On the linear convergence of the forward-backward splitting algorithm

B. K. Le

Abstract In this paper, we establish a linear convergence result for the *forward-backward* splitting algorithm in the finding a zero of the sum of two maximal monotone operators, where one of them is set-valued strongly monotone and the other is Lipschitz continuous. We show that our convergence rate is better than Douglas–Rachford splitting algorithm's rate used by Moursi–Vandenberghe (J Optim Theory Appl 183, 179–198, 2019) under the same assumptions in most of important subcases. In addition, the *forward-backward* splitting requires only one resolvent while it is necessary to compute the resolvents of both operators in Douglas–Rachford splitting, which means the *forward-backward* splitting is more preferable. We also compare the linear convergence rate of the *forward-backward* splitting to the rate of Douglas–Rachford splitting acquired by Giselsson (J Fixed Point Theory Appl 19, 2241–2270, 2017) when the Lipschitz continuity of the single-valued operator is strengthened with the cocoercivity.

Keywords Maximal monotone operators, forward-backward algorithm, Douglas-Rachford algorithm, linear convergence
Mathematics Subject Classification: 47H09, 47H05, 47J25

1 Introduction

The *forward-backward* splitting, firstly introduced by Passty [1], is one of the fundamental methods in solving optimization problems, particularly in minimizing the sum of two proper convex functions, which is equivalent to finding a zero of the sum of two maximal monotone operators A and B under some constraint qualifications. It is specially devoted to the case when one of the operators is single-valued and requires the computation of only the resolvent of the set-valued mapping. On the other hand, Douglas–Rachford splitting, proposed by Lions and Mercier [2], is also a powerful tool which can allow both set-valued operators but needs two resolvents consequently. The literature has witnessed numerous weak, strong and

B. K. Le
Faculty of Mathematics and Statistics, Ton Duc Thang University, Ho Chi Minh City, Vietnam
(lebakhiet@tdtu.edu.vn)

linear convergence results for the *forward-backward* algorithm (see, e.g., [3–8]) as well as Douglas–Rachford algorithm (see, e.g., [2, 9–14]).

In particular, linear convergence of splitting algorithms is usually associated with the strong monotonicity, Lipschitz continuity and cocoercivity of the involved operators. In [13], Giselsson provided tight global linear convergence rate bounds for Douglas–Rachford splitting for three cases: (*i*) A is cocoercive and B is strongly monotone; (*ii*) A is strongly monotone and Lipschitz continuous; and (*iii*) A is strongly monotone and cocoercive. It is known that cocoercivity is strictly stronger than Lipschitz continuity and it is not easy to verify the cocoercivity in general.

Recently, Moursi–Vandenberghe fulfilled this gap by relaxing the cocoercivity of A in the case (*i*) by only Lipschitz continuity [14]. The analysis is based on the averagedness of operators as in [13]. However the linear convergence rate of Douglas–Rachford used in [14] is not very nice and difficult to control. Our main contribution is to prove the linear convergence of the *forward-backward* algorithm under the same assumptions with better rate than Douglas–Rachford's used by Moursi–Vandenberghe in most of important subcases. Furthermore the *forward-backward* splitting does not have to compute two resolvents as Douglas–Rachford splitting does, which means the *forward-backward* splitting is more preferable. We also compare our convergence rate to the rate of Douglas–Rachford splitting considered in [13] when the Lipschitz continuity of A is strengthened with the cocoercivity.

The following is the organization of the paper. In Section 2, we recall some useful results related to monotone operators and the structure of the *forward-backward* algorithm as well as Douglas–Rachford algorithm. In Section 3, we prove the linear convergence of the *forward-backward* algorithm and provide a comparison with Douglas–Rachford algorithm used in [14] and [13]. We end the paper with some conclusions in Section 4.

2 Notations and preliminaries

Let H be a given Hilbert space with its norm $\| \cdot \|$ and scalar product $\langle \cdot, \cdot \rangle$. A mapping $A : H \to H$ is called *μ-strongly monotone* ($\mu > 0$) provided

$$\langle Ax - Ay, x - y \rangle \geq \mu \|x - y\|^2 \quad \forall \, x, y \in H.$$

It is called *L-Lipschitz continuous* ($L > 0$) if

$$\|Ax - Ay\| \leq L \|x - y\| \quad \forall \, x, y \in H.$$

If it is δ-Lipschitz continuous for some positive number $\delta \leq 1$, it is called *nonexpansive*. In addition, if $\delta < 1$, then A is called *contractive*.

It is called *α-cocoercive* if

$$\langle Ax - Ay, x - y \rangle \geq \alpha \|Ax - Ay\|^2, \quad \forall x, y \in H.$$

The domain, the range and the graph of a set-valued mapping $B : H \rightrightarrows H$ are defined respectively by

$$\text{dom}(B) = \{x \in H : \ B(x) \neq \emptyset\}, \ \ \text{rge}(B) = \bigcup_{x \in H} B(x)$$

and

$$\text{gph}(B) = \{(x,y) : x \in H, y \in B(x)\}.$$

It is called *monotone* provided

$$\langle x^* - y^*, x - y \rangle \geq 0 \ \ \forall \ x, y \in H, x^* \in B(x) \text{ and } y^* \in B(y).$$

In addition, it is called *maximal monotone* if there is no monotone operator B' such that the graph of B is strictly contained in graph of B'.

The *resolvent* and the *reflected resolvent* of B are defined respectively as follows

$$J_B := (Id + B)^{-1} \ \text{ and } \ R_B := 2J_B - Id.$$

Remark 1 The resolvent and the reflected resolvent of a maximal monotone operator are nonexpansive [15].

Let $A, B : H \rightrightarrows H$ be two maximal monotone operators. We want to find x numerically such that

$$0 \in (A + B)(x). \tag{1}$$

Douglas–Rachford splitting has the following form

$$x_0 \in H, \ \ x_{k+1} = [(1 - \alpha)Id + \alpha R_{\gamma A} R_{\gamma B}](x_k), \ k \in \mathbb{N} \tag{2}$$

for some $\alpha \in (0, 1]$ and $\gamma > 0$. It is known that $0 \in (A + B)(x^*)$ if and only if $x^* = J_{\gamma B}(z^*)$ where z^* is a solution of the fixed point problem $z = R_{\gamma A} R_{\gamma B}(z)$ (see, e.g., [2]).

When A is single-valued, the *forward-backward* splitting is given by

$$x_0 \in H, \ \ x_{k+1} = J_{\gamma B} L_{\gamma A}(x_k), \ k \in \mathbb{N} \tag{3}$$

where $L_{\gamma A} := Id - \gamma A$ and Id denotes the identity operator. It comes from the fact that $0 \in (A + B)(x^*) \Leftrightarrow (Id - \gamma A)x^* \in (Id + \gamma B)x^* \Leftrightarrow x^* = J_{\gamma B} L_{\gamma A}(x^*)$.

3 Main results

First let us suppose that the assumptions below are satisfied.

Assumption ($\mathbf{H_A}$) : The operator $A : H \to H$ is monotone and L-Lipschitz continuous.

Assumption ($\mathbf{H_B}$) : The operator $B : H \rightrightarrows H$ is maximal monotone and μ-strongly monotone.

We have the following linear convergence result for the *forward-backward* algorithm.

Theorem 1 *Let Assumptions $(H_A), (H_B)$ hold and let $\gamma^* := \frac{\mu}{L^2}$. Then the forward-backward algorithm converges to the unique solution x^* of (1) with linear rate $r_{FB} := \frac{L}{\sqrt{L^2+\mu^2}}$, i. e.,*

$$\|x_k - x^*\| \le r_{FB}^k \|x_0 - x^*\|, \tag{4}$$

where (x_k) is the sequence generated in (3) with $\gamma = \gamma^*$.

Proof Using Assumption (H_A), for all $x, y \in H$, we have

$$\|L_{\gamma A}x - L_{\gamma A}y\|^2 = \|x - y\|^2 - 2\gamma\langle x - y, Ax - Ay\rangle + \gamma^2\|Ax - Ay\|^2$$
$$\le (1 + \gamma^2 L^2)\|x - y\|^2,$$

which implies that $L_{\gamma A}$ is $\sqrt{1 + \gamma^2 L^2}$-Lipschitz continuous.

On the other hand, Assumption (H_B) deduces that

$$\langle x-y, \gamma Bx - \gamma By\rangle \ge \gamma\mu\|x-y\|^2 \Leftrightarrow \langle x-y, (\gamma B + I)x - (\gamma B + I)y\rangle \ge (1+\gamma\mu)\|x-y\|^2.$$

Let $u := (\gamma B + I)x, v := (\gamma B + I)y$ then $x = J_{\gamma B}u, y = J_{\gamma B}v$ and

$$\langle J_{\gamma B}u - J_{\gamma B}v, u - v\rangle \ge (1 + \gamma\mu)\|J_{\gamma B}u - J_{\gamma B}v\|^2,$$

which infers that $J_{\gamma B}$ is $\frac{1}{1+\gamma\mu}$-Lipschitz continuous. Consequently, $J_{\gamma B}L_{\gamma A}$ is $\frac{\sqrt{1+\gamma^2 L^2}}{1+\gamma\mu}$-Lipschitz continuous. The value $\gamma = \gamma^* = \frac{\mu}{L^2}$ optimizes the Lipschitz modulus of $J_{\gamma B}L_{\gamma A}$ and $J_{\gamma^* B}L_{\gamma^* A}$ is $\frac{L}{\sqrt{L^2+\mu^2}}$-Lipschitz continuous. The conclusion follows. $\square$

Remark 2 i) From the fact that $0 \in Ax + Bx \Leftrightarrow 0 \in \gamma Ax + \gamma Bx$, it is natural to see that the linear convergence rate r_{FB} of the forward-backward algorithm depends on the ratio $\frac{\mu}{L}$ (see, e. g., [13]).

ii) Under the same assumptions, Douglas-Rachford algorithm was used by Moursi–Vandenberghe in [14] to obtain the linear convergence rate

$$r_{DR} := \frac{1}{2(1+\mu)}\left(\sqrt{2\mu^2 + 2\mu + 1 + 2\left(1 - \frac{1}{(1+L)^2} - \frac{1}{1+L^2}\right)\mu(1+\mu)} + 1\right). \tag{5}$$

We can see that the form of r_{DR} is not very nice and difficult to control. In particular, if L is closed to zero and very small to μ then r_{FB} is closed to zero while $r_{DR} > \frac{1}{2(1+\mu)}(1+1) = \frac{1}{1+\mu}$, which means that the forward-backward algorithm is extremely faster than Douglas-Rachford algorithm used in [14].

Example 1 Let us consider in $\mathbb{R}^2$ with

$$A : \mathbb{R}^2 \to \mathbb{R}^2, \; x = (x_1, x_2) \mapsto Ax = \left(\frac{0.01}{2}x_1 + \frac{0.01}{3}\sin x_1, \frac{0.01}{2}x_2 - \frac{0.01}{2}\cos x_2\right),$$

and $B = N_C + Id$ where N_C denotes the normal cone operator to a closed convex set C in $\mathbb{R}^2$. It's easy to check that $J_{\gamma B}(x) = \text{proj}_C(\frac{x}{1+\gamma})$, where proj_C denotes the projection operator onto C. We have $L = 0.01, \mu = 1, r_{FB} < 0.01$ while $r_{DR} > 0.5$. Note that one can also apply the case (ii) established by Giselsson in

[13] (see the Introduction) with $A' = A + Id$ and $B' = N_C$. Then A' is $\mu' := 1$-strongly monotone and $L' := 1.01$-Lipschitz continuous. Using [13, Theorem 6.5], the optimal linear rate of Douglas–Rachford algorithm is $\sqrt{\frac{L'-\mu'}{L'+\mu'}} > 0.07$, which means that our approach is more efficient.

Next we show that r_{FB} is less than r_{DR} in most of important subcases.

Proposition 1 *We have*
$$r_{FB} < r_{DR}, \tag{6}$$

if

  a) $L \geq \mu \geq 1$, or

  b) $L < \mu$ and $\mu \geq 1$, or

  c) $L \leq \mu < 1$.

Proof

In order to compare r_{FB} to r_{DR}, we estimate $1 - r_{FB}$ and $1 - r_{DR}$. One has

$$1 - r_{DR} = \frac{1}{2(1+\mu)} \frac{2(\frac{1}{(1+L)^2} + \frac{1}{1+L^2})\mu(1+\mu)}{2\mu + 1 + \sqrt{2\mu^2 + 2\mu + 1 + 2(1 - \frac{1}{(1+L)^2} - \frac{1}{1+L^2})\mu(1+\mu)}}$$

$$= \frac{(\frac{1}{(1+L)^2} + \frac{1}{1+L^2})\mu}{2\mu + 1 + \sqrt{2\mu^2 + 2\mu + 1 + 2(1 - \frac{1}{(1+L)^2} - \frac{1}{1+L^2})\mu(1+\mu)}},$$

and

$$1 - r_{FB} = \frac{\mu^2}{\sqrt{L^2 + \mu^2}(\sqrt{L^2 + \mu^2} + L)}.$$

a) We have

$$2\mu + 1 + \sqrt{2\mu^2 + 2\mu + 1 + 2(1 - \frac{1}{(1+L)^2} - \frac{1}{1+L^2})\mu(1+\mu)}$$

$$= 2\mu + 1 + \sqrt{4\mu^2 + 4\mu + 1 - 2(\frac{1}{(1+L)^2} + \frac{1}{1+L^2})\mu(1+\mu)}$$

$$\geq 2\mu + 1 + \sqrt{4\mu^2 + 4\mu + 1 - 4\mu} \text{ (since } L \geq \mu \geq 1)$$

$$\geq 2\mu + 1 + 2\mu - 1 = 4\mu.$$

Consequently

$$1 - r_{DR} \leq \frac{\frac{2\mu}{1+L^2}}{4\mu} \leq \frac{1}{2(1+L^2)}. \tag{7}$$

On the other hand

$$1 - r_{FB} = 1 - \frac{L}{\sqrt{L^2 + \mu^2}} \geq 1 - \frac{L}{\sqrt{L^2 + 1}} = \frac{1}{\sqrt{L^2 + 1}(\sqrt{L^2 + 1} + L)}$$

$$> \frac{1}{2(1 + L^2)}. \tag{8}$$

From (7) and (8), one implies that $r_{FB} < r_{DR}$.

b) One has

$$1 - r_{DR} \leq \frac{\frac{2\mu}{1+L^2}}{2\mu + 2} = \frac{\mu}{(\mu + 1)(1 + L^2)}.$$

In addition

$$\frac{\mu}{(\mu + 1)(1 + L^2)} < 1 - r_{FB} \Leftrightarrow \frac{1}{(\mu + 1)(1 + L^2)} < \frac{\mu}{\sqrt{L^2 + \mu^2}(\sqrt{L^2 + \mu^2} + L)}$$

$$\Leftrightarrow L^2 + \mu^2 + L\sqrt{L^2 + \mu^2} < \mu^2 + \mu + \mu^2 L^2 + \mu L^2$$

$$\Leftrightarrow L^2 + L\sqrt{L^2 + \mu^2} < \mu + \mu^2 L^2 + \mu L^2. \tag{9}$$

The last inequality is true since $\mu > 1, L < \mu, L^2(\mu^2 - 1) \geq 0$ and $\mu(L^2 + 1) - L\sqrt{L^2 + \mu^2} \geq 2L\mu - \sqrt{2}L\mu > 0$. Consequently, $1 - r_{FB} > 1 - r_{DR}$.

c) It is sufficient to prove the inequality (9). Indeed, we have

$$(L^2 - \mu^2)(\mu^2 - 1) \geq 0,$$

which implies that

$$L^2\mu^2 + \mu^2 \geq \mu^4 + L^2. \tag{10}$$

Using Cauchy's inequality and the fact that $L \leq \mu$, one obtains

$$\frac{1}{2}\mu + \mu L^2 \geq \sqrt{2}L\mu \geq L\sqrt{L^2 + \mu^2}, \tag{11}$$

and

$$\mu(\mu^3 + \frac{1}{8} + \frac{3}{8}) \geq \frac{3\sqrt[3]{3}}{4}\mu^2 > \mu^2. \tag{12}$$

Adding (10), (11) and (12), the conclusion follows.

Remark 3   We cannot expect that r_{FB} is always less than r_{DR}. Indeed for fixed L, we have

$$\lim_{\mu \to 0} \frac{1 - r_{FB}}{1 - r_{DR}} = \lim_{\mu \to 0} \frac{\mu(2\mu + 1 + \sqrt{2\mu^2 + 2\mu + 1 + 2(1 - \frac{1}{(1+L)^2} - \frac{1}{1+L^2})\mu(1 + \mu)})}{(\frac{1}{(1+L)^2} + \frac{1}{1+L^2})\sqrt{L^2 + \mu^2}(\sqrt{L^2 + \mu^2} + L)}$$

$$= 0.$$

Thus r_{DR} is less than r_{FB} if μ is closed to zero. However in this case r_{FB} is also closed to r_{DR} (and 1) since $\lim_{\mu \to 0} r_{FB} = \lim_{\mu \to 0} r_{DR} = 1$. In addition, the forward-backward algorithm does not have to compute the resolvent of A as Douglas–Rachford algorithm does. It means that the forward-backward algorithm is also a good selection for this case.

Finally let us consider the case when the Lipschitz continuity of A is strengthened with the cocoercivity, i. e., A is $\frac{1}{L}$-cocoercive (which implies that A is L-Lipschitz continuous).

Theorem 2 *Let Assumptions $(H_A), (H_B)$ hold and suppose that A is $\frac{1}{L}$-cocoercive. Let $\gamma^* = \frac{2}{L}$. Then the forward-backward algorithm with $\gamma = \gamma^*$ converges to the unique solution x^* of (1) with linear rate $\tilde{r}_{FB} := \frac{1}{1+2\mu/L}$.*

Proof We have

$$\|L_{\gamma^*A}x - L_{\gamma^*A}y\|^2 = \|x - y\|^2 - 2\gamma^*\langle x - y, Ax - Ay\rangle + \gamma^{*2}\|Ax - Ay\|^2$$
$$\leq \|x - y\|^2 + (-2\gamma^*/L + \gamma^{*2})\|Ax - Ay\|^2$$
$$\leq \|x - y\|^2,$$

which deduces that L_{γ^*A} is non-expansive. On the other hand, we know that J_{γ^*B} is $\frac{1}{1+\gamma^*\mu}$-Lipschitz continuous (see the proof of Theorem 1). Thus the conclusion follows.

Remark 4 Under the same assumption of Theorem 2, Douglas–Rachford algorithm was used by Giselsson in [13], with the linear rate $\tilde{r}_{DR} = \frac{1}{1+\sqrt{\mu/L}}$. It is easy to see that if $4\mu \geq L$ then $\tilde{r}_{FB} < \tilde{r}_{DR}$ and else. Consequently, we totally recommend to use the forward-backward splitting instead of Douglas–Rachford splitting when $\mu \geq L/4$.

4 Conclusions

The paper provides a linear convergence result for the *forward-backward* splitting algorithm applied to finding a zero of the sum of two maximal monotone operator, where one of them is single-valued. We point out the cases when the *forward-backward* algorithm is totally better than Douglas–Rachford algorithm.

References

1. G. B. PASSTY, *Ergodic convergence to a zero of the sum of monotone operators in Hilbert space*, J Math Anal Appl 72(2), 383–390, 1979.
2. P. L. LIONS, B. MERCIER, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J Numer Anal 16, 964–979, 1979.
3. A. RENAUD, *Algorithmes de Régularisation et Décomposition pour les Problèmes Variationnels Monotones*, doctoral thesis, L'école Nationale Supérieure des Mines de Paris, 1993.
4. G. H. G. CHEN, R. T. ROCKAFELLAR, *Convergence Rates in Forward-Backward Splitting*, SIAM J Optim, 7(2), 421–444, 1997.
5. P. TSENG, *A modified forward-backward splitting method for maximal monotone mappings*, SIAM J Control Optim 38(2), 431–446, 2006.
6. H. ATTOUCH, A. CABOT, *Convergence Rates of Inertial Forward-Backward Algorithms*, SIAM J Optim, 28(1), 849–874, 2018.
7. A. GIBALI, D. V. THONG, *Tseng type methods for solving inclusion problems and its applications*, Calcolo 55(4), 55–49, 2018.
8. V. APIDOPOULOS, J. AUJOL, C. DOSSAL, *Convergence rate of inertial Forward-Backward algorithm beyond Nesterov's rule*, Math Program 180, 137–156, 2020.
9. J. ECKSTEIN, *Splitting methods for monotone operators with applications to parallel optimization*, PhD thesis, MIT, 1989.
10. B. F. SVAITER, *On weak convergence of the Douglas-Rachford method*, SIAM J Control Optim 49, 280–287, 2011.
11. B. HE, X. YUAN, *On the o(1/n) convergence rate of the Douglas-Rachford alternating direction method*, SIAM J Numer Anal 50(2), 700–709, 2012.
12. H. H. BAUSCHKE, W. M. MOURSI, *On the Douglas-Rachford algorithm,* Math Program, 164–263, 2017.
13. P. GISELSSON, *Tight global linear convergence rate bounds for Douglas-Rachford splitting*, J Fixed Point Theory Appl 19, 2241–2270, 2017.
14. W. M. MOURSI, L. VANDENBERGHE, *Douglas-Rachford Splitting for the Sum of a Lipschitz Continuous and a Strongly Monotone Operator*, J Optim Theory Appl 183, 179–198, 2019.
15. H. H. BAUSCHKE, P. L. COMBETTES, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer Berlin, 2011.