Noname manuscript No.
(will be inserted by the editor)

Weak convergence of an extended splitting method for monotone inclusions

Yunda Dong

Received: date / Accepted: date

Abstract In this article, we consider the problem of finding zeros of two-operator monotone inclusions in real Hilbert spaces, and the second operator has been linearly composed. We suggest an extended splitting method: At each iteration, it mainly solves one resolvent for each operator, respectively. For these two resolvents, the involved two scaling factors can be different from one to the other. Our suggested splitting method generates both primal sequence and dual sequence. By using a direct, not convoluted discussion, under the weakest possible conditions we prove the former's weak convergence to an element of the associated solution set. This method contains one parameter ranging from zero and one, and it recovers an equivalent version of some known method when the parameter is equal to zero. Furthermore, we via a numerical example clarify the necessity of introducing such a parameter.

Keywords Real Hilbert spaces · Monotone inclusions · Splitting method · Different scaling factors · Weak convergence

Mathematics Subject Classification (2000) 58E35 · 65K15

1 Introduction

Let $\mathcal{H}$, $\mathcal{G}$ be infinite-dimensional Hilbert spaces. In this article, we are mainly concerned with the following problem of finding an x in $\mathcal{H}$ such that

$$0 \in A(x) + L^* B(Lx - r), \tag{1}$$

Yunda Dong (✉)
School of Mathematics and Statistics, Zhengzhou University, Zhengzhou 450001, PR China
E-mail: ydong@zzu.edu.cn

where the operator $A : \mathcal{H} \rightrightarrows \mathcal{H}$ and the operator $B : \mathcal{G} \rightrightarrows \mathcal{G}$ are maximal monotone, and $L : \mathcal{H} \to \mathcal{G}$ is nonzero bounded linear with its adjoint L^*, and $r \in \mathcal{G}$. Throughout this article, the problem's solution set Z is always assumed to be nonempty.

As shown in [1], the monotone inclusion above provides a simple but powerful framework of reformulating practical problems. In particular, it includes convex minimization, complementarity problem, monotone variational inequality problem and so on.

In the following special case of two operators:

$$0 \in A(x) + B(x), \tag{2}$$

there have been three main splitting methods. The first one is the forward-backward splitting method [2, 3], whose recursive formula (in the case of A being single-valued) reads

$$x^{k+1} = (I + \mu_k B)^{-1}(x^k - \mu_k A(x^k)),$$

where $\mu_k > 0$. Through the simplest in form, its weak convergence usually requires that the forward operator A be Lipschitz continuous and monotone and the sum operator $A + B$ be strongly monotone [4]. Since these assumptions are rather restrictive, one may resort to the second splitting method [2] proposed in the year 1979

$$\begin{aligned} x^k &= (I + \mu A)^{-1}(z^k), \\ y^k &= (I + \mu B)^{-1}(2x^k - z^k), \\ z^{k+1} &= z^k - \gamma(x^k - y^k), \end{aligned}$$

where $\mu > 0$ is scaling factor and the parameter $\gamma \in (0, 2)$. Lions and Mercier [2] analyzed weak convergence of auxiliary sequence $\{z^k\}$ and [5] proved weak convergence of the main sequence $\{x^k\}$ to the solution point of the problem (2) above. In the $\gamma = 1$ case, Lions and Mercier [2] called it Douglas-Rachford algorithm. This is because that it has root in the alternating-direction implicit iterative method for solving special systems of linear equations, due to the work of Douglas and Rachford [6] in the year 1956. Sometimes, other name "the Douglas-Rachford splitting method" is also popular, even is referred to as the case of $\gamma \in (0, 2)$. In the forbidden case of $\gamma = 2$, it corresponds to Peaceman-Rachford algorithm [7]. The third one is Tseng splitting method [1] proposed in the year 2000

$$\begin{aligned} y^k &:= (I + \mu_k B)^{-1}(x^k - \mu_k A(x^k)), \\ x^{k+1} &= y^k - \mu_k A(y^k) + \mu_k A(x^k), \end{aligned}$$

where $\mu_k > 0$. If the forward operator A is (locally) Lipschitz continuous and monotone and the backward operator B is maximal monotone, and if μ_k is chosen in some proper way. then its weak convergence can be guaranteed. For a practical relaxed version, we refer to [8] for detailed discussions.

Note that, in the context of the three dominating splitting methods just mentioned above, the scaling factors for A and B are identical. To overcome this drawback, [9, Algorithm 2] suggested a novel splitting method and its main recursive formulae include: Choose $t \in [0,1]$. For the current primal iterate x^k and the current dual iterate v^k, compute

$$\begin{aligned} y^k &= (\alpha I + A)^{-1}(\alpha x^k - v^k), \\ \hat{y}^k &= (1-t)x^k + ty^k, \\ u^k &= (\beta I + B)^{-1}(\beta \hat{y}^k + v^k), \end{aligned}$$

where $\alpha > 0$ and $\beta > 0$ are scaling factors. Then, make use of all these information to update x^k and v^k to get new iterates, respectively; see Sect. 3 for more details. For other discussions of splitting methods for solving various types of monotone inclusions, we refer to [10–17], to cite a few.

As an iterative scheme for finding zeros of the sum of two maximal monotone operators, [9, Algorithm 2] has a nice property: at each iteration, it computes either resolvent with possibly different scaling factors. This is a sharp contrast to the three dominating splitting methods above.

In this article, we aim at extending [9, Algorithm 2] for solving (2) to solve general monotone inclusion (1), where the second operator has been linearly composed. As a result, we get an extended version of [9, Algorithm 2]. Its main recursive formulae include: Choose $t \in [0,1]$. For known primal iterate x^k and dual iterate v^k, compute

$$\begin{aligned} y^k &= (\alpha I + A)^{-1}(\alpha x^k - L^* v^k), \\ \hat{y}^k &:= (1-t)x^k + ty^k, \\ u^k &= (\beta I + B)^{-1}(\beta(L\hat{y}^k - r) + v^k), \end{aligned}$$

where $\alpha > 0$ and $\beta > 0$ are scaling factors. Then, we explain the reason why

$$-\Big(\alpha(x^k - y^k) + \beta L^*(L\hat{y}^k - r - u^k),\ u^k - Ly^k + r\Big)$$

can be viewed as a descent direction of the distance function $\|(x,v) - (x^*,v^*)\|^2$ with respect to primal-dual variables (x,v) at the current point (x^k,v^k), even if the primal-dual solution point (x^*,v^*) is unknown. Thus, it becomes no longer difficult to design the corresponding splitting method as done in Sect. 3. As to its weak convergence of the primal sequence to an element of the associated solution set, our proof mainly depends on known results such as Lemma 2.1 and Lemma 4.1, and thus looks direct and not convoluted.

The rest of this article is organized as follows. In Sect. 2, we give some useful concepts and preliminary results. In Sect. 3, we derive our full extension of [9, Algorithm 2] in Hilbert spaces. When specialized to the $t = 0$ case, it recovers an equivalent version of a splitting method described in [14, Proposition 3.5]. In contrast, our extension can take full advantage of the latest information on iterates, thus becomes more desirable in some cases. In Sect. 4, we analyze weak convergence of primal sequence generated by our extended splitting method in a different way from existing ones. In Sect. 5, we did numerical experiments to clarify the necessity of introducing the parameter. In Sect. 6, we close this article by some concluding remarks.

2 Preliminary Results

In this section, we first give some basic definitions and then provide some auxiliary results for later use.

Let $\mathcal{H}$ be an infinite-dimensional Hilbert space, in which $\langle x, y\rangle$ stands for the usual inner product and $\|x\| := \sqrt{\langle x, x\rangle}$ for the induced norm for any $x, y \in \mathcal{H}$. I stands for the identity operator, i.e., $Ix = x$ for all $x \in \mathcal{H}$. $\mathrm{dom}T$ stands for the effective domain of T, i.e., $\mathrm{dom}T := \{x \in \mathcal{H} : Tx \neq \emptyset\}$.

Definition 2.1 Let $T : \mathcal{H} \to \mathcal{H}$ be a single-valued operator. If there exists some constant number $\kappa > 0$ such that

$$\|T(x) - T(y)\| \leq \kappa\|x - y\|, \quad \forall x, y \in \mathcal{H},$$

then T is called Lipschitz continuous.

To concisely give the following definition, we agree on that the notation $(x, w) \in T$ and $x \in \mathcal{H}$, $w \in T(x)$ have the same meaning. Moreover, $w \in Tx$ if and only if $x \in T^{-1}w$, where T^{-1} stands for the inverse of T.

Definition 2.2 Let $T : \mathcal{H} \rightrightarrows \mathcal{H}$ be an operator. It is called monotone if and only if

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

maximal monotone if and only if it is monotone and for given $\hat{x} \in \mathcal{H}$ and $\hat{w} \in \mathcal{H}$ the following implication relation holds

$$\langle x - \hat{x}, \ w - \hat{w}\rangle \geq 0, \quad \forall (x, w) \in T \quad \Rightarrow \quad (\hat{x}, \hat{w}) \in T.$$

Definition 2.3 Let $T : \mathcal{H} \rightrightarrows \mathcal{H}$ be an operator. It is called μ-strongly monotone if and only if there exists $\mu > 0$ such that

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

Definition 2.4 Let $f : \mathcal{H} \to (-\infty, +\infty]$ be a closed proper convex function. Then for any given $x \in \mathcal{H}$ the sub-differential of f at x is defined by

$$\partial f(x) := \{s \in \mathcal{H} : f(y) - f(x) \geq \langle s, y - x \rangle, \ \forall y \in \mathcal{H}\}.$$

Each s is called a sub-gradient of f at x. Moreover, if f is further continuously differentiable, then $\partial f(x) = \{\nabla f(x)\}$, where $\nabla f(x)$ is the gradient of f at x.

As is well known, the sub-differential of any closed proper convex function in an infinite-dimensional Hilbert space is maximal monotone as well. An example is the following indicator function

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

where $\mathcal{C}$ is some nonempty closed convex set in $\mathcal{R}^n$, and its sub-differential must be closed, proper convex. Furthermore, for any given positive number $\lambda > 0$, we have

$$P_{\mathcal{C}} = (I + \lambda \partial \delta_{\mathcal{C}})^{-1},$$

where $P_{\mathcal{C}}$ is usual projection onto C and is defined by

$$P_{\mathcal{C}}(u) = \operatorname{argmin}\{\|u - x\| : x \in \mathcal{C}\}.$$

For any given maximal monotone operator $T : \mathcal{H} \rightrightarrows \mathcal{H}$, it is Minty [18] who proved that there must exist a unique $y \in \mathcal{H}$ such that $(I + \lambda T)(y) \ni x$ for all $x \in \mathcal{H}$ and $\lambda > 0$. This implies that the corresponding operator $(I + \lambda T)^{-1}$ is single-valued.

For any given maximal monotone operator $T : \mathcal{H} \rightrightarrows \mathcal{H}$, there are other related properties. (i) For all $x \in \mathcal{H}$, the set $T(x)$ must be either empty or nonempty closed convex; see [19, Proposition 3, § 6.7]. (ii) The solution set $\{x : 0 \in T(x)\}$ is either empty or nonempty closed convex.

An important instance of the problem (1) above is $A = \partial f, B = \partial g$, where $f : \mathcal{H} \to (-\infty, +\infty]$, $g : \mathcal{G} \to (-\infty, +\infty]$ are closed proper convex functions. Under suitable conditions, (1) corresponds to the following convex minimization

$$\min_{x \in \mathcal{H}} \ f(x) + g(Lx - r). \tag{3}$$

Note that the problem (1) results in the associated Kuhn-Tucker set

$$Z := \left\{(x, v) \in (\mathcal{H}, \mathcal{G}) : -L^* v \in A(x), \ \ Lx - r \in B^{-1}(v)\right\}, \tag{4}$$

where x and v are called primal variable and dual variable, respectively. Next, we will show that such a set must be closed convex according to the following lemma

Lemma 2.1 *Let A, B, L be operators defined in the problem (1). Then the resulting operator by Attouch-Thera duality principle*

$$T(x,v)=\begin{pmatrix}A & 0\\ 0 & B^{-1}\end{pmatrix}\begin{pmatrix}x\\ v\end{pmatrix}+\begin{pmatrix}0 & L^*\\ -L & 0\end{pmatrix}\begin{pmatrix}x\\ v\end{pmatrix}+\begin{pmatrix}0\\ r\end{pmatrix} \tag{5}$$

must be maximal monotone.

Proof Note that both A and B are maximal monotone. Thus, the first operator on the right-hand side must be maximal monotone. Meanwhile, the linearity of L means that the second must be maximal monotone as well [18]. Maximality of T follows from [20, Theorem 1].

Although Lemma 2.1 is a known result, we would like to follow [21, Proposition 2] or [22, Corollary 4.2] to give such a short proof for completeness.

Remark 2.1 In his PhD dissertation, the author [10, Chapter 4] suggested using some methods such as [10, Algorithm 4.2.1] (i.e., [11, Algorithm 3.0]) to solve the monotone inclusion (5) indirectly instead of its original problem (1) directly. Such idea was exploited in [23]. Notice that this dissertation can be found in the author's Researchgate.

Next, we would like to mention the Douglas-Rachford splitting method once again. Its main recursive formulae may be:

$$\begin{aligned}&(I+\mu A)(x^k)\ni z^k,\\ &(I+\mu B)(y^k)\ni 2x^k-z^k,\\ &z^{k+1}=z^k-\gamma_k(x^k-y^k),\end{aligned}$$

where the parameter γ_k may vary from iteration to iteration and it satisfies

$$0<\gamma_k<2,\quad \textstyle\sum_{k=0}^{+\infty}\gamma_k(2-\gamma_k)=+\infty. \tag{6}$$

For weak convergence of auxiliary sequence $\{z^k\}$, we refer to [24] for detailed discussions. On the other hand, [25] suggested an equivalent version: for known $x^k\in\mathcal{H}$, $a^k\in A(x^k)$, it has the following form

$$\begin{aligned}&(I+\mu B)(y^k)\ni x^k-\mu a^k,\\ &(I+\mu A)(x^{k+1})\ni x^k+\mu a^k-\gamma_k(x^k-y^k),\end{aligned}$$

where γ_k satisfies (6). Recently, [5] proved weak convergence of the main sequence $\{x^k\}$ generated by the Douglas-Rachford splitting method. Impressively, the case of $\gamma_k\geq 2$ was discussed there for the first time. In the year 2012, the author also confirmed that, if the forward operator A is further Lipschitz continuous, then the set sequence $\{(A+B)y^k\}$ asymptotically includes the origin and the speed of inclusion is at $o(1/k)$ in a sense. This result was first cited as [26, Theorem 2.2.4] in Zhou's master's thesis, completed in March of 2012. For an explanation, we refer to the manuscript entitled "An asymptotic inclusion speed for the Douglas-Rachford splitting method in Hilbert spaces", accepted by Optimization Online in December of 2014. For pertinent discussions, we refer to [27–31] for more details.

In such a two-operator case, the Spingarn splitting method [32, Sect. 5] reads

$$
\begin{aligned}
&(I+\mu A)(x^k) \ni z^k - \mu v^k, \;\; a^k \in A(x^k),\\
&(I+\mu B)(y^k) \ni z^k + \mu v^k, \;\; b^k \in B(x^k),\\
&z^{k+1} = \frac{1}{2}(x^k+y^k), \quad v^{k+1} = v^k + \frac{1}{2\mu}(x^k - y^k).
\end{aligned}
$$

See [9, Sect. 4.2] for a detailed explanation of why it can be viewed as an instance of the Douglas-Rachford splitting method for monotone inclusion.

From these splitting methods, we can see that the scaling factors for A and B are identical. In the next section, we will find that the splitting method described in [9, Sect. 4.3] and its extensions bypass this limitation.

3 Splitting methods

In this section, we describe an iterative schemes for approximating the Kuhn-Tucker set. It involves the parameter $t \in [0,1]$ and recovers an equivalent version of [14, Proposition 3.5] in the $t = 0$ case.

Next, we will give a detailed description of our extension of [9, Algorithm 2] to solving the monotone inclusion (1) under consideration.

Algorithm 3.1

Step 0. Choose $x^0 \in \mathcal{H}$, $v^0 \in \mathcal{G}$. Choose $t \in [0,1]$ and $\theta \in (0,2)$. Choose $\alpha > 0,\ \beta > 0$ such that

$$4\alpha > \beta t^2 \|L\|^2. \tag{7}$$

Set $k := 0$.

Step 1. For known x^k, v^k, compute

$$(\alpha I + A)(y^k) \ni \alpha x^k - L^* v^k, \tag{8}$$

$$\hat{y}^k := (1-t)x^k + t y^k, \tag{9}$$

$$(\beta I + B)(u^k) \ni \beta(L\hat{y}^k - r) + v^k, \;\; u^k := Lz^k - r. \tag{10}$$

Compute $\gamma_k = \theta t_{1,k}/t_{2,k}$, where $t_{1,k}$ $t_{2,k}$ are given by

$$\begin{aligned} t_{1,k} &= \alpha\|x^k - y^k\|^2 + \beta\langle Lx^k - Lz^k, L\hat{y}^k - Lz^k\rangle \\ &= \alpha\|x^k - y^k\|^2 + \beta\langle Lx^k - r - u^k, L\hat{y}^k - r - u^k\rangle, \\ t_{2,k} &= \|\alpha(x^k - y^k) + \beta L^*(L\hat{y}^k - Lz^k)\|^2 + \|Lz^k - Ly^k\|^2 \\ &= \|\alpha(x^k - y^k) + \beta L^*(L\hat{y}^k - r - u^k)\|^2 + \|u^k - Ly^k + r\|^2. \end{aligned}$$

Step 2. Compute

$$\begin{aligned} x^{k+1} &= x^k - \gamma_k \left(\alpha(x^k - y^k) + \beta L^*(L\hat{y}^k - Lz^k)\right) \qquad (11) \\ &= x^k - \gamma_k \left(\alpha(x^k - y^k) + \beta L^*(L\hat{y}^k - r - u^k)\right), \\ v^{k+1} &= v^k - \gamma_k(Lz^k - Ly^k) \qquad (12) \\ &= v^k - \gamma_k(u^k - Ly^k + r). \end{aligned}$$

Set $k := k + 1$.

Remark 3.1 In the case of $L = I$ and $r = 0$, Algorithm 3.1 reduces to the aforementioned splitting method [9, Algorithm 2]. Moreover, the condition (7) is reminiscent of the counterpart [9, relation (14)], in our notation, which reads $4\alpha > \beta t^2$. Thus, these two conditions are identical. Of course, if $t = 0$, then the condition (7) holds automatically. Furthermore, in the $t = 0$ case, elementary calculations can indicate that Algorithm 3.1 is equivalent to the one in [14, Proposition 3.5] in theory, and the latter further subsumes [11, Algorithm 3.1].

Remark 3.2 Now let us make some remarks on the condition (7). In some practical applications, the setting becomes in finite-dimensional Euclidean spaces and L corresponds to an $m \times n$ matrix, say (l_{ij}). Since it follows from [33, Sect. 2.3] that

$$\|L\|^2 \le \max_{j=1,\dots,n} \sum_{i=1}^{m} |l_{ij}| \cdot \max_{i=1,\dots,m} \sum_{j=1}^{n} |l_{ij}|,$$

the condition (7) can be replaced by

$$4\alpha > \beta t^2 \max_{j=1,\dots,n} \sum_{i=1}^{m} |l_{ij}| \cdot \max_{i=1,\dots,m} \sum_{j=1}^{n} |l_{ij}|. \qquad (13)$$

Obviously, the larger the coefficient t is, the better information Algorithm 3.1 can take full advantage of.

Remark 3.3 Be care of that the scaling factors in Algorithm 3.1 may vary from iteration to iteration. For example, we may replace them by α_k and β_k and the latter two satisfy

$$0 < \alpha_{\min} \le \alpha_k \le \alpha_{\max} < +\infty, \quad 0 < \beta_{\min} \le \beta_k \le \beta_{\max} < +\infty,$$

and satisfy a condition that corresponds to (7)

$$\lim \inf_{k \to +\infty} \left\{ 1 - \frac{\beta_k t^2 \|L\|^2}{4\alpha_k} \right\} > 0.$$

Yet, for notational simplicity, we fix them throughout this article.

Recently, another iterative scheme was analyzed in, e.g., [12, 13]: At k-th iteration, its main steps read

$$\begin{aligned}
&(\alpha I + A)(\bar{x}^k) \ni \alpha x^k - L^* v^k, \\
&(\beta I + B^{-1})(\bar{v}^k) \ni \beta v^k + L(2\bar{x}^k - x^k) - r, \\
&x^{k+1} = x^k - \gamma(x^k - \bar{x}^k), \\
&v^{k+1} = v^k - \gamma(v^k - \bar{v}^k),
\end{aligned}$$

where $\gamma \in (0, 2)$. If $\gamma = 1$ and $\alpha\beta > \|L\|^2$, then it can be viewed as a special of the proximal point algorithm [34, 35] (also see [36–41, 27, 28] for further discussion) in a sense. In contrast, Algorithm 3.1 seems beyond such framework. In addition, our condition (7) is milder than $\alpha\beta > \|L\|^2$ as required in their method. This is because that it indicates a restriction $\alpha = \beta > \|L\|$ whenever two scaling factors are identical.

We just give several splitting methods for solving monotone inclusion (1) mentioned above, and as we see, each has a nice feature: decouple the second operator B from linearly-composed part L successfully. Therefore, each can solve the problem such as (1) in a desirable way.

4 Convergence

In this section, for the primal sequence generated by Algorithm 3.1, we analyze its weak convergence to an element of the associated solution set. Our proof techniques are different from existing ones [9, 14].

The following lemma is a well-known result, we also refer to [42] for a discussion of a special case.

Lemma 4.1 *Consider any maximal monotone mapping $T : \mathcal{H} \rightrightarrows \mathcal{H}$. Assume that the sequence $\{w^k\}$ in $\mathcal{H}$ converges weakly to w, and the sequence $\{s^k\}$ on dom$T$ converges strongly to s. If $T(w^k) \ni s^k$ for all k, then the relation $T(w) \ni s$ must hold.*

Lemma 4.2 *Let $L : \mathcal{H} \to \mathcal{G}$ be nonzero, bounded and linear operator, and let $\alpha > 0$. If $\beta > \|L\|^2/(4\alpha)$, then the following*

$$\alpha\|x\|^2 + \langle Lx, v\rangle + \beta\|v\|^2 \geq \frac{1}{2}\left(\alpha + \beta - \sqrt{\|L\|^2 + (\alpha - \beta)^2}\right)\left(\|x\|^2 + \|v\|^2\right)$$

holds for all $x \in \mathcal{H}$ and all $v \in \mathcal{G}$.

To the author's best knowledge, Lemma 4.2 seems due to [15, Lemma 5.1]. Very recently, such a nice result was used in [17] and generalized in [16].

Theorem 4.1 *Let $\{x^k\}$, $\{v^k\}$ be the primal and dual sequences generated by Algorithm 3.1. Then*

$$\begin{aligned}&\|x^{k+1}-x^*\|^2+\|v^{k+1}-v^*\|^2\\ \le\ &\|x^k-x^*\|^2+\|v^k-v^*\|^2-2\gamma_k\left(\alpha\|x^k-y^k\|^2+\beta\langle Lx^k-Lz^k,L\hat{y}^k-Lz^k\rangle\right)\\ &+\gamma_k^2\left(\|\alpha(x^k-y^k)+\beta L^*(L\hat{y}^k-Lz^k)\|^2+\|Lz^k-Ly^k\|^2\right).\end{aligned} \tag{14}$$

Furthermore, when $4\alpha>\beta t^2\|L\|^2$, the primal sequence $\{x^k\}$ weakly converges to an element of the solution to the problem (1).

Proof Since there have been x^*, v^* such that

$$-L^*v^*\in A(x^*),\quad Lx^*-r\in B^{-1}(v^*), \tag{15}$$

it follows from (15) and (8) that

$$A(y^k)\ni\alpha(x^k-y^k)-L^*v^k,\quad A(x^*)\ni -L^*v^*,$$

which, together with monotonicity of A, imply

$$\langle y^k-x^*,\ \alpha(x^k-y^k)-L^*(v^k-v^*)\rangle\ge 0,$$

namely

$$\langle x^k-x^*,\ \alpha(x^k-y^k)\rangle-\langle Ly^k-Lx^*,\ v^k-v^*\rangle\ge\alpha\|x^k-y^k\|^2. \tag{16}$$

On the other hand, it follows from (15) and (10) that

$$B(Lz^k-r)\ni\beta(L\hat{y}^k-Lz^k)+v^k,\quad B(Lx^*-r)\ni v^*,$$

which, together with monotonicity of B, imply

$$\langle Lz^k-Lx^*,\ \beta(L\hat{y}^k-Lz^k)+v^k-v^*\rangle\ge 0,$$

namely

$$\langle x^k-x^*,\ \beta L^*(L\hat{y}^k-Lz^k)\rangle+\langle Lz^k-Lx^*,\ v^k-v^*\rangle\ge\beta\langle Lx^k-Lz^k,\ L\hat{y}^k-Lz^k\rangle. \tag{17}$$

Summing up (16) and (17), together with Lemma 4.2 yields

$$\begin{aligned}&\langle x^k-x^*,\ \alpha(x^k-y^k)+\beta L^*(L\hat{y}^k-Lz^k)\rangle+\langle v^k-v^*,\ Lz^k-Ly^k\rangle\\ \ge\ &\alpha\|x^k-y^k\|^2+\beta\langle Lx^k-Lz^k,L\hat{y}^k-Lz^k\rangle \qquad (18)\\ =\ &\alpha\|x^k-y^k\|^2+\beta\langle Lx^k-Lz^k,L((1-t)x^k+ty^k)-Lz^k\rangle\\ =\ &\alpha\|x^k-y^k\|^2+\beta\|Lx^k-Lz^k\|^2-t\beta\langle Lx^k-Lz^k,Lx^k-Ly^k\rangle\\ \ge\ &\frac{1}{2}\left(\alpha+\beta-\sqrt{t^2\beta^2\|L\|^2+(\alpha-\beta)^2}\right)\left(\|x^k-y^k\|^2+\|Lx^k-Lz^k\|^2\right), \qquad (19)\end{aligned}$$

where the first equality makes use of the fact (9). By (11) and (12), we have

$$\|x^{k+1} - x^*\|^2 = \|x^k - x^* - \gamma_k \left(\alpha(x^k - y^k) + \beta L^*(L\hat{y}^k - Lz^k)\right) \|^2,$$
$$\|v^{k+1} - v^*\|^2 = \|v^k - v^* - \gamma_k(Lz^k - Ly^k)\|^2.$$

From these two relations, we further have

$$\begin{aligned}
&\|x^{k+1} - x^*\|^2 + \|v^{k+1} - v^*\|^2 \\
= &\|x^k - x^*\|^2 + \|v^k - v^*\|^2 \\
&- 2\gamma_k \left(\langle x^k - x^*,\ \alpha(x^k - y^k) + \beta L^*(L\hat{y}^k - Lz^k)\rangle + \langle v^k - v^*,\ Lz^k - Ly^k\rangle\right) \\
&+ \gamma_k^2 \left(\|\alpha(x^k - y^k) + \beta L^*(L\hat{y}^k - Lz^k)\|^2 + \|Lz^k - Ly^k\|^2\right) \\
\le &\|x^k - x^*\|^2 + \|v^k - v^*\|^2 - 2\gamma_k \left(\alpha\|x^k - y^k\|^2 + \beta\langle Lx^k - Lz^k, L\hat{y}^k - Lz^k\rangle\right) \\
&+ \gamma_k^2 \left(\|\alpha(x^k - y^k) + \beta L^*(L\hat{y}^k - Lz^k)\|^2 + \|Lz^k - Ly^k\|^2\right) \\
= &\|x^k - x^*\|^2 + \|v^k - v^*\|^2 - (2-\theta)\gamma_k \left(\alpha\|x^k - y^k\|^2 + \beta\langle Lx^k - Lz^k, L\hat{y}^k - Lz^k\rangle\right),
\end{aligned}$$

which, together with (18) and (19) and the formula for γ_k in Algorithm 3.1, implies

$$\begin{aligned}
&\|x^{k+1} - x^*\|^2 + \|v^{k+1} - v^*\|^2 \\
\le &\|x^k - x^*\|^2 + \|v^k - v^*\|^2 - \frac{2-\theta}{2}\gamma_k \left(\alpha + \beta - \sqrt{t^2\beta^2\|L\|^2 + (\alpha-\beta)^2}\right) \left(\|x^k - y^k\|^2 + \|Lx^k - Lz^k\|^2\right)
\end{aligned}$$

Moreover, the denominator of γ_k can be bounded by

$$\begin{aligned}
&\|\alpha(x^k - y^k) + \beta L^*(L\hat{y}^k - Lz^k)\|^2 + \|Lz^k - Ly^k\|^2 \\
\le &2\left(\|\alpha(x^k - y^k)\|^2 + \|\beta L^*(L\hat{y}^k - Lz^k)\|^2\right) + 2\left(\|Lx^k - Ly^k\|^2 + \|Lx^k - Lz^k\|^2\right),
\end{aligned}$$

where we use the inequality $\|a+b\|^2 \le 2(\|a\|^2 + \|b\|^2)$ for all a, b in $\mathcal{H}$, and it will be used below several times. Meanwhile, we also have

$$\begin{aligned}
\|\beta L^*(L\hat{y}^k - Lz^k)\|^2 &= \|\beta L^*(Lx^k - Lz^k) - t\beta L^*(Lx^k - Ly^k)\|^2 \\
&\le 2\left(\|\beta L^*(Lx^k - Lz^k)\|^2 + \|t\beta L^*(Lx^k - Ly^k)\|^2\right) \\
&\le 2\left(\beta^2\|L^*\|^2\|Lx^k - Lz^k\|^2 + t^2\beta^2\|\ L^*L\|^2\|x^k - y^k\|^2\right).
\end{aligned}$$

Thus, an upper bound of the denominator of γ_k is given by

$$(2\alpha^2 + 4t^2\beta^2\|L^*L\|^2 + 2\|L\|^2)\,\|x^k - y^k\|^2 + (4\beta^2\|L^*\|^2 + 2)\,\|Lx^k - Lz^k\|^2.$$

Combining this with (18) and (19) and the formula for γ_k yields

$$\gamma_k \ge \frac{1}{2}\left(\alpha + \beta - \sqrt{t^2\beta^2\|L\|^2 + (\alpha-\beta)^2}\right)\sigma^{-1},$$

where the constant σ is given by

$$\sigma := \max\left\{2\alpha^2 + 4t^2\beta^2\|L^*L\|^2 + 2\|L\|^2,\ 4\beta^2\|L^*\|^2 + 2\right\}.$$

Thus, when $4\alpha > \beta t^2\|L\|^2$, the following statements hold

$$(i)\ \{(x^k, v^k) - (x^*, v^*)\} \text{ converges in norm, thus } \{(x^k, v^k)\} \text{ is bounded in norm;} \tag{20}$$

$$(ii)\ \|x^k - y^k\| \to 0,\ \ \|Lx^k - Lz^k\| \to 0. \tag{21}$$

On the other hand, it follows from (5) that

$$T(y^k,\ \beta(L\hat{y}^k - Lz^k) + v^k) = \begin{pmatrix} Ay^k + \beta L^*(L\hat{y}^k - Lz^k) + L^*v^k \\ B^{-1}(\beta(L\hat{y}^k - Lz^k) + v^k) - Ly^k + r \end{pmatrix},$$

which, together with (8) and (10), i.e.,

$$\begin{aligned} &Ay^k \ni \alpha(x^k - y^k) - L^*v^k, \\ &B^{-1}(\beta(L\hat{y}^k - Lz^k) + v^k) \ni Lz^k - r, \end{aligned}$$

implies

$$T(y^k,\ \beta(L\hat{y}^k - Lz^k) + v^k) \ni \begin{pmatrix} \alpha(x^k - y^k) + \beta L^*(L\hat{y}^k - Lz^k) \\ Lz^k - Ly^k \end{pmatrix}.$$

This is equivalent to saying that

$$T(y^k,\ \beta(L\hat{y}^k - Lz^k) + v^k) \ni \begin{pmatrix} \alpha(x^k - y^k) + \beta L^*(Lx^k - Lz^k - t(Lx^k - Ly^k)) \\ Lx^k - Ly^k + Lz^k - Lx^k \end{pmatrix}. \tag{22}$$

The property above (20) tells us that $\{(x^k, v^k)\}$ has at least one weak cluster point, say $\{(x^\infty, v^\infty)\}$. This means that there exists some subsequence of $\{(x^k, v^k)\}$ such that

$$x^{k_j} \rightharpoonup x^\infty, \quad v^{k_j} \rightharpoonup v^\infty, \quad \text{as } k_j \to +\infty,$$

where the notation "$\rightharpoonup$" stands for weak convergence. It follows from this and (21) that

$$y^{k_j} \rightharpoonup x^\infty, \quad \text{as } k_j \to +\infty.$$

Since the operator L is bounded and linear, a similar discussion yields

$$\beta(L\hat{y}^{k_j} - Lz^{k_j}) + v^{k_j} \rightharpoonup v^\infty, \quad \text{as } k_j \to +\infty.$$

Moreover, by (21) and boundedness and linearity of L, we can see that the upper term and the lower term on the right-hand side of (22) converge in norm strongly to zero, respectively, even as $k \to +\infty$. Combining these facts and Lemma 4.1 implies that, if $\{(x^\infty, v^\infty)\}$ is any weak cluster point of $\{(x^k, v^k)\}$, then it must be zero of T. Meanwhile, the proof of uniqueness of weak cluster point is standard [35], thus is omitted. So the whole sequence $\{x^k\}$ weakly converges to x^∞, which is the solution to the problem (1).

Obviously, it follows from (10) and (19) that, if the condition (7) holds, then

$$-\Big(\alpha(x^k - y^k) + \beta L^*(L\hat{y}^k - r - u^k),\ u^k - Ly^k + r\Big)$$

can serve as a descent direction of the distance function $\|(x,v) - (x^*,v^*)\|^2$ with respect to (x,v) at the current point (x^k,v^k), even if the primal-dual solution point (x^*,v^*) is unknown.

Remark 4.1 For the primal sequence generated by Algorithm 3.1 addressed in real Hilbert spaces, here we have made use of Lemma 2.1 and Lemma 4.1 to analyze weak convergence to an element of the associated solution set. Of course, we may invoke an alternative proof technique [14, Proposition 2.4] as well. Yet, we feel that our doing is not merely for self-contained style and our proof looks direct and not convoluted indeed.

5 Rudimentary numerical experiments

In this section, we did numerical tests to demonstrate potential advantages of Algorithm 3.1 with $t = 0.7$ over Algorithm 3.1 with $t = 0$ for some problems when other parameters had been properly chosen.

All numerical experiments were run in MATLAB R2014a (8.3.0.532) with 32-bit (win32) on a desktop computer with an Intel(R) Core(TM) i3-2120 CPU 3.30 GHz and 2 GB of RAM. The operating system is Windows XP Professional.

Our test problem is to solve the following monotone inclusion

$$0 \in x - p + L^T \partial\delta_{\mathcal{C}}(Lx),$$

where

$$p = (2,0,...,0)^T \in R^n,\ L = diag(1,1/2,...,1/n),\ \mathcal{C} = \{w \in R^n : \|w\| \le 1\}.$$

Obviously, its unique solution is $x^* = (1,0,...,0)^T$. The stopping criterion is

$$\|x^k - x^*\| \le 10^{-4}.$$

In practical implementations, we chose $n = 10000$ and chose the starting points as $x^0 = (1,...,1)^T,\ v^0 = (0,...,0)^T$.

After many times of trials, we found out that

$$\theta = 1.8,\quad \alpha \in \{0.7,0.8,0.9,1.0\},\quad \beta \in \{0.8,0.9,1.0,1.1,1.2\}$$

are good choices.

The corresponding numerical results were reported in the following Table 1, where the part above (below) stands for the numbers of iteration of Algorithm 3.1 (the elapsed time using tic and toc in seconds) in each of the four rows in the bottom of this table. Moreover, the notation ”–” corresponds to the case that the number of iteration is strictly larger than 9.

Table 1: Numerical results of Algorithm 3.1 (t=0 vs t=0.7)

β / α	0.8		0.9		1.0		1.1		1.2	
0.7	–	–	–	9	–	8	–	9	–	9
	–	–	–	17.25	–	15.41	–	17.27	–	17.47
0.8	–	8	–	8	–	7	–	7	–	8
	–	15.48	–	15.49	–	13.40	–	13.55	–	15.60
0.9	–	–	–	8	9	9	–	9	–	–
	–	–	–	15.50	17.41	17.37	–	17.27	–	–
1.0	–	–	–	9	–	–	–	–	–	–
	–	–	–	17.40	–	–	–	–	–	–

From Table 1, we can see that Algorithm 3.1 with $t = 0$ was outperformed by Algorithm 3.1 with $t = 0.7$ for our test problem when other parameters had been properly chosen. Specifically speaking, the fewest number of iteration of the former is 9 and the elapsed time is 17.41 seconds. This corresponds to only one case : $\alpha = 0.9$ and $\beta = 1.0$. In contrast, there are 13 cases in which the latter's numbers of iteration are 9, even 8 or 7. In particular, in the case of $\alpha = 0.8$ and $\beta = 1.0$, the fewest number of iteration of the latter is 7 and the elapsed time is 13.40 seconds.

6 Conclusions

In this article, we have suggested an extended splitting method for linearly-composed monotone inclusions in real Hilbert spaces, with a nice feature of different scaling factors. This method generates both primal sequence and dual sequence, and we have analyzed the former's weak convergence to an element of the associated solution set. Furthermore, we have shown that it subsumes several known splitting methods. We have stressed that our extended splitting method can take full advantage of the latest information on iterates both in theoretical analysis and in numerical results on an example.

Acknowledgement

The author is very grateful to Xiao Zhu for careful reading of this manuscript.

References

1. Tseng P (2000) A modified forward-backward splitting method for maximal monotone mapping. SIAM J Control Optim 38: 431–446
2. Lions PL, Mercier B (1979) Splitting algorithms for the sum of two nonlinear operators. SIAM J Numer Anal 16: 964–979
3. Passty GB (1979) Ergodic convergence to a zero of the sum of monotone oprators in Hilbert space. J Math Anal Appl 72: 383–390
4. Chen H, Rockafellar RT (1997) Convergence rates in forward-backward splitting. SIAM J Optim 7:421–444
5. Dong YD, Fischer A (2010) A family of operator splitting methods revisited. Nonlinear Anal 72: 4307–4315
6. Douglas J, Rachford HH (1956) On the numerical solution of the heat conduction problems in two and three space variables. Trans Am Math Soc 421–439
7. Peaceman D, Rachford HH (1955) The numerical solution of parabolic and elliptic differential equations. J Soc Indust Appl Math 28–41
8. Huang YY, Dong YD (2014) New properties of forward-backward splitting and a practical proximal-descent algorithm. Appl Math Comput 237: 60–68
9. Eckstein J, Svaiter BF (2008) A family of projective splitting methods for the sum of two maximal monotone operators. Math Program 111: 173-199
10. Dong YD (2003) Splitting Methods for Monotone Inclusions. PhD Dissertation, Nanjing University
11. Dong YD (2005) An LS-free splitting method for composite mappings. Appl Math Letters 18(8): 843–848
12. Vu BC (2013) A splitting algorithm for dual monotone inclusions involving cocoercive operators. Adv Comput Math 38(3): 667–681
13. Condat L (2013) A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. J Optim Theory Appl 158(2): 460–479
14. Alotaibi A, Combettes PL, Shahzad N (2014) Solving coupled composite monotone inclusions by successive Fejr approximations of their Kuhn-Tucker set. SIAM J Optim 24(4): 2076–2095
15. Latafat P, Patrinos P (2017) Asymmetric forward-backward-adjoint splitting for solving monotone inclusions involving three operators. Comput Optim Appl 68(1) 57–93
16. Dong YD (2017) A new splitting method for systems of monotone inclusions in Hilbert spaces. Manuscript

17. Dong YD, Yu XH (2019) A new splitting method for monotone inclusions of three operators. Calcolo 56: 3. https://doi.org/10.1007/s10092-018-0299-7
18. Minty GJ (1962) Monotone (nonlinear) operators in Hilbert space. Duke Math J 29: 341–346
19. Aubin JP, Ekeland I (1984) Applied Nonlinear Analysis, Interscience Publications. John Wiley, New York
20. Rockafellar RT (1970) On the maximality of sums of nonlinear monotone operators. T Am Math Soc 149(1): 75–88
21. Eckstein J, Ferris MC (1999) Smooth methods of multipliers for complementarity problems. Math Program 86:65–90
22. Pennanen T (2000) Dualization of generalized equations of maximal monotone type. SIAM J Optim 10(3): 809–835
23. Briceno-Arias L, Combettes PL (2011) A monotone+skew splitting model for composite monotone inclusions in duality. SIAM J Optim 21(4): 1230–1250
24. Combettes PL (2004) Solving monotone inclusions via compositions of nonexpansive averaged operators. Optim 53: 475–504
25. Eckstein J, Bertsekas DP (1992) On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math Program 55(3): 293–318
26. Zhou SF (2012) Operator splitting algorithms for ℓ_1-ℓ_2 problem and interior point method for its duality problem. Dissertation, Zhengzhou University
27. Dong YD (2014) The proximal point algorithm revisited. J Optim Theory Appl 161: 478–489
28. Dong YD (2015) Comments on "the proximal point algorithm revisited". J Optim Theory Appl 166: 343–349
29. Dong YD (2016) Douglas-Rachford splitting method for semi-definite programming. J Appl Math Comput 51: 569–591
30. Bello Cruz JY, Nghia Tran TA (2016) On the convergence of the forward-backward splitting method with linesearches. Optim Method Softw 31(6): 1209–1238
31. Matsushita SY (2017) On the convergence rate of the Krasnosel'skiĭ-Mann iteration. B Austr Math Soc. 96(1)(2017), pp. 162–170.
32. Spingarn JE (1983) Partial inverse of a monotone operator. Appl Math Optim 10(1): 247–265
33. Golub GH, Van Loan CF (1996) Matrix computations. 3rd ed. Baltimore Maryland, Johns Hopkins University Press
34. Moreau JJ (1965) Proximité et dualité dans un espace Hilbertien. Bull Soc Math Fr 93: 273–299
35. Rockafellar RT (1976) Monotone operators and the proximal point algorithm. SIAM J Control Optim 14: 877–898

36. Brézis H, Lions PL (1978) Produits infinis de resolvantes. Israel J Math 29: 329–345
37. Luque FJ (1984) Asymptotic convergence analysis of the proximal point algorithm. SIAM J Control Optim 22: 277–293
38. Güler O (1991) On the convergence of the proximal point algorithm for convex minimization. SIAM J Control Optim 29: 403–419
39. Pennanen T (2002) Local convergence of the proximal point algorithm and multiplier methods without monotonicity. Math Oper Res 27: 193–202
40. Dong YD (2006) A new relative error criterion for the proximal point algorithm. Nonlinear Anal 64: 2143–2148
41. Zaslavski AJ (2011) Maximal monotone operators and the proximal point algorithm in the presence of computational errors. J Optim Theory Appl 150: 20–32
42. Browder FE (1967) Convergence theorems for sequences of nonlinear operators in Banach spaces. Math Zeitschr 100: 201–225.