Chapter 1

Monotone Mappings

In this chapter, our goal is to review some important results regarding monotone mappings; most of them will be used in later chapters.

A mapping (also called operator) $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is called set-valued if it corresponds each point $x \in \mathcal{H}$ to a (possibly empty) subset $T(x)$ of $\mathcal{H}$. The inverse of T, denoted by $T^{-1}(y) := \{x \mid y \in T(x)\}$, always exists and, we always have $(T^{-1})^{-1} = T$. The domain of T is defined by

$$\mathrm{dom}T := \{x \mid T(x) \neq \emptyset\}$$

and the range of T is defined by

$$\mathrm{rge}T := \{y \mid \exists\, x\colon\ y \in T(x)\} = \mathrm{dom}T^{-1}$$

When T is single valued, we may write $T(x) = y$ instead of $T(x) = \{y\}$.

1.1. Maximality of Monotone Mappings. Let's begin with the definition of (maximal) monotonicity.

Definition 1.1.1. *A mapping $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is called monotone if*

$$\langle x - x',\ y - y'\ \rangle \geq 0 \quad \forall y \in T(x),\ \forall y' \in T(x').$$

It is called maximal monotone if its graph set $\{(x, y) \in \mathcal{H} \times \mathcal{H} \mid y \in T(x)\}$ can not be enlarged without destroying monotonicity.

Of course, if the mapping T is single-valued, monotone then the relation above may be of the form $\langle x - x',\ T(x) - T(x')\,\rangle \geq 0$ for all x and x'.

An important example of monotone mappings is the subdifferential ∂f of a proper convex function f. If in addition f is closed then ∂f is maximal. Of particular interest is the indicator function δ_C of a convex subset C in $\mathcal{H}$:

$$\delta_C(x) := \begin{cases} 0 & \text{if } x \in C, \\ \infty & \text{if } x \notin C. \end{cases}$$

The closedness of δ_C is equivalent to the closedness of C. Thus, the *normal cone* mapping $N_C := \partial \delta_C$ to C is maximal monotone when C is closed convex.

Another important example of monotone mappings is continuous, monotone mapping. If a monotone mapping is continuous in all of $\mathcal{H}$, then it is maximal monotone.

The next is the maximality criterion for the sum of two maximal monotone mappings; see [38, p. 557] for details.

Proposition 1.1.1. *Let $B, F : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximal monotone. If*

$$\mathit{ri(domB)} \cap \mathit{ri(domF)} \neq \emptyset,$$

then $T := B + F$ is maximal monotone.

Deep results on this were obtained by Pennanen most recently; see [35] and the references cited therein.

1.2. Resolvants of Monotone Mappings. We now introduce nonexpansive mappings, which play a fundamental role in the study of monotone mappings. A single-valued mapping N is called *nonexpansive* if

$$\|N(x) - N(x')\| \leq \|x - x'\| \quad \forall x, x' \in \mathcal{H}.$$

In his landmark paper [29], Minty established the one-to-one correspondence between monotone and nonexpansive mappings and used it to prove the following important facts.

Proposition 1.2.1. *Let $T : \mathcal{H} \rightrightarrows \mathcal{H}$ be monotone, and let $\lambda > 0$. Then the mapping $(I + \lambda T)^{-1}$ is monotone and nonexpansive. Moreover, T is maximal if and only if $dom(I + \lambda T)^{-1} = \mathcal{H}$, equivalently, $rge(I + \lambda T) = \mathcal{H}$.*

The mappings $J_{\lambda T} := (I + \lambda T)^{-1}$ for $\lambda > 0$ above are called the resolvants of T. Of course, this definition is also applicable to general mappings.

An interesting and important facts is that the resolvant of the normal cone mapping to a nonempty closed convex set C is equivalent to nothing but the orthogonal projection onto C: $P_C = (I + \lambda N_C)^{-1}$.

For the inverse mappings T^{-1}, their resolvants, also called *the Yosida regularizations of T,* are related to the resolvants of T by the following inverse-resolvant identity [38, p. 540]

$$(I + \lambda T^{-1})^{-1} = \lambda^{-1}(I - (I + \lambda T)^{-1}) \quad \forall \lambda > 0.$$

This identity is not only beautiful in form but also useful in practice. For example, in numerical tests, when $T = N_C$, we can evaluate $I - P_C$ rather than $(I + N_C^{-1})^{-1}$.

The following result is due to Lions and Mercier [1979].

Proposition 1.2.2. *Let B and F be two maximal monotone mappings from all of $\mathcal{H}$ into itself. Then the mapping*

$$N_{\lambda,B,F} := J_{\lambda B} \circ (2J_{\lambda F} - I) + (I - J_{\lambda F})$$

is nonexpansive. Furthermore, it has the following property for all x, $x' \in \mathcal{H}$

$$\|N_{\lambda,B,F}(x) - N_{\lambda,B,F}(x')\|^2 \leq \langle x - x',\, N_{\lambda,B,F}(x) - N_{\lambda,B,F}(x')\rangle.$$

1.3. Strong Monotonicity and Lipschitz Continuity. Strong Monotonicity of set-valued mappings dates back to Zarantonello [45]. It plays a crucial role in earlier analysis of convergence rate of a class of numerical methods, also in convergence proof of some methods. Lipschitz continuity also plays a similar role in numerical analysis.

Definition 1.3.1. *A mapping $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is called strongly monotone with modulus $\mu > 0$ if*

$$\langle x - x', y - y'\rangle \geq \mu \, \|x - x'\|^2 \quad \forall y \in T(x),\ \forall y' \in T(x').$$

Indeed, it is equivalent to the condition that $T - \mu I$ is monotone.

Definition 1.3.2. *Let $T : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximal monotone. The inverse mapping T^{-1} is called strongly monotone with modulus ν if there exists some $\nu > 0$ such that*

$$\langle x - x', y - y'\rangle \geq \nu \, \|y - y'\|^2 \quad \forall y \in T(x),\ \forall y' \in T(x').$$

The strong monotonicity of the inverse mapping is also called the *Dunn property* or the *co-coercivity* of T in the literature.

Definition 1.3.3. *A mapping $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is called Lipschitz continuous with modulus l if there exists some $l > 0$ such that*

$$\|y - y'\| \leq l \, \|x - x'\| \quad \forall y \in T(x), \ \forall y' \in T(x').$$

It is immediate from the case $x = x'$ that Lipschitz continuous mappings must be single-valued, and the relation above may be simplified to

$$\|T(x) - T(x')\| \leq l \, \|x - x'\| \quad \forall x, \ \forall x'.$$

For the case $l \leq 1$, T is called a nonexpansive mapping, just mentioned above. For the case $l < 1$, it is called a contractive mapping.

Proposition 1.3.1. *Let $T : \mathcal{H} \rightrightarrows \mathcal{H}$ be a mapping. Define*

$$\mathcal{N}[T] := \{(x + y, \, x - y) \mid y \in T(x)\}.$$

Then any of the following statements holds

(1) T is monotone if and only if $\mathcal{N}[T]$ is nonexpansive;

(2) T is maximal monotone if and only if $\mathcal{N}[T]$ is a nonexpansive mapping defined on all of $\mathcal{H}$;

This two statements describe the the one-to-one correspondence between monotone and nonexpansive mappings established by Minty, just mentioned above.

1.4. Inclusions of Monotone Mappings. Let $T : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximal monotone mapping. A relevant fundamental problem is to find $x \in \mathcal{H}$ such that

$$0 \in T(x). \tag{1.1}$$

The monotone inclusions of this type include as special cases convex minimization, complementarity problem, monotone variational inequalities.

The solution sets Z of such inclusions have a nice property. As proved by Minty [1964], they must be (possibly empty) closed convex sets.

The next concepts are related to the solution sets Z.

Definition 1.4.1. *Let $T : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximal monotone mapping. T^{-1} is called Lipschitz continuous at the origin if $0 \in T(x)$ has the unique solution z and, there exist $\epsilon > 0$, $\tau > 0$ such that*

$$\forall y \in \mathbb{B}\,(0, \epsilon), \ \forall x \in T^{-1}(y) \quad \|x - z\| \leq \tau \, \|y\|.$$

It is easy to check that, when T is strongly monotone with modulus μ, this relation holds automatically with $\epsilon := +\infty$ and $\tau := \mu^{-1}$.

This concept above was first used by Rockafellar to analyze convergence rate and finite convergence of the PPA. Still, it requires the solution-uniqueness of $0 \in T(x)$ and, thus appears to be restrictive. Later, it was meaningfully generalized by Luque [1984] to the following growth condition.

Definition 1.4.2. *Let $T : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximal monotone mapping. we simply call that T^{-1} satisfies growth condition if there exist $\epsilon > 0$, $\tau > 0$ such that*

$$\forall y \in \mathbb{B}\,(0, \epsilon),\ \forall x \in T^{-1}(y) \quad |x - Z| \leq \tau \,\|\, y \,\|.$$

In contrast, this growth condition is a weaker one. It at least does not require the solution-uniqueness of $0 \in T(x)$ and is known to hold when T is polyhedral; see Ronbison [1981]. When specialized in variational inequalities, as shown by Tseng [2000], it also may be inferred from results on error bounds for monotone variational inequalities.

This chapter ends with a well known result [2, p. 27]; see also [34, 44]. This result in general plays a key role in convergence analysis of some methods for monotone inclusions.

Proposition 1.4.1. *Let $T : \mathcal{H} \rightrightarrows \mathcal{H}$ be any maximal monotone mapping. If $\{x^k\}$ is a sequence in $\mathcal{H}$ bounded in norm and converging weakly to some x, and $\{\omega^k\}$ is a sequence in $\mathcal{H}$ converging strongly to some ω and $\omega^k \in T(x^k)$ for all k, then $\omega \in T(x)$.*

Chapter 2

The Proximal Point Algorithm

2.1. Introduction. In this chapter, we discuss the proximal point algorithm for solving (1.1):

$$x^{k+1} = (I + c_k T)^{-1}(x^k) \quad \forall k \geq 0. \tag{2.1}$$

This algorithm is first proposed by Martinet [27, 28], and then generalized by Rockafellar [37]. Its dual version in the context of convex programming and their variants have been extensively studied; see for example [5, 12, 14, 23, 24, 43] and the references cited therein.

In [37], Rockafellar studied convergence behavior of the following approximate version:

$$x^{k+1} = (I + c_k T)^{-1}(x^k) + \bar{e}^{k+1} \quad \forall k \geq 0, \tag{2.2}$$

where $c_k \geq c > 0$ and $\bar{e}^{k+1}$ is an error. Later, Luque [24] considerably improved results on rate of convergence and finite convergence obtained by Rockafellar.

However, it appears to be impossible to evaluate the resolvant of T in many cases. Therefore, we turn to consider its recently popular approximate version:

$$x^{k+1} + c_k\, T(x^{k+1}) \ni x^k + e^{k+1} \quad \forall k \geq 0,$$

which is equivalent to

$$x^{k+1} = (I + c_k T)^{-1}(x^k + e^{k+1}) \quad \forall k \geq 0, \tag{2.3}$$

where $c_k > c > 0$ and e^{k+1} is an error. Furthermore, for convenience, we henceforth abbreviate the proximal point algorithm described by (2.3) as the PPA.

In this chapter, one major goal is to show that when the PPA is implemented with the error criterion described by (2.17), *super*linear convergence can be guaranteed under the growth condition. The result itself on this is *not new.* However, its error criterion is better than the one suggested by Luque and possibly best. Especially, its proof techniques are new and powerful. Aside from this, we also discuss other two error criteria:

Criterion 2.1.1. $\|e^{k+1}\| \le \varepsilon_k\,, \quad \sum_{k=0}^{\infty} \varepsilon_k^2 < \infty\,;$

Criterion 2.1.2. $\|e^{k+1}\| \le \delta\,\|x^k - x^{k+1}\|,\ \delta > 0.$

Assumption. The solution of $T(x) \ni 0$, say Z, is assumed to be nonempty in this chapter.

2.2. Convergence. This section mainly addresses the issue of weak convergence of the PPA with Criterion 2.1.1. To this end, we need to make the following two assumptions.

Assumption 2.2.1. There exist $\epsilon > 0$, $\tau > 0$ such that

$$\forall y \in \mathbb{B}\,(0, \epsilon),\ \forall x \in T^{-1} y \quad |x - Z| \le \tau\,\|\,y\,\|. \tag{2.4}$$

Remark. This assumption is reasonably mild. For instance, for the case of monotone variational inequalities, corresponding to $T = F + N_C$ and F single-valued, continuous and monotone on C, it follows from [44, p. 440] that this assumption can be inferred from well known results on local error bounds for variational inequalities; see [11, 26, 33, 36]. For the case of convex minimization, corresponding to $T := \partial f$, where f is closed proper convex, Assumption 2.2.1 can be judged by Rockafellar [37, Proposition 6, Proposition 7]. Certainly, if $f : \mathcal{R}^n \to \mathcal{R}$ is, in addition, strongly convex with modulus $\alpha > 0$, then we always have

$$\|x - z\| \le \alpha^{-1}\|\,y\,\|, \quad \forall x \in \mathcal{R}^n,\ \forall y \in T(x),$$

where z is the unique minimizer of f. That is to say, this assumption is globally valid for the case of strong convexity of f, which, certainly, is entirely equivalent to strong monotonicity of $T := \partial f$. See [37, p. 891] for a detailed exposition. All these suffice to show that Assumption 2.2.1 is not rather restrictive but mild. More importantly, this assumption is weaker than the one [37] as it does not require the solution set to be a singleton. It has been applied by Luque [24] to analyze asymptotic convergence of the PPA.

Assumption 2.2.2. There exist $C > 0$, $C' > 0$ such that for all $k \ge 0$,

$$\|x^k - x^{k+1}\| \le C,\ \|e^k\| \le C'. \tag{2.5}$$

Remark. At first glance, it may appear rather strange that this assumption involves x^k, which can not be determined before using the algorithm. Still, one may knows a priori that it will hold, for example if dom$T$ is bounded; see [37, Remark 1]. Furthermore, when the PPA is implemented approximately with summable errors, the resulting sequence satisfies this assumption; see [37, Thm. 1]. This fact will be used in Corollary 2.3.1 below.

Remark. The common goal of both Assumption 2.2.1 and Assumption 2.2.2 is nothing but to guarantee the validity of (2.10). Since, when T is strongly monotone, it is valid globally, as showed in both Corollary 2.2.1 and Corollary 2.2.2 below, the two assumptions are removed for analysis of the convergence and the rate of convergence in this setting.

The result below is due to Eckstein [7, Lemma 2].

Lemma 2.2.1. *Let $\{x^k\}$ be the sequence generated by (2.3). Then for any $z \in Z$ and all $k \geq 0$:*

$$\|x^{k+1} - z\|^2 \leq \|x^k - z\|^2 - \|x^k - x^{k+1}\|^2 + 2\,\langle e^{k+1}, x^{k+1} - z\rangle. \tag{2.6}$$

Note that when all errors are zero the inequality above can be viewed as a special case of [37, part (c) of Proposition 1, p. 881].

Lemma 2.2.2. *Suppose that Assumptions 2.2.1–2.2.2 hold. Choose any starting point $x^0 \in \mathcal{H}$. Let $\{x^k\}$ be the sequence generated by (2.3) with either Criterion* 2.1.1 *or Criterion* 2.1.2*. Suppose that there exists an index K such that $c_k \geq c \geq \epsilon^{-1}(C + C')$ for all $k \geq K$. Then*

$$\|x^{k+1} - z^k\| \leq (1 + \tau c^{-1})(\|x^k - x^{k+1}\| + \|e^{k+1}\|), \quad \forall k \geq K, \tag{2.7}$$

where $z^k = argmin_{z\in Z}\|x^k - z\|$.

Proof. Assumption 2.2.2 says that for all $k \geq 0$

$$\|x^k - x^{k+1}\| \leq C, \quad \|e^{k+1}\| \leq C'. \tag{2.8}$$

Since, by hypothesis, $c_k \geq \epsilon^{-1}(C + C')$ for all $k \geq K$, then it is easy to check that

$$\|c_k^{-1}(x^k - x^{k+1} + e^{k+1})\| \leq c_k^{-1}(\|x^k - x^{k+1}\| + \|e^{k+1}\|) \leq \epsilon,\ \forall k \geq K. \tag{2.9}$$

This relation, together with $c_k^{-1}(x^k - x^{k+1} + e^{k+1}) \in T(x^{k+1})$ and Assumption 2.2.1, implies that for all $k \geq K$

$$\begin{aligned} |x^{k+1} - Z| &\leq \tau\,\|c_k^{-1}(x^k - x^{k+1} + e^{k+1})\| \\ &\leq \tau c_k^{-1}(\|x^k - x^{k+1}\| + \|e^{k+1}\|), \end{aligned} \tag{2.10}$$

where z^k satisfies $z^k = \operatorname{argmin}_{z\in Z}\|x^k - z\|$.
On the other hand, since T is maximal monotone, Z is nonempty closed convex [30, 31], hence it follows from the definition of z^k and the nonexpansive property of the orthogonal projection [45, Eq. (1.8)] that we have

$$\|z^{k+1} - z^k\| \le \|x^k - x^{k+1}\|,$$

which, together with (2.10), implies that for all $k \ge K$

$$\begin{aligned}
\|x^{k+1} - z^k\| &\le \|x^{k+1} - z^{k+1}\| + \|z^{k+1} - z^k\| \\
&= |x^{k+1} - Z| + \|z^{k+1} - z^k\| \\
&\le \tau c_k^{-1}(\|x^k - x^{k+1}\| + \|e^{k+1}\|) + \|x^k - x^{k+1}\| \\
&\le (1 + \tau c_k^{-1})\|x^k - x^{k+1}\| + \tau c_k^{-1}\|e^{k+1}\|.
\end{aligned}$$

Since $c_k \ge c$ for all $k \ge K$, we have

$$\|x^{k+1} - z^k\| \le (1 + \tau c^{-1})(\|x^k - x^{k+1}\| + \|e^{k+1}\|), \quad \forall k \ge K. \tag{2.11}$$

This completes the proof of Lemma 2.2.2. We now proceed with the main convergence results.

Theorem 2.2.1. *Suppose that Assumptions 2.2.1–2.2.2 hold. Choose any starting point $x^0 \in \mathcal{H}$. Let $\{x^k\}$ be the sequence generated by the PPA with Criterion 2.1.1. Suppose that there exists an index K such that $c_k \ge c \ge \epsilon^{-1}(C + C')$ for all $k \ge K$. Then $\{|x^k - Z|\}$ converges to zero.*

Proof. It follows from (2.11) from that for all $k \ge K$ and $0 < \lambda < 1$

$$\begin{aligned}
& \langle e^{k+1},\ x^{k+1} - z^k\rangle \\
\le\ & \|e^{k+1}\|\|x^{k+1} - z^k\| \\
\le\ & (1 + \tau c^{-1})\|e^{k+1}\|(\|x^k - x^{k+1}\| + \|e^{k+1}\|) \\
\le\ & \frac{1}{2}(1-\lambda)\|x^k - x^{k+1}\|^2 + \frac{1}{2}(1-\lambda)^{-1}(1 + \tau c^{-1})^2\|e^{k+1}\|^2 + (1 + \tau c^{-1})\|e^{k+1}\|^2 \\
=\ & \frac{1}{2}(1-\lambda)\|x^k - x^{k+1}\|^2 + (\frac{1}{2}(1-\lambda)^{-1}(1 + \tau c^{-1})^2 + (1 + \tau c^{-1}))\|e^{k+1}\|^2.
\end{aligned}$$

Putting this relation into (3.7) yields

$$\begin{aligned}
& \|x^{k+1} - z^k\|^2 \\
\le\ & \|x^k - z^k\|^2 - \|x^k - x^{k+1}\|^2 + 2\,\langle e^{k+1},\ x^{k+1} - z^k\rangle \\
\le\ & \|x^k - z^k\|^2 - \lambda\|x^k - x^{k+1}\|^2 + ((1-\lambda)^{-1}(1 + \tau c^{-1})^2 + 2(1 + \tau c^{-1}))\|e^{k+1}\|^2,
\end{aligned}$$

which further implies

$$|x^{k+1} - Z|^2 \leq |x^k - Z|^2 - \lambda \|x^k - x^{k+1}\|^2 + \kappa \, \|e^{k+1}\|^2, \tag{2.12}$$

where $\kappa := (1-\lambda)^{-1}(1+\tau c^{-1})^2 + 2(1+\tau c^{-1})$.
In view of the inequality $(a+b)^2 \leq 2(a^2+b^2)$ for $a, b \geq 0$, it follows from (2.10) that

$$\begin{aligned} |x^{k+1} - Z|^2 &\leq \tau^2 c_k^{-2} \, (\|x^k - x^{k+1}\| + \|e^{k+1}\|)^2 \\ &\leq 2\tau^2 c_k^{-2} \, (\|x^k - x^{k+1}\|^2 + \|e^{k+1}\|^2). \end{aligned}$$

Combining this relation with (4.23) and eliminating the term $\|x^k - x^{k+1}\|^2$ yield

$$(1 + \frac{1}{2}\tau^{-2}\lambda c_k^2)|x^{k+1} - Z|^2 \leq |x^k - Z|^2 + (\lambda + \kappa)\|e^{k+1}\|^2.$$

Since $c_k \geq c$ for all $k \geq K$, we have

$$|x^{k+1} - Z|^2 \leq \theta \, |x^k - Z|^2 + \kappa' \, \|e^{k+1}\|^2, \tag{2.13}$$

where

$$\theta := (1 + \frac{1}{2}\tau^{-2}\lambda c^2)^{-1} < 1,$$

and

$$\begin{aligned} \kappa' : &= (\lambda + \kappa)(1 + \frac{1}{2}\tau^{-2}\lambda c^2)^{-1} \\ &= (\lambda + (1-\lambda)^{-1}(1+\tau c^{-1})^2 + 2(1+\tau c^{-1}))(1 + \frac{1}{2}\tau^{-2}\lambda c^2)^{-1}. \end{aligned}$$

Summing up the two sides of (2.13) and rearranging the terms, for any given $L \geq 1$ we have

$$\begin{aligned} &(1-\lambda) \sum_{i=K}^{K+L} |x^i - Z|^2 \\ \leq \; & |x^K - Z|^2 - \theta \, |x^{K+L} - Z|^2 + \kappa' \sum_{i=K+1}^{K+L} \|e^i\|^2 \\ \leq \; & |x^K - Z|^2 + \kappa' \sum_{i=K+1}^{K+L} \|e^i\|^2, \end{aligned}$$

which, together with Criterion 2.1.1, implies that $\sum_{i=K}^{K+L} |x^i - Z|^2$ is bounded for any given $L \geq 1$, and thus we have $\{|x^k - Z|\}$ converges to zero. Furthermore, if z is the unique solution of $T(x) \ni 0$, then $\{\|x^k - z\|\}$ converges to zero, i.e., the iterate sequence $\{x^k\}$ converges strongly to the unique solution z. $\square$

For the case of strong monotonicity, we can get to the following stronger results.

Corollary 2.2.1. *Choose any starting point $x^0 \in \mathcal{H}$. Let $\{x^k\}$ be the sequence generated by the PPA with Criterion 2.1.1. Suppose that T is strongly monotone and lim $\inf_{k\to\infty} c_k > 0$. Then $\{x^k\}$ converges strongly to the unique solution of $T(x) \ni 0$.*

Proof. Let T be strongly monotone with modulus $\mu > 0$. Then, in view of [37, Prop. 5], for the unique solution of $T(x) \ni 0$, say z, we have

$$\|x - z\| \le \mu^{-1}\|y\|, \quad \forall x,\ \forall y \in T(x),$$

which, together with $c_k^{-1}(x^k - x^{k+1} + e^{k+1}) \in T(x^{k+1})$, implies

$$\begin{aligned} \|x^{k+1} - z\| &\le \mu^{-1}\|c_k^{-1}(x^k - x^{k+1} + e^{k+1})\| \\ &\le \mu^{-1}c_k^{-1}(\|x^k - x^{k+1}\| + \|e^{k+1}\|) \\ &\le \mu^{-1}c\,(\|x^k - x^{k+1}\| + \|e^{k+1}\|), \end{aligned}$$

where $c := \liminf_{k\to\infty} c_k > 0$. Note that this relation can play the same role as (2.10) in the proof of Theorem 2.2.1, so the proof of this corollary can be proceeded in the same way as that of Theorem 2.2.1 as far as the corresponding remainder is concerned. □

2.3. Rate of Convergence. This section mainly addresses the issue of *super*linear convergence of the PPA with Criterion 2.1.2.

Theorem 2.3.1 *Suppose that Assumptions 2.2.1–2.2.2 hold. Choose any starting point $x^0 \in \mathcal{H}$. Let $\{x^k\}$ be the sequence generated by the PPA with Criterion 2.1.2. Suppose that there exists an index K such that $c_k \ge c \ge \epsilon^{-1}(C + C')$ for all $k \ge K$. Let $\delta \le \frac{1}{3}(1 + \tau c^{-1})^{-1}$. Then $\{|x^k - Z|\}$ converges to zero linearly. If in addition $c_k \uparrow +\infty$, then the convergence is superlinear.*

Proof. By Lemma 2.2.1, we have

$$\|x^{k+1} - z^k\|^2 \le \|x^k - z^k\|^2 - \|x^k - x^{k+1}\|^2 + 2\,\langle e^{k+1},\ x^{k+1} - z^k\rangle.$$

Since Assumptions 2.2.1–2.2.2 hold, it follows from Lemma 2.2.2 that $c_k \ge c \ge \epsilon^{-1}(C + C')$ for all $k \ge K$ implies that

$$\|x^{k+1} - z^k\| \le (1 + \tau c^{-1})(\|x^k - x^{k+1}\| + \|e^{k+1}\|), \quad \forall k \ge K.$$

Combining the two relations with Criterion 2.2 yields

$$\begin{aligned}
& \|x^{k+1} - z^k\|^2 \\
\leq\ & \|x^k - z^k\|^2 - \|x^k - x^{k+1}\|^2 + 2\,\langle e^{k+1},\ x^{k+1} - z^k\rangle \\
\leq\ & \|x^k - z^k\|^2 - \|x^k - x^{k+1}\|^2 + 2\,\|e^{k+1}\|\|x^{k+1} - z^k\| \\
\leq\ & \|x^k - z^k\|^2 - \|x^k - x^{k+1}\|^2 + 2\delta(1+\delta)(1+\tau c^{-1})\|x^k - x^{k+1}\|^2 \\
=\ & \|x^k - z^k\|^2 - \tilde{\nu}\,\|x^k - x^{k+1}\|^2,
\end{aligned}$$

where $\tilde{\nu} := 1 - 2\delta(1+\delta)(1+\tau c^{-1})$. (Since, by hypothesis, $\delta \leq \frac{1}{3}(1+\tau c^{-1})^{-1}$, then $\tilde{\nu}$ is a positive number.)

So, we further have

$$|x^{k+1} - Z|^2 \leq |x^k - Z|^2 - \tilde{\nu}\,\|x^k - x^{k+1}\|^2. \tag{2.14}$$

This relation guarantees a sufficient descent of the distance of the iterate to the solution set at each step.

On the other hand, from (2.10) and Criterion 2.1.2, we have

$$|x^{k+1} - Z| \leq \tau(1+\delta)c^{-1}\|x^k - x^{k+1}\|.$$

Combining this relation with (2.14) and eliminating the term $\|x^k - x^{k+1}\|$ yield

$$\tau^{-2}(1+\delta)^{-2}c^2\ \tilde{\nu}\,|x^{k+1} - Z|^2 \leq |x^k - Z|^2 - |x^{k+1} - Z|^2.$$

We further have

$$|x^{k+1} - Z| \leq \theta\,|x^k - Z|, \qquad \forall k \geq K,$$

where

$$\theta := (1+\nu c^2)^{-\frac{1}{2}}, \tag{2.15}$$

$$\nu := \tau^{-2}(1+\delta)^{-2}\tilde{\nu} = \tau^{-2}(1+\delta)^{-2}(1 - 2\delta(1+\delta)(1+\tau c^{-1})) \tag{2.16}$$

with

$$\delta \leq \frac{1}{3}(1+\tau c^{-1})^{-1}.$$

Obviously, if $c_k \uparrow +\infty$ then the distance sequence converges $\{|x^k - Z|\}$ superlinearly to zero. If, in addition, the solution of $T(x) \ni 0$ is unique then the sequence $\{x^k\}$ converges strongly to this solution linearly. □

As a direct consequence, we have

Corollary 2.3.1. *Choose any starting point $x^0 \in \mathcal{H}$. Let $\{x^k\}$ be the sequence generated by the PPA with the error criterion*

$$\|e^k\| \leq \min\{\varepsilon_k,\, \delta\,\|x^{k+1} - x^k\|\}, \tag{2.17}$$

and $\delta \leq \frac{1}{3}(1 + \tau c^{-1})^{-1}$. Suppose that Assumption 2.2.1 holds and there exists an index K such that $c_k \geq c \geq \epsilon^{-1}(C + C')$ for all $k \geq K$. Then $\{|x^k - Z|\}$ converges to zero at a linear rate bounded from above by θ described by (2.15). If in addition $c_k \uparrow +\infty$, then the convergence is superlinear.

Proof. Since the error criterion (2.17) implies that all errors are summable, which implies [37, Thm. 1] that the iterate sequence is bounded, then, the conditions of Theorem 2.3.1 subsume that of this corollary, and thus the conclusions in that theorem are in force. □

Remark. When the exact version (2.1) (i.e. $\delta = 0$) is implemented, this bound for linear rate, which is $\theta := (1 + \tau^{-2}c^2)^{-\frac{1}{2}}$, can be achieved; see [24, p. 282-283]. Furthermore, Assumption 2.2.1 is not necessary; see [20, 24].

Below we discuss an auxiliary error criterion:

$$\|e^{k+1}\| \leq \min\{\varepsilon_k,\, \eta_k\,\|x^{k+1} - x^k\|\}, \quad \sum \varepsilon_k < +\infty,\ \varepsilon_k \geq 0,\ \forall k \geq 0,$$

where $\{\eta_k\}$ is any given nonnegative sequence such that $\eta_k \to 0$. Obviously, it removes the traditional assumption that $\sum \eta_k < +\infty$ and, can also guarantee *super*linear convergence of the PPA. In fact, by hypothesis, since $\eta_k \to 0$, then there exists K' such that for all $k \geq K'$ we have $\eta_k \leq \eta$. Therefore, when $k \geq K' + K$, the conclusion of Corollary 2.3.1 is in force.

For the case of strong monotonicity, we can get to the following stronger results on convergence rate of the PPA with Criterion 2.1.2.

Corollary 2.3.2. *Choose any starting point $x^0 \in \mathcal{H}$. Let $\{x^k\}$ be the sequence generated by the PPA with Criterion 2.1.2. Suppose that T is strongly monotone and lim $\inf_{k\to\infty} c_k > 0$. Then $\{x^k\}$ converges strongly to the unique solution of $T(x) \ni 0$ linearly. If in addition $c_k \uparrow +\infty$, then the convergence is superlinear.*

Note that for this two corollary, Assumptions 2.2.1–2.2.2 are removed since the Lipschitz property of T^{-1} is globally valid.

Very interestingly, for such a special case of strong monotonicity, when we follow the line of argument pioneered by Rockafellar (see [37, p. 879-880]), we can obtain a stronger result than Corollary 2.3.2.

Let us begin with a technical lemma.

Lemma 2.3.1. *Suppose that T is strongly monotone with modulus $\mu > 0$. Let $\{x^k\}$ be the sequence generated by the PPA. Then for all $k \geq 0$*

$$\|x^{k+1} - z\| \leq (1 + c_k\mu)^{-1}(\|x^k - z\| + \|e^{k+1}\|), \tag{2.18}$$

where z is the unique solution of $T(x) \ni 0$.

Proof. Denote $T' := T - \mu I$. By (2.3), we have

$$\begin{aligned} x^{k+1} &= (I + c_kT)^{-1}(x^k + e^{k+1}) \\ &= ((1 + c_k\mu)I + c_kT')^{-1}(x^k + e^{k+1}) \\ &= (1 + c_k\mu)^{-1}(I + (1 + c_k\mu)^{-1}c_kT')^{-1}(x^k + e^{k+1}). \end{aligned}$$

Since T is strongly monotone with modulus $\mu > 0$, then, in view of [37, p. 879], $T' := T - \mu I$ is monotone and $J_{\beta_k,\,T'} := (I + (1 + c_k\mu)^{-1}c_kT')^{-1}$ is nonexpansive for $\beta_k := (1 + c_k\mu)^{-1}c_k > 0$.
On the other hand, since z is the solution of $T(x) \ni 0$, then it is easy to see that

$$z = (1 + c_k\mu)^{-1}\, J_{\beta_k,\,T'}\,(z).$$

So, in view of the nonexpansive property of this resolvant $J_{\beta_k,\,T'}$, we have

$$\begin{aligned} & \|x^{k+1} - z\| \\ =\; & (1 + c_k\mu)^{-1}\|J_{\beta_k,\,T'}(x^k + e^{k+1}) - J_{\beta_k,\,T'}(z)\| \\ \leq\; & (1 + c_k\mu)^{-1}\|x^k - z + e^{k+1}\| \\ \leq\; & (1 + c_k\mu)^{-1}(\|x^k - z\| + \|e^{k+1}\|). \end{aligned}$$

This completes the proof of Lemma 2.3.1. □

Theorem 2.3.2. *Suppose that T is strongly monotone with modulus $\mu > 0$. Let $\{x^k\}$ be the sequence generated by the PPA with $\|e^{k+1}\| \leq \delta'\,\|x^{k+1} - x^k\|$, $0 \leq \delta' < 1$. Then for all $k \geq 0$*

$$\|x^{k+1} - z\| \leq (1 + c_k\alpha)^{-1}(1 + 2\delta'(1 - \delta')^{-1})\|x^k - z\|,$$

where z is the unique solution of $T(x) \ni 0$.
Furthermore, if there exists an index K such that $c_k \geq c > 2\mu^{-1}\delta'(1 - \delta')^{-1}$ for all $k \geq K$ then

$$\|x^{k+1} - z\| \leq \theta'\,\|x^k - z\|,$$

where $\theta' := (1+c\mu)^{-1}(1+2\delta'(1-\delta')^{-1}) < 1.$
In particular, if $c_k \uparrow \infty$*, then*

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

Proof. Since $\{x^k\}$ is the sequence generated by the PPA, we have

$$x^{k+1} = (I + c_kT)^{-1}(x^k + e^{k+1}).$$

Thus it follows from the nonexpansive property of $(I + cT)^{-1}$ and this criterion that

$$\begin{aligned}
&\|x^{k+1} - x^k\| \\
=\ & \|(I + c_kT)^{-1}(x^k + e^{k+1}) - x^k\| \\
\le\ & \|(I + c_kT)^{-1}(x^k + e^{k+1}) - (I + c_kT)^{-1}(x^k)\| + \|(I + c_kT)^{-1}(x^k) - x^k\| \\
\le\ & \|(I + c_kT)^{-1}(x^k) - x^k\| + \|e^{k+1}\| \\
\le\ & \|(I + c_kT)^{-1}(x^k) - x^k\| + \delta'\,\|x^{k+1} - x^k\|. \qquad (2.19)
\end{aligned}$$

So, we have

$$\|x^{k+1} - x^k\| \le (1-\delta')^{-1}\|(I + c_kT)^{-1}(x^k) - x^k\|.$$

This fact shows that

$$\|e^{k+1}\| \le \delta'\,\|x^{k+1} - x^k\| \le \delta'(1-\delta')^{-1}\|(I + c_kT)^{-1}(x^k) - x^k\|,$$

which, together with

$$\begin{aligned}
\|(I + c_kT)^{-1}(x^k) - x^k\| &= \|(I + c_kT)^{-1}(x^k) - (I + c_kT)^{-1}(z) - (x^k - z)\| \\
&\le \|(I + c_kT)^{-1}(x^k) - (I + c_kT)^{-1}(z)\| + \|x^k - z\| \\
&\le 2\,\|x^k - z\|,
\end{aligned}$$

implies that

$$\|e^{k+1}\| \le 2\,\delta'(1-\delta')^{-1}\|x^k - z\|.$$

Combining this relation with Lemma 2.3.1, we have

$$\|x^{k+1} - z\| \le (1 + c_k\mu)^{-1}(1 + 2\,\delta'\,(1-\delta')^{-1})\,\|x^k - z\|.$$

Thus, it is not very difficult to check that the conclusions of this theorem is valid.

This completes the proof of this theorem. □

Theorem 2.3.2 shows that for the case of strong monotonicity, if c_k is sufficiently large the iterate sequence converges to the unique solution of $T(x) \ni 0$, not only strongly, but also at least as fast as the linear rate. Furthermore, if $c_k \uparrow \infty$, the convergence is *super*linear.

This section ends with a partial converse to Corollary 2.3.1. Moreover, a partial converse to Theorem 2.3.1 can be easily obtained in a similar way, and thus is omitted.

Theorem 2.3.3. *Choose any starting point $x^0 \in \mathcal{H}$. Let $\{x^k\}$ be the sequence generated by the PPA with (2.17) with $\delta < 1$. Suppose that $c_k \uparrow c_\infty < +\infty$, and suppose that there exist $\tau > 0$, $\epsilon > 0$ such that*

$$\forall y \in \mathbb{B}(0, \epsilon),\ \forall x \in T^{-1}y \quad |x - Z| \geq \tau \, \|y\|. \tag{2.20}$$

If $\{x^k\}$ does not converge to Z in a finite steps. Then

$$\liminf_{k\to\infty} \frac{|x^{k+1} - Z|}{|x^k - Z|} = 1.$$

Proof. Choose some fixed $\tau > 0$. A discussion similar to the proof of Corollary 2.3.1 yields that $\{x^k - x^{k+1}\}$ and $\{e^{k+1}\}$ are bounded. Thus, $c_k \uparrow c_\infty < +\infty$ shows that there exists some K such that

$$\|c_k^{-1}(x^k - x^{k+1} + e^{k+1}\| \leq \epsilon, \quad \forall k \geq K.$$

In view of (2.20) and $c_k^{-1}(x^k - x^{k+1} + e^{k+1}) \in T(x^{k+1})$, we have

$$|x^{k+1} - Z| \geq \tau \, \|c_k^{-1}(x^k - x^{k+1} + e^{k+1})\|,$$

which, together with $\|e^{k+1}\| \leq \delta \|x^k - x^{k+1}\|$, $\delta < 1$, implies that

$$|x^{k+1} - Z| \geq \tau c_k^{-1}(1-\delta)\|x^k - x^{k+1}\|. \tag{2.21}$$

On the other hand, it is easy to check that

$$\|x^k - x^{k+1}\| \geq |x^k - Z| - |x^{k+1} - Z|.$$

Combining this relation with (2.21) and eliminating the term $\|x^k - x^{k+1}\|$ yield

$$c_k \tau^{-1}(1-\delta)^{-1}|x^{k+1} - Z| \geq |x^k - Z| - |x^{k+1} - Z|.$$

So, we further have

$$\liminf_{k\to\infty} \frac{|x^{k+1} - Z|}{|x^k - Z|} \geq \lim_{k\to\infty} \frac{1}{1 + c_k \tau^{-1}(1-\delta)^{-1}} = \frac{1}{1 + c_\infty \tau^{-1}(1-\delta)^{-1}} \, .$$

Since τ can be arbitrarily large, the theorem follows. $\square$

This theorem says that if such hypothesis hold then $\{|x^k - Z|\}$ can not converges to zero faster than sublinearly; see [24, Theorem 4.1] for a related result.

2.4. Applications. This section mainly exemplifies applications of the PPA.

Example 2.4.1. When specialized in the case of convex minimization where $T := \partial f$, subdifferential of some closed proper convex function f, the PPA corresponds to

$$0 \in \lambda_k \partial f(x^{k+1}) + x^{k+1} - x^k,$$

which is equivalent to

$$x^{k+1} = \text{argmin}\{\lambda_k f(x) + \frac{1}{2}\|x - x^k\|^2\}.$$

The method described above is the very method proposed by Martinet for convex minimization.

Example 2.4.2. When specialized in the case of monotone variational inequalities where $T := N_C + F$, the PPA corresponds to

$$0 \in \lambda_k(N_C(x^{k+1}) + F(x^{k+1})) + x^{k+1} - x^k,$$

which is equivalent to

$$x^k - \lambda_k F(x^{k+1}) \in (I + \lambda_k N_C)x^{k+1} \quad \Leftrightarrow \quad x^{k+1} = (I + \lambda_k N_C)^{-1}[x^k - \lambda_k F(x^{k+1})].$$

So, we have

$$x^{k+1} = P_C[x^k - \lambda_k F(x^{k+1})].$$

Example 2.4.3. We consider the case of evolution equations:

$$0 \in \frac{\partial u(t)}{\partial t} + Tu(t) \qquad u(0) = x^0,$$

where $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is maximal monotone.

Discretizing this partial differential equation yields

$$0 \in \frac{x^{k+1} - x^k}{\lambda_k} + Tx^{k+1}$$

and thus we obtain

$$0 \in (I + \lambda_k T)x^{k+1} - x^k \quad \Leftrightarrow \quad x^{k+1} = (I + \lambda_k T)^{-1}x^k.$$

That is, we can use the PPA to solve the evolution equations above.

Chapter 3

Classical Splitting Methods

3.1. Introduction. In the last chapter, we discussed the PPA for maximal monotone mappings, and showed under standard conditions its *super*linear convergence. However, such theoretically nice convergence properties do not guarantee that the PPA is practically implementable method. This is because that in many cases the resolvants involved in the PPA are not very easy to evaluate. As such, when the PPA is implemented, the number of iterations is small, but, in general, the computational cost per iteration is too much expansive !

One alternative is to resort to so-called *splitting methods* for monotone mappings. Precisely, for the following inclusions

$$0 \in T(x),$$

we do not directly use the PPA while we consider the possible equivalent inclusions

$$0 \in B(x) + F(x), \tag{3.1}$$

where $T := B + F$ with B, F maximal monotone, and $J_{\lambda B}$ and/or $J_{\lambda F}$ are/is relatively easier to evaluate than $J_{\lambda T}$. We can then devise a method that uses only $J_{\lambda B}$ and/or $J_{\lambda F}$, instead of $J_{\lambda T}$. All methods of such procedure are called splitting methods.

3.2. Forward-Backward Splitting — Error Bounds. We first discuss the simplest splitting method – forward-backward splitting method, which may be of the following form:

$$x^{k+1} \in (I + \lambda B)^{-1}(I - \lambda F)x^k \quad \forall k \geq 0.$$

When F is not single-valued, the forward-backward splitting method described above may fail to converge; see [7]. Therefore, we merely consider the case where F is single-valued. Of course, in this case, the forward-backward splitting method for solving (3.1) is of the form:

$$x^{k+1} = (I + \lambda_k B)^{-1}(I - \lambda_k F)x^k \quad \forall k \geq 0, \tag{3.2}$$

where $\lambda_k \geq \lambda > 0$.

The following convergence theorem is due to Chen and Rockafellar [3].

Proposition 3.2.1. *Suppose that $F : \mathcal{H} \rightrightarrows \mathcal{H}$ is Lipschitz continuous, monotone and $T = B + F : \mathcal{H} \rightrightarrows \mathcal{H}$ is strongly monotone. Then any sequence generated by (3.2) strongly converges to the unique solution at a linear rate.*

Now our goal is to study the convergence behavior of the resulting iterates when (3.2) is implemented approximately. To this end, we first introduce the concept of error bounds on forward-backward splitting.

Theorem 3.2.1. *Consider any maximal monotone mappings $B, F : \mathcal{H} \rightrightarrows \mathcal{H}$ such that F is Lipschitz continuous. Let $T := B+F$. Then the following statements are equivalent.*

(a) *There exist $\epsilon' > 0$, $\tau' > 0$ such that*

$$\forall y \in \mathbb{B}\,(0, \epsilon'),\ \forall x \in T^{-1}y \quad |x - Z| \leq \tau' \,\|\, y \,\| \tag{3.3}$$

(b) *There exist $\epsilon > 0$, $\tau > 0$ such that*

$$\forall x \in Z_\lambda(\epsilon) \quad |x - Z| \leq \tau \,\|x - J_\lambda(x)\| \tag{3.4}$$

where $Z_\lambda(\epsilon) := \{x \mid \|x - J_\lambda(x)\| \leq \epsilon\}$ with $J_\lambda := (I + \lambda B)^{-1}(I - \lambda F)$.

Proof. $(b) \Rightarrow (a)$. In fact, for any given $\epsilon' > 0$, it follows from [2, 44] that

$$\forall y \in \mathbb{B}\,(0, \epsilon'),\ \forall x \in T^{-1}(y) \qquad \|x - J_\lambda(x)\| \leq \lambda \,\|\, y \,\|. \tag{3.5}$$

Thus, if in addition $\epsilon' \leq \lambda^{-1}\epsilon$ then

$$\|x - J_\lambda(x)\| \leq \lambda \,\|y\| \leq \lambda\, \epsilon' \leq \epsilon.$$

This shows that for all $y \in \mathbb{B}\,(0, \epsilon')$ and for all $x \in T^{-1}(y)$,

$$\epsilon' \leq \lambda^{-1}\epsilon \quad \Rightarrow \quad x \in Z_\lambda(\epsilon).$$

On the other hand, it follows from (b) that

$$x \in Z_\lambda(\epsilon) \quad \Rightarrow \quad |x - Z| \leq \tau \,\|x - J_\lambda(x)\|.$$

Therefore, combining the results above yields

$$|x - Z| \leq \tau \|x - J_\lambda(x)\| \leq \tau \, \|\, y \,\| \leq \tau' \, \|\, y \,\|,$$

where we let $\tau' \geq \tau\lambda$.

$(a) \Rightarrow (b)$. Since $J_\lambda(x) := (I + \lambda B)^{-1}(I - \lambda F)(x)$,

$$J_\lambda(x) + \lambda B(J_\lambda(x)) \ni x - \lambda F(x).$$

That is,

$$x - J_\lambda(x) - \lambda F(x) \in \lambda B(J_\lambda(x)).$$

Summing up $\lambda F(J_\lambda(x))$ and then multiplying by λ^{-1} both sides of the relation above yield

$$\lambda^{-1}(x - J_\lambda(x)) - F(x) + F(J_\lambda(x)) \in T(J_\lambda(x)). \tag{3.6}$$

Let $y := \lambda^{-1}(x - J_\lambda(x)) - F(x) + F(J_\lambda(x))$. The Lipschitz continuity of F implies that for all $x \in Z_\lambda(\epsilon)$ with $\epsilon \leq (\lambda^{-1} + l)^{-1}\epsilon'$ the following relations hold

$$\|\, y \,\| \leq (\lambda^{-1} + l)\|x - J_\lambda(x)\| \leq (\lambda^{-1} + l)\epsilon \leq \epsilon'.$$

This shows that if $\epsilon \leq (\lambda^{-1} + l)^{-1}\epsilon'$ then $y \in \mathbb{B}\,(0, \epsilon')$. In the meanwhile, (4.8) implies that $J_\lambda(x) \in T^{-1}(y)$. The two facts, together with (b), imply that if $\epsilon \leq (\lambda^{-1} + l)^{-1}\epsilon'$ then

$$\begin{aligned} |J_\lambda(x) - Z| &\leq \tau' \|\, y \,\| \\ &\leq \tau' \|\lambda^{-1}(x - J_\lambda(x)) - F(x) + F(J_\lambda(x))\| \\ &\leq \tau'(\lambda^{-1} + l)\|x - J_\lambda(x)\|. \end{aligned} \tag{3.7}$$

On the other hand, since Z is a nonempty closed convex set [30] and the orthogonal projection P_Z is nonexpansive [45, Eq. (1.8)], then

$$\begin{aligned} & \|x - P_Z(x)\| \\ \leq \;& \|x - J_\lambda(x)\| + \|J_\lambda(x) - P_Z(J_\lambda(x))\| + \|P_Z(J_\lambda(x)) - P_Z(x)\| \\ \leq \;& 2\|x - J_\lambda(x)\| + \|J_\lambda(x) - P_Z(J_\lambda(x))\| \end{aligned}$$

Consequently, we have

$$|x - Z| \leq 2\, \|x - J_\lambda(x)\| + |J_\lambda(x) - Z|. \tag{3.8}$$

Combining (3.8) and (3.7) yields that for all $x \in Z_\lambda(\epsilon)$ with $\epsilon \leq (\lambda^{-1} + l)^{-1}\epsilon'$

$$|x - Z| \leq (\tau'\lambda^{-1} + \tau' l + 2)\|x - J_\lambda(x)\| \leq \tau\, \|x - J_\lambda(x)\|, \tag{3.9}$$

where we let $\tau \geq (\tau'\lambda^{-1} + \tau' l + 2)$.

Note that the statement (a) is known to hold when T is polyhedral [36]. Furthermore, the relation $(b) \Rightarrow (a)$ has been reported by Tseng in the context of variational inequalities; see [44, p. 440] for more details.

Corollary 3.2.1. *If in addition T is strongly monotone with moludus $\mu > 0$. Then the following statements are equivalent.*

$$(a) \quad \forall x,\ \forall y \in T(x) \quad \|x - z\| \leq \mu^{-1} \|\, y \,\|;$$

$$(b) \quad \|x - z\| \leq (\mu^{-1}\lambda^{-1} + \mu^{-1} l + 1)\|x - J_\lambda(x)\|,$$

where z is the unique solution to $0 \in T(x)$.

Proof. For any given x, since $0 \in T(z)$ and T is μ-strongly monotone then for all $y \in T(x)$ we have

$$\langle x - z, y - 0 \rangle \geq \mu \, \|x - y\|^2.$$

And it is easily seen from the relation above that

$$\forall x,\ \forall y \in T(x) \quad \|x - z\| \leq \mu^{-1} \|y\|.$$

A discussion similar to Theorem 3.2.1 yields

$$\|x - z\| \leq (\mu^{-1}\lambda^{-1} + \mu^{-1} l + 1)\|x - J_\lambda(x)\|.$$

Note that the condition number is relaxed to $\mu^{-1}\lambda + \mu^{-1} l + 1$ rather than $\mu^{-1}\lambda + \mu^{-1} l + 2$ (cf. (3.9)) in that the solution set is a singleton.

As a consequence, it is immediate that

Corollary 3.2.2. *If in addition $B = N_C$, where C is a nonempty closed convex subset in the Euclidean space. Then*

$$\|x - z\| \leq (\mu^{-1}\lambda^{-1} + \mu^{-1} l + 1)\|x - P_C[x - F(x)]\|.$$

Note that this global error bound is similar to global error bound for strongly monotone variational inequalities [33]; an entirely equivalent result easily follows from (4.7). In addition, for the case (b) of Theorem 3.2.1 in the context of variational inequalities, the reader can see [26, 36] for a related discussion.

3.3. Forward-Backward Splitting — Two Error Criteria. In this section, we propose an approximate version of the forward-backward splitting method described by (3.2) with two error criteria. Under standard assumptions, we prove global convergence of this approximate version with one criterion (see Criterion 3.3.1), and apply error bounds for forward-backward splitting to establish linear convergence with the other one (see Criterion 3.3.2).

Consider the following approximate forward-backward splitting method

$$x^{k+1} = (I + \lambda_k B)^{-1}(I - \lambda_k F)x^k + e^{k+1} \quad \forall k \geq 0. \tag{3.10}$$

Two error criteria are treated in this paper

Criterion 3.3.1. $\|e^{k+1}\| \leq \varepsilon_k \,, \quad \sum_{k=0}^{\infty} \varepsilon_k^2 < \infty \,;$

Criterion 3.3.2. $\|e^{k+1}\| \leq \eta \, \|x^k - x^{k+1}\|, \; \eta > 0.$

Assumption 3.3.1. Assume that $B, F : \mathcal{H} \rightrightarrows \mathcal{H}$ are maximal monotone mappings such that F is l-Lipschitz continuous and μ-strongly monotone.

To simplify the proofs of Theorem 3.3.1 and Theorem 3.3.2, we first introduce some technical results.

Lemma 3.3.1. *Suppose that Assumption 3.3.1 holds. Let $\{x^k\}$ be the sequence generated by (3.10) and z be the unique solution of (3.1). Then*

(a) *there exists $\beta > 0$ independent of $\lambda < 2\mu/l^2$ such that*

$$\|\phi^k\| \leq \beta \, (\|x^{k+1} - x^k\| + \|e^{k+1}\|),$$

where $\phi^k := x^k - z + 2(x^{k+1} - x^k) + \lambda F(x^k) - \lambda F(z) - e^{k+1}$.

(b) *Furthermore,*

$$\begin{aligned} & \langle x^{k+1} - z, x^{k+1} - x^k \rangle \\ \leq \;\; & \lambda l \, \|x^{k+1} - x^k\| \|x^k - z\| - \lambda\mu \|x^k - z\|^2 + \beta \|x^{k+1} - x^k\| \|e^{k+1}\| + \beta \|e^{k+1}\|^2, \end{aligned}$$

Proof. Let us first prove the first part of this lemma. It follows from (3.10) that

$$x^{k+1} = J_\lambda(x^k) + e^{k+1}.$$

Consequently,

$$\|x^k - J_\lambda(x^k)\| \leq \|x^{k+1} - x^k\| + \|e^{k+1}\|. \tag{3.11}$$

On the other hand, since Assumption 3.3.1 holds, then it easily follows from Corollary 3.2.1 and (3.11) that

$$\begin{aligned}
& \|\phi^k\| \\
= \; & \|x^k - z + 2(x^{k+1} - x^k) + \lambda F(x^k) - \lambda F(z) - e^{k+1}\| \\
\leq \; & 2\,\|x^{k+1} - x^k\| + (\lambda\, l + 1)\|x^k - z\| + \|e^{k+1}\| \\
\leq \; & 2\,\|x^{k+1} - x^k\| + (\lambda\, l + 1)(\mu^{-1}\lambda^{-1} + \mu^{-1} l + 1)\|x^k - J_\lambda(x^k)\| + \|e^{k+1}\| \\
\leq \; & 2\,\|x^{k+1} - x^k\| + (\lambda\, l + 1)(\mu^{-1}\lambda^{-1} + \mu^{-1} l + 1)\|x^{k+1} - x^k\| \\
& +((\lambda l + 1)(\mu^{-1}\lambda^{-1} + \mu^{-1} l + 1) + 1)\|e^{k+1}\|.
\end{aligned}$$

Then there exists $\beta > 0$ independent of $\lambda < 2\mu/l^2$ such that

$$\|\phi^k\| \leq \beta\,(\|x^{k+1} - x^k\| + \|e^{k+1}\|). \tag{3.12}$$

Note that it is entirely reasonable to *a priori* require that $\lambda < 2\mu/l^2$ in that this constant corresponds to the upper bound of stepsize for the exact case.

Now we are in a position to prove the second part of this lemma.

By (3.10), one has

$$(I + \lambda B)(x^{k+1} - e^{k+1}) \ni (I - \lambda F)(x^k),$$

which implies

$$\lambda^{-1}(x^k - x^{k+1} + e^{k+1}) - F(x^k) \in B(x^{k+1} - e^{k+1}). \tag{3.13}$$

On the other hand, $0 \in B(z) + F(z)$ yields

$$-F(z) \in B(z). \tag{3.14}$$

Since B is (maximal) monotone, then it follows from (3.13) and (3.14) that

$$\langle x^{k+1} - e^{k+1} - z, \lambda^{-1}(x^k - x^{k+1} + e^{k+1}) - F(x^k) + F(z)\rangle \geq 0. \tag{3.15}$$

Rearranging the terms of this relation above yields

$$\begin{aligned}
& \langle x^{k+1} - z, x^{k+1} - x^k\rangle \\
\leq \; & -\lambda\langle x^{k+1} - z, F(x^k) - F(z)\rangle + \langle e^{k+1}, \phi^k\rangle \\
\leq \; & -\lambda\langle x^{k+1} - x^k, F(x^k) - F(z)\rangle - \lambda\langle x^k - z, F(x^k) - F(z)\rangle + \langle e^{k+1}, \phi^k\rangle
\end{aligned}$$

where ϕ^k is defined in (a) of this lemma.
Since F is l-Lipschitz continuous and μ-strongly monotone, then this relation above, together with (3.12), implies that

$$\begin{aligned} & \langle x^{k+1} - z, x^{k+1} - x^k \rangle \\ \leq\ & \lambda l \,\|x^{k+1} - x^k\| \|x^k - z\| - \lambda\mu \|x^k - z\|^2 + \|e^{k+1}\| \|\phi^k\| \\ \leq\ & \lambda l \,\|x^{k+1} - x^k\| \|x^k - z\| - \lambda\mu \|x^k - z\|^2 + \beta \|x^{k+1} - x^k\| \|e^{k+1}\| + \beta \|e^{k+1}\|^2. \end{aligned}$$

The proof of this lemma is complete. $\square$

Theorem 3.3.1. *Suppose that Assumption 3.3.1 holds. Choose any starting point $x^0 \in \mathcal{H}$. Let $\{x^k\}$ be the sequence generated by (3.10) with Criterion 3.3.1. Suppose that $\lambda < 2(1-\theta)\mu/l^2$, where $\theta \in (0,1)$. Then $\{x^k\}$ converges strongly to the unique solution to (3.1), say z.*

Proof. According to the equality $\|u+v\|^2 = \|u\|^2 - \|v\|^2 + 2\langle v, u+v \rangle$ for all $u, v \in \mathcal{H}$, one has

$$\|x^{k+1} - z\|^2 = \|x^k - z\|^2 - \|x^{k+1} - x^k\|^2 + 2\langle x^{k+1} - z, x^{k+1} - x^k \rangle. \tag{3.16}$$

On the other hand, it follows from (b) of Lemma 3.3.1 that

$$\begin{aligned} & 2\langle x^{k+1} - z, x^{k+1} - x^k \rangle \\ \leq\ & 2\lambda l \,\|x^{k+1} - x^k\| \|x^k - z\| - 2\lambda\mu \|x^k - z\|^2 + 2\|e^{k+1}\| \|\phi^k\| \\ \leq\ & (1-\theta)\|x^{k+1} - x^k\|^2 + (1-\theta)^{-1}\lambda^2 l^2 \|x^k - z\|^2 - 2\lambda\mu \|x^k - z\|^2 \\ & +2\beta \|e^{k+1}\| \|x^{k+1} - x^k\| + 2\beta \|e^{k+1}\|^2 \\ \leq\ & (1-\theta)\|x^{k+1} - x^k\|^2 + (1-\theta)^{-1}\lambda^2 l^2 \|x^k - z\|^2 - 2\lambda\mu \|x^k - z\|^2 \\ & +\theta \|x^{k+1} - x^k\|^2 + \theta^{-1}\beta^2 \|e^{k+1}\|^2 + 2\beta \|e^{k+1}\|^2 \\ \leq\ & \|x^{k+1} - x^k\|^2 + ((1-\theta)^{-1}\lambda^2 l^2 - 2\lambda\mu)\|x^k - z\|^2 \\ & +(\theta^{-1}\beta^2 + 2\beta)\|e^{k+1}\|^2. \end{aligned}$$

which, together with (3.16), implies

$$\|x^{k+1} - z\|^2 \leq \lambda \|x^k - z\|^2 + (\theta^{-1}\beta^2 + 2\beta)\|e^{k+1}\|^2. \tag{3.17}$$

where

$$\kappa := (1-\theta)^{-1}\lambda^2 l^2 - 2\lambda\mu + 1. \tag{3.18}$$

Summing up, rearranging both sides of (3.17) and considering the nonnegativeness of $\|x^{k+1} - z\|^2$ yield

$$(1-\kappa)\sum_{i=0}^{k} \|x^i - z\|^2 \leq \|x^0 - z\|^2 + \sum_{i=0}^{k} (\theta^{-1}\beta^2 + 2\beta)\|e^{i+1}\|^2. \tag{3.19}$$

Since $\lambda < 2(1-\theta)\mu/l^2$, it follows from (3.18) that $\kappa < 1$. Moreover, in view of Criterion 3.1, we further have that $\sum_{i=0}^{k} \|e^{i+1}\|^2 < +\infty$. These show that $\sum_{i=0}^{k} \|x^i - z\|^2 < +\infty$ implying that $\{x^k\}$ converges strongly to z.

The proof of this theorem is complete. □

Theorem 3.3.2. *Suppose that Assumption 3.3.1 holds. Suppose that the step-size* $\lambda < 2(1-\theta)\mu/l^2$ *, where* $\theta \in (0,1)$ *is any given sufficiently small positive number. Choose any starting point* $x^0 \in \mathcal{H}$*. Let* $\{x^k\}$ *be the sequence generated by (3.10) with Criterion 3.3.2, where* $\eta := (\sqrt{1+2\beta^{-1}\theta} - 1)/2$*, and* z *be the unique solution. Then* $\{x^k\}$ *is strongly convergent at a linear rate no larger than* $\sqrt{1+(1-\theta)^{-1}\lambda^2 l^2 - 2\lambda\mu}$*.*

Proof. Criterion 3.3.2 says

$$\|e^{k+1}\| \le \eta \|x^{k+1} - x^k\|. \tag{3.20}$$

On the other hand, according to the proof of Theorem 3.3.1, the following relations hold

$$\begin{aligned} & \|x^{k+1} - z\|^2 \\ \le\ & \|x^k - z\|^2 - \|x^{k+1} - x^k\|^2 + 2c\, l \|x^{k+1} - x^k\| \|x^k - z\| \\ & -2c\mu \|x^k - z\|^2 + 2\beta \|e^{k+1}\| (\|x^{k+1} - x^k\| + \|e^{k+1}\|). \end{aligned}$$

Consequently, it follows from the relation above and (3.20) that

$$\begin{aligned} & \|x^{k+1} - z\|^2 \\ \le\ & \|x^k - z\|^2 - \|x^{k+1} - x^k\|^2 + (1-\theta)\|x^{k+1} - x^k\|^2 \\ & +(1-\theta)^{-1}\lambda^2 l^2 \|x^k - z\|^2 - 2\lambda\mu \|x^k - z\|^2 + 2\beta\eta(1+\eta)\|x^{k+1} - x^k\|^2. \end{aligned}$$

Since $\eta := (\sqrt{1+2\beta^{-1}\theta} - 1)/2$, then it is easy to check that $2\beta\eta(1+\eta) = \theta$. Therefore, this relation above implies that

$$\begin{aligned} & \|x^{k+1} - z\|^2 \\ \le\ & (1+(1-\theta)^{-1}\lambda^2 l^2 - 2\lambda\mu)\|x^k - z\|^2 - (\theta - 2\beta\eta(1+\eta))\|x^{k+1} - x^k\|^2 \\ =\ & (1+(1-\theta)^{-1}\lambda^2 l^2 - 2\lambda\mu)\|x^k - z\|^2. \end{aligned}$$

This shows that the iterate sequence $\{x^k\}$ is strongly convergent at linear convergence rate no larger than $\sqrt{1+(1-\theta)^{-1}\lambda^2 l^2 - 2\lambda\mu}$.

Note that some key techniques in the proofs of Lemma 3.3.1, Theorems 3.3.1–3.3.2 can be founded in [26, 22, 13].

3.4. Peaceman- and Douglas-Rachford Splitting. We now review the Peaceman- and Douglas-Rachford family of splitting methods for solving the monotone inclusions (3.1); each involves in resolvants of both the forward mapping and the backward mapping.

The corresponding Peaceman-Rachford splitting method is:

$$x^{k+1} = (2(I+\lambda B)^{-1} - I)\,(2(I+\lambda F)^{-1} - I)\,x^k \quad \forall k \geq 0.$$

The corresponding Douglas-Rachford splitting method is:

$$x^{k+1} = (I+\lambda F)^{-1}[(I+\lambda B)^{-1}(I-\lambda F) + \lambda F]\,x^k \quad \forall k \geq 0.$$

Furthermore, they can be implemented efficiently in the following ways.

Let x^0 be any starting point. For any given $k \geq 0$, find the unique y^k such that

$$y^k + \lambda B(y^k) \ni x^k - \lambda F(x^k).$$

Then find x^{k+1} such that

$$x^{k+1} + \lambda F(x^{k+1}) = (1-\gamma)x^k + \gamma y^k + \lambda F(x^k).$$

The $\gamma = 1$ case corresponds to the Douglas-Rachford splitting, whereas the $\gamma = 2$ case corresponds to the Peaceman-Rachford splitting [23]. Furthermore, the $\gamma \in (0,2)$ cases have already been well studied in [7, 8].

Chapter 4

New Splitting Methods

In this chapter and the next chapter, we are mainly concerned with (1.1) with such splitting of $T := B + F$ where B is also maximal monotone and F is continuous monotone. Then, we fully exploit this kind of special structure and develop a class of new and effective splitting methods for solving (3.1).

4.1. Basic Inequalities. Now we aim to develop new splitting methods for solving (3.1) in the context above. To this end, we first give some basic inequalities.

Lemma 4.1.1. *Let $B, F : \mathcal{H} \rightrightarrows \mathcal{H}$ be two maximal monotone mappings. If F is single-valued, then for any zero of (2.15), say z, we have*

$$\langle x - J_\lambda(x), x + \lambda F(x) - z - \lambda F(z)\rangle \geq \|x - J_\lambda(x)\|^2. \tag{4.1}$$

If in addition $F(x) := Mx + q$, where M is a linear mapping from $\mathcal{H}$ to $\mathcal{H}$ and $q \in \mathcal{H}$, then

$$\langle x-z,\ (I+M^T)(x-(I+B)^{-1}(x-Mx-q))\rangle \geq \|x-(I+B)^{-1}(x-Mx-q)\|^2. \tag{4.2}$$

Proof. Since $J_\lambda(x) := (I + \lambda B)^{-1}(I - \lambda F)(x)$, we have

$$J_\lambda(x) + \lambda B(J_\lambda(x)) \ni x - \lambda F(x).$$

That is,

$$\lambda^{-1}(x - J_\lambda(x)) - F(x) \in B(J_\lambda(x)). \tag{4.3}$$

On the other hand, since z is a zero of $T := B + F$ then we have

$$-F(z) \in B(z) \tag{4.4}$$

It follows from (4.3), (4.4) and monotonicity of B that

$$\langle \lambda^{-1}(x - J_\lambda(x)) - (F(x) - F(z)),\ J_\lambda(x) - z\rangle \geq 0.$$

i.e.,

$$\langle \lambda^{-1}(x - J_\lambda(x)) - (F(x) - F(z)),\ x - z - (x - J_\lambda(x))\rangle \geq 0. \tag{4.5}$$

Rearranging the terms of (4.5) yields

$$\langle x - J_\lambda(x),\ x + \lambda F(x) - z - \lambda F(z)\rangle \geq \|x - J_\lambda(x)\|^2 + \lambda\langle F(x) - F(z),\ x - z\rangle. \tag{4.6}$$

Consequently, it follows from (4.7) and monotonicity of F that

$$\langle x - J_\lambda(x),\ x + \lambda F(x) - z - \lambda F(z)\rangle \geq \|x - J_\lambda(x)\|^2. \tag{4.7}$$

Furthermore, the relation (4.2) similarly follows. □

Lemma 4.1.2. *Let $B, F : \mathcal{H} \rightrightarrows \mathcal{H}$ be two maximal monotone mappings. If F is single-valued, then for any zero of (2.15), say z, we have*

$$\begin{aligned} & \langle x - z,\ x - J_\lambda(x) - \lambda F(x) + \lambda F(J_\lambda(x))\rangle \\ \geq\ & \|x - J_\lambda(x)\|^2 - \lambda\langle x - J_\lambda(x),\ F(x) - F(J_\lambda(x))\rangle. \end{aligned} \tag{4.8}$$

Proof. By (4.5), we have

$$\langle \lambda^{-1}(x - J_\lambda(x)) - (F(x) - F(z)),\ x - z - (x - J_\lambda(x))\rangle \geq 0,$$

which, together with monotonicity of F, i.e.,

$$\langle F(J_\lambda(x)) - F(z),\ x - z - (x - J_\lambda(x))\rangle \geq 0,$$

implies that

$$\langle \lambda^{-1}(x - J_\lambda(x)) - F(x) + F(J_\lambda(x))),\ \ x - z - (x - J_\lambda(x))\rangle \geq 0.$$

Rearranging the terms of this relation yields

$$\begin{aligned} & \langle x - z, x - J_\lambda(x) - \lambda F(x) + \lambda F(J_\lambda(x))\rangle \\ \geq\ & \|x - J_\lambda(x)\|^2 - \lambda\langle x - J_\lambda(x), F(x) - F(J_\lambda(x))\rangle. \end{aligned}$$

4.2. New Splitting Methods. We now make use of the above-mentioned inequalities to develop some new splitting methods for solving (3.1).

Algorithm 4.2.1. *Let $F(x) = Mx + q$. Choose any starting point $x^0 \in \mathcal{H}$. For any given $k \geq 0$, computing x^{k+1} from x^k according to*

$$x^{k+1} = x^k - \gamma_k (I + M^T) r(x^k) \quad \forall k \geq 0. \tag{4.9}$$

where

$$\begin{aligned} r(x) &= x - (I+B)^{-1}(x - Mx - q), \\ \gamma_k &= \|(I+M^T)r(x^k)\|^{-2}\|r(x^k)\|^2. \end{aligned} \tag{4.10}$$

That deserving much attention is that dom$B$ is convex but not necessarily closed [38, Chapter 12] and all the resulting iterates do not necessarily lie in dom$B$, but the solution set Z is a nonempty *closed* convex set contained in dom$B$. So, as proven below, the corresponding iterates must weakly converge to some element of Z in an asymptotic way.

The convergence proof of Algorithm 4.2.1 is similar to that of [40, Theorem 2.1] except using (4.2) instead of the two inequalities in line -7 and line -9 in [40, p. 1817]; see [17] for an earlier discussion on convergence. Furthermore, the proof of convergence rate is similar to that of [40, Theorem 2.1], but using (4.11) instead of [40, Eq. (1.3)]. It is also similar in spirit to those in [42, 25, 26]. However, we include its proof here for completeness.

Theorem 4.2.1. *Any sequence $\{x^k\}$ generated by Algorithm 4.2.1 is weakly convergent. Let $Z(\epsilon) = \{x \mid r(x) \leq \epsilon\}$. If there exist $\epsilon > 0$, $\tau > 0$ such that*

$$|x - Z| \leq \tau \, \|r(x)\|, \quad \forall x \in Z(\epsilon). \tag{4.11}$$

Then the convergence is R-linear.

Proof. Let $z \in Z$. For any given $k \geq 0$, we have from (4.19) that

$$\begin{aligned} & \|x^{k+1} - z\|^2 \\ = \; & \|x^k - z - \gamma_k (I+M^T) r(x^k)\|^2 \\ = \; & \|x^k - z\|^2 - 2\gamma_k \langle x^k - z, (I+M^T)r(x^k)\rangle + \gamma_k^2 \, \|(I+M^T)r(x^k)\|^2 \end{aligned} \tag{4.12}$$

On the other hand, in view of (4.2) in Lemma 4.1.1, we have

$$\langle x^k - z, \; (I+M^T)r(x^k)\rangle \geq \|r(x^k)\|^2.$$

Using this to bound the next-to-last term in (4.12) yields the following relation

$$\begin{aligned} \|x^{k+1} - z\|^2 \;\leq\; & \|x^k - z\|^2 - 2\gamma_k \|r(x^k)\|^2 + \gamma_k^2 \, \|(I+M^T)r(x^k)\|^2 \\ = \; & \|x^k - z\|^2 - \|(I+M^T)r(x^k)\|^{-2}\|r(x^k)\|^4 & (4.13) \\ \leq \; & \|x^k - z\|^2 - \|(I+M^T)\|^{-2}\|r(x^k)\|^2, & (4.14) \end{aligned}$$

where the equality follows from (4.10). The remaining argument is patterned after the proof of [37, Thm. 1] and of [42, Thm. 1] and of [40, Thm. 2.1].

The relation (4.14) above shows that $\{\|x^k - z\|\}$ is also bounded above. Thus, $\{x^k\}$ must be bounded and at least has one weak cluster point, say x^∞. Let $\{x^{k(j)}\}$ be some subsequence converging weakly to x^∞. Let $J(x) = (I+B)^{-1}(x - Mx - q)$. It is easy to check that $r(x^k) = x^k - J(x^k) \to 0$ strongly, thus $\{J(x^{k(j)})\}$ converges weakly to x^∞ as well. Furthermore, it can be seen that $J(x^{k(j)}) \in \mathrm{dom}B$ is bounded in norm for all $k(j)$.

On the other hand, since $J(x^k) = (I+B)^{-1}(x^k - Mx^k - q)$, we have

$$x^k - J(x^k) - Mx^k - q \in B(J(x^k)).$$

Furthermore, it is immediate that when $k(j) \to +\infty$ we have

$$x^{k(j)} - J(x^{k(j)}) - Mx^{k(j)} - q \;\to\; -Mx^\infty - q \quad \text{strongly.} \tag{4.15}$$

So, in view of Proposition 1.4.1 with $T = B$, we have

$$-Mx^\infty - q \in B(x^\infty).$$

This implies that x^∞ is a zero of (3.1).

The uniqueness argument of weak cluster point easily follows from the one of Martinet [27]; see also [37, 5]. Thus is omitted.

The next goal is to show its local linear convergence. Since (4.13) holds for all k and all $z \in Z$, we obtain for all k

$$\begin{aligned} & \|x^{k+1} - z\|^2 \\ \le\; & \|x^k - z\|^2 - \|(I+M^T)r(x^k)\|^{-2}\|r(x^k)\|^4 \\ =\; & \|x^k - z\|^2 - \|(I+M^T)r(x^k)\|^{-2}\|r(x^k)\|^4 \\ \le\; & \|x^k - z\|^2 - \eta\|r(x^k)\|^2, \end{aligned}$$

where we let $\eta = \|(I+M^T)\|^{-2}$. So, we have

$$|x^{k+1} - Z|^2 \le |x^k - Z|^2 - \|(I+M^T)r(x^k)\|^{-2}\|r(x^k)\|^4, \tag{4.16}$$

and

$$|x^{k+1} - Z|^2 \le |x^k - Z|^2 - \eta\|r(x^k)\|^2. \tag{4.17}$$

On the other hand, since $\|r(x^k)\| \to 0$, we have $\|r(x^k)\| \le \epsilon$ for all k greater than some $\bar{k}$, in which case (4.11) yields $|x^k - Z| \le \tau\|r(x^k)\|$. Combing this and (4.17) yields

$$|x^{k+1} - Z|^2 \le |x^k - Z|^2 - \frac{\eta}{\tau^2}|x^k - Z|^2 \tag{4.18}$$

for all $k > \bar{k}$, so $\{|x^k - Z|\}$ converges to zero in the quotient sense and, by (4.17), $\{r(x^k)\}$ converges to R-linearly to zero. Since by (4.9), (4.10) and (4.16) we have

$$\|x^k - x^{k+1}\| = \|(I + M^T)r(x^k)\|^{-1}\|r(x^k)\|^2 \leq (|x^k - Z|^2 - |x^{k+1} - Z|^2)^{1/2}$$

for all k, it follows from $\{|x^k - Z|\} \to 0$ in the quotient sense that $\{\|x^k - x^{k+1}\|\}$ converges R-linearly to zero and hence $\{x^k\}$ converges R-linearly. $\square$

Algorithm 4.2.1 can be further extended by replacing the term $(I + B)^{-1}(x - Mx - q)$ in the definition of $r(x)$ with a more general matrix-splitting term. In particular, consider the following algorithm.

Algorithm 4.2.2. *Choose any starting point $x^0 \in \mathcal{H}$ and any positive definite matrix M'. For any given $k \geq 0$, computing x^{k+1} from x^k according to*

$$x^{k+1} = x^k - \gamma_k(M' + M^T)(x^k - u^k) \quad \forall k \geq 0, \tag{4.19}$$

where u^k is the unique solution of the nonlinear equations

$$u^k = (I + B)^{-1}(u^k - Mx^k - q + M'(x^k - u^k))$$

and

$$\gamma_k = \|(M' + M^T)(x^k - u^k)\|^{-2}\langle x^k - u^k,\ M'(x^k - u^k)\rangle. \tag{4.20}$$

Note that if we choose $M' = I$, then Algorithm 4.2.2 reduces to Algorithm 4.2.1. We have the following result whose proof is similar to that of Theorem 4.2.1 and thus is omitted.

Theorem 4.2.2. *Any sequence $\{x^k\}$ generated by Algorithm 4.2.2 is convergent. If in addition (4.11) holds. Then the convergence is R-linear.*

Remark. Impressively, we are unable to propose a generalized version of [15, Algorithm PC] or [40, Algorithm 2.3]. This mainly results from the absence of an appropriate generalized form of the inequality:

$$\langle x - P_C(x - Mx - q),\, Mx + q\rangle \geq \|x - P_C(x - Mx - q)\|^2$$

while [15, Algorithm PC] or [40, Algorithm 2.3] for affine complementarity problems or variational inequalities depends heavily on this inequality in a way that does not seem to carry over to the case of (3.1).

Algorithm 4.2.3. *Choose any starting point $x^0 \in \mathcal{H}$ and $\lambda \in (0,\, 1/l)$, where l is a constant satisfying*

$$\langle x' - x,\ F(x') - F(x)\rangle \leq l\,\|x' - x\|^2 \quad \forall x',\ \forall x.$$

For any given $k \geq 0$, computing x^{k+1} from x^k according to

$$x^{k+1} = x^k - \gamma_k (x^k - J_\lambda(x^k) - \lambda F(x^k) + \lambda F(J_\lambda(x^k)) \quad \forall k \geq 0, \tag{4.21}$$

where

$$\gamma_k = (1 - \lambda\, l)\|x^k - J_\lambda(x^k) - \lambda F(x^k) + \lambda F(J_\lambda(x^k)\|^{-2}\|x^k - J_\lambda(x^k)\|^2. \tag{4.22}$$

The proof below is similar to that of [40, Theorem 3.1] except using (4.8) instead of the inequalities in lines 3-12 in [40, p. 1822] for convergence (see also [18]), whereas using (4.23) instead of [40, Eq. (1.3)] for rate of convergence.

Theorem 4.2.3. *Let $F(x)$ is monotone and Lipschitiz continuous. Then any sequence $\{x^k\}$ generated by Algorithm 4.2.3 is convergent. If in addition there exist $\tau > 0$, $\epsilon > 0$ such that*

$$|x - Z| \leq \tau \, \|x - J_\lambda(x)\|, \quad \forall x \in Z_\lambda(\epsilon). \tag{4.23}$$

Then for the choice of $\lambda = 1$ (it can be met by trivially scaling Lipschitz constant of F), the convergence is R-linear.

Proof. Let $z \in Z$. For any given $k \geq 0$, we have from (4.21) that

$$\begin{aligned}
& \|x^{k+1} - z\|^2 \\
= \; & \|x^k - z - \gamma_k(x^k - J_\lambda(x^k) - \lambda F(x^k) + \lambda F(J_\lambda(x^k))\|^2 \\
= \; & \|x^k - z\|^2 - 2\gamma_k \langle x^k - z, x^k - J_\lambda(x^k) - \lambda F(x^k) + \lambda F(J_\lambda(x^k))\rangle \\
& +\gamma_k^2 \, \|x^k - J_\lambda(x^k) - \lambda F(x^k) + \lambda F(J_\lambda(x^k))\|^2.
\end{aligned} \tag{4.24}$$

On the other hand, it follows from (4.8) that

$$\begin{aligned}
& \langle x^k - z, x^k - J_\lambda(x^k) - J_\lambda(x^k) F(x^k) + \lambda F(J_\lambda(x^k))\rangle \\
\geq \; & \|x^k - J_\lambda(x^k)\|^2 - \lambda \langle x^k - J_\lambda(x^k), F(x^k) - F(J_\lambda(x^k))\rangle.
\end{aligned}$$

Using this to bound the next-to-last term in (4.24) yields

$$\begin{aligned}
\|x^{k+1} - z\|^2 \;\leq\; & \|x^k - z\|^2 - 2\gamma_k(1 - \lambda\, l)\|x^k - J_\lambda(x^k)\|^2 + \\
& \gamma_k^2 \, \|x^k - J_\lambda(x^k) - \lambda F(x^k) + \lambda F(J_\lambda(x^k))\|^2 \\
= \; & \|x^k - z\|^2 - (1 - \lambda\, l)^2 \|x^k - J_\lambda(x^k) - \lambda F(x^k) + \\
& \lambda F(J_\lambda(x^k))\|^{-2}\|x^k - J_\lambda(x^k)\|^4,
\end{aligned} \tag{4.25}$$

where the equality follows from (4.22).

The remainder of the proof is similar to that of Theorem 4.2.1, but using (4.23) instead of (4.11). In fact, since we assume, in addition, that $\lambda = 1$, in view of (4.21) and (4.26), we have

$$\begin{aligned}\|x^{k+1} - x^k\|^2 &= (1-l)\|x^k - J(x^k) - F(x^k) + F(J(x^k))\|^{-1}\|x^k - J(x^k)\|^2 \\ &\geq (1-l)(1+l')^{-1}\|x^k - J(x^k)\|\end{aligned}$$

for all k, where l' denotes Lipschitz constant of F. Thus, the rightmost term in (4.25) is bounded above by a positive constant times $-\|x^k - J(x^k)\|^2$ and, when this term converges R-linearly to zero as $k \to \infty$, so does $\|x^{k+1} - x^k\|^2$; hence $\{x^k\}$ converges R-linearly.

Algorithm 4.2.4. *Choose any starting point $x^0 \in \mathcal{H}$ and $\lambda_{-1} \in (0, +\infty)$. Also choose $\rho \in (0,1)$ and $\beta \in (0,1)$. For any given $k \geq 0$, computing (x^{k+1}, λ_k) from (x^k, λ_{k-1}) where λ_k is the largest $\lambda \in \{\lambda_{k-1}, \lambda_{k-1}\beta, \lambda_{k-1}\beta^2, ...\}$ satisfying*

$$\lambda\langle x^k - J_\lambda(x^k),\ F(x^k) - F(J_\lambda(x^k))\rangle \leq (1-\rho)\|x^k - J_\lambda(x^k)\|^2,$$

and let

$$x^{k+1} = x^k - \gamma_k(x^k - J_{\lambda_k}(x^k) - \lambda_k F(x^k) + \lambda_k F(J_{\lambda_k}(x^k))) \quad \forall k \geq 0,$$

where

$$\gamma_k = \rho\|(x^k - J_{\lambda_k}(x^k)) - \lambda_k F(x^k) + \lambda_k F(J_{\lambda_k}(x^k))\|^{-2}\|x^k - J_{\lambda_k}(x^k)\|^2. \qquad (4.26)$$

Below we present the convergence results of Algorithm 4.2.4. The proof is patterned after that of Algorithm 4.2.3 and thus is omitted; see [40, 41] for its special versions in the context of monotone variational inequalities.

Theorem 4.2.4. *Let $F(x)$ is monotone and continuous. Then any sequence $\{x^k\}$ generated by Algorithm 4.2.4 is convergent. If in addition (4.23) holds and F is Lipschitiz continuous on $Z + \mathbb{B}(0, \epsilon)$, then the convergence is R-linear.*

These methods above are a class of new splitting methods for (3.1). Compared with a method proposed by Tseng [44], our proposed splitting methods remove projection step involved whenever dom$B \neq \mathcal{H}$. (Note that in the case where dom$B = \mathcal{H}$, Tseng's method may be viewed as an instance of a method proposed by Solodov and Svaiter [39] by choosing the projection set to be $\mathcal{H}$.) For a comparison with the Peaceman- and Douglas-Rachford family of splitting methods, we will detail it later.

Now we give the following splitting method:

Algorithm 4.2.5. *Choose any starting point $x^0 \in \mathcal{H}$. For any given $k \geq 0$, computing x^{k+1} from x^k according to*

$$x^{k+1} = x^k - \gamma\,(I+F)^{-1}(x^k - J(x^k)) \quad \forall k \geq 0.$$

Furthermore, when $F(x) = Mx + q$, the iterative formula reduces to

$$x^{k+1} = x^k - \gamma\,(I+M)^{-1}(x^k - (I+B)^{-1}(x^k - Mx^k - q)) \quad \forall k \geq 0,$$

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

The convergence theorem of this algorithm above is given below. However, we do not include its proof here in that it, as noted later, is can be viewed as a special case of the Peaceman- and Douglas-Rachford family of splitting methods. Thus, its convergence can be implied by that of the latter. Furthermore, it is also straightforward from deep analytical machinery of [19, 4, 5] especially for finite dimensional case.

Theorem 4.2.5. *Let $B, F : \mathcal{H} \rightrightarrows \mathcal{H}$ be any given maximal monotone mappings with F single-valued. Choose any starting point $x^0 \in \mathcal{H}$. Let $\{x^k\}$ be the sequence generated by Algorithm* 4.2.5. *Then if T has at least one zero, $\{x^k\}$ converges to a zero of $0 \in T(x)$. If $0 \in T(x)$ has no any zero, $\{x^k\}$ is an unbounded sequence.*

We turn to study Algorithm 4.2.5's intimate connections with the Peaceman- and Douglas-Rachford family of splitting methods. In fact, these can be easily seen when it is implemented as follows.

Let x^0 be any starting point. For any given $k \geq 0$, find the unique y^k such that

$$y^k + \lambda B(y^k) \ni x^k - \lambda F(x^k).$$

Then find x^{k+1} such that

$$x^{k+1} + \lambda F(x^{k+1}) = (1-\gamma)\,x^k + \gamma\, y^k + \lambda F(x^k).$$

This shows that the Peaceman-Rachford splitting method and the Douglas-Rachford splitting method correspond to the $\gamma = 2$ case and the $\gamma = 1$ case of Algorithm 4.2.5, respectively. Therefore, we can conclude in this sense that Algorithm 4.2.5 is basically a special expression of the Peaceman- and Douglas-Rachford family of splitting methods when F is single-valued. Furthermore, it is also direct generalization of a method proposed by He [19] for monotone variational inequalities, and also that of a generalized proximal point algorithm originally proposed by Gol'shtein and Tret'yakov [10] when F vanishes; see also [4, 5].

We hereby wish to point out that approximate versions of these splitting methods are not under further consideration here. They may be implied by analysis similar to previous related machinery. Furthermore, for any sequence $\{c_k\}$ satisfying $c_k \geq c > 0$, replacing F, B by $c_k F$, $c_k B$, respectively, is also applicable to our proposed splitting methods.

4.3. Role of Error Bounds for Forward-Backward Splitting. We hereby wishes to comment on role of error bounds for forward and backward splitting in analysis of convergence of these splitting methods.

Let's take Algorithm 4.2.1 as an example to show this. In fact, when we exploit analytical structure of its convergence behavior, it can be seen that if related error bound holds then it follows from (4.18) that

$$|x^{k+1} - Z| \leq \sqrt{1 - \tau^{-2}\,\eta}\ \ |x^k - Z|.$$

This relation above implies that

$$|x^k - Z| \leq \sqrt{1 - \tau^{-2}\,\eta}\ ^k\, |x^0 - Z|.$$

This in turn shows that for any given accuracy requirement ε, it can be met in less than

$$\frac{2\,\ln\left(\varepsilon\ \|x^0 - Z\|^{-1}\right)}{\ln\left(1 - \tau^{-2}\,\eta\right)} \tag{4.27}$$

steps when Algorithm 4.2.1 is implemented. In other words, theoretically, error bound condition can guarantee finite convergence of Algorithm 4.2.1 whereas finite convergence property appears to be crucial for any given numerical method. However, numerically, if its condition number τ is larger (even error bound holds) then Algorithm 4.2.1 is still possibly ill-behaved in that tendency to zero of $\ln(1 - \tau^{-2}\,\eta)$ leads to tendency to infinity of possible implementation steps described by (4.27). Once such circumstances occur, this algorithm will be in a dilemma: The stopping criterion is fairly good whereas such fairly good criterion may fail to guarantee real closeness of the last iterate to the solution set. Of course, if error bound does not hold at all, then the circumstance may become worse.

4.4. Applications. Now we show that the methods proposed are applicable to a class of general monotone inclusion frameworks and also demonstrate their considerable potentials in practical applications.

Example 4.4.1. Consider the following monotone inclusion

$$0 \in B(x,y,z) + F(x,y,z) \tag{4.28}$$

with

$$B(x,y,z) = U(x) \times V(y) \times \{b\}, \quad F(x,y,z) = (D^T z,\, E^T z,\, -Dx - Ey)$$

where U and V are maximal monotone mappings on $\mathcal{R}^m$ and $\mathcal{R}^n$, respectively, and $D \in \mathcal{R}^{l\times m}$, $E \in \mathcal{R}^{l\times n}$ and $b \in \mathcal{R}^l$. Then B, F are maximal monotone and F is linear on $\mathcal{R}^{l+m+n}$ with Lipschitz constant

$$l = \sqrt{\|D^T\|^2 + \|D\|^2 + \|E^T\|^2 + \|E\|^2}.$$

As is shown [44, Example 4] that under standard conditions this inclusion provides a most powerful framework and includes as a special case the following convex program:

$$\begin{array}{ll} \text{minimize} & f(x) + g(y) \\ \text{subject to} & Dx + Ey = b \end{array}$$

where f, g are closed proper convex functions on $\mathcal{R}^m$ and $\mathcal{R}^n$, respectively, and $D \in \mathcal{R}^{l\times m}$, $E \in \mathcal{R}^{l\times n}$, $b \in \mathcal{R}^l$. Furthermore, when $m = n$, $D = -E = I$ and $b = 0$, it reduces to

$$0 \in U(x) + V(x).$$

Of course, if in addition $U = N_C$ and $V = F$ then it corresponds to the standard monotone variational inequalities.

Applying Algorithm 4.2.1 with such choice of B and F, we can get a new method to solve the inclusion (4.28) above.

Let $u^0 = (x^0, y^0, z^0)$ be any starting point. For any given $k \geq 0$, find the unique $\omega^k := (\omega_x^k,\, \omega_y^k,\, \omega_z^k)$ such that

$$\begin{aligned} \omega_x^k + U(\omega_x^k) &\ni x^k - D^T z^k, \\ \omega_y^k + V(\omega_y^k) &\ni y^k - E^T z^k, \\ \omega_z^k &= z^k + Dx^k + Ey^k - b. \end{aligned}$$

Let $u^k := (x^k,\, y^k,\, z^k)$ and $r(u^k) := u^k - \omega^k$. Then we obtain the next iterate from the following iterative formula:

$$u^{k+1} = u^k - \gamma_k (I + M^T)\, r(u^k),$$

where

$$\gamma_k = \|(I + M^T)r(u^k)\|^{-2}\|r(u^k)\|^2$$

and

$$M = \begin{pmatrix} 0 & 0 & D^T \\ 0 & 0 & E^T \\ -D & -E & 0 \end{pmatrix}$$

Of course, in this context, Algorithm 4.2.5 corresponds to

$$u^{k+1} = u^k - \gamma\,(I + M)^{-1} r(u^k) \quad \forall k \geq 0,$$

where $\gamma \in (0,\, 2)$, M and $r(u^k)$ are identical to the expressions above.

Example 4.4.2. Let U be a maximal monotone mapping on a Hilbert space X and V be a maximal monotone mapping on a Hilbert space Y. Let $L : U \to V$ be a linear mapping.

A primal formulation associated with such a triple $(U,\, V,\, L)$ is to find $x \in X$ such that

$$0 \in U(x) + L^* V(L\,x) \tag{4.29}$$

where L^* denote the adjoint of L.

A dual formulation associated with the triple $(U,\, V,\, L)$ is to find $y \in Y$ such that

$$0 \in -LU^{-1}(-L^* y) + V^{-1}(y).$$

Note that an earlier version of this duality framework is due to [1] with the restrictions $X = Y$, $L = I$.

A primal-dual formulation associated with the triple $(U,\, V,\, L)$ is to find $(x, y) \in X \times Y$ such that

$$0 \in U(x) + L^* y \qquad 0 \in -L\,x + V^{-1}(y)$$

or equivalently

$$0 \in B(x,\, y) + F(x,\, y), \tag{4.30}$$

where $B = U \times V^{-1}$ and F is defined by

$$F(u) = \begin{pmatrix} 0 & L^* \\ -L & 0 \end{pmatrix} u \qquad u = \begin{pmatrix} x \\ y \end{pmatrix}$$

It is clear that F is a linear, monotone mapping.

As is shown [9] that of the three formulations, the primal-dual formulation is in some sense the best behaved in the sense that $B+F$ is maximal monotone whenever maximal whenever U, V are. However, the other two are not necessarily this case; see [9, 35] for related maximality criteria and the references cited therein. In addition, the reader may consult [9] for their corresponding versions in the context of convex minimization.

Now we apply Algorithm 4.2.1 with such choice of B and F to solve the inclusion (4.30) above.

Let $u^0 = (x^0, y^0)$ be any starting point. For any given $k \geq 0$, find the unique $\omega^k := (\omega_x^k,\, \omega_y^k)$ such that

$$\begin{aligned} \omega_x^k + U(\omega_x^k) &\ni x^k - L^* y^k, \\ \omega_y^k + V^{-1}(\omega_y^k) &\ni y^k + L\, x^k. \end{aligned} \tag{4.31}$$

Let $u^k := (x^k,\, y^k)$ and $r(u^k) := u^k - \omega^k$. Then we obtain the next iterate from the following iterative formula:

$$u^{k+1} = u^k - \gamma_k\, (I + M^T)\, r(u^k),$$

where

$$\gamma_k = \|(I + M^T) r(u^k)\|^{-2} \|r(u^k)\|^2$$

and

$$M = \begin{pmatrix} 0 & L^* \\ -L & 0 \end{pmatrix}$$

Of course, in this context, Algorithm 4.2.5 corresponds to

$$u^{k+1} = u^k - \gamma\, (I + M)^{-1} r(u^k) \quad \forall k \geq 0,$$

where $\gamma \in (0,\, 2)$, M and $r(u^k)$ are identical to the expressions above.

Note that (4.31) is equivalent to

$$\omega_y^k = (I + V^{-1})^{-1}(y^k + L\, x^k).$$

If the resolvant of V^{-1} is not easy to evaluate then we can alternatively consider the following inverse-resolvant identity [38, p. 540]:

$$(I + V^{-1})^{-1} = I - (I + V)^{-1}.$$

For example, when $V = N_C$, we can evaluate the equivalent expression $I - P_C$ rather than $(I + N_C^{-1})^{-1}$ itself in that the latter is in general is not very easy to express explicitly; see [9] for an expression of N_C^{-1} when C is a *special* box-constrained closed set.

In [9], Eckstein and Ferris proposed smooth methods of multipliers for dual and primal-dual formulations in the context of monotone variational inequalities and complementarity problems. Still, they did not discuss the case of general choice L. But, our proposed method can work well in this case.

Remark. As far as the two monotone inclusion frameworks above are concerned, the underlying matrices of the forward mappings are the very larger scale *sparse* matrices desired by us. Numerically, such sparsity properties speed convergence of Algorithm 4.2.1 whereas such share taken by Algorithm 4.2.5 appears to be relatively limited, even negligible. As shown in the next chapter, sparsity properties sometimes may constitute one major reason for why the former can overrun the latter in some cases.

Chapter 5

Numerical Tests

To numerically understand the convergence behavior of new splitting methods in practice, we implemented Algorithm 4.2.1 and Algorithm 4.2.5 with the choice of $\gamma = 1.9$ via the following test problem in finite dimensional Euclidean space. All codes were compiled by MatLab and run on a P-Ⅲ 667 PC under MatLab Version 6.1.

The test problem is to find $x \in \mathcal{R}^{m+n}$ such that

LCP $\qquad x \geq 0, \quad Mx + q \geq 0, \quad \langle\, x,\, Mx + q\,\rangle = 0,$

where possible nonzero elements of M: $a_{ij} = -a_{ji}, i = 1, 2, ..., m, j = m+1, ..., m+n$ are randomly generated in $(0, 1)$, $q_j = -1, j = 1, ..., m$ and $q_j = 1, j = m+1, ..., m+n$.

The LCP above corresponds to the case of Example 2.4.2 with

$$C = \mathcal{R}^n_+, \quad F(x) = Mx + q.$$

For this LCP, Algorithm 4.2.1 may be expressed as

$$x^{k+1} = x^k - \gamma_k\,(I + M^T)\,r(x^k),$$

where $\gamma_k = \|(I + M^T)r(x^k)\|^{-2}\|r(x^k)\|^2$ and $r(x) := \min\{x, Mx + q\}$.

And, Algorithm 4.2.5 may be of the form

$$x^{k+1} = x^k - \gamma\,(I + M)^{-1}r(x^k),$$

where $r(x) := \min\{x, Mx + q\}$ with $\gamma = 1.9$.

Remark. Algorithm 4.2.5 also corresponds to the $\gamma = 1.9$ case of the Peaceman- and Douglas- Rachford family of splitting methods. As far as this algorithm is concerned, the $\gamma = 1.9$ choice is an empirically good one. This phenomenon has been observed and reported by He [19] who suggested to choose $\gamma \in (1.5, 2)$ deliberately to accelerate convergence.

The results of these implementations of Algorithm 4.2.1 and Algorithm 4.2.5 via this LCP have been reported in the following table. The implementation procedures invariantly started with the origin and stopped as soon as $\|r(x^k)\|_\infty \leq 10^{-6}$.

Table: Numerical results of Algorithm 4.2.1 and Algorithm 4.2.5 on LCP

Dimesion	Algorithm 4.2.1		Algorithm 4.2.5	
(m, n)	Number of Iter.	CPU Time (sec.)	Number of Iter.	CPU Time (sec.)
(5, 10)	830	0.3300	1604	0.8300
(5, 20)	1430	0.7100	2474	2.2500
(5, 30)	1690	1.2100	4454	6.2000
(5, 40)	2531	2.3100	2627	5.5500
(5, 50)	3771	4.3400	7258	21.4700
(10, 5)	493	0.1600	1530	0.7700
(10, 10)	2919	1.2100	4877	3.3500
(10, 15)	6620	3.1300	11705	10.6100
(10, 20)	3146	1.8100	7659	8.1300
(10,25)	2856	2.0300	6146	8.5700
(10, 30)	6996	5.4400	4383	7.2000
(15, 5)	674	0.3300	1471	0.9800
(15, 10)	1634	0.8800	3027	2.6900
(15, 15)	16187	9.2300	22022	23.3400
(20, 5)	1312	0.7100	2065	1.7500
(20, 10)	2062	1.2100	5564	6.1000
(30, 5)	10155	7.1400	20255	28.0100
(30, 10)	2691	2.0900	4068	6.7066

From the table above, we can see that when the underlying matrices were of the same sparsity structures as that of Examples 4.4.1-4.4.2, Algorithm 4.2.1 in general needed less numbers of iterations and less CPU time than Algorithm 4.2.5.

The reasons for why are at least threefold: The first reason is that in general Algorithm 4.2.5 needs $O((m+n)^3)$ operations per iteration whereas Algorithm 4.2.1 only needs $O((m+n)^2)$ operations per iteration. The second one is that the stepsizes in Algorithm 4.2.5 are the same constant whereas Algorithm 4.2.1

adopts adaptive stepsizes, which can take full advantage of information on the current iterates. The last, perhaps more important is that Algorithm 4.2.1 can exploit *sparsity* properties of the underlying matrices whereas Algorithm 4.2.5 fails to better them. Therefore, Algorithm 4.2.1 in our test problem consumed much less CPU time especially when the scale of the LCP above became larger.

Based on these observations, we can conclude that our proposed splitting methods may become the practical alternatives to the classical Peaceman- and Douglas-Rachford family of splitting methods, especially for the two classes of monotone inclusion frameworks described by Examples 4.4.1-4.4.2.

Bibliography

[1] H. Attouch and M. Théra [1996], A general duality principle for the sum of two operators, Journal of Convex Analysis, 3, 1-24.

[2] H. Bréis [1973], Opéateurs Maximal Monotones, North-Holland, Amsterdam.

[3] H.-G. Chen and R.T. Rockafellar [1997], Convergence rates in forward-backward splitting, SIAM Journal on Optimization, 7, 421-444.

[4] J. Eckstein [1989], Splitting methods for monotone operators with applications to parallel optimization, Doctoral dissertation, M.I.T.

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

[6] J. Eckstein [1993], Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming, Mathematics of Operations Research, 18, 202–226.

[7] J. Eckstein [1998], Approximate iterations in Bregman-function-based proximal algorithms, Mathematical Programming, 83, 113-123.

[8] J. Eckstein and M.C. Ferris [1998], Operator-splitting methods for monotone affine variational inequalities, with a parallel application to optimal control, INFORMS Journal on Computing, 10, 218-235.

[9] J. Eckstein and M.C. Ferris [1999], Smooth methods of multipliers for complementarity problems, Mathematical Programming, 86, 65-90.

[10] E. G. Gol'shtein and N. V. Tret'yakov [1979], Modified Lagrangians in convex programming and their generalizations, Mathematical Programming Study, 10, 86-97.

[11] M. S. Gowda and R. Sznajder [1996], On the Lipschitzian properties of polyhedral multifunctions, Mathematical Programming, 74, 267-278.

[12] O. Güler [1991], On the convergence of the proximal point algorithm for convex minimization, SIAM Journal on Control and Optimization, 29, 403-419.

[13] D.R. Han and B.S. He [2001], A new accuracy criterion for approximate proximal point algorithms, Journal of Mathematical Analysis and Applications, 263, 343-354.

[14] S. P. Han and G. Lou [1988], A parallel algorithm for a class of convex programs, SIAM Journal on Control and Optimization, 26, 345-355.

[15] B. S. He [1992], A class of projection and contraction methods for a class of linear complementarity problems and its application in convex quadratic programming, Applied Mathematics and Optimization, 25, 247-262.

[16] B. S. He [1994], A new method for a class of variational inequalities, Mathematical Programming, 66, 137-144.

[17] B. S. He [1994], Solving a class of linear projection equations, Numerische Mathematik, 68, 71-80.

[18] B. S. He [1997], A class of projection and contraction methods for monotone variational inequalities, Applied Mathematics and Optimization, 35, 69-76.

[19] B. S. He [1999], Inexact implicit methods for monotone general variational inequalities, Mathematical Programming, 86, 199-217.

[20] A. V. Kryanev [1973], The solution of incorrectly posed problems by means of sucessive approximations, Soviet Mathematics Doklady, 14, 673-676.

[21] B.W. Kort and D.P. Bertsekas [1976], Combined primal-dual and penalty methods for convex programming, SIAM Journal on Control and Optimization, 14, 268-294.

[22] W. Li [1993], Remarks on convergence of the matrix splitting algorithm for the symmetric linear complementarity problem, SIAM Journal on Optimization, 3, 155-163.

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

[24] F. J. Luque [1984], Asymptotic convergence analysis of the proximal point algorithm, SIAM Journal on Control and Optimization, 22, 277-293.

[25] Z. Q. Luo and P. Tseng [1992], On the linear convergence rate of descent methods for convex essentially smooth minimization, SIAM Journal on Control and Optimization, 30, 408-425.

[26] Z.Q. Luo and P. Tseng [1992], Error bounds and the convergence analysis of matrix splitting algorithms for the affine variational inequality problem, SIAM Journal on Optimization, 2, 43-54.

[27] B. Martinet [1970], Regularisation d'inéquations variationelles par approximations successives, Revue Francaise d'Informatique et de Recherche Operationelle, 4, 154-158.

[28] B. Martinet [1972], Determination approchée d'un application pseudo-contractante. Cas de I'application prox, Comptes Rendus de I'Academie des Sciences, Paris, 274, 163-165.

[29] G. J. Minty [1962], Monotone (nonlinear) operators in Hilbert space, Duke Mathematics Journal, 29, 341-346.

[30] G. J. Minty [1964a], On the monotonicity of the gradient of a convex function, Pacific Journal of Mathematics, 14, 243-247.

[31] J. J. Moreau [1965], Proximité et dualité dans un espace Hilbertien, Bulletin de la Société mathématique de France, 93, 273-299.

[32] A. Moudafi and M. Théra [1997], Finding a zero of the sum of two maximal monotone operators, Journal of Optimization Theory and Applications, 94, 425-448.

[33] J. -S. Pang [1987], A posteriori error bound for the linearly-constrained variational inequality problem, Mathematics of Operations Research, 12, 474-484.

[34] D. Pascali and S. Sburlan [1978], Nonlinear Mappings of Monotone Type, Edittura Academiei, Bucharest.

[35] T. Pennanen [2000], Dualization of generalized equations of maximal monotone type, SIAM Journal on Optimization, 10, 809-835.

[36] S. M. Robinson [1981], Some continuity properties of polyhedral multifunctions, Mathematical Programming Study, 14, 206-214.

[37] R. T. Rockafellar [1976], Monotone operators and proximal algorithm, SIAM Journal on Control and Optimization, 14, 877-898.

[38] R. T. Rockafellar and R. J.-B. Wets [1998], Variational Analysis, Springer-Verlag, Berlin.

[39] M. V. Solodov and B. F. Svaiter [1999], A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator, Set-Valued Analysis, 7, 323-345.

[40] M.V. Solodov and P. Tseng [1996], Modified projection-type methods for monotone variational inequalities, SIAM Journal on Control and Optimization, 34, 1814-1830.

[41] D. Sun [1996], A class of iterative methods for solving nonlinear projection equations, Journal of Optimization Theorey and Applications, 91, 123-140.

[42] P. Tseng [1995], On linear convergence of iterative methods for the variational inequality problem, Journal of Computational and Applied Mathematics, 60, 237-252.

[43] P. Tseng [1997], Alternating projection-proximal methods for convex programming and variational inequalities, SIAM Journal on Optimization, 7, 951-965.

[44] P. Tseng [2000], A modified forward-backward splitting method for maximal monotone mapping, SIAM Journal on Control and Optimization, 38, 431-446.

[45] E. H. Zarantonello [1971], Projections on convex sets in Hilbert space and spectral theory. In Zarantonello, E.H., Contributions to Nonlinear Functional Analysis, 237-424. Academic Press, New York.