. Global Convergence. First, we show the global convergence of (1.7). Note that we do not need the assumption “$T^{-1}$ is Lipschitz continuous at 0 with positive modulus” for proving the global convergence. The next theorem shows that the sequence $\{z^k\}$ generalized by (1.7) with $\gamma \in (0,2)$ is strictly contractive with respect to the solution set of (1.1), an important property ensuring its global convergence.

**Theorem 3.1. (Strict contraction)** Let $\{z^k\}$ be the sequence generated by the exact version of the generalized PPA scheme (1.7) with $\gamma \in (0,2)$ and $\{c_k\}$ bounded away from 0; let $z^*$ be a solution point of (1.1). We have

$$
\|z^{k+1} - z^*\|^2 \leq \|z^k - z^*\|^2 - \gamma(2 - \gamma) \|z^k - z^*\|^2.
$$

(3.1)

**Proof.** First, applying the property (ii) in Lemma 2.6 with $z = z^k$ and $\tilde{z} = z^*$, we get

$$
\|z^k - z^*\|^2 \geq \|\tilde{z}^k - z^*\|^2 + \|z^k - z^*\|^2.
$$

(3.2)

We thus have

$$
\langle \tilde{z}^k - z^*, z^k - \tilde{z}^k \rangle \geq 0,
$$

and furthermore

$$
\langle z^k - z^*, z^k - \tilde{z}^k \rangle \geq \|z^k - \tilde{z}^k\|^2.
$$

(3.3)

Consequently, we have

$$
\|z^{k+1} - z^*\|^2 = \|z^k - \gamma(z^k - \tilde{z}^k) - z^*\|^2
$$

$$
= \|z^k - z^*\|^2 - 2\gamma\langle z^k - z^*, z^k - \tilde{z}^k \rangle + \gamma^2\|z^k - z^*\|^2
$$

$$
\leq \|z^k - z^*\|^2 - \gamma(2 - \gamma)\|z^k - z^*\|^2,
$$

where the inequality follows from (3.3). Thus, the assertion (3.1) is proved.

Based on Theorem 3.1, the convergence of (1.7) can be easily established.

**Theorem 3.2. (Global convergence)** Let $\{z^k\}$ be the sequence generated by the exact version of the generalized PPA scheme (1.7) with $\gamma \in (0,2)$ and $\{c_k\}$ bounded away from 0. Then it globally converges to a solution point of (1.7).

**Proof.** According to (3.1), the sequence $\{z^k\}$ is bounded, and it has at least one accumulation point, say $z^\infty$. Let $\{\tilde{z}^k\}$ be the subsequence converging to $z^\infty$. Recall the notation $\tilde{z}^k = J_{c_kT}(z^k)$ and the definition of $J_{c_kT}$ in (1.4). We thus have

$$
c^{-1}_{\tilde{k}}(z^k - \tilde{z}^k) \in T(\tilde{z}^k).
$$

Then, using the monotonicity of $T$, for an integer $k_j$, it holds that

$$
\langle z - \tilde{z}^k, w - c^{-1}_{\tilde{k}}(z^k - \tilde{z}^k) \rangle \geq 0, \text{ for all } z, w \text{ satisfying } w \in T(z).
$$

(3.4)

Again, it follows from (3.1) that $\lim_{j \to \infty} \|z^{k_j} - \tilde{z}^{k_j}\| = 0$. Combining this fact with $\lim_{j \to \infty} \|z^{k_j} - z^\infty\| = 0$, we get $\lim_{j \to \infty} \|\tilde{z}^{k_j} - z^\infty\| = 0$. Recall $\{c_k\}$ is bounded away from 0. Then, taking $j \to \infty$ in (3.4), we obtain

$$
\langle z - z^\infty, w \rangle \geq 0 \text{ for all } z, w \text{ satisfying } w \in T(z).
$$

In view of the maximality of $T$, this inequality implies that $z^\infty$ is a solution point of (1.1), see, e.g. [33]. It is easy to see from Theorem 3.1 that the sequence $\{z^k\}$ cannot have more than one accumulation point. Thus, $\{z^k\}$ converges to $z^\infty$ which a solution point of (1.1). The proof is complete.
3.2. Linear Convergence. Now, under the assumption “$T^{-1}$ is Lipschitz continuous at 0 with positive modulus”, we prove the linear convergence of (1.7). First, two lemmas are presented.

Lemma 3.3. Let $T : H \to 2^H$ be maximal monotone and $z^*$ be a solution point of (1.7); let $c_k > 0$. If $T^{-1}$ is Lipschitz continuous at 0 with modulus $a > 0$, then there exists a positive $\tau$ such that

$$
\|J_{c_kT}(z) - z^*\| \leq \frac{\alpha}{\sqrt{a^2 + c_k^2}}\|z - z^*\| \quad \text{when } \|c_k^{-1}(z - J_{c_kT}(z))\| \leq \tau, \forall z \in H.
$$

(3.5)

Proof. Applying Property (ii) in Lemma 2.6 with $z = z^*$ and $c = c_k$, we get

$$
\|z - z^*\|^2 \geq \|J_{c_kT}(z) - z^*\|^2 + \|(I - J_{c_kT})(z)\|^2.
$$

(3.6)

Recall the definition of $J_{c_kT}$ in (1.4). We have

$$
c_k^{-1}(I - J_{c_kT})(z) \in T(J_{c_kT}(z)).
$$

Since $T^{-1}$ is Lipschitz continuous at 0 with modulus $a > 0$, it follows from Definition 2.4 that there exists a positive parameter $\tau$ such that

$$
\|J_{c_kT}(z) - z^*\| \leq a\|c_k^{-1}(I - J_{c_kT})(z)\| \quad \text{when } \|c_k^{-1}(I - J_{c_kT})(z)\| \leq \tau.
$$

Substituting this inequality into (3.6), we obtain (3.5). The proof is complete.

Lemma 3.4. Let $\{z^k\}$ be the sequence generated by the exact version of the generalized PPA scheme (1.7) with $\gamma \in (0, 2)$, and $z^*$ be a solution point of (1.7). If $T^{-1}$ is Lipschitz continuous at 0 with modulus $a > 0$, and the proximal parameter sequence $\{c_k\}$ is bounded away from zero ($c_k \geq \kappa > 0$ for any $k$), then there exists an integer $\hat{k}$ such that

$$
\|z^k - z^*\| \leq \frac{\alpha}{\sqrt{a^2 + c_k^2}}\|z^k - z^*\| \forall k > \hat{k}.
$$

(3.7)

Proof. Applying Lemma 3.3 with $z = z^k$, we know there exists $\tau > 0$ such that

$$
\|J_{c_kT}(z^k) - z^*\| \leq \frac{\alpha}{\sqrt{a^2 + c_k^2}}\|z^k - z^*\| \quad \text{when } \|c_k^{-1}(z^k - J_{c_kT}(z^k))\| \leq \tau.
$$

Using the notation $J_{c_kT}(z^k) = \tilde{z}^k$, this inequality can be rewritten as

$$
\|\tilde{z}^k - z^*\| \leq \frac{\alpha}{\sqrt{a^2 + c_k^2}}\|z^k - z^*\| \quad \text{when } \|c_k^{-1}(z^k - \tilde{z}^k)\| \leq \tau.
$$

It follows from Theorem 3.1 that $\lim_{k \to \infty} \|z^k - \tilde{z}^k\| = 0$. Then, there exists an integer $\hat{k}$ such that,

$$
c_k^{-1}\|z^k - \tilde{z}^k\| \leq \kappa^{-1}\|z^k - \tilde{z}^k\| \leq \tau \quad \text{when } k > \hat{k}.
$$

Thus, the assertion (3.7) is implied by the two inequalities above. The proof is complete.

Now, we prove the linear convergence rate of (1.7) in the following theorem.

Theorem 3.5. If $T^{-1}$ is Lipschitz continuous at 0 with modulus $a > 0$ and the proximal parameter $\{c_k\}$ is bounded away from zero ($c_k \geq \kappa > 0$), then the sequence $\{z^k\}$ generated by the exact version of the generalized PPA scheme (1.7) with $\gamma \in (0, 2)$ converges linearly to a solution point of (1.7).
The inequality (3.9) implies the linear convergence rate of the sequence
\[ \parallel z^k - z^* \parallel \geq (1 - \frac{a}{\sqrt{a^2 + c_k^2}}) \parallel z^{k-1} - z^* \parallel, \quad k > \hat{k}. \] (3.8)

Then, together with Theorem 3.3, the assertion (3.8) enables us to derive
\[
\parallel z^{k+1} - z^* \parallel^2 
\leq \parallel z^k - z^* \parallel^2 - \gamma(2 - \gamma) \parallel z^k - \hat{z}^k \parallel^2
\leq \parallel z^k - z^* \parallel^2 - \gamma(2 - \gamma)(1 - \frac{a}{\sqrt{a^2 + c_k^2}})^2 \parallel z^k - z^* \parallel^2
= \left( 1 - (2\gamma - \gamma^2)(1 - \frac{a}{\sqrt{a^2 + c_k^2}})^2 \right) \parallel z^k - z^* \parallel^2. \] (3.9)

Since $\gamma \in (0, 2)$ and $c_k \geq \kappa > 0$ for any $k$, we have
\[ 0 \leq 1 - (2\gamma - \gamma^2) \leq 1 - (2\gamma - \gamma^2)(1 - \frac{a}{\sqrt{a^2 + c_k^2}})^2 \leq 1 - (2\gamma - \gamma^2)(1 - \frac{a}{\sqrt{a^2 + \kappa^2}})^2 < 1. \]

Thus, the inequality (3.9) implies the linear convergence rate of the sequence $\{z^k\}$. The proof is complete. \[ \square \]

**Remark 1.** Similarly as Definition 2.4, if a sequence $\{z^k\}$ converges to $z^*$, we can define “$T^{-1}$ is Lipschitz continuous with modulus $a \geq 0$ at the sequence $\{z^k\}$” if there exists some $\tau > 0$ such that
\[ \parallel z^k - z^* \parallel \leq a\parallel w^k \parallel \text{ whenever } z^k \in T^{-1}(w^k) \text{ and } \parallel w^k \parallel \leq \tau. \]

Then, it can be easily seen that the linear convergence of $\{z^k\}$ generated by (1.7) can be guaranteed under the less strengthen assumption “$T^{-1}$ is Lipschitz continuous at the iterates $\hat{z}^k$ with positive modulus when $k$ is sufficiently large”. Recall the fact $\hat{z}^k \in T^{-1}(c_k^{-1}(z^k - \hat{z}^k))$ and $z^* \in T^{-1}(0)$. Then, this less strengthen assumption is equivalent to saying that there exists an integer $\hat{k}$ such that
\[ \parallel \hat{z}^k - z^* \parallel \leq a\parallel c_k^{-1}(z^k - \hat{z}^k) \parallel \text{ when } k > \hat{k}, \] (3.10)

where $\hat{k}$ is large enough such that $\parallel c_k^{-1}(z^k - z^*) \parallel \leq \tau$. Note that $\parallel c_k^{-1}(z^k - \hat{z}^k) \parallel \leq \tau$ can be ensured by the fact \[ \lim_{k \to \infty} \parallel z^k - \hat{z}^k \parallel = 0 \text{ implied in (3.1) and that } \{c_k\} \text{ is bounded away from } 0. \]

More discussion is referred to Section \[ \square \]

**4. The Convergence of the Inexact Version (1.8).** In this section, we specify the inexactness criterion for (1.8) and show its linear convergence under the same assumption of “$T^{-1}$ is Lipschitz continuous at 0 with positive modulus”. This is a generalization of the inexact version (1.6) considered in [3,3]. More specifically, we consider the scheme
\[
\begin{align*}
{z}^{k+1} &= (1 - \gamma)z^k + \gamma \hat{z}^k, \\
\parallel z^k - J_{c_kT}(z^k) \parallel &\leq \delta_k \parallel z^k - \hat{z}^{k+1} \parallel,
\end{align*}
\] (4.1)

where $\gamma \in (0, 2)$, $c_k \geq \kappa > 0$ for any $k$, and $\{\delta_k\}$ is a sequence of positive real numbers satisfying $\sum_k \delta_k < +\infty$.

Note that in (4.1), we consider using relative errors to control the accuracy in (1.8); thus it is different from the inexact version in [5] which uses absolute errors. We still use the notation $\hat{z}^k = J_{c_kT}(z^k)$ in the upcoming analysis.
4.1. Global Convergence. Again, we first show the global convergence for the sequence \( \{z^k\} \) generated by (4.1). Note that we do not need the assumption “\( T^{-1} \) is Lipschitz continuous at 0 with positive modulus” for proving the global convergence. We first prove several lemmas for this purpose. Their proofs are elementary; but we still include them for completeness.

**Lemma 4.1.** Let \( \{\alpha_k\} \) be a positive sequence satisfying \( \sum_{k=1}^{\infty} \alpha_k < +\infty \). Then, we have

\[
\prod_{k=1}^{\infty} (1 + \alpha_k) < +\infty.
\]

**Proof.** Obviously, it holds that \( \log(1 + x) \leq x \) when \( 0 < x < 1 \). Hence, we have

\[
\sum_{k=1}^{\infty} \log(1 + \alpha_k) \leq \sum_{k=1}^{\infty} \alpha_k < +\infty,
\]

which implies the assertion immediately.

**Lemma 4.2.** Let \( \{\delta_k\} \) be a positive sequence satisfying \( \sum_{k=1}^{\infty} \delta_k < +\infty \) and \( \gamma > 0 \) be a constant. Then, we have

\[
\prod_{k=1}^{\infty} \frac{1 + \gamma \delta_k}{1 - \gamma \delta_k} < +\infty.
\]

**Proof.** Since \( \sum_{k=1}^{\infty} \delta_k < +\infty \), we have \( \delta_k \to 0 \) when \( k \to \infty \). Thus, there exists an integer \( \hat{k} \) such that

\[1 - \gamma \delta_k \geq \frac{1}{2} \quad \text{when} \quad k \geq \hat{k}. \tag{4.2}\]

Hence, we have

\[
\prod_{k=1}^{\infty} \frac{1 + \gamma \delta_k}{1 - \gamma \delta_k} = \prod_{k=1}^{k-1} \frac{1 + \gamma \delta_k}{1 - \gamma \delta_k} \cdot \prod_{k=\hat{k}}^{\infty} \frac{1 + \gamma \delta_k}{1 - \gamma \delta_k} \leq \prod_{k=1}^{k-1} \frac{1 + \gamma \delta_k}{1 - \gamma \delta_k} \cdot 2 \prod_{k=\hat{k}}^{\infty} (1 + \gamma \delta_k) < +\infty.
\]

The proof is complete.

**Lemma 4.3.** Let \( \{a_k\} \) and \( \{b_k\} \) be positive sequences; \( \sum_{k=1}^{\infty} b_k < +\infty \); and

\[a_{k+1} \leq a_k + b_k, \quad \forall k. \tag{4.3}\]

Then, the sequence \( \{a_k\} \) is convergent.

**Proof.** First, it follows from (4.3) that

\[a_{k+1} \leq a_1 + \sum_{i=1}^{k} b_i \leq a_1 + \sum_{i=1}^{\infty} b_i, \quad \forall k.
\]

Since \( \sum_{k=1}^{\infty} b_k < +\infty \), the sequence \( \{a_k\} \) is bounded. Thus, it has at least one accumulation point, say \( a^*_1 \). That is, there exists a subsequence \( \{a_{k_i}\} \) converging to \( a^*_1 \). Suppose that the sequence \( \{a_k\} \) is not convergent. Then, there exists another subsequence \( \{a_{k_i}\} \) converging to another point, say \( a^*_2 \). Obviously, \( a^*_1 \neq a^*_2 \). Without loss of generality, let us assume \( a^*_2 > a^*_1 \). Define \( \epsilon = \frac{1}{2}(a^*_2 - a^*_1) > 0 \). There exists an integer \( J_2 \) such that

\[\sum_{i=J_2}^{\infty} b_i < \epsilon, \quad \forall k.
\]
where $\epsilon > 0$ is a given scalar. On the other hand, for the given $\epsilon$, there exists integers $J_1 > J_2$ such that

$$|a_{k, J_1} - a_1^*| < \epsilon.$$ 

Then, we get

$$a_{k_t} \leq a_{k, J_1} + \sum_{i=k, J_1+1}^{\infty} b_i < a_1^* + \epsilon + \epsilon = a_2^*, \ \forall k_t > k, J_1.$$ 

It contradicts with the fact $a_{k_t} \rightarrow a_2^*$ when $t \rightarrow \infty$. Hence, the sequence $\{a_k\}$ is convergent. The proof is complete.

Now we start to prove the global convergence of (4.1). The key is the sequence generated by the inexact version (4.1) is asymptotically emerged with the sequence by the generalized PPA (1.7). With this fact, the convergence of (4.1) can be established easily.

**Theorem 4.4.** Let $\{z^k\}$ be the sequence generated by the inexact version of the generalized PPA scheme (4.1). Then, we have

1. The sequence $\{z^k\}$ is bounded.
2. It holds that

$$\lim_{k \rightarrow \infty} \|z^k - \tilde{z}^k\| = 0. \quad (4.4)$$

**Proof.** Recall we use $\tilde{z}^k = J_{\epsilon_t T}(z^k)$ for easier notation. Let us use one more notation

$$\hat{z}^{k+1} := (1 - \gamma)z^k + \gamma\tilde{z}^k.$$ 

Indeed, $\hat{z}^{k+1}$ denotes the iterate generated by the exact version (1.7) from the given $z^k$. Thus, for an arbitrary solution point $z^*$ of (1.1), it follows from (3.1) that

$$\|\hat{z}^{k+1} - z^*\|^2 \leq \|z^k - z^*\|^2 - \gamma\|z^{k-1} - \tilde{z}^k\|. \quad (4.5)$$

Recall the definition of $z^{k+1}$ in (4.1). We have

$$z^{k+1} - z^{k+1} = \gamma(z^k - \tilde{z}^k), \quad (4.6)$$

where $z^k$ is also given in (4.1). Thus, for any solution point $z^*$ of (1.1), we have

$$\|z^{k+1} - z^*\| \leq \|z^{k+1} - \hat{z}^{k+1}\| + \|\hat{z}^{k+1} - z^*\|$$

$$\leq \gamma\delta_k\|z^k - \hat{z}^{k+1}\| + \|zh^{k+1} - z^*\|$$

$$\leq \gamma\delta_k(\|z^k - z^*\| + \|z^{k+1} - z^*\|) + \|\hat{z}^{k+1} - z^*\|$$

$$\leq \gamma\delta_k(\|z^k - z^*\| + \|z^{k+1} - z^*\|) + \|z^k - z^*\|, \quad (4.7)$$

where the second inequality results from the inexact criterion in (4.1) and the last inequality follows from (4.5). Then, we get

$$\|z^{k+1} - z^*\| \leq \frac{1 + \gamma\delta_k}{1 - \gamma\delta_k}\|z^k - z^*\| \leq \cdots \leq \prod_{i=1}^{k} \frac{1 + \gamma\delta_i}{1 - \gamma\delta_i}\|z^0 - z^*\|.$$ 

Using Lemma 4.2, the sequence $\{z^k\}$ is bounded. The first assertion is proved.

Now we prove the second assertion. Again, for an arbitrary solution point $z^*$ of (1.1), since $\{z^k\}$ is bounded and because of (4.5), there exists a positive scalar $R$ such that

$$\|z^k - z^*\| < R, \ \forall k \quad (4.8)$$
We thus have
\[
\|z^{k+1} - z^*\|^2 = \|z^{k+1} - z^* + (z^{k+1} - \tilde{z}^{k+1})\|^2
\]
\[
= \|z^{k+1} - z^*\|^2 + \|z^{k+1} - \tilde{z}^{k+1}\|^2 + 2(z^{k+1} - z^*, z^{k+1} - \tilde{z}^{k+1})
\]
\[
\leq \|z^{k+1} - z^*\|^2 + 2\|z^{k+1} - \tilde{z}^{k+1}\| \|z^{k+1} - \tilde{z}^{k+1}\| + \|z^{k+1} - \tilde{z}^{k+1}\|^2
\]
\[
\leq \|z^{k+1} - z^*\|^2 + 2R\gamma\delta_k \|z^k - z^{k+1}\| + \gamma^2\delta^2_k \|z^k - z^{k+1}\|^2
\]
\[
\leq \|z^k - z^*\|^2 - \gamma(2 - \gamma)\|z^k - z^*\|^2 + 4R^2\gamma\delta_k + 4R^2\gamma^2\delta^2_k,
\]
(4.10)
where the second inequality follows from (4.9) and (4.1); and the last inequality is because of (4.5) and (4.8).
Moreover, since \(\gamma \in (0,2)\), we have
\[
\|z^{k+1} - z^*\|^2 \leq \|z^k - z^*\|^2 + 4R^2\gamma\delta_k + 4R^2\gamma^2\delta^2_k,
\]
and
\[
\sum_k (4R^2\gamma\delta_k + 4R^2\gamma^2\delta^2_k) < +\infty.
\]
Now, using Lemma 4.3 with \(a_k := \|z^k - z^*\|^2\) and \(b_k := 4R^2\gamma\delta_k + 4R^2\gamma^2\delta^2_k\), we obtain
\[
\lim_{k \to \infty} \|z^k - z^*\| = A,
\]
(4.11)
where \(A\) is a positive scalar. On the other hand, recall that \(\{\delta_k\}\) is summable, so is \(\{\delta^2_k\}\). We thus have \(\sum_k \delta^2_k < \infty\). Then, it follows from (4.10) that
\[
\gamma(2 - \gamma)\|z^k - z^*\|^2 \leq \|z^k - z^*\|^2 - \|z^{k+1} - z^*\|^2 + 4R^2\gamma\delta_k + 4R^2\gamma^2\delta^2_k, \forall k.
\]
Then, we have \(\sum_k \|z^k - z^k\| < \infty\) and thus \(\lim_{k \to \infty} \|z^k - z^k\| = 0\). The proof is complete.

Theorem 4.5 shows that the accuracy of iterates generated by the inexact version (4.1) is iteratively increased, which essentially implies the convergence of the sequence of (4.1). We provide the rigorous proof in the following theorem.

THEOREM 4.5. (Global convergence) Let \(\{z^k\}\) be the sequence generated by the inexact version of the generalized PPA scheme (4.7). Then, it converges to a solution point of (4.7).

Proof. Since the sequence \(\{z^k\}\) is bounded, it has an accumulation point \(z^\infty\). Let \(\{z_{k_j}\}\) be the subsequence converging to \(z^\infty\). That is, \(\lim_{j \to \infty} \|z_{k_j} - z^\infty\| = 0\). Using (4.4) with \(k = k_j\), we have
\[
\|z^{k_j} - z^{k_j}\| \to 0, \text{ when } j \to \infty.
\]
(4.12)
Then, combining with \(\|z^{k_j} - z^\infty\| \to 0\), we get
\[
\|J_{c_{k_j}}^* (z^{k_j}) - z^\infty\| \to 0, \text{ when } j \to \infty.
\]
(4.13)
Also, it follows from (4.12) that
\[
\|z^{k_j} - J_{c_{k_j}}^* (z^{k_j})\| \to 0, \text{ when } j \to \infty.
\]
(4.14)
Note that
\[ c^{-1}_k(z^k - J_{c_k} T(z^k)) \in T(J_{c_k} T(z^k)). \]

Thus, using the monotonicity of \( T \), for any \( k \), we have
\[ (z - J_{c_k} T(z^k), w - c^{-1}_k(z^k - J_{c_k} T(z^k))) \geq 0, \quad \text{for all } z, w \text{ satisfying } w \in T(z). \]

Let \( k = k_j \) in the above inequality, take \( j \to \infty \), and combine it with (4.13) and (4.14). We thus have
\[ (z - z^\infty, w) \geq 0, \quad \text{for all } z, w \text{ satisfying } w \in T(z). \]

which, together with the monotonicity of \( T \), means that \( z^\infty \) is a solution point of (1.1).

Finally, since \( z^\infty \) is a solution point of (1.1), (4.1) can be written as \( \lim_{k \to \infty} \| z^k - z^\infty \| = A \) and indeed we have \( A = 0 \) because \( z^{k_j} \to z^\infty \). Thus, the sequence \( \{z^k\} \) converges to \( z^\infty \) which is a solution point of (1.1). The proof is complete.

4.2. Linear Convergence. In this subsection, under the assumption “\( T^{-1} \) is Lipschitz continuous at 0 with positive modulus”, we prove the linear convergence for the sequence \( \{z^k\} \) generated by (4.1). Let \( k \in \mathbb{Z} \) be the sequence generated by the inexact version of the generalized PPA scheme (4.1). Then, there exists an integer \( k \) such that
\[ \| z^k - z^* \| \leq \frac{a}{\sqrt{a^2 + c^2_k}} \| z^k - z^* \| \quad \forall k > k_1. \]

**Proof.** From Theorem 4.4 we know that \( \lim_{k \to 0} \| z^k - z^k \| = 0 \). Then, the conclusion follows immediately from the proof of Lemma 4.4.

**Theorem 4.7.** Assume \( T^{-1} \) is Lipschitz continuous at 0 with modulus \( a > 0 \) and the proximal parameter sequence \( \{ \gamma_k \} \) is bounded away from zero \( (c_k \geq \kappa > 0) \). Let \( \{z^k\} \) be the sequence generated by the inexact version of the generalized PPA scheme (4.7). Then, there exist an integer \( \hat{k} \) such that
\[ \| z^{k+1} - z^* \| \leq \theta_k \| z^k - z^* \| \quad \text{when } k > \hat{k}, \]

where \( z^* \) is a solution point of (1.1) and
\[ 0 < \theta_k := \frac{\sqrt{1 - (2\gamma - \gamma^2)(1 - \frac{a}{\sqrt{a^2 + c^2_k}})^2 + \gamma \delta_k}}{1 - \gamma \delta_k} < 1, \quad \text{when } k > \hat{k}. \]

That is, \( \{z^k\} \) converges linearly to \( z^* \).

**Proof.** Recall in Theorem 4.3 it is proved that the sequence \( \{z^k\} \) converges to a solution point \( z^* \) of (1.1). First, it is easy to see that there exists an integer \( k_1 \) such that
\[ \| z^k - z^k \| = \| z^k - z^* + z^* - z^k \| \]
\[ \geq \| z^k - z^* \| - \| z^* - z^k \| \]
\[ \geq (1 - \frac{a}{\sqrt{a^2 + c^2_k}}) \| z^k - z^* \|, \quad k > k_1, \]
where the last inequality follows from Lemma 4.6. Then, using (4.5), we have

\[
\|z^{k+1} - z^*\|^2 \leq \|z^k - z^*\|^2 - \gamma (2 - \gamma) \|z^k - z^*\|^2
\]

\[
\leq \|z^k - z^*\|^2 - \gamma (2 - \gamma) \left(1 - \frac{a}{\sqrt{a^2 + c_k^2}}\right)^2 \|z^k - z^*\|^2
\]

\[
= \left(1 - (2\gamma - \gamma^2) \left(1 - \frac{a}{\sqrt{a^2 + c_k^2}}\right)^2\right) \|z^k - z^*\|^2, \quad k > k_1. \tag{4.15}
\]

In addition, it follows from (4.7) that

\[
\|z^{k+1} - z^*\| \leq \gamma \delta_k (\|z^k - z^*\| + \|z^{k+1} - z^*\|) + \|z^k - z^*\|
\]

\[
\leq \gamma \delta_k (\|z^k - z^*\| + \|z^{k+1} - z^*\|) + \sqrt{\left(1 - (2\gamma - \gamma^2) \left(1 - \frac{a}{\sqrt{a^2 + c_k^2}}\right)^2\right)} \|z^k - z^*\|, \quad k > k_1,
\]

where the last inequality follows from (4.15). Accordingly, we have

\[
\|z^{k+1} - z^*\| \leq \frac{\sqrt{1 - (2\gamma - \gamma^2) \left(1 - \frac{a}{\sqrt{a^2 + c_k^2}}\right)^2} + \gamma \delta_k}{1 - \gamma \delta_k} \|z^k - z^*\| \quad \text{when } k > k_1.
\]

Note that \(\delta_k \to 0\) and \(c_k \geq \kappa > 0\). Then, there exists an integer \(\hat{k}\), without loss of generality, assuming \(\hat{k} > k_1\), such that

\[
\theta_k := \frac{\sqrt{1 - (2\gamma - \gamma^2) \left(1 - \frac{a}{\sqrt{a^2 + c_k^2}}\right)^2} + \gamma \delta_k}{1 - \gamma \delta_k} < 1, \quad \text{when } k > \hat{k}.
\]

Hence, \(\{z^k\}\) converges linearly to \(z^\ast\), a solution point of (1.1). The proof is complete.

**Remark 2.** Similarly as Section 3.2, it can be seen from the proofs of Lemma 4.6 and Theorem 4.7 that the linear convergence of the sequence \(\{z^k\}\) generated by (4.4) can be guaranteed under the less strengthened condition “\(T^{-1}\) is Lipschitz continuous at the iterates \(\{z^k\}\) with positive modulus when \(k\) is large enough”.

### 5. Further Study on the Assumption

Under the assumption “\(T^{-1}\) is Lipschitz continuous at 0 with positive modulus”, we have shown the linear convergence for both the exact version (1.7) and inexact version (4.1) of the generalized PPA. Recall that the generalized PPA (1.7) include the PPA (1.5) as a special case with \(\gamma = 1\) and our analysis extends the result in [33] for (1.5). In [8], the linear convergence of the generalized PPA (1.7) with \(c_k \equiv c\) has been studied under the assumption that \(T\) is \(\alpha\)-strongly monotone, which is stronger than “\(T^{-1}\) is Lipschitz continuous at 0 with positive modulus”.

In the following, we show that although we restrict our analysis under the assumption “\(T^{-1}\) is Lipschitz continuous at 0 with positive modulus”, theoretically this assumption can be further relaxed in order to ensure the linear convergence of (1.7) and (4.1). Note that the assertion in the following lemma does not depend on any specific iterative scheme.

**Theorem 5.1.** Let \(z^\ast\) be a solution point of (1.1) and the sequence \(\{c_k\}\) be bounded below away from 0, i.e., \(c_k \geq \kappa > 0\) for all \(k\). If \(T^{-1}\) is Lipschitz continuous at 0 with positive modulus \(a\), then \(J_{c_kT}\) defined in (1.4) is Lipschitz continuous at \(z^\ast\) and \(\sup_k \{L_k\} < 1\), where \(L_k\) is the Lipschitz constant of \(J_{c_kT}\).

**Proof.** It follows from Lemma 3.3 that there exists \(\tau > 0\) such that

\[
\|J_{c_kT}(z) - z^\ast\| \leq \frac{a}{\sqrt{a^2 + c_k^2}} \|z - z^\ast\|, \quad \text{when } \|c_k^{-1}(z - J_{c_kT}(z))\| \leq \tau. \tag{5.1}
\]

Recall that \(z^\ast \in T^{-1}(0)\), \(J_{c_kT}(z) \in T^{-1}(c_k^{-1}(z - J_{c_kT}(z)))\) and \(T^{-1}\) is Lipschitz continuous at 0, it yields that

\[
\|J_{c_kT}(z) - z^\ast\| \to 0, \quad \text{when } \|c_k^{-1}(z - J_{c_kT}(z))\| \to 0.
\]
Since $c_k \geq \kappa > 0$ for all $k$, we have $\|J_{\kappa T}(z) - z^*\| \leq \|J_{\kappa T}(z) - z^*\| \to 0$. Thus, $J_{\kappa T}$ is Lipschitz continuous at $z^*$ with the constant $L_k := \frac{a}{\sqrt{a^2 + \kappa^2}} \leq \frac{a}{\sqrt{a^2 + \kappa^2}} < 1$ for any $k$, according to [5.1]. The proof is complete. \[\blacksquare\]

**Theorem 5.2.** Suppose the sequence $\{c_k\}$ is both upper and lower bounded, that is, there exists constants $\varsigma$ and $\kappa$ such that $0 < \kappa \leq c_k \leq \varsigma$ for all $k$. Let $\{z^k\}$ be the sequence generated by the exact version of the generalized PPA (1.7) or the inexact version (4.1). If $J_{\kappa T}$ is Lipschitz continuous at $z^*$ with the constant $L_k$, and $L_G := \sup_k \{L_k\} < 1$, then

1. $T^{-1}$ is Lipschitz continuous at all the iterates $\{z^k\}$ with positive modulus when $k$ is sufficiently large.
2. $\{z^k\}$ converges linearly to a solution point of (1.1).

**Proof.** For a solution point of (1.1), $z^*$, we have $z^* = J_{\kappa T}(z^*)$. Recall the notation $\tilde{z}^k = J_{\kappa T}(z^k)$. Thus, it holds that

$$\|z^k - \tilde{z}^k\| = \|(z^k - z^*) - (J_{\kappa T}(z^k)) - J_{\kappa T}(z^*)\| \geq \|z^k - z^*\| - \|J_{\kappa T}(z^k) - J_{\kappa T}(z^*)\| \geq (1 - L_G)\|z^k - z^*\|, \quad (5.2)$$

which implies

$$\frac{1}{1 - L_G^2}\|z^k - z^*\|^2 \geq \|z^k - z^*\|^2 \geq \|z^k - z^*\|^2 + \|z^k - \tilde{z}^k\|^2. \quad (5.3)$$

Then, it follows from the above inequality and $0 < c_k \leq \varsigma$ that

$$\|z^k - z^*\|^2 \leq \frac{2L_G - L_G^2}{(1 - L_G)^2}\|z^k - \tilde{z}^k\|^2 \leq \frac{2L_G - L_G^2}{(1 - L_G)^2}\|c_k^{-1}(z^k - \tilde{z}^k)\|^2. \quad (5.4)$$

According to Theorems 3.1 and 4.4, for the sequence $\{z^k\}$ generated by either the exact version (1.7) or the inexact version (4.1), we have $\lim_{k \to \infty} \|z^k - \tilde{z}^k\| = 0$. Since $c_k \geq \kappa > 0$, there exists an integer $\hat{k}$ such that

$$\|c_k^{-1}(z^k - \tilde{z}^k)\| \leq \kappa^{-1}\|z^k - \tilde{z}^k\| \leq \tau \quad \text{when} \quad k > \hat{k},$$

where $\tau > 0$ is a given constant. Note the facts $\tilde{z}^k \in T^{-1}(c_k^{-1}(z^k - \tilde{z}^k))$ and $z^* \in T^{-1}(0)$. Consequently, it follows from (5.4) that $T^{-1}$ is Lipschitz continuous at all the iterates $\{z^k\}$ with modulus $\alpha := \frac{\sqrt{2L_G - L_G^2}}{1 - L_G}$ when $k$ is large enough.

Now, we prove (2). Indeed, as commented in Remarks 1 and 2, the linear convergence of the schemes (1.7) and (4.1) can be ensured since $T^{-1}$ is Lipschitz continuous at all the iterates $\{z^k\}$ with positive modulus when $k$ is sufficiently large and $\{c_k\}$ is below bounded. Thus, the assertion (2) is proved. The proof is complete. \[\blacksquare\]

Therefore, theoretically it is still possible to relax the condition “$T^{-1}$ is Lipschitz continuous at 0 with positive modulus” in [33] for studying the linear convergence rate of PPA-based schemes, though such an analysis is not covered in this paper.

So far, we have mentioned various conditions including the coercivity in [8], the assumption in [33] and the one in Theorems 5.1 and 5.2 to ensure the linear convergence of the schemes (1.7) and (4.1). In Figure 5.1 we show their relationships for the special case where $c_k \equiv c$ for all $k$, which is clearly an interesting choice for implementing the schemes (1.7) and (4.1).

**6. Discuss on the Superlinear Convergence.** In [33], under the assumption that “$T^{-1}$ is Lipschitz continuous at 0 with positive modulus”, it was shown that the special case of (4.1) with $\gamma = 1$ is superlinearly convergent if the proximal parameter $c_k \to \infty$. See Theorem 2 in [33]. One may ask if we can extend the same superlinear convergence result to (1.1) with a general $\gamma$ in (0, 2). In this section, we take a closer look at this issue and give a negative answer to this question. It is sufficient to just analyze the exact version (1.7) to answer this question.
We reconsider (3.9) as follows:

\[ \|z^{k+1} - z^*\|^2 \leq \|z^k - z^*\|^2 - \gamma (2 - \gamma) \|z^k - z^*\|^2 \]
\[ \leq \|z^k - z^*\|^2 - \gamma (2 - \gamma) (1 - \frac{a}{\sqrt{a^2 + c_k}})^2 \|z^k - z^*\|^2 \]
\[ = \left( 1 - (2\gamma - \gamma^2)(1 - \frac{a}{\sqrt{a^2 + c_k}})^2 \right) \|z^k - z^*\|^2. \] (6.1)

Note that \( 1 - \frac{a}{\sqrt{a^2 + c_k}} \to 1 \) when \( c_k \to \infty \). Moreover, we have

\[ 0 < 2\gamma - \gamma^2 = 1 - (\gamma - 1)^2 \leq 1 \]

when \( \gamma \in (0, 2) \) and \( \gamma \neq 1 \). Thus, if \( c_k \to \infty \), the coefficient in (6.1) goes to 0 only when \( \gamma = 1 \). This excludes the hope of establishing the superlinear convergence for the exact version of the generalized PPA (1.7) with \( \gamma \neq 1 \) even when \( c_k \to \infty \).

7. Application to ALM. Previously, we have discussed the linear convergence rates for the generalized PPA schemes (1.7) and (1.8) in the generic setting of (1.1) where \( T \) is an abstract maximal operator. In this and next sections, we specify our discussion to some special convex minimization models and discuss the linear convergence rates for two important algorithms which can be obtained by specifying the exact version of the generalized PPA (1.7). For succinctness, discussions for their inexact counterparts stemming from the inexact version (1.8) are omitted.

Let us first recall some known results and summarize them in the following two lemmas. The proof of the first lemma can be found in [32], and the proof of the second is trivial.

**Lemma 7.1.** Let \( f : \mathbb{R}^n \to \mathbb{R} \) be closed, proper and convex. Then, we have

i) If \( f \) is \( \mu_f \)-strongly convex, then \( f^* \) is differentiable and \( \nabla f^* \) is \( (1/\mu_f) \)-Lipschitz continuous.

ii) If \( f \) is differentiable and \( \nabla f \) is \( L_f \)-Lipschitz continuous, then \( f^* \) is \( 1/L_f \)-strongly convex.

**Lemma 7.2.** Let \( f : \mathbb{R}^n \to \mathbb{R} \) be closed, proper and strongly convex; let \( \partial f \) be the subdifferential of \( f \). Then, \( \partial f \) is coercive.

7.1. Preliminaries of ALM. First, we consider a canonical convex minimization model with linear constraints:

\[
\min \ f(x) \\
\text{s.t.} \ Ax = b. \] (7.1)
where \( f : \mathbb{R}^n \to (-\infty, \infty] \) is a closed and convex function, \( A \in \mathbb{R}^{m \times n} \) and \( b \in \mathbb{R}^m \).

For solving \((7.1)\), a benchmark is the augmented Lagrangian method (ALM) originally proposed in \([21, 31]\). Its iterative scheme reads as

\[
\begin{align*}
x^{k+1} &= \arg \min_x \{ f(x) - \langle p^k, Ax \rangle + \frac{1}{2}c_k \|Ax - b\|^2 \}, \\
p^{k+1} &= p^k - c_k(\mathcal{R}J(x^{k+1}) - b),
\end{align*}
\tag{7.2}
\]

where \( p^k \) is the Lagrange multiplier and \( c_k > 0 \) is the penalty parameter of the linear constraints. As analyzed in \([33]\), the dual problem of \((7.1)\) is

\[
\max_p -f^*(A^\top p) + \langle b, p \rangle,
\tag{7.3}
\]

where \(^*\) denotes the conjugate of a function, see \([33]\). Thus, solving \((7.3)\) is equivalent to

\[
0 \in \mathcal{S}_A(p) := \mathcal{A} \cdot \partial f^* \cdot (A^\top p) - b,
\tag{7.4}
\]

which is a specific application of the generic setting of \((1.1)\) with \( T = \mathcal{S}_A \). In \([33]\), it was precisely analyzed that the ALM scheme \((7.2)\) is an application of the PPA \((1.7)\) to the dual problem \((7.4)\). Also in \([33]\), it was mentioned that the mapping \( \mathcal{S}_A(p) \) defined in \((7.4)\) is maximal monotone.

### 7.2. A Generalized ALM

Following the analysis in \([33]\), it is easy to see that if we apply the generalized PPA scheme \((1.7)\) to \((7.4)\), we can obtain a generalized ALM scheme as follows

\[
\begin{align*}
x^{k+1} &= \arg \min_x \{ f(x) - \langle p^k, Ax \rangle + \frac{1}{2}c_k \|Ax - b\|^2 \}, \\
p^{k+1} &= p^k - \gamma c_k(\mathcal{R}J(x^{k+1}) - b),
\end{align*}
\tag{7.5}
\]

which differs from the original ALM \((7.2)\) in that there is a parameter \( \gamma \in (0, 2) \) for updating the Lagrange multiplier \( p^{k+1} \). The details are presented in the following theorem.

**Theorem 7.3.** The generalized ALM scheme \((7.5)\) is an application of the exact version of the generalized PPA \((7.4)\) to \((7.4)\).

**Proof.** First, the generalized ALM \((7.5)\) can be rewritten as

\[
\begin{align*}
x^{k+1} &= \arg \min_x \{ f(x) - \langle p^k, Ax \rangle + \frac{1}{2}c_k \|Ax - b\|^2 \}, \\
p^k &= p^k - c_k(\mathcal{R}J(x^{k+1}) - b), \\
p^{k+1} &= p^k - \gamma(p^k - \bar{p}^k).
\end{align*}
\tag{7.6}
\]

Since the first-order optimality condition of the \( x \)-subproblem in \((7.6)\) is

\[
A^\top (p^k - c_k(\mathcal{R}J(x^{k+1}) - b)) = \partial f(x^{k+1}),
\]

it follows from the second equation in \((7.6)\) that \( A^\top \bar{p}^k = \partial f(x^{k+1}). \) Then, we have

\[
\mathcal{R}J(x^{k+1}) - b \in \mathcal{A} \cdot \partial f^* \cdot (A^\top p^k) - b.
\]

We thus conclude that

\[
\bar{p}^k = p^k - c_k(\mathcal{R}J(x^{k+1}) - b) = p^k - c_k(\mathcal{A} \cdot \partial f^* \cdot (A^\top \bar{p}^k) - b) = p^k - c_k \mathcal{S}_A(p^k),
\]

which implies that \( \bar{p}^k = J_{c_k \mathcal{S}_A}(p^k) \). Then, it follows from the last equation in \((7.6)\) that

\[
p^{k+1} = p^k - \gamma(p^k - J_{c_k \mathcal{S}_A}(p^k)),
\]

meaning that the generalized ALM scheme \((7.5)\) is an application of \((1.7)\) to \((7.4)\). The proof is complete. \( \blacksquare \)
7.3. Linear Convergence of ALM schemes. Below, we show some conditions that can sufficiently ensure that the mapping $S_A^{-1} (S_A$ defined in (7.3)) is Lipschitz continuous at 0 with positive modulus, and thus guarantee the linear convergence rate of the generalized ALM (7.5) (also the original ALM (7.2) if taking $\gamma = 1$ in (7.3)).

**Theorem 7.4.** Let $S_A$ be defined in (7.4) and $\{p^k\}$ be the sequence generated by the generalized ALM scheme (7.3). For the model (7.1), if $f$ is convex and differentiable, $\nabla f$ is $L_f$-Lipschitz continuous, and the matrix $A$ is full row rank. Then, we have

1. The mapping $S_A$ is coercive.
2. The mapping $S_A^{-1}$ exists and it is Lipschitz continuous at 0 with positive modulus.
3. The sequence $\{p^k\}$ converges linearly to a zero point of $S_A$.

**Proof.** (1) Since $f$ is differentiable and $\nabla f$ is $L_f$-Lipschitz continuous, it follows from Property (ii) of Lemma 7.1 that $f^*$ is $\frac{1}{L_f}$-strongly convex. Then, it follows from Lemma 7.2 that $\partial f^*$ is $\frac{1}{L_f}$-coercive. For any $p,p' \in \mathbb{R}^m$; $w \in S_A(p)$ and $w' \in S_A(p')$; there exist $u \in \partial f^* \cdot (A^\top p)$ and $u' \in \partial f^* \cdot (A^\top p')$ such that $w = Au$ and $w' = Au'$. We thus have

$$
\langle w - w', p - p' \rangle = \langle A(u - u'), p - p' \rangle \\
= \langle u - u', A^\top (p - p') \rangle \\
\geq \frac{1}{L_f} \|A^\top (p - p')\|^2 \\
\geq \frac{1}{L_f} \lambda_{\min}(AA^\top) \|p - p'\|^2,
$$

where the first inequality is because of the $\frac{1}{L_f}$-coercivity of $\partial f^*$, and $\lambda_{\min}(AA^\top)$ is the minimal eigenvalue of $AA^\top$ with $\lambda_{\min}(AA^\top) > 0$ because $A$ is assumed to be full row rank. Thus, it follows from Definition 2.3 that the mapping $S_A$ is coercive.

(2) This is obvious based on Definitions 2.3 and 2.4. (3) This is just a conclusion of Theorem 7.5 with $T = S_A$ and the second assertion. The proof is complete. \qed

8. Application to ADMM. In this section, we consider another convex minimization model with a separable objective function:

$$
\min_x \; f(x) + g(Mx) \tag{8.1}
$$

where $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ and $g : \mathbb{R}^m \rightarrow (-\infty, \infty]$ are closed and convex functions, and the matrix $M \in \mathbb{R}^{m \times n}$. Again, we only focus on the specification of the exact version of the generalized PPA (1.7) and discuss how to ensure its linear convergence rate for the particular convex minimization context (8.1).

8.1. Preliminaries of ADMM. One particular case of (8.1) with a wide range of applications is where the functions $f$ and $g$ have their own properties and it is necessary to treat them individually in algorithmic design. For this purpose, we can reformulate (8.1) as

$$
\min_x \; f(x) + g(w) \\
\text{s.t.} \; Mx = w, \tag{8.2}
$$

where $w \in \mathbb{R}^m$ is an auxiliary variable. For solving (8.2), a benchmark is the ADMM scheme originally proposed in [10]. The ADMM scheme for (8.2) reads as

$$
\begin{align*}
\begin{cases}
x^{k+1} &= \arg \min_x \{f(x) + \langle p^k, Mx \rangle + \frac{1}{2\lambda} \|Mx - w^k\|^2\}, \\
w^{k+1} &= \arg \min_y \{g(w) - \langle p^k, w \rangle + \frac{1}{2\lambda} \|Mx^{k+1} - w\|^2\}, \\
p^{k+1} &= p^k - \lambda(Mx^{k+1} - w^{k+1}),
\end{cases}
\end{align*}
$$

(8.3)
where $p^k$ is the Lagrange multiplier and $\lambda > 0$ is a penalty parameter of the linear constraints in (8.2).

Next, we recall some results in [12, 15] to demonstrate that the ADMM is indeed a special case of the PPA (1.5). All the details can be found in [12]. First, the dual of (8.2) is

$$\max_{p \in \mathbb{R}^n} -(f^*(-M^Tp) + g^*(p))$$

(8.4)

where “$f^*$” and “$g^*$” denote the conjugate of the convex functions $f$ and $g$, respectively. Let

$$A := \partial[f^*(-M^T)]$$

and

$$B := \partial[g^*].$$

(8.5)

As shown in [32], both $A$ and $B$ defined in (8.5) are maximal monotone. Then, (8.4) can be written as

$$0 \in A(p) + B(p).$$

(8.6)

We use $J_{\lambda A}$ and $J_{\lambda B}$ to denote the resolvent operators of $A$ and $B$, respectively. Moreover, we denote

$$G_{\lambda, A, B} = J_{\lambda A}(2J_{\lambda B} - I) + (I - J_{\lambda B})$$

(8.7)

and

$$S_{\lambda, A, B} := G_{\lambda, A, B}^{-1} - I.$$  

(8.8)

As shown in [12], $S_{\lambda, A, B}$ is maximal monotone when $A$ and $B$ are both maximal monotone. Indeed, the definition of $S_{\lambda, A, B}$ can be expressed as

$$S_{\lambda, A, B} = \{(v + \lambda b, u - v) | (u, b) \in B, (v, a) \in A, v + \lambda a = u - \lambda b\},$$

(8.9)

where $A$ and $B$ are defined in (8.5). Moreover, let $p^*$ be an solution point of (8.6) and $z^*$ a solution point of

$$0 \in S_{\lambda, A, B}(z).$$

(8.10)

and let the sequence $\{z^k\}$ be iteratively represented by

$$z^{k+1} = J_{\lambda A}((2J_{\lambda B} - I)(z^k)) + (I - J_{\lambda B})(z^k).$$

(8.11)

Indeed, (8.11) is exactly the application of the Douglas-Rachford splitting method (DRSM) in [11, 23] to (8.6).

According to [12], we know some conclusions such as: (1) If $z^*$ is a solution point of (8.10), then we have $p^* := J_{\lambda B}(z^*)$ is a solution point of (8.6); and (2) If $p^*$ is a solution point of (8.6) and $(x^*, w^*)$ is a solution point of (8.2), then we have $x^* \in \partial f^*(-M^Tp^*)$ and $w^* \in \partial g^*(p^*)$.

Applying the scheme (1.7) with $T = S_{\lambda, A, B}$, we obtain the exact version of the generalized PPA scheme

$$z^{k+1} = z^k - \gamma(z^k - J_{\lambda, A, B}(z^k)) \text{ with } \gamma \in (0, 2).$$

(8.12)

Indeed, via (8.12), the following exact version of the generalized ADMM scheme proposed in [12] can be recovered

$$\begin{align*}
  x^{k+1} &= \arg \min_x \{ f(x) + \langle p^k, Mx \rangle + \frac{1}{2\lambda} \| Mx - w^k \|^2 \}, \\
  w^{k+1} &= \arg \min_w \{ g(w) - \langle p^k, w \rangle + \frac{1}{2\lambda} \| \gamma Mx^{k+1} + (1 - \gamma)w^k - w \|^2 \}, \\
  p^{k+1} &= p^k + \lambda(\gamma Mx^{k+1} + (1 - \gamma)w^k - w^{k+1}).
\end{align*}$$

(8.13)

In the following, we elucidate the relationship between the sequence $\{(x^k, w^k, p^k)\}$ generated by the generalized ADMM (8.13) and $\{z^k\}$ represented by (8.12); and demonstrate that the generalized ADMM (8.13) can be
written compactly as (8.12). The following lemma also clearly shows that the generalized ADMM (8.13) is an application of the generalized PPA (1.7) with $T = S_{\lambda,\mathcal{A},\mathcal{B}}$ and $c_k \equiv 1$ to (8.10).

**Lemma 8.1.** Let $\{(x^k, w^k, p^k)\}$ be generated by the generalized ADMM (8.13) and $\{z^k\}$ be represented by (8.12); the operator $S_{\lambda,\mathcal{A},\mathcal{B}}$ be defined in (8.3). Assume that the initial points satisfy with $z^0 = p^0 + \lambda w^0$ and $p^0 = J_{\mathcal{B}}(z^0)$. Then, it holds that $z^k = p^k + \lambda w^k$ and $p^k = J_{\mathcal{B}}(z^k)$ for all iterates.

**Proof.** The proof is mainly inspired by Theorem 8 in [12]. We provide the proof for completeness. First, we introduce an auxiliary sequence $\{q^k\}$ as
\[ q^k = p^k + \lambda(Mx^{k+1} - w^k). \] (8.14)
Assume that $z^k = p^k + \lambda w^k$, in the following we show that $z^{k+1} = p^{k+1} + \lambda w^{k+1}$. First, denote $z^k = J_{\mathcal{A},\mathcal{B}}(z^k)$. Since $J_{\mathcal{B}}(z^k) = p^k$, it yields that $z^k = p^k + \lambda w^k$ and $w^k \in \mathcal{B}(p^k)$. Then, we have
\[ J_{\mathcal{A}}(2J_{\mathcal{B}} - I)(z^k) = J_{\mathcal{A}}(2p^k - z^k) = J_{\mathcal{A}}(p^k - \lambda w^k) = J_{\mathcal{A}}(q^k - Mx^{k+1}) = q^k, \]
where the second equality follows from $z^k = p^k + \lambda w^k$, the third follows from the definition of $q^k$ (8.14), and the last comes from $-Mx^{k+1} \in \mathcal{A}(q^k)$. We thus have
\[ z^{k+1} = J_{\mathcal{A}}(2J_{\mathcal{B}} - I)(z^k) + (I - J_{\mathcal{B}})(z^k) = q^k + z^k - p^k = q^k + \lambda w^k. \]
Then, we have
\[ z^{k+1} = z^k - \gamma(z^k - z^k) = p^k + \lambda w^k - \gamma(p^k - q^k) = p^k + \lambda w^k - \gamma(-\lambda Mx^{k+1} + \lambda w^k) = p^k + \lambda(\gamma Mx^{k+1} + (1 - \gamma)w^k) = p^{k+1} + \lambda w^{k+1}, \]
where the second equality follows from the fact $z^k = p^k + \lambda w^k$, the third is because of the definition of $q^k$ (8.14) and the last comes from the update scheme of $p^{k+1}$ in (8.13). The proof is complete.

Finally, let us first present a lemma; its proof can be found in [12].

**Lemma 8.2.** The operator $G_{\lambda,\mathcal{A},\mathcal{B}}$ defined in (8.7) is firmly nonexpansive and it satisfies
\[ \langle G_{\lambda,\mathcal{A},\mathcal{B}}(z) - G_{\lambda,\mathcal{A},\mathcal{B}}(z'), z - z' \rangle \geq \|G_{\lambda,\mathcal{A},\mathcal{B}}(z) - G_{\lambda,\mathcal{A},\mathcal{B}}(z')\|^2 \]
\[ + \langle (I - J_{\mathcal{B}})(z) - (I - J_{\mathcal{B}})(z'), J_{\mathcal{B}}(z) - J_{\mathcal{B}}(z') \rangle; \forall z, z' \in H. \] (8.15)

**8.2. When Does the Assumption Hold?** Based on our previous analysis, it is clear that the linear convergence of the generalized ADMM (8.13) can be ensured by the assumption “The mapping $S_{\lambda,\mathcal{A},\mathcal{B}}$ (where $\mathcal{A}$ and $\mathcal{B}$ are specific model (8.1) and thus guarantee the linear convergence of the sequence $\{z^k\}$ represented by (8.12)).

**Theorem 8.3.** For the model (8.1), if the function $g$ is differentiable and strongly convex, and $\nabla g$ is Lipschitz continuous near a solution point, then we have
(1) The operator $\mathcal{B} := \nabla g^*$ is both coercive and Lipschitz continuous near the solution point.

(2) The mapping $S_{\lambda,A,B}^{-1}$ is Lipschitz continuous at the iterate $z^k := J_{S_{\lambda,A,B}}(z^k)$ with positive modulus when $k$ stays large enough.

(3) The sequence $\{z^k\}$ represented by (8.12) converges linearly to a solution point of (8.10).

Proof. (1) According to Lemma 7.1, we know that $g^*$ is differentiable. Thus, $\mathcal{B} := \nabla g^*$ is both coercive and Lipschitz continuous near the solution point of (8.4). According to Lemmas 7.1 and 7.2, Thus, the first conclusion is proved.

(2) Next, we show that the Lipschitz constant of the operator $G_{\lambda,A,B}$ is less than 1. Note that $\frac{1}{\lambda}(I - J_{\lambda AB})(z) \in \mathcal{B}(J_{\lambda AB}(z))$. Let us assume that the coercivity modulus of $\mathcal{B}$ is $\alpha$. That is,

$$
\langle (I - J_{\lambda AB})(z) - (I - J_{\lambda AB})(z'), J_{\lambda AB}(z) - J_{\lambda AB}(z') \rangle \geq \lambda \alpha \|J_{\lambda AB}(z) - J_{\lambda AB}(z')\|^2, \quad \forall z, z' \in H.
$$

Moreover, let us assume that Lipschitz continuous constant of $\mathcal{B}$ is $\beta$. Then, we have

$$
\|z - z'\|^2 = \|J_{\lambda AB}(z) - J_{\lambda AB}(z') + \lambda \mathcal{B}(J_{\lambda AB}(z)) - \lambda \mathcal{B}(J_{\lambda AB}(z'))\|^2 \leq (1 + \lambda \beta)^2 \|J_{\lambda AB}(z) - J_{\lambda AB}(z')\|^2, \quad \forall z, z' \in H.
$$

Combining these two inequalities, we get

$$
\langle (I - J_{\lambda AB})(z) - (I - J_{\lambda AB})(z'), J_{\lambda AB}(z) - J_{\lambda AB}(z') \rangle \geq \frac{\lambda \alpha}{(1 + \lambda \beta)^2} \|z - z'\|^2, \quad \forall z, z' \in H. \quad (8.16)
$$

Then, it follows from Lemma S.2 and S.10 that

$$
\|z - z'\|^2 \geq \langle G_{\lambda,A,B}(z) - G_{\lambda,A,B}(z'), z - z' \rangle \\
\geq \|G_{\lambda,A,B}(z) - G_{\lambda,A,B}(z')\|^2 \|z - z'\|^2 + \langle (I - J_{\lambda AB})(z) - (I - J_{\lambda AB})(z'), J_{\lambda AB}(z) - J_{\lambda AB}(z') \rangle \\
\geq \|G_{\lambda,A,B}(z) - G_{\lambda,A,B}(z')\|^2 \|z - z'\|^2 + \frac{\lambda \alpha}{(1 + \lambda \beta)^2} \|z - z'\|^2, \quad \forall z, z' \in H,
$$

where the first inequality follows from the non-expansiveness of the operator $G_{\lambda,A,B}$; the second inequality is because of (8.15) and the last inequality holds because of (8.16). Consequently, we prove that

$$
\|G_{\lambda,A,B}(z) - G_{\lambda,A,B}(z')\| \leq \sqrt{1 - \frac{\lambda \alpha}{(1 + \lambda \beta)^2}} \|z - z'\|, \quad \forall z, z' \in H.
$$

Recall the definitions of the coercivity and the Lipschitz continuity of $\mathcal{B}$. We have $\alpha \leq \beta$ and thus the above inequality means the fact that the Lipschitz continuity constant of the operator $G_{\lambda,A,B}$ is less than 1. Finally, it follows from Corollary S.2 with $T = S_{\lambda,A,B}$, $c_k \equiv 1$ and $G = G_{\lambda,A,B}$ that the mapping $S_{\lambda,A,B}^{-1}$ is Lipschitz continuous at the iterate $\{z^k\}$ with positive modulus when $k$ stays large enough, where $\tilde{z}^k := J_{S_{\lambda,A,B}}(z^k)$.

(3) Finally, the linear convergence of the sequence $\{z^k\}$ follows Remark 4 and assertion (2) immediately. The proof is complete.

Note that the linear convergence of $\{z^k\}$ represented by (8.12) can be easily specified as the linear convergence of the generalized ADMM scheme (8.13) in terms of the variables in (8.2) and its dual. We summarize the specifications in the following corollary and omit the proofs.

**Corollary 8.4.** When the sequence $\{z^k\}$ represented by (8.12) converges linearly to a solution point of (8.10), we have

(1) The sequence $\{p^k\}$ converges $R$-linearly to a solution point $p^*$ of the dual problem (8.4).

(2) The sequence $\{w^k\}$ converges $R$-linearly to a solution point $w^*$ of the primal problem (8.2).

(3) The sequence $\{Mx^k\}$ converges $R$-linearly to $Mx^*$, where $x^*$ is a solution point of the primal problem (8.1). Moreover, if $M$ is full column rank, then the sequence $\{x^k\}$ converges $R$-linearly to $x^*$, where $x^* = (M^TM)^{-1}M^T(Mx^*)$. 

Remark 3. Under one of the following conditions, we can also prove the conclusion “The mapping $S^{-1}_{\lambda, A, B}$ ($S_{\lambda, A, B}$ defined in (5.3)) exists and it is Lipschitz continuous at 0 with positive modulus”. We omit the proof because it is analogous to that of Theorem 8.3.

1. The matrix $M$ is full row rank, the function $f$ is strongly convex and $\nabla f$ is Lipschitz continuous near $x^*$, where $x^*$ is a solution point of (5.1).
2. The matrix $M$ is full row rank, the function $f$ is convex and $g$ is strongly convex near $M x^*$, and $\nabla f$ is Lipschitz continuous near $x^*$, where $x^*$ is a solution point of (5.1).
3. The matrix $M$ is full rank, the function $f$ is strongly convex near $x^*$ and $g$ is convex, and $\nabla g$ are Lipschitz continuous near $M x^*$, where $x^*$ is a solution point of (5.1).

Together with the condition in Theorem 8.3 these conditions coincide with the conditions in [9] (when $B = -I$ and $b = 0$ in the model (2) therein) to ensure the linear convergence of the generalized ADMM (8.13) for solving (8.1). In other words, the assumption “The mapping $S^{-1}_{\lambda, A, B}$ ($S_{\lambda, A, B}$ defined in (5.3)) exists and it is Lipschitz continuous at 0 with positive modulus” is weaker than these conditions.

Remark 4. In [8], the linear convergence of the generalized ADMM (8.13) for solving (8.1) is ensured under the following assumptions: (1). $M$ is full rank, $f$ is convex and differentiable, $\nabla f$ is Lipschitz continuous, and $g$ is strongly convex; (2). $f$ is strongly convex, $g$ is convex and differentiable, and $\nabla g$ is Lipschitz continuous. We here give some less strengthen conditions.

9. Conclusion. In this paper, we extend the condition in [33] that can ensure the linear convergence of the proximal point algorithm (PPA) to a generalized PPA scheme. Both the exact and inexact versions of the generalized PPA are studied, and their linear convergence rates are established under the same condition as the original PPA in [33]. We specifically consider two convex optimization models and study the linear convergence rates for generalized versions of the benchmark augmented Lagrangian method (ALM) and the alternating direction method of multipliers (ADMM), both are special cases of the proposed generalized PPA. Some concrete conditions are specified in the convex optimization contexts. It is interesting to find that the condition in [33] turns out to be still weaker than most of the existing conditions in the literature that were proposed to ensure the linear convergence for various specific forms of the PPA. This study provides a unified understanding of the linear convergence of a family of operator splitting methods which have found a board spectrum of applications in various areas. These methods include the mentioned ALM, ADMM, their generalized and inexact versions, the Douglas-Rachford splitting method, the Peaceman-Rachford splitting method, and their generalized versions.

REFERENCES