Linear convergence of the generalized PPA and several splitting methods for the composite inclusion problem

Li Shen* and Shaohua Pan†

August 16, 2015; Revised: April 6, 2016

Abstract
For the inclusion problem involving two maximal monotone operators, under the metric subregularity of the composite operator, we derive the linear convergence of the generalized proximal point algorithm and several splitting algorithms, which include the over-relaxed forward-backward splitting algorithm, the generalized Douglas-Rachford splitting algorithm and Davis’ three-operator splitting algorithm. To the best of our knowledge, this linear convergence condition is weaker than the existing ones that almost all require the strong monotonicity of the composite operator. Withal, we give some sufficient conditions to ensure the metric subregularity of the composite operator. At last, the preliminary numerical performances on some toy examples support the theoretical results.

Keywords: Linear convergence, metric subregularity, generalized PPA, over-relaxed FBS algorithm, generalized DRS algorithm, three-operator splitting algorithm

1 Introduction
Let $X$, $Y$ and $Z$ be the finite dimensional linear spaces endowed with the inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$. Given the maximal monotone operators $A, B: Z \rightrightarrows Z$ and $\theta$-cocoercive operator $C: Z \rightrightarrows Z$, we focus on the composite operator inclusion problem

$$0 \in F_z \quad \text{with} \quad F := A + B + C.$$ 

We are interested in the case that one of $A, B$ is single valued and Lipschitz continuous. Unless otherwise stated, we always assume that $F^{-1}(0) \neq \emptyset$ for problems (1).

The inclusion problem (1) has many applications such as the variational inequality problem [25], the problem of finding a common point of closed convex sets [3, 4] and covers

---

*Department of Mathematics, South China University of Technology, Tianhe District of Guangzhou, 510641, China (shen.li@mail.scut.edu.cn).
†Corresponding author. Department of Mathematics, South China University of Technology, Tianhe District of Guangzhou, China(shhpan@scut.edu.cn).
many classes of convex optimization problems. Specifically, we consider the following unconstrained nonsmooth composite convex minimization problem with the form that

$$\min f(x) + g(Dx) + h(x)$$

(2)

where $f : \mathbb{Z} \to [-\infty, +\infty]$ and $g : \mathbb{Y} \to [-\infty, +\infty]$ are low semicontinuous convex functions, $g : \mathbb{Z} \to (-\infty, +\infty)$ is continuous differentiable and gradient Lipschitz convex function and $D : \mathbb{Z} \to \mathbb{Y}$ is linear operators. Involved in the dual variable $y \in \mathbb{Y}$, it is easy to check that solving the optimization problem (2) is equivalent to solve the inclusion

$$(0, 0) \in (A + B + C)(x, y)$$

(3)

with $C(x, y) = (\nabla h(x), 0)$ being the cocoercive operator, $A(x, y) = (D^*y, -Dx)$ being the single valued, Lipschitz continuous operator and $B(x, y) = (\partial f(x), \partial g^*(y))$ being the maximal monotone operator where $\partial g^*$ is the conjugate function of $g$, which indicates that the composite convex problem (2) can be reformulated as a special case of (1) with the above specified operators $A, B, C$. Furthermore, when $g$ is also continuous differentiable and gradient Lipschitz, the problem (2) can be directly represented as the form of (1) with $A = \partial f$, $B = D^* \circ \nabla g \circ D$ and $C = \nabla h$. Additionally, the inclusion problem (1) is highly related to the linearly constrained two-block separable convex minimization problem which has many applications as listed in [6] and takes the following form that

$$\min_{u \in \mathbb{X}, v \in \mathbb{Y}} f(u) + g(v)$$

\text{s.t. } Gu + Hv = c.

In addition, it has a dual problem falling into framework of (2) with the following form

$$\min_{z \in \mathbb{Z}} \Phi(z) := f^*(G^*z) + g^*(H^*z) - \langle c, z \rangle,$$

which also can be reformulated as a special case of inclusion (1). Here $f : \mathbb{X} \to (-\infty, +\infty]$ and $g : \mathbb{Y} \to (-\infty, +\infty]$ are closed proper convex functions whose conjugate function are written as $f^*$ and $g^*$, respectively, $G : \mathbb{X} \to \mathbb{Z}$ and $H : \mathbb{Y} \to \mathbb{Z}$ are linear operators whose adjoint are $G^*$ and $H^*$, respectively, and $c \in \mathbb{Z}$ is a vector.

Let $J_{r\mathcal{T}}$ denote the resolvent of an operator $\mathcal{T}$ of index $\gamma > 0$, i.e., $J_{r\mathcal{T}} = (\mathcal{I} + r\mathcal{T})^{-1}$. If by chance the calculation of $J_{r\mathcal{T}}$ is easy, then the classical proximal point algorithm [36] or its over-relaxed version (the generalized PPA [19]) is a desirable solver for the inclusion problem (1). However, in practice the estimation of $J_{r\mathcal{T}}$ is usually much more difficult than that of $J_{rA}$, $J_{rB}$ and $J_{rC}$. Motivated by this, Davis and Yin [16] proposed a three operator splitting method for the inclusion (1) by using $J_{rA}$ and $J_{rB}$. Moreover, when one of $B$ and $C$ vanishes, several two operator splitting algorithms, including the forward-backward splitting (FBS) algorithm [34, 22, 10, 11], the Peaceman-Rachford splitting (PRS) algorithm [30, 35], and the Douglas-Rachford splitting (DRS) algorithm [30, 18], have been developed by using $J_{rA}$ or $J_{rB}$, $J_{rC}$. The convergence of these splitting algorithms have been well studied. In view of the strong assumption required
by the linear convergence, some authors recently focus on the iteration complexity of these algorithms [26, 14, 29]. By contrast, the study for their convergence rate is quite deficient. To the best of our knowledge, several existing linear convergence rate results (see [30, 10, 23, 15, 16]) all require the strong monotonicity of one of the operators $A, B$ and $C$ and single valued and Lipschitz continuous property of $B$. It is well known that the strong monotonicity assumption of $A, B$ or $C$ is too stringent. Recently, Liang et al. [29] and Bauschke et al. [5] establish the local linear convergence rate of substantial splitting algorithms based on the Krasnosel’skii-Mann fixed point iteration [27, 32] scheme with the metric subregularity assumption and the (bounded) linear regularity assumption on the fixed point operator at a point of its graph, respectively. The condition used by Liang et al. [29] is shown to be equivalent to the metric subregularity condition of $F$ at a point $(x^*, 0) \in \text{gph}F$ for the generalized PPA algorithm and over-relaxed FBS algorithm according to lemma 3.3 in the following.

The main contribution of this paper is to derive the linear convergence rate of the above several splitting algorithms and Davis’ splitting algorithm [16] under the metric subregularity of the operator $F$ at a point $(x, 0)$ of its graph. In addition, as will be shown in the section 3, the metric subregularity of an operator at a point of its graph is weaker than some existing regularization conditions such as the strongly monotone and the projective type error bound [40] on the fixed point operator.

2 Preliminaries

This section recalls some necessary concepts and lemmas that will be used in the subsequent analysis. Firstly, we introduce some concepts associated to an operator $T : X \rightrightarrows Y$, for which we make no difference from its graph $\text{gph} \, T := \{(x, y) \in X \times Y \mid y \in T(x)\}$. The domain and range of an operator $T : X \rightrightarrows Y$ are respectively defined as

$$\text{dom} \, T = \{x \in X \mid T \, x \neq \emptyset\} \quad \text{and} \quad \text{ran} \, T = \{y \in Y \mid \exists x \in X \text{ such that } (x, y) \in T\}.$$ 

The inverse of $T$ is given by $T^{-1} := \{(y, x) \in Y \times X \mid (x, y) \in T\}$. For any $c \in \mathbb{R}$, we let $cT = \{(x, cy) \mid (x, y) \in T\}$, and if $G$ and $H$ are any operators from $X$ to $Y$, we let

$$\mathcal{G} + \mathcal{H} = \{(x, y + z) \mid (x, y) \in \mathcal{G}, (x, z) \in \mathcal{H}\}.$$ 

An operator $T : X \rightrightarrows X$ is said to be firmly nonexpansive if $\langle x - y, u - v \rangle \geq \|u - v\|^2$ for $(x, u), (y, v) \in T$. Moreover, $T$ is said to be nonexpansive if $\|Tu - Tv\| \leq \|x - y\|$. By [2, Proposition 5.14] we have the following result for a nonexpansive operator.

**Lemma 2.1** Let $T : X \rightrightarrows X$ be a nonexpansive operator with $\text{Fix} \, T \neq \emptyset$, and $\{\mu_k\}$ be a sequence in $[0, 1]$ satisfying $\sum_{k=0}^{\infty} \mu_k(1 - \mu_k) = +\infty$. Let $\{w^k\}$ be generated by

$$w^{k+1} = w^k + \mu_k(Tw^k - w^k) \quad \text{with} \quad w^0 \in X.$$ 

Then, the sequences $\{w^k\}$ and $\{Tw^k\}$ converge to a point in $\text{Fix} \, T$ and for any $w \in \text{Fix} \, T$,

$$\|w^{k+1} - w\|^2 \leq \|w^k - w\|^2 - \mu_k(1 - \mu_k)\|Tw^k - w^k\|^2 \quad \text{for all} \quad k \in \mathbb{N}.$$ 


Next we recall from the monograph [2] the concept of the $\alpha$-averaged operator.

**Definition 2.1** Let $D$ be a nonempty subset of $X$, $T: D \to X$ be a nonexpansive operator, and $\alpha \in (0, 1)$ be a constant. Then the operator $T$ is said to be $\alpha$-averaged if there exists a nonexpansive operator $R: D \to X$ such that $T = (1 - \alpha)\text{Id} + \alpha R$.

By [2, Prop. 5.15] we have the following result for an $\alpha$-averaged operator $T: X \to X$.

**Lemma 2.2** Let $T: X \to X$ be an $\alpha$-averaged operator of $\alpha \in (0, 1)$ with $\text{Fix } T \neq \emptyset$, and $\{\mu_k\} \subseteq [0, \frac{1}{\alpha}]$ be a sequence satisfying $\sum_{k=0}^{\infty} \mu_k (\frac{1}{\alpha} - \mu_k) = +\infty$. Let $\{w^k\}$ be generated by (4). Then, $\{w^k\}$ and $\{Tw^k\}$ converge to a point in $\text{Fix } T$ and for any $w \in \text{Fix } T$,

$$\|w^{k+1} - w\|^2 \leq \|w^k - w\|^2 - \mu_k (\alpha^{-1} - \mu_k) \|Tw^k - w^k\|^2$$

for all $k \in \mathbb{N}$.

The following definition is about the metric subregularity [17] of $T$ at $(\overline{x}, \overline{y}) \in gph T$.

**Definition 2.2** An operator $T: X \rightrightarrows Y$ is metrically subregular at $(\overline{x}, \overline{y}) \in gph T$ with constant $\kappa > 0$ if there exists a neighborhood $U$ of $\overline{x}$ such that

$$\text{dist}(x, T^{-1}(\overline{y})) \leq \kappa \text{dist}(\overline{y}, Tx)$$

for all $x \in U$.

## 3 Linear convergence of several splitting algorithms

In the first four subsections, we derive the linear convergence rate of the generalized PPA with both $B$ and $C$ vanishing, the over-relaxed FBS and the generalized DRS algorithm for problem (1) with the corresponding $B$ and $C$ vanishing, and Davis-Yin’s three-operator splitting algorithm for problem (1) under assumption that $F$ is metric subregular at a point $(z^*, 0) \in gph F$. In the last subsection, we will discuss the equivalence on metric subregularity condition between the inclusion operator $F$ in (1) and fixed point operator in [29] for generalized PPA and over-relaxed FBS algorithm. Some sufficient conditions are also given to ensure the metric subregular of $F$ at $(z^*, 0) \in gph F$.

### 3.1 Linear convergence of generalized PPA

The generalized PPA [19] for problem (1) with $B$ and $C$ vanishing takes the iteration step

$$z^{k+1} = z^k + \lambda_k (J_{y^k}z^k - z^k) \quad \text{for some } \gamma > 0. \quad (5)$$

When $\lambda_k \equiv 1$, the iteration (5) reduces to that of the classical PPA [31, 36]. The linear convergence rate of the classical PPA is first established in [36] under the Lipschitz continuity of $F^{-1}$ at 0. Later, Artacho et al. [1] and Leventhal [28] derived the linear convergence rate of the classical PPA under the metric regularity of $F$ at a point $(\overline{x}, 0) \in gph F$ and the metric subregularity of $F$ at a point $(\overline{x}, 0) \in gph F$, respectively. Besides, the latter is weaker than the metric regularity of $F$ at a point $(\overline{x}, 0) \in gph F$ and the Lipschitz continuity of $F^{-1}$ near 0 in the sense of [36]. We next establish the linear convergence rate of the generalized PPA under the same assumption as in [28].
**Theorem 3.1** Assume that the operator $\mathcal{F}$ is maximally monotone. Let $\{z^k\}$ be given by the generalized PPA with $\{\lambda_k\} \subseteq [0, 2]$ satisfying $\sum_{k=0}^{\infty} \lambda_k (2 - \lambda_k) = +\infty$. Then,

(a) the sequence $\{z^k\}$ converges to a point $z^* \in \mathcal{F}^{-1}(0)$, and moreover, it holds that
$$\|z^{k+1} - z\|^2 \leq \|z^k - z\|^2 - \lambda_k (2 - \lambda_k) \|J_{\gamma_{F}} z^k - z\|^2 \quad \forall z \in \mathcal{F}^{-1}(0) \text{ and } \forall k \in \mathbb{N}. \quad (6)$$

(b) If in addition $\mathcal{F}$ is metrically subregular at $(z^*, 0) \in \text{gph} \mathcal{F}$ with constant $\kappa > 0$, then there exists $\overline{k} \in \mathbb{N}$ such that
$$\text{dist}(z^{k+1}, \mathcal{F}^{-1}(0)) \leq \sqrt{1 - \frac{\lambda_k (2 - \lambda_k) \gamma^2}{(\gamma + \kappa)^2}} \text{dist}(z^k, \mathcal{F}^{-1}(0)) \quad \text{for } k \geq \overline{k}.\quad \text{Proof:}$$

(a) Since $\mathcal{F}$ is maximally monotone, $J_{\gamma_{F}}$ is firmly nonexpansive (see [37, Theorem 12.12]), and so is $(1/2)$-averaged by [2, Remark 4.24]. The result follows by Lemma 2.2.

(b) Let $x^k = J_{\gamma_{F}} z^k$. Notice that $z^* = J_{\gamma_{F}} z^*$. We have $\|x^k - z^*\| \leq \|z^k - z^*\|$ by the nonexpansiveness of $J_{\gamma_{F}}$, which by part (a) implies that $x^k \to z^*$. Since $\mathcal{F}$ is metrically subregular at $(z^*, 0) \in \text{gph} \mathcal{F}$ with constant $\kappa > 0$, there exists $\overline{k} \in \mathbb{N}$ such that
$$\text{dist}(x^k, \mathcal{F}^{-1}(0)) \leq \kappa \text{dist}(0, \mathcal{F}(x^k)) \leq \frac{\kappa}{\gamma} \|z^k - x^k\| \quad \text{for all } k \geq \overline{k},$$
where the last inequality is due to $\gamma^{-1}(z^k - x^k) \in \mathcal{F} x^k$ implied by $x^k = J_{\gamma_{F}} z^k$. Then,
$$\text{dist}(z^k, \mathcal{F}^{-1}(0)) \leq \text{dist}(x^k, \mathcal{F}^{-1}(0)) + \|z^k - x^k\| \leq (1 + \frac{\kappa}{\gamma}) \|z^k - x^k\| \quad \text{for } k \geq \overline{k}.$$

Combining the last inequality with inequality (6) yields that for all $k \geq \overline{k}$,
$$\text{dist}(z^{k+1}, \mathcal{F}^{-1}(0))^2 \leq \|z^{k+1} - \Pi_{\mathcal{F}^{-1}(0)}(z^k)\|^2$$
$$\leq \|z^k - \Pi_{\mathcal{F}^{-1}(0)}(z^k)\|^2 - \lambda_k (2 - \lambda_k) \|z^k - x^k\|^2$$
$$= \text{dist}(z^k, \mathcal{F}^{-1}(0))^2 - \lambda_k (2 - \lambda_k) \|z^k - x^k\|^2$$
$$\leq \left[ 1 - \frac{\lambda_k (2 - \lambda_k) \gamma^2}{(\gamma + \kappa)^2} \right] \text{dist}(z^k, \mathcal{F}^{-1}(0))^2,$$
where $\Pi_{\mathcal{F}^{-1}(0)}(\cdot)$ is the projection operator onto $\mathcal{F}^{-1}(0)$. The proof is completed. \(\square\)

It is worthwhile to point out that Corman and Yuan [13] derived the linear convergence rate of the generalized PPA under the strong monotonicity of $\mathcal{F}$, which is more stringent than the metric subregularity of $\mathcal{F}$ at $(\mathcal{F}, \mathcal{F}) \in \text{gph} \mathcal{F}$. We also observe that Liang et al. [29] establish the similar local linear convergence rate under the metric subregularity of $(\mathcal{I} - J_{\gamma_{F}})$ at $(z^*, 0) \in \text{gph} (\mathcal{I} - J_{\gamma_{F}})$. By the Lemma 3.3 in subsection 3.5, we will show that this condition is equivalent to the metric subregularity of $\mathcal{F}$ at $(z^*, 0) \in \text{gph} \mathcal{F}$. Very recently, Tao and Yuan [39] also established the linear convergence rate of the generalized PPA under the Lipschitz continuity of $\mathcal{F}^{-1}$ near 0 in the sense of [36], which is stronger than the metric subregularity of $\mathcal{F}$ at $(z, 0) \in \text{gph} \mathcal{F}$. 

5
3.2 Over-relaxed forward-backward splitting algorithm

The over-relaxed FBS algorithm for (1) with $\mathcal{B}$ vanishing takes the iteration steps:

$$z^{k+1} = z^k + \lambda_k (\mathcal{J}_{\gamma\mathcal{A}}(I - \gamma \mathcal{C}) z^k - z^k), \tag{7}$$

which takes the form of equation (4) with $\mathcal{T} := \mathcal{J}_{\gamma\mathcal{A}}(I - \gamma \mathcal{C})$ where $\gamma > 0$ is the stepsize. When $\lambda_k \equiv 1$, the iteration (7) reduces to the FBS algorithm studied in [22, 34, 11].

For the sequence $\{z^k\}$ generated by (7), we have the following linear convergence result under the metric subregularity of $\mathcal{F} := \mathcal{A} + \mathcal{C}$ at a point $(z^*, 0) \in \text{gph} \mathcal{F}$.

**Theorem 3.2** Let $\mathcal{B}$ be $\beta$-cocoercive and $\{z^k\}$ be the sequence generated by (7) with $\gamma \in (0, 2\beta)$ and $\lambda_k \in [0, \delta)$ such that $\sum_{k=0}^{\infty} \lambda_k (\delta - \lambda_k) = +\infty$, where $\delta = \min(1, \frac{\beta}{\gamma}) + \frac{1}{2}$. Then,

(a) the sequences $\{z^k\}$ and $\{T z^k\}$ converge to a point $z^* \in \text{Fix} \mathcal{T} = \mathcal{F}^{-1}(0)$, and

$$\|z^{k+1} - z\| \leq \|z^k - z\| - \lambda_k (\delta - \lambda_k) \|T z^k - z^k\| \quad \forall z \in \mathcal{F}^{-1}(0) \text{ and } \forall k \in \mathbb{N}. \tag{8}$$

(b) If in addition $\mathcal{F}$ is metrically subregular at the point $(z^*, 0) \in \text{gph} \mathcal{F}$ with constant $\kappa > 0$, then there exists $\overline{k} \in \mathbb{N}$ such that for all $k \geq \overline{k}$,

$$\text{dist}(z^{k+1}, \mathcal{F}^{-1}(0)) \leq \sqrt{1 - \frac{\gamma^2 \lambda_k (\delta - \lambda_k)}{\gamma^2 + \kappa^2}} \text{dist}(z^k, \mathcal{F}^{-1}(0)).$$

**Proof:** (a) From the proof of [2, Theorem 25.8], it follows that $\mathcal{T}$ is $(1/\delta)$-averaged. Thus, the result of part (a) follows directly from Lemma 2.2.

(b) Let $x^k = T z^k$. From the definition of $\mathcal{T}$ and the single-valuedness of $\mathcal{C}$, we have $z^k - \gamma C z^k \in x^k + \gamma A x^k$. Hence, $\gamma^{-1}(z^k - x^k) + C x^k - C z^k \in A x^k + C x^k = \mathcal{F} x^k$. In addition, from part (a) it follows that $x^k \to z^*$. Now by the metric subregularity of $\mathcal{F}$ at $(z^*, 0) \in \text{gph} \mathcal{F}$, there exists $\overline{k} \in \mathbb{N}$ such that for $k \geq \overline{k}$,

$$\text{dist}(x^k, \mathcal{F}^{-1}(0)) \leq \kappa \text{dist}(0, \mathcal{F}(x^k)) \leq \kappa \|\gamma^{-1}(z^k - x^k) + C x^k - C z^k\|$$

$$\leq \kappa \sqrt{\gamma^{-2} \|z^k - x^k\|^2 + (1 - \frac{2\beta}{\gamma}) \|C x^k - C z^k\|^2} \leq (\kappa/\gamma) \|x^k - z^k\|$$

where the third inequality is using the cocoercivity of $\mathcal{C}$, and the last one is due to $\gamma \in (0, 2\beta)$. The last inequality immediately implies that for $k \geq \overline{k}$,

$$\text{dist}(z^k, \mathcal{F}^{-1}(0)) \leq \text{dist}(x^k, \mathcal{F}^{-1}(0)) + \|z^k - x^k\| \leq (1 + \kappa/\gamma) \|x^k - z^k\|. \tag{9}$$

Combining inequality (9) with inequality (8), we obtain that for all $k \geq \overline{k}$,

$$\text{dist}(z^{k+1}, \mathcal{F}^{-1}(0))^2 \leq \|z^{k+1} - \Pi_{\mathcal{F}^{-1}(0)}(z^k)\|^2$$

$$\leq \|z^k - \Pi_{\mathcal{F}^{-1}(0)}(z^k)\|^2 - \lambda_k (\delta - \lambda_k) \|z^k - x^k\|^2$$

$$\leq \left[1 - \frac{\gamma^2 \lambda_k (\delta - \lambda_k)}{\gamma^2 + \kappa^2}\right] \text{dist}(z^k, \mathcal{F}^{-1}(0))^2.$$
This implies the desired result of part (b). The proof is then completed. □

By Theorem 3.2, one may see that the linear convergence rate coefficient is smallest when $\lambda_k = \frac{1}{\delta}$. Recall that Chen and Rockafellar [11] derived the linear convergence of the FBS algorithm under the strong monotonicity of $F := A + C$, which implies the single-valuedness and Lipschitz continuity of $F^{-1}$, and then the metric subregularity of $F$ at $(z^*, 0) \in \text{gph} F$. Notice that the linear convergence is also derived in the work of Liang et al. [29] under the metric suregular of $(I - T)$ at the point $(z^*, 0) \in \text{gph}(I - T)$. In the following lemma 3.3 in subsection 3.5, we show that the metric subregularity of $(I - T)$ is equivalent to the one of $F$ at $(z^*, 0) \in \text{gph}(F)$.

### 3.3 Generalized Douglas-Rachford splitting algorithm

Given $\gamma > 0$, the generalized DRS method for (1) with $C$ vanishing takes the iterations:

$$
\begin{align*}
&\left\{ \begin{array}{l}
z^k = J_{\gamma B} x^k, \\
y^k = J_{\gamma A}(2z^k - x^k), \\
x^{k+1} = x^k + \lambda_k (y^k - z^k),
\end{array} \right.
\end{align*}
(10)
$$

which can be rewritten as the form of (4) with $T := \frac{1}{2}((2J_{\gamma A} - I)(2J_{\gamma B} - I) + I)$, i.e.,

$$
x^{k+1} = x^k + \lambda_k (Tx^k - x^k).
(11)
$$

When $\lambda_k \equiv 1$, equation (10) gives the DRS method [30, 18], and when $\lambda_k \equiv 2$ it gives the PRS method [30, 35]. Before stating the linear convergence rate of the generalized DRS method, we establish the relationship between the set $F^{-1}(0)$ and the set $\text{Fix} T$.

**Lemma 3.1** The set $F^{-1}(0)$ has the following relations with the fixed-point set $\text{Fix} T$:

(a) $F^{-1}(0) = J_{\gamma B}(\text{Fix} T)$. If $B$ is single-valued, $\text{Fix} T = (I + \gamma B)(F^{-1}(0))$.

(b) $\text{Fix} T \subseteq (I - \gamma A)(F^{-1}(0))$. If $A$ is single-valued, then $\text{Fix} T = (I - \gamma A)(F^{-1}(0))$.

The proof of the above lemma is given in the Appendix. Next, we show the DRS method converges linearly under the metric subregularity of $F$ at $(z^*, 0) \in \text{gph} F$.

**Theorem 3.3** Let $\{x^k\}, \{y^k\}$ and $\{z^k\}$ be generated by the generalized DRS method with $\{\lambda_k\} \subseteq [0, 2]$ such that $\sum_k^\infty \lambda_k(2 - \lambda_k) = +\infty$. Then, the following statements hold.

(a) $\{y^k\}$ and $\{z^k\}$ converge to $z^* \in F^{-1}(0)$, and $\{x^k\}$ converges to $x^* \in \text{Fix} T$ and

$$
\|x^{k+1} - x\|^2 \leq \|x^k - x\|^2 - \lambda_k(2 - \lambda_k)\|Tx^k - x^k\|^2 \quad \forall x \in \text{Fix} T \text{ and } \forall k \in \mathbb{N}.
(12)
$$

(b) If $A$ is single-valued and Lipschitz continuous with modulus $\frac{\beta}{2}$, and $F$ is metrically subregular at $(z^*, 0) \in \text{gph} F$ with constant $\kappa > 0$, then there exists $\bar{k} \in \mathbb{N}$ such that

$$
\text{dist}(x^{k+1}, \text{Fix} T) \leq \sqrt{1 - \frac{\lambda_k(2 - \lambda_k)}{2 + \gamma^2 \beta^2 - 2(1 + \kappa(\gamma^{-1} + \beta^{-1}))^2}} \text{dist}(x^k, \text{Fix} T), \ \forall k \geq \bar{k}.
$$

7
(c) If B is single-valued and Lipschitz continuous with modulus \( \frac{1}{\beta} \), and \( F \) is metric subregular at \((z^*, 0)\) \( \in \text{gph} \ F \) with constant \( \kappa > 0 \), then there exists \( T \in \mathbb{N} \) such that
\[
\text{dist}(x^{k+1}, \text{Fix} \ T) \leq \sqrt{1 - \frac{\beta^2 \lambda_k (2 - \lambda_k)}{(\gamma + \beta)^2 (1 + \kappa \sqrt{\gamma - 2 + \beta^2})^2}} \text{dist}(x^k, \text{Fix} \ T), \quad \forall k \geq T.
\]

**Proof:** (a) It is easy to check that \( T \) is nonexpansive. The result follows directly from Lemma 2.1 and the first equality of Lemma 3.1(a).

(b) By the iteration steps (10), it follows that \( \gamma^{-1}(x^k - z^k) \in B z^k \), \( \gamma^{-1}(z^k - x^k + z^k - y^k) = A y^k \) and \( \gamma^{-1}(z^k - y^k) \in A y^k + B z^k \). This, along with the single-valuedness of \( A \), means that
\[
\gamma^{-1}(z^k - y^k) + A z^k - A y^k \in A z^k + B z^k = F z^k.
\]

By part (a), using the metric subregularity of \( F \) at the point \((z^*, 0)\) and the Lipschitz continuity of \( A \), it follows that there exists \( T \in \mathbb{N} \) such that for all \( k \geq T \),
\[
\text{dist}(z^k, F^{-1}(0)) \leq \kappa \text{dist}(0, F z^k) \leq \kappa \|\gamma^{-1}(z^k - y^k) + A z^k - A y^k\| \\
\leq \kappa (\gamma^{-1} + \beta^{-1}) \|z^k - y^k\|,
\]
which further implies that
\[
\text{dist}(y^k, F^{-1}(0)) \leq \text{dist}(z^k, F^{-1}(0)) + \|z^k - y^k\| \leq \left[1 + \kappa (\gamma^{-1} + \beta^{-1})\right]\|y^k - z^k\|. \quad (13)
\]
Let \( \pi^k = \Pi_{F^{-1}(0)}(y^k) - \gamma A(\Pi_{F^{-1}(0)}(y^k)) \). From Lemma 3.1(b), \( \pi^k \in \text{Fix} \ T \). In addition, notice that \( 2 z^k - x^k = y^k + \gamma A y^k \) by the second equality in (10). Thus, for all \( k \geq T \),
\[
\text{dist}(x^k, \text{Fix} \ T) \leq \|x^k - \Pi_{F^{-1}(0)}(y^k) + \gamma A(\Pi_{F^{-1}(0)}(y^k))\| \\
= \|2 z^k - y^k - \gamma A y^k - \Pi_{F^{-1}(0)}(y^k) + \gamma A(\Pi_{F^{-1}(0)}(y^k))\| \\
\leq \|2 z^k - 2 y^k\| + \|y^k - \Pi_{F^{-1}(0)}(y^k)\| - \gamma A y^k + \gamma A(\Pi_{F^{-1}(0)}(y^k))\| \\
\leq 2 \|z^k - y^k\| + \sqrt{1 + \gamma^2 \beta^{-2}} \text{dist}(y^k, F^{-1}(0)) \\
\leq \left[2 + \sqrt{1 + \gamma^2 \beta^{-2}}(1 + \kappa (\gamma^{-1} + \beta^{-1}))\right] \|z^k - y^k\|,
\]
where the third inequality is using the Lipschitz continuity and monotonicity of \( A \), and the last one is due to (13). In addition, from equation (11) and the last equality of (10), we have \( T x^k - x^k = \frac{(z^k - x^k)}{\lambda_k} = (y^k - z^k) \), which together with (12) implies that
\[
\text{dist}(x^{k+1}, \text{Fix} \ T)^2 \leq \|x^{k+1} - \Pi_{F \text{Fix} \ T}(x^k)\|^2 \\
\leq \|x^k - \Pi_{F \text{Fix} \ T}(x^k)\|^2 - \lambda_k (2 - \lambda_k) \|T x^k - x^k\|^2 \\
= \text{dist}(x^k, \text{Fix} \ T)^2 - \lambda_k (2 - \lambda_k) \|T x^k - x^k\|^2 \\
= \text{dist}(x^k, \text{Fix} \ T)^2 - \lambda_k (2 - \lambda_k) \|y^k - z^k\|^2. \quad (14)
\]
The desired result of part (b) follows directly from the last two inequalities.
(c) Notice that $\mathcal{B}$ is single-valued and Lipschitzian. Thanks to part (a) and (b), we have
\[ \gamma^{-1}(z^k - y^k) + By^k - Bz^k \in A y^k + B y^k = F y^k. \]
From the metric subregularity of $\mathcal{F}$ at the point $(z^*, 0)$, there exists $k \geq \overline{k}$ such that
\[ \text{dist}(y^k, \mathcal{F}^{-1}(0)) \leq \kappa \text{dist}(0, \mathcal{F} y^k) \leq \kappa \| \gamma^{-1}(z^k - y^k) + By^k - Bz^k \|
\leq \kappa \sqrt{\gamma^{-2} + \beta^{-2}} \| z^k - y^k \| \quad \text{for } k \geq \overline{k}, \]
where the last inequality is using the monotonicity of $\mathcal{B}$. Consequently, for all $k \geq \overline{k}$,
\[ \text{dist}(z^k, \mathcal{F}^{-1}(0)) \leq \text{dist}(y^k, \mathcal{F}^{-1}(0)) + \| z^k - y^k \| \leq [1 + \kappa \sqrt{\gamma^{-2} + \beta^{-2}}] \| y^k - z^k \|. \]
Let $z^k = B(\Pi_{\mathcal{F}^{-1}(0)}(z^k))$. By Lemma 3.1(a), clearly, $\Pi_{\mathcal{F}^{-1}(0)}(z^k) + \gamma z^k \in \text{Fix } \mathcal{T}$. Moreover, using the Lipschitzian property of $\mathcal{B}$, we have $\| z^k - B z^k \| \leq \beta^{-1} \| \Pi_{\mathcal{F}^{-1}(0)}(z^k) - z^k \|$. In addition, according to $x^k = z^k + \gamma B z^k \in \mathcal{J}^{-1}_{\gamma \mathcal{B}}(z^k)$, we further obtain
\[ \text{dist}(x^k, \text{Fix } \mathcal{T}) \leq \| x^k - \Pi_{\mathcal{F}^{-1}(0)}(z^k) - \gamma z^k \|
\leq \| z^k - \Pi_{\mathcal{F}^{-1}(0)}(z^k) \| + \| \gamma z^k - x^k \|
\leq (1 + \gamma \beta^{-1}) \| z^k - \Pi_{\mathcal{F}^{-1}(0)}(z^k) \|
\leq (1 + \gamma \beta^{-1}) \sqrt{1 + \kappa \sqrt{\gamma^{-2} + \beta^{-2}}} \| y^k - z^k \| \quad \text{for } k \geq \overline{k}. \]
Combining this inequality with (14) yields the desired result. The proof is completed. □

**Remark 3.1** Giselsson [23] and Davis et al. [15] recently derived the linear convergence rate of the generalized DRS method under the assumption that $\mathcal{A}$ is strongly monotone and $\mathcal{B}$ is $\beta$-Lipschitz continuous, which is stronger than the assumption of Theorem 3.3(c).

It is well known that the generalized DRS method is a generalized PPA associated with operator $\mathcal{S}_{\gamma \mathcal{A}, \mathcal{B}}$ in the sense that $\mathcal{T} = (I + \mathcal{S}_{\gamma \mathcal{A}, \mathcal{B}})^{-1} = \mathcal{J}_{\mathcal{S}_{\gamma \mathcal{A}, \mathcal{B}}}$ by [19, Theorem 5]. That is, the sequence $\{x^k\}$ in (10) can be generated by the following iteration step
\[ x^{k+1} = x^k + \lambda_k (\mathcal{J}_{\mathcal{S}_{\gamma \mathcal{A}, \mathcal{B}}} x^k - x^k). \]
By Theorem 3.1, we also have the linear convergence rate of the generalized DRS method under the metric subregularity of $\mathcal{S}_{\gamma \mathcal{A}, \mathcal{B}}$ at $(\overline{T}, 0) \in \text{gph } \mathcal{S}_{\gamma \mathcal{A}, \mathcal{B}}$, stated as follows.

**Theorem 3.4** Let $\{x^k\}$ be the sequence generated by equation (15) with $\{\lambda_k\} \subseteq [0, 2]$ and $\sum_{k=0}^{\infty} \lambda_k (2 - \lambda_k) = +\infty$. Then, the following statements hold.

(a) $\{z^k\}$ and $\{x^k\}$ converge to $z^* \in \mathcal{F}^{-1}(0)$ and $x^* \in \mathcal{S}_{\gamma \mathcal{A}, \mathcal{B}}^{-1}(0)$, respectively, and
\[ \| x^{k+1} - x \|^2 \leq \| x^k - x \|^2 - \lambda_k (2 - \lambda_k) \| \mathcal{J}_{\mathcal{S}_{\gamma \mathcal{A}, \mathcal{B}}} x^k - x^k \|^2 \quad \forall x \in \mathcal{S}_{\gamma \mathcal{A}, \mathcal{B}}^{-1}(0) \text{ and } \forall k \in \mathbb{N}. \]
(b) If $S_{\gamma,A,B}$ is metrically subregular at $(x^*,0)$ with constant $\kappa > 0$, then there exists $k \in \mathbb{N}$ such that for all $k \geq k$,$$
dist(x^{k+1}, S_{\gamma,A,B}^{-1}(0)) \leq \sqrt{1 - \frac{\lambda_k(2 - \lambda_k)}{(1 + \kappa)^2}} \ndist(x^k, S_{\gamma,A,B}^{-1}(0)). \quad (16)$$

Corman and Yuan [13] derived the linear convergence rate of the generalized DRS method under the assumption that $S_{\gamma,A,B}$ is strongly monotone (implied by the strong monotonicity of $F$ and one of $A$ and $B$ is firmly nonexpansive), which is stronger than the metric subregularity of $S_{\gamma,A,B}$ by Proposition 3.2 in subsection 3.5. More recently, Liang et al. [29] establish its local linear convergence rate like (16) under the metric subregularity of $(I - T)$ at a point $(z^*,0) \in \text{gph}(I - T)$ which is equivalent to the metric subregular of $S_{\gamma,A,B}$ at $(z^*,0) \in \text{gph}S_{\gamma,A,B}$ according to lemma 3.3 in the subsection 3.5.

Although, the linear convergence of the generalized DRS algorithm can be derived under the metric subregularity of $S_{\gamma,A,B}$ or $(I - T)$ at a point of its graph, this regular condition may be too difficult to be certified since that $S_{\gamma,A,B}$ is highly compound of $A$ and $B$. On the contrast, the metric subregularity of $F := A + B$ at the point $(z^*,0) \in \text{gph}F$ may be slightly easier to check due to its simple formulation. In the last subsection, we will give some sufficient conditions to ensure the metric subregularity of $F = A + B$.

### 3.4 Davis’ three-operator splitting algorithm

Davis’s splitting method [16] for the inclusion problem (1) takes the following iterations

\[
\begin{align*}
  z^k &= J_{\gamma B}(x^k), \\
  y^k &= J_{\gamma A}(2z^k - x^k - \gamma C z^k), \\
  x^{k+1} &= x^k + \lambda_k (y^k - z^k). 
\end{align*} \quad (17)
\]

Let $T := I - J_{\gamma B} + J_{\gamma A} \circ (2J_{\gamma B} - I - \gamma C \circ J_{\gamma B})$. Then, with this operator, the iterations in equation (17) can be compactly written as $x^{k+1} = x^k + \lambda_k (T x^k - x^k)$.

The following lemma present the relation between the solution set $F^{-1}(0)$ and the fixed-point set $\text{Fix } T$. For the sake of coherence, its proof is given in the appendix.

**Lemma 3.2** The set $F^{-1}(0)$ has the following relations with the fixed-point set $\text{Fix } T$:

(a) $F^{-1}(0) = J_{\gamma B}(\text{Fix } T)$. If $B$ is single-valued, $\text{Fix } T = (I + \gamma B)(F^{-1}(0))$.

(b) $\text{Fix } T \subseteq (I - \gamma (A + C))(F^{-1}(0))$. If $A$ is single-valued, $\text{Fix } T = (I - \gamma (A + C))(F^{-1}(0))$.

**Theorem 3.5** Let $\{x^k\}$ be the sequence generated by (17) with $\{\lambda_k\} \subseteq [0, \frac{4\theta - \gamma}{2\theta}]$ such that $\sum_{k=0}^{\infty} \lambda_k \left(\frac{4\theta - \gamma}{2\theta} - \lambda_k\right) = +\infty$ for $\gamma \in (0,2\theta)$. Then, the following statements hold.

(a) $\{y^k\}$ and $\{z^k\}$ converge to $z^* \in F^{-1}(0)$, and $\{x^k\}$ converges to $x^* \in \text{Fix } T$, and

$$\|x^{k+1} - x\|^2 \leq \|x^k - x\|^2 - \lambda_k \left(\frac{4\theta - \gamma}{2\theta} - \lambda_k\right) \|T x^k - x^k\|^2 \quad \forall x \in \text{Fix } T \text{ and } \forall k \in \mathbb{N}.$$
(b) If $\mathcal{A}$ is single-valued and Lipschitz continuous with modulus $1/\beta$ and $\mathcal{F}$ is metric subregular with constant $\kappa > 0$ at $(z^*,0)$, then there exists $\overline{k} \in \mathbb{N}$ such that
\[
\text{dist}(x^{k+1}, \text{Fix } T) \leq (1 - \varrho) \text{dist}(x^k, \text{Fix } T) \quad \text{for } k \geq \overline{k}
\] (18)
with
\[
\varrho = \frac{\lambda_k(4\vartheta - \gamma - 2\theta \lambda_k)}{2\vartheta[(2 + \gamma \vartheta^{-1}) + (\gamma \vartheta^{-1} + \sqrt{1 + \gamma^2 \beta^{-2}})(1 + \kappa(\gamma^{-1} + \beta^{-1}))]^2}.
\]

(c) If $\mathcal{B}$ is single-valued and Lipschitz continuous with modulus $1/\beta$ and $\mathcal{F}$ is metric subregular with constant $\kappa > 0$ at $(z^*,0)$, then there exists $\overline{k} \in \mathbb{N}$ such that (18) holds for all $k \geq \overline{k}$ with
\[
\varrho = \frac{\lambda_k(4\vartheta - \gamma - 2\theta \lambda_k)}{2\vartheta(1 + \gamma \beta^{-1})(1 + \kappa(\gamma^{-1} + \beta^{-1}))^2}.
\]

**Proof:** (a) By [16, Proposition 3.1], $T$ is $\alpha$-averaged with $\alpha = \frac{2\vartheta}{4(\vartheta - 1)}$ for $\gamma \in (0, 2\vartheta)$. By Lemma 3.2(a), $F^{-1}(0) = \mathcal{J}_{\gamma \mathcal{B}}(\text{Fix } T)$. Thus, the result directly follows by Lemma 2.2.

(b) From the iteration step (17), $\gamma^{-1}(z^k - y^k) \in Bz^k$ and $\gamma^{-1}(2z^k - x^k - \gamma Cz^k - y^k) \in A \mathcal{F}$. Hence, we have $\gamma^{-1}(z^k - y^k) = A y^k + Bz^k + Cz^k$, which further implies that
\[
\gamma^{-1}(z^k - y^k) + A z^k - A y^k \in A z^k + Bz^k + Cz^k \in \mathcal{F}z^k.
\]
Since $\mathcal{F}$ is metrically subregular at $(z^*,0)$ with constant $\kappa$ and $z^k \rightarrow z^*$ by part (a), there exists $\overline{k} \in \mathbb{N}$ such that for all $k \geq \overline{k}$, the latter inequalities hold
\[
\text{dist}(z^k, \mathcal{F}^{-1}(0)) \leq \kappa\text{dist}(0, \mathcal{F}(z^k)) \leq \kappa\|\gamma^{-1}(z^k - y^k) + A z^k - A y^k\| \leq \kappa(\gamma^{-1} + \beta^{-1})\|z^k - y^k\|
\]
where the last inequality is due to the Lipschitz continuity of $\mathcal{A}$. Thus, for all $k \geq \overline{k},$
\[
\text{dist}(y^k, \mathcal{F}^{-1}(0)) \leq \text{dist}(z^k, \mathcal{F}^{-1}(0)) + \|z^k - y^k\| \leq [1 + \kappa(\gamma^{-1} + \beta^{-1})]\|z^k - y^k\|. \quad (19)
\]
Let $\overline{x}^k = \Pi_{\mathcal{F}^{-1}(0)}(y^k) = \gamma(\mathcal{A} + C)(\Pi_{\mathcal{F}^{-1}(0)}(y^k))$. From Lemma 3.2(b), $\overline{x}^k \in \text{Fix } T$. In addition, notice that $x^k = 2z^k - y^k - \gamma A y^k - \gamma C z^k$ by (17). Hence, for all $k \geq \overline{k},$
\[
\text{dist}(x^k, \text{Fix } T) \leq \|x^k - \Pi_{\mathcal{F}^{-1}(0)}(y^k) + \gamma(\mathcal{A} + C)(\Pi_{\mathcal{F}^{-1}(0)}(y^k))\|
= \|2z^k - y^k - \gamma A y^k - \gamma C z^k - \Pi_{\mathcal{F}^{-1}(0)}(y^k) + \gamma(\mathcal{A} + C)(\Pi_{\mathcal{F}^{-1}(0)}(y^k))\|
\leq \|2z^k - 2y^k\| + \|y^k - \Pi_{\mathcal{F}^{-1}(0)}(y^k) - \gamma A y^k + \gamma A(\Pi_{\mathcal{F}^{-1}(0)}(y^k))\|
\quad + \|\gamma C y^k - \gamma C z^k + \gamma C(\Pi_{\mathcal{F}^{-1}(0)}(y^k)) - \gamma C y^k\|
\leq (2 + \gamma \vartheta^{-1})\|z^k - y^k\| + (\gamma \vartheta^{-1} + \sqrt{1 + \gamma^2 \beta^{-2}})\text{dist}(y^k, \mathcal{F}^{-1}(0))
\leq [(2 + \gamma \vartheta^{-1}) + (\gamma \vartheta^{-1} + \sqrt{1 + \gamma^2 \beta^{-2}})(1 + \kappa(\gamma^{-1} + \beta^{-1}))]\|z^k - y^k\|,
\]
\[11\]
where the third inequality is using the monotonicity of \( \mathcal{A} \) and the last one is due to (19). In addition, using part (a) and the same arguments as for Theorem 3.3(b) yields

\[
\text{dist}(x^{k+1}, \text{Fix } \mathcal{T})^2 \leq \text{dist}(x^k, \text{Fix } \mathcal{T})^2 - \lambda_k \left( \frac{4\theta - \gamma}{2\theta} - \lambda_k \right) \| y^k - z^k \|^2.
\]

The desired result then follows from the last two inequalities.

(c) By the proof of part (b), we have \( \gamma^{-1}(z^k - y^k) \in A y^k + B z^k + C z^k \). This along with the single-value \( \mathcal{B} \) yield \( \gamma^{-1}(z^k - y^k) + (B y^k - B z^k) \in A y^k + B y^k + C z^k \). By the cocoercivity of \( \mathcal{C} \), we obtain that \( \gamma^{-1}(z^k - y^k) + (\eta^k - \xi^k) + C y^k - C z^k \in A y^k + B y^k + C y^k = \mathcal{F} y^k \). Since \( \mathcal{F} \) is metrically subregular at \((z^*, 0)\) with constant \( \kappa \), from part (a) and the Lipschitzian of \( \mathcal{B} \) and the cocoercivity of \( \mathcal{C} \), it follows that there exists \( \kappa \) such that for all \( k \geq \kappa \),

\[
\text{dist}(y^k, \mathcal{F}^{-1}(0)) \leq \kappa \text{dist}(0, \mathcal{F}(y^k)) \leq \kappa \| \gamma^{-1}(z^k - y^k) + (\eta^k - \xi^k) + C y^k - C z^k \| \\
\leq \kappa \left[ \| \eta^k - \xi^k \| + \| \gamma^{-1}(z^k - y^k) + C y^k - C z^k \| \right] \\
\leq \kappa \left[ \frac{1}{\beta} + \sqrt{\frac{1}{\gamma^2} + \max \left( \frac{\gamma^2 - 2\gamma\theta}{\theta^2}, 0 \right)} \right] \| z^k - y^k \|.
\]

Combining this inequality with \( \text{dist}(z^k, \mathcal{F}^{-1}(0)) \leq \text{dist}(y^k, \mathcal{F}^{-1}(0)) + \| z^k - y^k \| \) implies that

\[
\text{dist}(z^k, \mathcal{F}^{-1}(0)) \leq \left[ 1 + \kappa \left( \frac{1}{\beta} + \frac{1}{\gamma} \sqrt{1 + \max \left( \frac{\gamma^2 - 2\gamma\theta}{\theta^2}, 0 \right)} \right) \right] \| z^k - y^k \|, \forall k \geq \kappa
\]

From \( \text{Fix } \mathcal{T} = \mathcal{F}^{-1}(0) \), it follows that \( \Pi_{\mathcal{F}^{-1}(0)}(z^k) + \gamma \mathcal{B}(\Pi_{\mathcal{F}^{-1}(0)}(z^k)) \in \text{Fix } \mathcal{T} \). Moreover, using the Lipschitzian of \( \mathcal{B} \) yields \( \| \mathcal{B}(\Pi_{\mathcal{F}^{-1}(0)}(z^k)) - B z^k \| \leq \beta^{-1} \| \Pi_{\mathcal{F}^{-1}(0)}(z^k) - z^k \| \). Combining this inequality with the facts that \( x^k = z^k + \gamma B z^k \) and \( \Pi_{\mathcal{F}^{-1}(0)}(z^k) \rightarrow z^* \), and using the same arguments as for those of Theorem 3.3(c) yield that for all \( k \geq \kappa \),

\[
\text{dist}(x^k, \text{Fix } \mathcal{T}) \leq (1 + \beta \beta^{-1}) \left[ 1 + \kappa \left( \frac{1}{\beta} + \frac{1}{\gamma} \sqrt{1 + \max \left( \frac{\gamma^2 - 2\gamma\theta}{\theta^2}, 0 \right)} \right) \right] \| z^k - y^k \|.
\]

Combining this inequality with (20) yields the desired result. The proof is completed. \( \Box \)

Davis and Yin [16] derived the linear convergence rate of their algorithm under the condition that one of \( \mathcal{A}, \mathcal{B} \) and \( \mathcal{C} \) is strongly monotone and one of \( \mathcal{A} \) and \( \mathcal{B} \) is single-valued and Lipschitz continuous, which is stronger than that of Theorem 3.5 (b) and (c).

### 3.5 Sufficient conditions for the metric subregularity

In this section, we give some sufficient conditions to ensure the metric subregularity of the maximal monotone operator \( \mathcal{F} := \mathcal{A} + \mathcal{B} \) under the condition that \( \mathcal{B} \) is single valued and Lipschitz continuous with modulus \( \frac{1}{\beta} \). Write \( \mathcal{R}(z) = z - J_{\gamma \mathcal{A}}(I - \gamma \mathcal{B})z \) which is clearly a single valued mapping. In the following lemma, we give an equivalent characterization on the metrically subregularity of \( \mathcal{F} \) at \( z^* \) for \( 0 \in \mathcal{F}(z^*) \).
Lemma 3.3 Let $\mathcal{F} := \mathcal{A} + \mathcal{B}$ where $\mathcal{A}$ is maximal monotone and $\mathcal{B}$ is single valued and Lipschitz continuous with modulus $\frac{1}{\beta}$. Then, the mapping $R$ is metrically subregular at $z^*$ for $0 \in \mathcal{R}(z^*)$ if and only if the operator $\mathcal{F}$ is metrically subregular at $z^*$ for $0 \in \mathcal{F}(z^*)$.

**Proof:** By the metric subregularity of $R$ at $z^*$ for $0 \in R(z^*)$, there exist a constant $\kappa' > 0$ and a sufficiently small $\delta' > 0$ such that

$$\text{dist}(z, R^{-1}(0)) \leq \kappa' \|R(z)\|, \quad \forall z \in B(z^*, \delta')$$

(21)

where $B(z^*, \delta')$ denotes the closed ball in the space $\mathbb{H}$ centered at $z^*$ with radius $\delta'$. Let $z$ be arbitrary point from $B(z^*, \delta')$. Take $y \in \mathcal{F}(z)$ as the point such that $\text{dist}(0, \mathcal{F}(z)) = \|y\|$. Notice that $\mathcal{F} = \mathcal{A} + \mathcal{B}$ and $\mathcal{B}$ is single valued and lipschitz continuous, it is easy to get that $z + \gamma y - \gamma Bz \in (I + \gamma A)z$ which in turn implies that $z = \mathcal{J}_{\gamma A}(z + \gamma y - \gamma Bz)$. Together with the last equation and the metric subregularity (21) of $\mathcal{R}$, we obtain that

$$\text{dist}(z, \mathcal{F}^{-1}(0)) = \text{dist}(z, \mathcal{R}^{-1}(0)) \leq \kappa' \|\mathcal{R}(z)\| = \kappa' \|\mathcal{J}_{\gamma A}(z + \gamma y - \gamma Bz) - \mathcal{J}_{\gamma A}(z - \gamma Bz)\|.$$ 

Notice that $\mathcal{J}_{\gamma A}$ is nonexpansive. This along with the equality $\text{dist}(0, \mathcal{F}(z)) = \|y\|$ yield

$$\text{dist}(z, \mathcal{F}^{-1}(0)) \leq \kappa' \gamma \text{dist}(0, \mathcal{F}(z)),$$

which shows that the operator $\mathcal{F}$ is metrically subregular at $z^*$ for $0 \in \mathcal{F}(z^*)$.

Conversely, suppose that $\mathcal{F}$ is metrically subregular at $z^*$ for $0 \in \mathcal{F}(z^*)$. Then, there exist a constant $\kappa > 0$ and a sufficiently small $\delta > 0$ such that

$$\text{dist}(z, \mathcal{F}^{-1}(0)) \leq \kappa \text{dist}(0, \mathcal{F}(z)) \quad \forall z \in B(z^*, \delta).$$

(22)

Take $\delta' = \frac{\delta}{(1 + \gamma \beta)^{-1}}$. Notice that the equation holds $\mathcal{J}_{\gamma A}(I - \gamma B)z^* = z^*$ since $z^* \in \mathcal{F}^{-1}(0)$. Then, $\mathcal{J}_{\gamma A}(I - \gamma B)z \in B(z^*, \delta)$ due to $\|\mathcal{J}_{\gamma A}(I - \gamma B)z - z^*\| \leq (1 + \gamma \beta^{-1})\|z - z^*\| \leq \delta$. Combine $z - \gamma Bz - \mathcal{J}_{\gamma A}(z - \gamma Bz) \in \gamma A(\mathcal{J}_{\gamma A}(z - \gamma Bz))$ and metric subregularity of $\mathcal{F}$ yield

$$\text{dist}(z, \mathcal{R}^{-1}(0)) = \text{dist}(z, \mathcal{F}^{-1}(0))$$

$$\leq \text{dist}(\mathcal{J}_{\gamma A}(z - \gamma Bz), \mathcal{F}^{-1}(0)) + \|z - \mathcal{J}_{\gamma A}(z - \gamma Bz)\|$$

$$\leq \kappa \text{dist}(0, \mathcal{F}(\mathcal{J}_{\gamma A}(z - \gamma Bz))) + \|z - \mathcal{J}_{\gamma A}(z - \gamma Bz)\|$$

$$= \kappa \text{dist}(-B(\mathcal{J}_{\gamma A}(z - \gamma Bz)), A(\mathcal{J}_{\gamma A}(z - \gamma Bz))) + \|z - \mathcal{J}_{\gamma A}(z - \gamma Bz)\|$$

$$\leq \kappa \| -B(\mathcal{J}_{\gamma A}(z - \gamma Bz)) - \gamma^{-1}[z - \mathcal{J}_{\gamma A}(z - \gamma Bz)] + Bz\| + \|z - \mathcal{J}_{\gamma A}(z - \gamma Bz)\|$$

$$\leq (1 + \gamma^{-1} \kappa + \beta^{-1} \kappa)\|z - \mathcal{J}_{\gamma A}(z - \gamma Bz)\| = (1 + \gamma^{-1} \kappa + \beta^{-1} \kappa)\|\mathcal{R}(z)\|$$

which implies $\mathcal{R}$ is metric subregular at $(z^*, 0) \in \text{gph} \mathcal{R}$. The proof is completed. $\Box$

By the above lemma and Theorem 3.3, the generalized DRS algorithm is linear convergence if $R$ is metric subregular at $z^*$ for $0 \in R(z^*)$. Moreover, when $B$ is reduced to $B = 0$, strengthened as a cocoercive operator and specified as $B = C + D$ with $C$ being cocoercive and $D$ being single valued and lipschitz, respectively. By Theorem 3.1 3.2, 3.5 and above Lemma 3.3, we know that the generalized PPA, the over-relaxed FBS algorithm, and the Davis-Yin’s three operator splitting method are linearly convergent with $\mathcal{R}$ is metric subregular at $z^*$ for $0 \in \mathcal{R}(z^*)$ accordingly. Next, we give a sufficient condition to ensure the metric subregularity of $\mathcal{R}$ at $z^*$ for $0 \in \mathcal{R}(z^*)$. 

13
Lemma 3.4 The single-valued mapping $R$ is metrically subregular at $z^*$ for $0 \in \mathcal{R}(z^*)$ if the following projection type error bound \cite[Eq. 5]{40} holds
\[ \text{dist}(z, R^{-1}(0)) \leq \kappa'' \| R(z) \|, \quad \forall \ z \text{ with } \| R(z) \| \leq \delta'' \] (23)

Proof: Let $\delta' = \frac{\delta''}{2 + \gamma \beta - 1}$. For any $z \in B(z^*, \delta')$, using the equality $z^* = \mathcal{J}_{\gamma A}(I - \gamma B)^* z = \mathcal{J}_{\gamma A}(I - \gamma B)z^*) \leq (2 + \gamma \beta^{-1}) \| z - z^* \| \leq \delta''$ (24)

Together with the above inequality and the projective bound (3.4), we get the desired results that $R$ is metrically subregular at $z^*$ for $0 \in \mathcal{R}(z^*)$. \hfill \Box

Next, we give certain instances with the projection type error bound (23) or metric subregularity of $R$ holding. Consequently, the metrically subregularity of $\mathcal{F} = A + B$ holds at $z^*$ for $0 \in \mathcal{F}(z^*)$. The proof of the following proposition is followed directly according to \cite{40, 43, 42}. Here, we omit the details.

Proposition 3.1 Let $\mathcal{F} := A + B$ where $A$ is maximal monotone and $B$ is single valued and Lipschitz continuous. Then, the operator $\mathcal{F}$ is metrically subregular at a point $z^*$ for $0 \in \mathcal{F}(z^*)$, i.e., there exists $\kappa, \delta$ such that inequality (22) holds whenever one of the following statements holds.

(C1) $\mathcal{F} := A + B$ is strongly monotone;

(C2) $B$ is affine operator and $A$ is polyhedron operator;

(C3) $B = \mathcal{E} \nabla f(\mathcal{E} \cdot) + C$ where $f$ is strongly convex and gradient Lipschitz $\mathcal{E}$ is linear operator and $C$ is a constant. $A$ is the subdifferential operator of $\ell_p$ norm with $p \in [1, 2] \cup \{\infty\}$ or polyhedral convex function;

(C4) $B = \mathcal{E} \nabla f(\mathcal{E} \cdot) + C$ where $f$ is strongly convex and gradient Lipschitz and $\mathcal{E}$ is linear operator and $C$ is a constant. $A$ is the subdifferential operator of the nuclear norm. In addition, $-B x \in \text{ri}(\partial \| x \|_p)$.

To end this subsection, we make some comments on the metric subregularity of $\mathcal{S}_{\gamma, A, B}$ at a point $x^*$ with $0 \in \mathcal{S}_{\gamma, A, B}(x^*)$. Up to now, we are not clear whether the metric subregularity of $\mathcal{S}_{\gamma, A, B}$ at $(x^*, 0) \in \text{gph} (\mathcal{S}_{\gamma, A, B})$ is weaker than that of $\mathcal{F} := A + B$ at $(z^*, 0) \in \text{gph} \mathcal{F}$ or not when $A$ or $B$ is single-valued and Lipschitz continuous. The following proposition gives a sufficient condition to guarantee the metric subregularity of $\mathcal{S}_{\gamma, A, B}$ at a point $(x^*, 0) \in \text{gph} (\mathcal{S}_{\gamma, A, B})$. Its proof is also provided in the appendix.

Proposition 3.2 If $\mathcal{F}$ is strongly monotone with constant $\alpha > 0$ and one of $A$ and $B$ is single-valued and Lipschitz continuous with modulus $\beta > 0$, then
\[ \mathcal{S}_{\gamma, A, B}^{-1}(w) \subseteq \mathcal{S}_{\gamma, A, B}^{-1}(0) + \kappa \| w \| B \quad \text{for } w \in \mathcal{X}, \] (25)

This implies that $\mathcal{S}_{\gamma, A, B}$ is metrically subregular at $(x^*, 0) \in \text{gph} (\mathcal{S}_{\gamma, A, B})$. 14
4 Toy examples

In this section, we first consider the following nonsmooth convex optimization problems

\[
\min f(x) + g(Dx)
\]  

(26)

where \( f : \mathbb{Z} \to [-\infty, +\infty] \) and \( g : \mathbb{Y} \to [-\infty, +\infty] \) are low semicontinuous convex function, and \( D : \mathbb{Z} \to \mathbb{Y} \) is linear operator. Notice that the problem (26) embodies an abundance of popular applications such as famous Rudin-Osher-Fatemi (ROF) denoising model [38], \( TV L_1 \) minimization model [7], the convex image segmentation model [8, 24] and the \( \ell_1/\ell_1 \)-regularization model [9]. It is obvious that any optimal solution of (26) satisfies the inclusion \( 0 \in \partial f(x) + D^* \partial g(Dx) \). Involved in the dual variable \( y \in \mathbb{Y} \), the above inclusion can be reformulated as the inclusion \((0, 0) \in (\mathcal{T}_1 + \mathcal{T}_2)(x, y)\) with

\[
\begin{align*}
\mathcal{T}_1(x, y) &= (D^* y, -Dx) \\
\mathcal{T}_2(x, y) &= (\partial f(x), \partial g^*(y))
\end{align*}
\]

(27)

It is obvious that \( \mathcal{T}_1 \) is single valued, Lipschitz continuous and affine operator. Hence, we can apply the generalized DRS algorithm for the inclusion (27) with following iterations

\[
\begin{align*}
x_1^k &= (I + \gamma^2 D^* D)^{-1} (z_1^k - \gamma D^* z_2^k) \\
y_1^k &= (I + \gamma^2 D D^*)^{-1} (\gamma D z_1^1 + z_2^k) \\
x_2^k &= \min f(x_2) + \|x_2 - (2x_1^k - z_1^k)/2\|/(2\gamma) \\
y_2^k &= \min g^*(y_2) + \|y_2 - (2y_1^k - z_2^k)/2\|/(2\gamma) \\
z_1^{k+1} &= z_1^k + \lambda_k (x_2^k - x_1^k) \\
z_2^{k+1} &= z_2^k + \lambda_k (y_2^k - y_1^k)
\end{align*}
\]

(28a)-(28f)

with \( \gamma > 0 \) and relaxation parameter \( \lambda_k \in (0, 2) \). In the following, we specify model (26) as the \( \ell_1/\ell_1 \)-regularization model proposed by Chan et al. [9] with the following form

\[
\min_x \|Ax - b\|_1 + \lambda \|x\|_1
\]

(29)

\[
\min_x \|Ax - b\|^2 + \lambda \|x\|_1
\]

Now, we apply the above generalized DRS algorithm (28a)-(28f) to the above \( \ell_1/\ell_1 \) regularization minimization (29). In this case, we get that \( \mathcal{T}_1(x, y) = (A^T y, b - Ax) \) is singled valued and Lipschitz continuous operator and \( \mathcal{T}_2(x, y) = (\partial \|x\|_1, N_{\|y\|_\infty \leq 1}(y)) \) is polyhedral operator. By condition (C2) in Proposition 3.1, we know that the algorithm (28a)-(28f) converges linearly when it is applied to the \( \ell_1/\ell_1 \) regularization minimization with random generated sensing matrix \( A \) and regularization parameter \( \lambda = 1 \). The Figure 1 shows that numerical performance of generalized DRS algorithm (28a)-(28f) when it is applied to the problem (29), which is coincided with Theorem 3.3 that the prime-dual points sequences \( \{(x^k, y^k)\} \) and sequences \( \{(z_1^k, z_2^k)\} \) converge linearly.
Next, we consider the $\ell_p$ norm regularization problem with $p \in [1, 2] \cup \{+\infty\}$.

$$\min f(Dx) + g(x)$$

where $f : \mathbb{Y} \to [-\infty, +\infty]$ is a strongly convex and gradient Lipschitz continuous function and $D : \mathbb{X} \to \mathbb{Y}$ is a linear operator, and $g(x) = \sum_{J \in \mathcal{J}} w_J \|x_J\|_p$ where $\mathcal{J}$ is a non-overlapping divisibility of index set $\{1, 2, \cdots, n\}$ and $w_J > 0$, and $\|x_J\|_p$ denotes the $l_p$ norm defined by

$$\|x_J\|_p = \left( \sum_{i=1}^l \|x_i\|^p \right)^{\frac{1}{p}}.$$ 

This problem (30) incorporates massive applications such as Group-lasso regularization [41], $\ell_{1,p}$-regularization regression [21, 20] and the referees in [42]. Here, we consider the following $\ell_p$ norm regularization problem

$$\min_x \frac{1}{2} \|Ax - b\|^2 + \lambda \|x\|_p$$

Now, we apply the over-relaxed forward backward splitting algorithm (7) and the generalized Douglas Rachford splitting algorithm (10) to the problem (31) with random generated sensing matrix $A$, regularized parameter $\lambda = 1$ and $p = 2$, respectively. The figure 2 shows performance of over-relaxed FBS and generalized DRS algorithm when they are applied to the toy example (31), which is coincided with Theorem 3.2, 3.3.

5 Conclusion

In this paper, for the inclusion problem (1), we have established the linear convergence rate of the generalized PPA and several popular splitting algorithms under the metric subregularity of the composite operator, which is much weaker than the existing ones that almost all require the strong monotonicity of the composite operator. Some sufficient condition are provided to ensure the metric subregularity of the composite operator $\mathcal{F}$ holds at a point $(z^*, 0) \in \text{gph } \mathcal{F}$. The preliminary numerical performances also support the theoretical results.
References

[1] F. J. A. Artacho, A. L. Dontchev and M. H. Geoffroy, Convergence of the proximal point method for metrically regular mappings, ESAIM: Proceedings, vol. 17, pp. 1-8, 2007.

[2] H. H. Bauschke and P. L. Combettes, Convex analysis and monotone operator theory in hilbert spaces, Springer, 2011.

[3] H. H. Bauschke and J. M. Borwein, On projection algorithms for solving convex feasibility problems, SIAM Review, vol. 38, pp. 367-426, 1995.

[4] H. H. Bauschke, P. L. Combettes and D. R. Luke, Finding best approximation pairs relative to two closed convex sets in Hilbert spaces, Journal of Approximation Theory, vol. 127, pp. 178-192, 2004.

[5] H. H. Bauschke, D. Noll and H. M. Phan, Linear and strong convergence of algorithms involving averaged nonexpansive operators, Journal of Mathematical Analysis and Applications, vol. 421, pp. 1-20, 2015.

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

[7] T. F. Chan and S. Esedoglu, Aspects of total variation regularized $\ell_1$ function approximation, SIAM Journal on Applied Mathematics, vol. 65, pp. 1817-1837, 2005.

[8] T. F. Chan, S. Esedoglu and M. Nikolova, Algorithms for finding global minimizers of image segmentation and denoising models, SIAM Journal on Applied Mathematics, vol. 66, pp. 1632-1648, 2006.

[9] R. H. Chan, M. Tao and X. M. Yuan, Constrained total variation deblurring models and fast algorithms based on alternating direction method of multipliers, SIAM Journal on Imaging Sciences, vol. 6, pp. 680-697, 2013.

[10] G. H-G. Chen and R. T. Rockafellar, Convergence rates in forward-backward splitting, SIAM Journal on Optimization, vol. 7, pp. 1-25, 1997.

[11] P. L. Combettes and V. R. Wajs, Signal recovery by proximal forward-backward splitting, Multiscale Modeling & Simulation, vol. 4, pp. 1168-1200, 2006.

[12] P. L. Combettes and J. C. Pesquet, Proximal splitting methods in signal processing, Fixed-Point Algorithms for Inverse Problems in Science and Engineering, vol. 49, pp. 185-212, 2011.

[13] E. Corman and X. M. Yuan, A generalized proximal point algorithm and its convergence rate, SIAM Journal on Optimization, vol. 24, pp. 1614-1638, 2014.
[14] D. Davis and W. T. Yin, *Convergence rate analysis of several splitting schemes*, arXiv preprint arXiv:1406.4834, 2014.

[15] D. Davis and W. T. Yin, *Faster convergence rates of relaxed Peaceman-Rachford and ADMM under regularity assumptions*, arXiv preprint arXiv:1407.5210, 2014.

[16] D. Davis and W. T. Yin, *A three-operator splitting scheme and its optimization application*, arXiv preprint arXiv:1504.01032, 2015.

[17] A. L. Dontchev and R. T. Rockafellar, *Implicit functions and solution mappings—a view from variational analysis*, Springer, 2009.

[18] J. Douglas and H. H. Rachford, *On the numerical solution of heat conduction problems in two and three space variables*, Transactions of the American Mathematical Society, vol. 82, pp. 421-439, 1956.

[19] J. Eckstein and D. P. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Mathematical Programming, vol. 55, pp. 293-318, 1992.

[20] Y. C. Eldar, P. Kupinger and H. Bölcskei, *Block-sparse signals: uncertainty relations and efficient recovery*, IEEE Transactions on Signal Processing, vol. 58, pp. 3042-3054, 2010.

[21] M. Fornasier and H. Rauhut, *Recovery algorithms for vector-valued data with joint sparsity constraints*, SIAM Journal on Numerical Analysis, vol. 46, pp. 577-613, 2008.

[22] D. Gabay, *Applications of the method of multipliers to variational inequalities*, in: M. Fortin and R. Glowinski, eds., Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems (North-Holland, Amsterdam, 1983).

[23] P. Giselsson, *Tight global linear convergence rate bounds for Douglas-Rachford splitting*, arXiv preprint arXiv:1506.01556, 2015.

[24] T. Goldstein, X. Bresson and S. Osher, *Geometric applications of the split Bregman method: segmentation and surface reconstruction*, Journal of Scientific Computing, vol. 45, pp. 272-293, 2010.

[25] P. T. Harker and J.-S. Pang, *Finite-dimensional variational inequality and non-linear complementarity problems: A survey of theory, algorithms and applications*, Mathematical Programming, vol. 48, pp. 161-220, 1990.

[26] B. S. He and X. M. Yuan, *On the convergence rate of Douglas-Rachford operator splitting method*, Mathematical Programming, DOI 10.1007/s10107-014-0805-x.

[27] M. Krasnosel’skii, *Two remarks on the method of successive approximations*, Uspekhi Matematicheskikh Nauk, vol. 10, pp. 123-127, 1955.
[28] D. Leventhal, *Metric subregularity and the proximal point method*, Journal of Mathematical Analysis and Applications, vol. 360, pp. 681-688, 2009.

[29] J. W. Liang, J. M. Fadili and G. Peyră, *Convergence rates with inexact nonexpansive operators*, Mathematical Programming, DOI 10.1007/s10107-015-0964, 2015.

[30] P. L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM Journal on Numerical Analysis, vol. 16, pp. 964-979, 1979.

[31] B. Martinet, *Regularisation d’inéquations variationnelles par approximations successives*, Rev. Française Informat. Recherche Opérationnelle, vol. 4, pp. 154-159, 1970.

[32] W. R. Mann, *Mean value methods in iteration*, Proceedings of the American Mathematical Society, vol. 4, pp. 506-510, 1953.

[33] Y. Nesterov, *Gradient methods for minimizing composite functions*, Mathematical Programming, vol. 140, pp. 125-161, 2012.

[34] G. B. Passty, *Ergodic convergence to a zero of the sum of monotone operators in Hilbert space*, Journal of Mathematical Analysis and Applications, vol. 72, pp. 383-390, 1979.

[35] D. W. Peaceman and H. H. Rachford, *The numerical solution of parabolic and elliptic differential equations*, Journal of the Society for Industrial and Applied Mathematics, vol. 3, pp. 28-41, 1954.

[36] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM Journal on Control and Optimization, vol. 14, pp. 877-898, 1976.

[37] R. T. Rockafellar and R. J-B. Wets, *Variational Analysis*, Springer, 1998.

[38] L. I. Rudin, S. Osher and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D: Nonlinear Phenomena, vol. 60, pp. 259-268, 1992.

[39] M. Tao and X. M. Yuan, *On the optimal linear convergence rate of a generalized proximal point algorithm*, ArXiv preprint, arXiv:1605.05474, 2016.

[40] Paul Tseng, *On linear convergence of iterative methods for the variational inequality problem*, Journal of Computational and Applied Mathematics, vol. 60, pp. 237-252, 1995.

[41] M. Yuan and Y. Lin, *Model selection and estimation in regression with grouped variables*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 68, pp. 49-67, 2006.

[42] Z. R. Zhou, Q. Zhang and A. M. C. So, *ℓ1,p-Norm Regularization: Error Bounds and Convergence Rate Analysis of First-Order Methods*, Proceedings of the 32nd International Conference on Machine Learning (ICML 2015), pp. 1501-1510, 2015.
Appendix

Proof of Lemma 3.1: Part (a) directly follows from [2, Proposition 25.1(ii)]. We next make use of $S_{\gamma,A,B}$ in [19] to prove the inclusion of part (b), where $S_{\gamma,A,B}$ is defined by

$$S_{\gamma,A,B} = \{(v + \gamma b, u-v) \mid (u,b) \in B, (v,a) \in A, v + \gamma a = u - \gamma b\}.$$  \hfill (32)

From [19, Theorem 5], it follows that $\text{Fix} T = S_{\gamma,A,B}^{-1}(0) = \bigcup_{x \in X} \left[u + \gamma(-Au \cap Bu)\right]$. Let $x$ be an arbitrary point from $\text{Fix} T$. Then, there exist $u \in X$ and $b \in (-Au \cap Bu)$ such that $x = u + \gamma b \in (I - \gamma A)(u)$. Clearly, $0 \in Au + Bu = Fu$, i.e., $u \in F^{-1}(0)$. Hence,

$$x \in \bigcup_{z \in F^{-1}(0)} (I - \gamma A)z = (I - \gamma A)(F^{-1}(0)).$$

The inclusion then follows from the arbitrariness of $x$ in the set $\text{Fix} T$.

Now assume that $A$ is single-valued. To establish the equality, it suffices to argue that $(I - \gamma A)(F^{-1}(0)) \subseteq \text{Fix} T$. Let $x$ be an arbitrary point from $(I - \gamma A)(F^{-1}(0))$. Then there exist $z \in F^{-1}(0)$ and $y = Az$ such that $x = z - \gamma y$. Since $0 \in Az + Bz$, we have $-Az = -y \in Bz$. Thus, $x \in z + \gamma Bz$, which by $z \in F^{-1}(0)$ and part (a) implies that $x \in \text{Fix} T$. The inclusion follows by the arbitrariness of $x$ in $(I - \gamma A)(F^{-1}(0))$. Hence, the proof is completed. \hfill $\square$

Proof of Lemma 3.2: Part (a) follows from [16, Lemma 3.2]. It suffices to prove part (b). From [16, Lemma 3.2], it follows that $\text{Fix} T = \bigcup_{u \in F^{-1}(0)} \left[u + \gamma((-A - C)u) \cap Bu\right]$.

Let $x$ be an arbitrary point from $\text{Fix} T$. Then, there exist $u \in F^{-1}(0)$ and $b \in (-A - C)u \cap Bu$ such that $x = u + \gamma b \in (I - \gamma (A + C))u$. This immediately implies that

$$x \in \bigcup_{z \in F^{-1}(0)} (I - \gamma (A + C))z = (I - \gamma (A + C))(F^{-1}(0)).$$

By the arbitrariness of $x$ in $\text{Fix} T$, the inclusion follows. Now assume that $A$ is single-valued. We only need to argue that $(I - \gamma (A + C))(F^{-1}(0)) \subseteq \text{Fix} T$. Let $x$ be an arbitrary point from $(I - \gamma (A + C))(F^{-1}(0))$. Then there exist $z \in F^{-1}(0)$ and $y = (A + C)z$ such that $x = z - \gamma y$. Since $0 \in Az + Bz + Cz$, we have $-(A + C)z = -y \in Bz$. Thus, $x \in z + \gamma Bz$. This, along with $z \in F^{-1}(0)$ and part (a), implies that $x \in \text{Fix} T$. By the arbitrariness of $x$ in $(I - \gamma (A + C))(F^{-1}(0))$, the inclusion follows. \hfill $\square$

Proof of Proposition 3.2: Let $x$ be an arbitrary point from $\text{dom} S_{\gamma,A,B}$. Since $S_{\gamma,A,B}$ is maximal monotone (see [19, Theorem 4]), the set $S_{\gamma,A,B}(x)$ is closed convex by [37, Exercise 12.8]. In the following, we proceed the arguments by two cases as shown below.

[43] Z. R. ZHOU AND A. M. C. SO, A Unified Approach to Error Bounds for Structured Convex Optimization Problems, arXiv preprint arXiv:1512.03518, 2015.
Case 1: \( A \) is single-valued and Lipschitz continuous with modulus \( \beta > 0 \). From the closed convexity of \( S_{\gamma,A,B}(x) \) and the expression of \( S_{\gamma,A,B} \), there exist \( u, b, v \in X \) with \((u, b) \in B \) and \( v + \gamma Av = u - \gamma b \) such that \( u - v = \text{dist}(0, S_{\gamma,A,B}(x)) \) for \( x = v + \gamma b \). So,

\[
\gamma^{-1}(u - v) + (Au - Av) \in F u.
\] (33)

By Lemma 3.1(b) and the remark after it, we have \( S_{\gamma,A,B}^{-1}(0) = (I - \gamma A)(F^{-1}(0)) \). Thus,

\[
dist(x, S_{\gamma,A,B}^{-1}(0)) = \text{dist}(u - \gamma Av, S_{\gamma,A,B}^{-1}(0)) \leq \| u - \gamma Av - z^* + \gamma Az^* \| \\
\leq \| u - z^* \| + \gamma \beta \| v - z^* \| \leq (1 + \gamma \beta) \| u - z^* \| + \gamma \beta \| u - v \|,
\] (34)

where \( z^* \) is an arbitrary point from \( F^{-1}(0) \). Since \( F = A + B \) is strongly monotone with constant \( \alpha \), \( F^{-1} \) is single-valued and Lipschitz continuous with modulus \( \alpha^{-1} \), i.e.,

\[
\| F^{-1}(w) - F^{-1}(0) \| \leq \alpha^{-1}\| w \| \quad \forall w \in X.
\] (35)

Let \( \overline{w} \in Fu \) be such that \( \| \overline{w} \| = \text{dist}(0, Fu) \). Then, by the last inequality, it follows that

\[
\| u - z^* \| = \| F^{-1}(\overline{w}) - F^{-1}(0) \| \leq \alpha^{-1}\| \overline{w} \| = \alpha^{-1}\text{dist}(0, Fu).
\]

Combining this inequality with inequalities (33) and (34), we have that

\[
dist(x, S_{\gamma,A,B}^{-1}(0)) \leq \alpha^{-1}(1 + \gamma \beta) \text{dist}(0, Fu) + \gamma \beta \| u - v \| \\
\leq \alpha^{-1}(1 + \gamma \beta) \| \gamma^{-1}(u - v) + (Au - Av) \| + \gamma \beta \| u - v \| \\
\leq \left[ \alpha^{-1}(1 + \gamma \beta)(\gamma^{-1} + \beta) + \gamma \beta \right] \| u - v \| \\
= \left[ \alpha^{-1}(1 + \gamma \beta)(\gamma^{-1} + \beta) + \gamma \beta \right] \text{dist}(0, S_{\gamma,A,B}(x)).
\]

This, along with the arbitrariness of \( x \) in \( \text{dom} S_{\gamma,A,B} \), is equivalent to saying that

\[
S_{\gamma,A,B}^{-1}(w) \subseteq S_{\gamma,A,B}^{-1}(0) + \kappa \| w \| B \quad \text{for} \ w \in X.
\]

Notice that \( S_{\gamma,A,B}^{-1}(0) \) is singleton due to the strong monotonicity of \( F \). Together with the last inclusion, \( S_{\gamma,A,B} \) is metrically subregular at \((x^*, 0) \in \text{gph}S_{\gamma,A,B} \).

Case 2: \( B \) is single-valued and Lipschitz continuous with modulus \( \beta > 0 \). From the closed convexity of \( S_{\gamma,A,B}(x) \) and the expression of \( S_{\gamma,A,B} \), there exist \( u, v, a \in X \) with \((v, a) \in A \) and \( v + \gamma a = u - \gamma Bu \) such that \( u - v = \text{dist}(0, S_{\gamma,A,B}(x)) \) for \( x = v + \gamma Bu \). So,

\[
\gamma^{-1}(u - v) + (Bu - Bu) \in Fu.
\] (36)

By Lemma 3.1(a) and the remark after it, we have \( S_{\gamma,A,B}^{-1}(0) = (I + \gamma B)(F^{-1}(0)) \). Thus,

\[
dist(x, S_{\gamma,A,B}^{-1}(0)) = \text{dist}(v + \gamma Bu, S_{\gamma,A,B}^{-1}(0)) \leq \| v + \gamma Bu - z^* - Bz^* \| \\
\leq \| v - z^* \| + \gamma \beta \| u - z^* \| \\
\leq (1 + \gamma \beta) \| v - z^* \| + \gamma \beta \| u - v \|.
\] (37)
where $z^*$ is an arbitrary point from $\mathcal{F}^{-1}(0)$. Since $\mathcal{F} = \mathcal{A} + \mathcal{B}$ is strongly monotone with constant $\alpha$, $\mathcal{F}^{-1}$ is single-valued and Lipschitz continuous with modulus $\alpha^{-1}$. Let $\overline{w} \in \mathcal{F}v$ be such that $\|\overline{w}\| = \text{dist}(0, \mathcal{F}v)$. Then, from (35) it follows that
\[
\|v - z^*\| = \|\mathcal{F}^{-1}(\overline{w}) - \mathcal{F}^{-1}(0)\| \leq \alpha^{-1}\|\overline{w}\| = \alpha^{-1}\text{dist}(0, \mathcal{F}v).
\]
Combining this inequality with inequalities (36) and (37), we have that
\[
\text{dist}(x, S_{\gamma,\mathcal{A},\mathcal{B}}^{-1}(0)) \leq \alpha^{-1}(1 + \gamma \beta)\text{dist}(0, \mathcal{F}v) + \gamma \beta\|u - v\|
\leq \alpha^{-1}(1 + \gamma \beta)\gamma^{-1}(u - v) + (\mathcal{B}v - \mathcal{B}u)\| + \gamma \beta\|u - v\|
\leq \left[\alpha^{-1}(1 + \gamma \beta)(\gamma^{-1} + \beta) + \gamma \beta\right]\|u - v\|
= \left[\alpha^{-1}(1 + \gamma \beta)(\gamma^{-1} + \beta) + \gamma \beta\right]\text{dist}(0, S_{\gamma,\mathcal{A},\mathcal{B}}(x)).
\]
The rest arguments are similar to those for Case 1, and we omit them. \qed