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

# An asymptotic inclusion speed for the Douglas-Rachford splitting method in Hilbert spaces

**Yunda Dong**

Received: date / Accepted: date

**Abstract** In this paper, we consider the Douglas-Rachford splitting method for monotone inclusion in Hilbert spaces. It can be implemented as follows: from the current iterate, first use forward-backward step to get the intermediate point, then to get the new iterate. Generally speaking, the sum operator involved in the Douglas-Rachford splitting takes the value of every intermediate point as a set. Our goal of this paper is to show that such generated set-valued sequence asymptotically includes the origin and the corresponding asymptotic inclusion speed remains desirable if the forward splitting is further Lipschitz continuous.

**Keywords** Monotone inclusion · Douglas-Rachford splitting · Proximal point algorithm · Asymptotic inclusion speed.

Yunda Dong

School of Mathematics and Statistics, Zhengzhou University, Zhengzhou 450001, P.R. China

E-mail: ydong@zzu.edu.cn

**Mathematics Subject Classification (2000)** **58E35 · 65K15**

## 1 Introduction

In this paper, we consider the following problem of finding an $x$ in a real Hilbert space $H$ such that

$$0 \in T(x), \quad T := A + B, \tag{1}$$

where both $A$ and $B$ are possibly set-valued, maximal monotone operators from $H$ to $H$. This problem serves as a simple, powerful and unified framework to describe definite linear systems, convex programs, monotone variational inequalities, and more general monotone inclusions, see Rockafellar [1] and Varga [2] for further discussions.

A known approach to the problem (1) above is the proximal point algorithm, which is based on the concept of proximal mapping [3] introduced by Moreau and Yosida. This algorithm was first proposed by Martinet [4] and generalized by Rockafellar [1] to get today's form: Choose a starting point $x^0 \in H$. For known $x^k \in H$, the new iterate $x^{k+1}$ is defined by

$$0 \in T(x^{k+1}) + \lambda_k^{-1}(x^{k+1} - x^k), \;\; k = 0, 1, ..., \tag{2}$$

where $\lambda_k > 0$ is called proximal parameter. See Brézis and Lions [6] and Güler [7] for some fundamental results. For related discussions, we refer to [8–11], to cite a few.

If one applies the proximal point algorithm to solving some problems arising in practice, then it can be not efficient. So, one may resort to its practical alternative: Choose a starting point $x^0 \in H$. For known $x^k \in H$, $a^k \in A(x^k)$, the new iterate $x^{k+1}$ is defined by

$$(I + \lambda B)(y^k) \ni x^k - \lambda a^k, \tag{3}$$

$$(I + \lambda A)(x^{k+1}) \ni x^k + \lambda a^k - \gamma(x^k - y^k). \tag{4}$$

where $\lambda$, $\gamma$ are positive numbers. Note that the $\gamma = 1$ case corresponds to the Douglas-Rachford (DR for short) splitting method of Lions and Mercier [5]. For this family of methods, the reader may consult [5, 12–16] for analysis of weak convergence and see [5, 14, 16] their rates of convergence.

Interestingly, Eckstein and Bertsekas [13] demonstrated that the proximal point algorithm and the DR splitting method are interrelated. Furthermore, they also made use of the former to characterize some properties of the DR splitting method. On the other hand, by means of an improvement a classical result of Brézis and Lions [6, Proposition 8] on the proximal point algorithm, the paper [17] recently discussed a speed at which the set sequence $\{T(x^k)\}$ asymptotically includes the origin; see [18] for a slight revision. So, we in this paper show how to extend this recently proposed result on the proximal point algorithm to the DR splitting method. As a result, we give a desirable speed at which the sequence $\{(A+B)(y^k)\}$ asymptotically includes the origin provided that the sum operator $A + B$ is maximal monotone as well and the forward operator $A$ is further Lipschitz continuous..

## 2 Preliminary Results

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

Let $H$ 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 $x, y \in H$. Let $U$ be a nonempty, closed and convex set in $H$. We use

$$|\cdot - U| := \min\{\|\cdot - u\| : u \in U\}$$

to stand for the usual distance between a point and a set $U$.

**Definition 2.1** Let $F : H \to H$ be continuous. It is called Lipschitz continuous if there exists some positive number $L > 0$ such that

$$\|F(x) - F(y)\| \leq L\|x - y\|, \quad \forall x \in H,\ \forall y \in H.$$

**Definition 2.2** Let $f : H \to ]-\infty, +\infty]$ be a closed, proper and convex function. Then, for any given $x \in H$, the sub-differential of $f$ at $x$ is defined by

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

Each element $s$ 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 $x$.

To concisely give the following definition, we agree on that the notation $(x, w) \in T$ and $x \in H$, $w \in T(x)$ have the same meaning.

**Definition 2.3** Let $T : H \rightrightarrows H$ be an operator. It is called monotone iff

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

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

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

As is well known, the sub-differential of any closed, proper and convex function in Hilbert spaces is maximal monotone as well. Furthermore, for any given maximal monotone operator $T : H \rightrightarrows H$, Minty [19] proved that there must exist a unique $y \in H$ such that $(I + \lambda T)(y) \ni x$ for all $x \in H$ and $\lambda > 0$, where $I$ stands for the identity operator, i.e., $I(x) = x$ for all $x \in H$. This implies that the corresponding operator $(I + \lambda T)^{-1}$ is single-valued.

For any given maximal monotone operator $T : H \rightrightarrows H$, there are other related properties. (i) For all $x \in H$, the set $T(x)$ must be either empty or nonempty, closed and convex; see [20]. (ii) The set $T^{-1}(0) := \{x : 0 \in T(x)\}$ is either empty or nonempty, closed and convex. Therefore, for all $x \in H$, the distance between the set $T(x)$ and the origin is uniquely determined, and so is the distance between the set $T^{-1}(0)$ and the origin.

For the problem (1), it is equivalent to finding an $x \in H$ such that

$$a \in A(x), \ \ -a \in B(x),$$

which is further equivalent to solving

$$x - (I + \lambda B)^{-1}(x - \lambda a) = 0,$$

where $\lambda$ is any given positive number. Some properties of the quantities of

$$\|x - (I + \lambda B)^{-1}(x - \lambda a)\|, \quad \|x - (I + \lambda B)^{-1}(x - \lambda a)\|/\lambda$$

with respect to the variable $\lambda$ can be found in [21, 22]. See [23] for a pertinent discussion.

At the end of this section, we would like to go to the recursive formula of the proximal point algorithm for minimizing a closed, proper and convex function $f : H \to ]-\infty, +\infty]$, which may read

$$0 \in \partial f(x^{k+1}) + \lambda_k^{-1}(x^{k+1} - x^k), \quad k = 0, 1, .... \tag{5}$$

In this setting, the following results are entirely attributed to Güler [7], and they are classical.

**Proposition 2.1** *Let* $\{(x^k, \lambda_k)\}$ *be the sequence in the proximal point algorithm for convex minimization. Assume that* $\lim_{k\to+\infty} \sum_{i=0}^{k-1} \lambda_i = +\infty$. *Then, we have*

$$(f(x^k) - \min f(x)) \sum_{i=0}^{k-1} \lambda_i \leq \frac{1}{2} |x^0 - \partial f^{-1}(0)|^2, \tag{6}$$

*where* $\partial f^{-1}(0)$ *stands for the set of all minimizers and is assumed to be nonempty. Furthermore, if the proximal point algorithm does converge strongly, then the associated convergence rate can be improved above to*

$$f(x^k) - \min f(x) = o\left(1/\sum_{i=0}^{k-1} \lambda_i\right), \tag{7}$$

*where the notation* $o$ *means that* $s_k = o(1/t_k)$ *if and only if* $\lim_{k\to+\infty} s_k t_k = 0$.

In fact, in the case of asymptotic rate, the assumption on algorithm's strong convergence can be removed by using [17, Lemma 2.2]; also see [18, Lemma 3.3]. This is because that

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - 2\lambda_k(f(x^{k+1}) - f(x^*))$$

and the sequence $\{f(x^k)\}$ is decreasing, see [7] for more details.

## 3 DR splitting

In this section, we mainly review the DR splitting method of Lions and Mercier and a deep understanding of Eckstein and Bertsekas as an instance of the proximal point algorithm.

For any given two maximal monotone operators $A$ and $B$, Lions and Mercier [5] defined DR splitting as follows:

$$G_{\lambda AB} := (I + \lambda B)^{-1} \circ (2(I + \lambda A)^{-1} - I) + I - (I + \lambda A)^{-1}. \tag{8}$$

where $\lambda > 0$ is a positive number and $\circ$ denotes functional composition. Moreover, they showed the basic property of $G_{\lambda AB}$, from which weak global convergence of the corresponding DR splitting method given by

$$z^{k+1} = G_{\lambda AB}(z^k) \tag{9}$$

was proved in an elegant way.

Note that, for the iterate $z^k$, there must exist the unique $x^k \in H$ such that

$$x^k + \lambda a^k = z^k, \;\; a^k \in A(x^k) \quad \Leftrightarrow \quad (I + \lambda A)^{-1}(z^k) := x^k. \tag{10}$$

Next, we have

$$(I-(I+\lambda A)^{-1})(z^k) = x^k + \lambda a^k - x^k = \lambda a^k$$

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

Set

$$(I+\lambda B)^{-1}(x^k - \lambda a^k) := y^k \quad \Leftrightarrow \quad y^k + \lambda b^k = x^k - \lambda a^k,\ \ b^k \in B(y^k).$$

Thus, in views of theses identities, one may follow the analysis of Eckstein and Bertsekas [13] to get the following alternative expression of the DR splitting method: For the current iterate $x^k \in H$, $a^k \in A(x^k)$, first compute the intermediate point $y^k$, and then get the new iterate $x^{k+1}$ by the recursion relation

$$(I+\lambda B)(y^k) \ni x^k - \lambda a^k, \tag{11}$$

$$(I+\lambda A)(x^{k+1}) \ni y^k + \lambda a^k. \tag{12}$$

Importantly, from these facts, Eckstein and Bertsekas [13] confirmed that (9) can be viewed as an instance of the proximal point algorithm. Specifically speaking, they first made use of the algorithmic description above to give the following expression for $G_{\lambda AB}$:

$$G_{\lambda AB} = \{(x+\lambda a, y+\lambda a)\ :\ (x,a)\in A,\ (y,b)\in B,\ x-\lambda a = y+\lambda b\}.$$

Then, they proved that $T_{\lambda AB} := (G_{\lambda AB})^{-1} - I$ is maximal monotone provided that both $A$ and $B$ are. Thus, (9) can be rewritten as

$$z^{k+1} = (I+T_{\lambda AB})^{-1}(z^k). \tag{13}$$

So, one may imagine finding a zero of $A + B$ by using the proximal point algorithm on $T_{\lambda AB}$ as done in (13) and taking the unique solution of

$$(I + \lambda A)(x) \ni z^k$$

as the desired iterate $x^k$, as designated above.

Note that, if we follow (4), then we may rewrite (12) as

$$(I + \lambda A)(x^{k+1}) \ni x^k + \lambda a^k - (x^k - y^k).$$

This corresponds to (4) in the case of $\gamma = 1$. For the family of splitting methods described by (3) and (4), under suitable conditions, convergence can be guaranteed even in the $\gamma > 2$ case. Furthermore, if the forward operator is further Lipschitz continuous and the other is strongly monotone, then linear rate of convergence was derived for the first time; see [16] for more details. When specialized to the context of finite-dimensional monotone variational inequality problems, the reader may consult [24–26] for related discussions.

## 4 Main Results

In this section, we derive a desirable speed at which the sequence $\{(A+B)(y^k)\}$ asymptotically includes the origin if we further assume that the involved sum operator is maximal monotone.

This section begins with some of fundamental properties of the proximal point algorithm.

**Proposition 4.1** *Let $\{(z^k, \lambda_k)\}$ be the corresponding iterate-parameter sequence in the proximal point algorithm (2). Assume that the problem (1) has at least one solution, and assume that*

$$\lim_{k\to+\infty}\sum_{i=0}^{k-1}\lambda_i^2 = +\infty. \tag{14}$$

*Then*

(a) $\{z^k\}$ *converges weakly, but not strongly in general*;

(b) $\|\lambda_k^{-1}(z^k - z^{k+1})\| \downarrow 0, \quad \text{as } k \to +\infty;$

(c) $|T(z^{k+1}) - 0|^2 \le \|\lambda_k^{-1}(z^k - z^{k+1})\|^2 \le o\left(1/\sum_{i=0}^{k}\lambda_i^2\right).$

Note that weak convergence in the statement (a) above was proved by Rockafellar [1] by assuming that the sequence of proximal parameters has a positive lower bound and was reproved by Brézis and Lions [6] by using a weaker assumption (14). The second half is attributed to Güler [7, Sect. 5] who constructed a famous example for which the proximal point algorithm does not converge strongly. The statement (b) was discussed by Brézis and Lions [6]. For the statement (c), it goes to [17, 18], in which a classical result of Brézis and Lions [6, Proposition 8] was improved from $O(\cdot)$ to $o(\cdot)$, without any other restrictions.

Below, we state the main results of this paper, and they are direct consequences of Proposition 4.1.

**Theorem 4.1** *Let $\{(x^k, y^k)\}$ be the sequence in the DR splitting method given by (11) and (12). Assume that the problem (1) has at least one solution, and* *assume that $A$, $B$ and $A+B$ are maximal monotone. Then, we have*

(a) $\{x^k\}$ *converges weakly, but not strongly in general*;

(b) $\|x^k - y^k\| \downarrow 0, \quad as \;\; k \to +\infty$;

(c) *If $A$ is further Lipschitz continuous, then* $|(A+B)(y^k) - 0|^2 \le o(1/k)$.

*Proof* Obviously, the statement (a) immediately follows from the corresponding one in Proposition 4.1. For the second statement, we may consider (10) and (12) and conclude that

$$x^k - y^k = x^k + \lambda a^k - (x^{k+1} + \lambda a^{k+1}) = z^k - z^{k+1}, \tag{15}$$

where $z^k$ can be viewed as the iterate generated by the proximal point algorithm (13), with $\lambda_k \equiv 1$. Thus, the statement (b) here easily follows. Now we prove the last statement. In fact, since $A$ is further Lipschitz continuous, (11) implies that

$$A(y^k) + B(y^k) \ni s^k := \lambda^{-1}(x^k - y^k) - (A(x^k) - A(y^k)).$$

and

$$\|s^k\| \le (\lambda^{-1} + L)\|x^k - y^k\|,$$

where $L > 0$ is the $A$'s Lipschitz constant. On the other hand, we have known (15). So, we can make use of the statement (c) in Proposition 4.1 to get

$$\|x^k - y^k\|^2 = \|z^k - z^{k+1}\|^2 = \|\lambda_k^{-1}(z^k - z^{k+1})\|^2 \le o(1/k),$$

where we have taken $\lambda_k \equiv 1$ as mentioned above. Therefore, we further have

$$|(A+B)(y^k) - 0|^2 \le \|s^k\|^2 \le (\lambda^{-1} + L)^2\|x^k - y^k\|^2 \le o(1/k).$$

Moreover, the set sequence $\{(A+B)(y^k)\}$ must asymptotically include the origin, as it is always nonempty, closed and convex. The proof is complete. $\square$

Note that our proof of the statement (b) in Theorem 4.1 is the shortest one because we take advantage of the known results presented in [5, 13]. Otherwise, the process of proof can be long, see [26, 27] for some instances.

## 5 Conclusions

Now we have studied the DR splitting method for monotone inclusion in Hilbert spaces. By combining the classical result of Eckstein and Bertsekas in 1992 with a recently proposed property of the proximal point algorithm, we have derived a desirable asymptotic inclusion speed for the DR splitting method if the forward splitting is further Lipschitz continuous. Thus, we have resolved the issue presented at the end of the paper [17]. Our next goal is to consider a self-adaptive version of the DR splitting method [16] and to check what will happen.

## References


1. Rockafellar R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 877-898 (1976)
2. Varga R.S.: Matrix Iterative Analysis. Prentice-Hall, Englewood Cliffs, NJ (1962)
3. Moreau J.J.: Proximité et dualité dans un espace Hilbertien. Bull. Soc. Math. Fr. **93**, 273-299 (1965).
4. Martinet B.: Regularisation d'inéquations variationelles par approximations successives. Rev. Fr. d'Informatique Recherche Opér. **4**, 154-158 (1970)

5. Lions P.L., Mercier B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**, 964-979 (1979)
6. Brézis H., Lions P.L.: Produits infinis de resolvantes. Isr. J. Math. **29**, 329-345 (1978)
7. Güler O.: On the convergence of the proximal point algorithm for convex minimization. SIAM J. Control Optim. **29**, 403-419 (1991)
8. Luque F.J.: Asymptotic convergence analysis of the proximal point algorithm. SIAM J. Control Optim. **22**, 277-293 (1984)
9. Pennanen T.: Local convergence of the proximal point algorithm and multiplier methods without monotonicity. Math. Oper. Res. **27**, 193-202 (2002)
10. Dong Y.D.: A new relative error criterion for the proximal point algorithm. Nonlinear Anal. **64**, 2143-2148 (2006)
11. Zaslavski A.J.: Maximal monotone operators and the proximal point algorithm in the presence of computational errors. J. Optim. Theory Appl. **150**, 20-32 (2011)
12. Lawrence J., Spingarn J.E.: On fixed points of non-expansive piecewise isometric mappings. P. Lond. Math. Soc. **55**, 605-624 (1987)
13. Eckstein J., Bertsekas D.P.: On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. **55(3)**, 293-318 (1992)
14. Eckstein J., Ferris M.C.: Operator-splitting methods for monotone affine variational inequalities, with a parallel application to optimal control. INFORMS J. Computing. **10**, (1998) 218-235.
15. Combettes P.L.: Solving monotone inclusions via compositions of nonexpansive averaged operators. Optim. **53**, 475-504 (2004)
16. Dong Y.D., Fischer A.: A family of operator splitting methods revisited. Nonlinear Anal. **72**, 4307-4315 (2010)
17. Dong Y.D.: The proximal point algorithm revisited. J. Optim. Theory Appl. **161**, 478-489 (2014)
18. Dong Y.D.: Comments on "The proximal point algorithm revisited". J. Optim. Theory Appl. http://dx.doi.org/10.1007/s10957-014-0685-5.

19. Minty G.J.: Monotone (nonlinear) operators in Hilbert space. Duke Math. J. **29**, 341-346 (1962)
20. Aubin J.P., Ekeland I.: Applied Nonlinear Analysis, Interscience Publications, John Wiley, New York (1984).
21. Tseng P.: A modified forward-backward splitting method for maximal monotone mappings. SIAM J. Control Optim. **38(2)**, 431-446 (2000)
22. Huang Y.Y., Dong Y.D.: New properties of forward-backward splitting and a practical proximal-descent algorithm. Appl. Math. Comput. **237**, 60-68 (2014)
23. Dong Y.D.: An extension of Luque's growth condition. Appl. Math. Lett. **22(9)**, 1390-1393 (2009)
24. He B.S.: Inexact implicit methods for monotone general variational inequalities. Math. Program. **86**, 199-227 (1999)
25. Huang N.J.: A new method for a class of nonlinear variational inequalities with fuzzy mappings. Appl. Math. Lett. **10**, 129-133 (1997)
26. Sun X.Z.: A descent property of implicit method for monotone variational inequality. Numer Math.: A J Chinese Universities. **32(1)**, 75-80 (2002)
27. He B.S., Yuan X.M.: On the convergence rate of Douglas-Rachford operator splitting method. Math. Prog. http://dx.doi.org/ 10.1007/s10107-014-0805-x.