Math. Program., Ser. A (2015) 153:715–722
DOI 10.1007/s10107-014-0805-x

SHORT COMMUNICATION

On the convergence rate of Douglas–Rachford operator splitting method

Bingsheng He · Xiaoming Yuan

Received: 13 December 2011 / Accepted: 1 August 2014 / Published online: 11 September 2014

Abstract This note provides a simple proof of a worst-case convergence rate measured by the iteration complexity for the Douglas–Rachford operator splitting method for finding a root of the sum of two maximal monotone set-valued operators. The accuracy of an iterate to the solution set is measured by the residual of a characterization of the original problem, which is different from conventional measures such as the distance to the solution set.

Keywords Douglas–Rachford operator splitting method · Convergence rate

Mathematics Subject Classification 90C25 · 65K10 · 65N12

1 Introduction

We consider the problem

$$0 \in A(u) + B(u), \tag{1.1}$$

Bingsheng He was supported by the NSFC Grant 91130007, and the grant of MOE of China 20110091110004. Xiaoming Yuan was partially supported by the General Research Fund from Hong Kong Research Grants Council: 203613.

B. He (✉)
International Centre of Management Science and Engineering, Department of Mathematics, Nanjing University, Nanjing 210093, China
e-mail: hebma@nju.edu.cn

X. Yuan
Department of Mathematics, Hong Kong Baptist University, Hong Kong, China
e-mail: xmyuan@hkbu.edu.hk

where A and B are continuous maximal monotone set-valued operators in $\Re^n$. As in [17], throughout the solution set of (1.1) is assumed to be nonempty. That is, there exists $u \in \Re^n$, $a \in A(u)$ and $b \in B(u)$ such that $a + b = 0$.

A fundamental method for solving (1.1) is the proximal point algorithm (PPA), which dates back to [18,20,22]. The iterative scheme of PPA for (1.1) is

$$0 \in A\left(u^{k+1}\right) + B\left(u^{k+1}\right) + \frac{1}{\lambda}\left(u^{k+1} - u^k\right), \tag{1.2}$$

where $\lambda > 0$ is the proximal parameter. For a monotone set-valued operator (say T) and a positive scalar λ, T's resolvent operator is defined as $J_T^\lambda := (I + \lambda T)^{-1}$. Recall that the resolvent operator of a monotone set-valued operator is single-valued. We refer to, e.g. [1,3,19], for the definition and more properties of resolvent operators. According to (1.2), applying PPA to (1.1) requires computing the resolvent operator of $A + B$ exactly. This task, however, could be as hard as the original problem (1.1). This difficulty thus has inspired many operator splitting methods in the literature, whose common idea is to alleviate the computation of J_{A+B}^λ to individual computation of J_A^λ and J_B^λ, see [9,13,17] to mention a few earlier articles.

This short note only discusses the Douglas–Rachford operator splitting method discussed in [17]. For the special case of (1.1) where both A and B are single-valued linear operators, the following scheme was proposed in [7] for solving heat conduction problems:

$$\begin{cases} \frac{1}{\lambda}\left(u^{k+\frac{1}{2}} - u^k\right) + A\left(u^{k+\frac{1}{2}}\right) + B\left(u^k\right) = 0, \\ \frac{1}{\lambda}\left(u^{k+1} - u^{k+\frac{1}{2}}\right) + B\left(u^{k+1}\right) - B\left(u^k\right) = 0. \end{cases} \tag{1.3}$$

Eliminating $u^{k+\frac{1}{2}}$ in the above scheme yields

$$\left(J_B^\lambda\right)^{-1} u^{k+1} = \left(J_A^\lambda(2J_B^\lambda - I) + (I - J_B^\lambda)\right)\left(J_B^\lambda\right)^{-1}\left(u^k\right). \tag{1.4}$$

Therefore, defining $z^k = (J_B^\lambda)^{-1}u^k$ which equivalently means $u^k = (J_B^\lambda)(z^k)$, the scheme (1.4) can be written as

$$z^{k+1} = J_A^\lambda\left(2J_B^\lambda - I\right)z^k + \left(I - J_B^\lambda\right)z^k. \tag{1.5}$$

In [17], Lions and Mercier extended the scheme (1.5) to the generic case where both A and B are set-valued nonlinear operators in (1.1). As elaborated in [17], when A and B are set-valued nonlinear operators, the scheme (1.5) should be understood in this way: Starting from an arbitrary iterate u^0 in the domain of B, choosing $b^0 \in B(u^0)$ and setting $z^0 = u^0 + \lambda b^0$, then $u^0 = J_B^\lambda(z^0)$ (the existence of the pair (u^0, z^0) is unique by the Representation Lemma, see Corollary 2.3 in [9]). Thus a sequence $\{z^k\}$ is generated by the Douglas–Rachford scheme (1.5); and consequently a sequence $\{u^k := J_B^\lambda(z^k)\}$ converging to a solution point of (1.1) can be generated (see Theorem 3.15 in [8]). We refer to [6] for the precise connection between (1.5) and the original

Douglas–Rachford scheme in [7] for heat conduction problems. It turns out that the scheme (1.5) is the root of a number of celebrated methods such as the alternating direction method of multipliers (ADMM) in [11,12] (see the analysis in [5]) and a primal Douglas–Rachford scheme in [10].

We focus on analyzing the convergence rate of the Douglas–Rachford scheme (1.5). An important reference is [9], where the Douglas–Rachford operator splitting method in [17] is shown to be a special application of the PPA. For some special cases of (1.1) such as the variational inequality problem where the mapping B is the normal cone of a nonempty closed convex set in $\Re^n$, and the convex minimization problem where the mapping B is zero and A is the subdifferential of a convex function, a worst-case convergence rate of the PPA has been analyzed in the literature, see. e.g. [14,21]. In [16], a worst-case convergence rate measured by the iteration complexity was established for ADMM, a special case of the Douglas–Rachford scheme (1.5) in the context of convex optimization with a separable objective function. We also refer to [2,4] for some relevant convergence rate results for projection methods.

The purpose of this note is to provide a novel and very simple proof to derive a worst-case $O(1/\sqrt{k})$ convergence rate for the Douglas–Rachford scheme (1.5). Here, a worst-case $O(1/\sqrt{k})$ convergence rate means the accuracy (measured by a certain criterion) to a solution is of order $O(1/\sqrt{k})$ after k iterations of an iterative scheme; or equivalently, it requires at most $O(1/\epsilon^2)$ iterations to achieve an approximate solution with an accuracy of ϵ. Note that we analyze the generic case of (1.1) without any assumption on the structure of A and B.

2 Preliminaries

First, let us make clear how to measure the accuracy of an iterate of (1.5) to a solution point of (1.1). Let

$$R_T^\lambda := 2J_T^\lambda - I$$

denote the nonexpansive "reflection" operator associated with a monotone set-valued operator T, see e.g. [1,19]. In [9], the scheme (1.5) was explained as an application of PPA to the maximal monotone operator $S_{\lambda,A,B}$ which is defined as

$$S_{\lambda,A,B} := \{(v + \lambda b, u - v) | (u, b) \in B, (v, a) \in A, v + \lambda a = u - \lambda b\}.$$

That is, the scheme (1.5) can be written as

$$z^{k+1} = \left(J_A^\lambda(2J_B^\lambda - I) + (I - J_B^\lambda)\right) z^k = (I + S_{\lambda,A,B})^{-1} z^k = J_{S_{\lambda,A,B}}(z^k).$$

Therefore, we have

$$S_{\lambda,A,B}(z^*) = 0 \Leftrightarrow z^* = J_{S_{\lambda,A,B}}(z^*) \Leftrightarrow z^* = \left(J_A^\lambda(2J_B^\lambda - I) + (I - J_B^\lambda)\right) z^* \Leftrightarrow z^* = R_A^\lambda \circ R_B^\lambda(z^*),$$

where "$\circ$" is the composition of two operators, see details in [1]. Moreover, according to Theorem 5 in [9], for any given root z^* of $S_{\lambda,A,B}$, $J_B^\lambda(z^*)$ is a root of $A+B$. Hence, $J_B^\lambda(z^*)$ is a solution point of (1.1) whenever z^* satisfies

$$z^* = R_A^\lambda \circ R_B^\lambda(z^*). \tag{2.1}$$

Thus, the accuracy of a vector $z \in \Re^n$ to a point satisfying (2.1) can be measured by $\|e(z,\lambda)\|$ where

$$e(z,\lambda) := \frac{1}{2}\big(z - R_A^\lambda \circ R_B^\lambda(z)\big) \tag{2.2}$$

and $\|\cdot\|$ is the Euclidean norm. Note that the measure (2.2) is different from conventional measures such as the distance of an iterate to the solution set. Furthermore, using the notation R_T^λ, the Douglas–Rachford scheme (1.5) can be written as

$$\begin{aligned} z^{k+1} &= J_A^\lambda(2J_B^\lambda - I)z^k + (I - J_B^\lambda)z^k \\ &= z^k + \frac{1}{2}\big(2J_A^\lambda(2J_B^\lambda(z^k) - z^k) - (2J_B^\lambda(z^k) - z^k) - z^k\big) \\ &= z^k + \frac{1}{2}\big(R_A^\lambda \circ R_B^\lambda(z^k) - z^k\big). \end{aligned} \tag{2.3}$$

Hence, using the notation $e(z,\lambda)$ in (2.2), we have

$$z^{k+1} = z^k - e(z^k,\lambda). \tag{2.4}$$

In our discussion, inspired by [23] (see pp. 240) and [15] for the variational inequality problem case of (1.1), our analysis will be conducted for a general version of (2.4):

$$z^{k+1} = z^k - \gamma e(z^k,\lambda), \tag{2.5}$$

with $\gamma \in (0,2)$. Obviously, the original Douglas–Rachford scheme (2.4) is recovered when $\gamma = 1$ in (2.5). In addition, if γ is allowed to vary by iterations, the scheme (2.5) is exactly the generalized Douglas–Rachford scheme discussed in [9]. Since we focus on the generic case of (1.1) where no specific structure of A and B is assumed, we do not discuss how to solve the subproblem (2.5).

Then, we establish some preliminary results which are useful for further analysis. Since the proof of Lemma 2.1 could be found easily in many literatures such as [1,19], we omit it (see Figure 1 in [9] for an illustration).

Lemma 2.1 *Let T be a monotone set-valued operator in $\Re^n$. Then, T's reflection operator R_T^λ is nonexpansive, i.e.,*

$$\|R_T^\lambda(v) - R_T^\lambda(\tilde{v})\| \le \|v - \tilde{v}\|, \quad \forall\, v, \tilde{v} \in \Re^n. \tag{2.6}$$

The following lemma and its corollary are essential tools for establishing the main result later.

Lemma 2.2 *Let $e(z,\lambda)$ be defined in (2.2) with $\lambda > 0$. Then, $e(z,\lambda)$ is firmly nonexpansive. That is,*

$$(z-\tilde{z})^T\big(e(z,\lambda)-e(\tilde{z},\lambda)\big) \geq \|e(z,\lambda)-e(\tilde{z},\lambda)\|^2,\ \forall z,\tilde{z}\in\Re^n. \tag{2.7}$$

Proof The proof can be regarded as a conclusion of Lemma 1 in [9]. It yields from Lemma 2.1 that $R_A^\lambda R_B^\lambda$ is nonexpansive, so is $-R_A^\lambda R_B^\lambda$, and hence $e(\cdot,\lambda)=\frac{1}{2}(I-R_A^\lambda \circ R_B^\lambda)$ must be firmly nonexpansive. The proof is complete. □

Corollary 2.1 *For any $z\in\Re^n$ and z^* satisfying (2.1), we have*

$$(z-z^*)^T e(z,\lambda) \geq \|e(z,\lambda)\|^2. \tag{2.8}$$

Proof By setting $\tilde{z}=z^*$ and using $e(z^*,\lambda)=0$, (2.8) is derived from (2.7) immediately. □

The proof of the next lemma is very simple, and it is relevant to Theorem 3 in [9].

Lemma 2.3 *Let $\{z^k\}$ be the sequence generated by (2.5). For any z^* satisfying (2.1), we have*

$$\|z^{k+1}-z^*\|^2 \leq \|z^k-z^*\|^2-\gamma(2-\gamma)\|e(z^k,\lambda)\|^2. \tag{2.9}$$

Proof Using (2.5), we get

$$\begin{aligned}\|z^{k+1}-z^*\|^2 &= \|(z^k-z^*)-\gamma e(z^k,\lambda)\|^2\\ &= \|z^k-z^*\|^2-2\gamma(z^k-z^*)^T e(z^k,\lambda)+\gamma^2\|e(z^k,\lambda)\|^2.\end{aligned} \tag{2.10}$$

Substituting (2.8) into the right-hand side of (2.10), we obtain (2.9). The proof is complete. □

Below we show that for the sequence $\{z^k\}$ generalized by (2.5), $\{\|e(z^k,\lambda)\|\}$ is monotonically decreasing.

Lemma 2.4 *Let $\{z^k\}$ be the sequence generated by (2.5). Then we have*

$$\|e(z^{k+1},\lambda)\|^2 \leq \|e(z^k,\lambda)\|^2-\frac{2-\gamma}{\gamma}\|e(z^k,\lambda)-e(z^{k+1},\lambda)\|^2. \tag{2.11}$$

Proof Setting $z=z^k$ and $\tilde{z}=z^{k+1}$ in (2.7), we get

$$(z^k-z^{k+1})^T\{e(z^k,\lambda)-e(z^{k+1},\lambda)\} \geq \|e(z^k,\lambda)-e(z^{k+1},\lambda)\|^2. \tag{2.12}$$

Note that [(see (2.5)]

$$z^k-z^{k+1}=\gamma e(z^k,\lambda).$$

Hence, it follows that

$$e(z^k,\lambda)^T\{e(z^k,\lambda)-e(z^{k+1},\lambda)\} \geq \frac{1}{\gamma}\|e(z^k,\lambda)-e(z^{k+1},\lambda)\|^2. \tag{2.13}$$

On the other hand, we have the identity

$$\begin{aligned}\|e(z^{k+1},\lambda)\|^2 = \|e(z^k,\lambda)\|^2 - 2e(z^k,\lambda)^T\{e(z^k,\lambda)-e(z^{k+1},\lambda)\}\\ + \|e(z^k,\lambda)-e(z^{k+1},\lambda)\|^2.\end{aligned}$$

Replacing the second term of the right-hand side of this identity with (2.13), we obtain the assertion (2.11). The proof is complete. □

3 Main result

In this section, we establish a worst-case $O(1/\sqrt{k})$ convergence rate for the generalized Douglas–Rachford scheme (2.5).

Theorem 3.1 *Let* $\{z^k\}$ *be the sequence generated by (2.5). For any integer* $k > 0$ *and* z^* *satisfying (2.1), we have*

$$\|e(z^k,\lambda)\|^2 \leq \frac{1}{\gamma(2-\gamma)(k+1)}\|z^0-z^*\|^2. \tag{3.1}$$

Proof First, it follows from (2.9) that

$$\sum_{i=0}^{\infty}\gamma(2-\gamma)\|e(z^i,\lambda)\|^2 \leq \|z^0-z^*\|^2 \tag{3.2}$$

for any z^* satisfying (2.1). In addition, it follows from (2.11) that the sequence $\{\|e(z^i,\lambda)\|^2\}$ is monotonically non-increasing. Therefore,we have

$$(k+1)\|e(z^k,\lambda)\|^2 \leq \sum_{i=0}^{k}\|e(z^i,\lambda)\|^2. \tag{3.3}$$

The assertion (3.1) follows from (3.2) and (3.3) directly. □

Since we use $\|e(z,\lambda)\|$ to measure the accuracy of $z \in \Re^n$ to a point satisfying (2.1) [see (2.2)], the assertion (3.1) indicates a worst-case $O(1/\sqrt{k})$ convergence rate measured by the iteration complexity of the sequence $\{z^k\}$. Also, recall $u^k := J_B^\lambda(z^k)$. It follows from Theorem 2 in [9] that an operator B is monotone if and only if its resolvent J_B^λ is firmly nonexpansive. Thus, J_B^λ is nonexpansive (see also Lemma 1 in [9]); and we have

$$\|J_B^\lambda(u_1)-J_B^\lambda(u_2)\| \leq \|u_1-u_2\|, \quad \forall\, u_1,u_2 \in \Re^n.$$

Therefore, together with (2.5) and the result in Theorem 3.1, we have

$$\begin{aligned}\|u^{k+1} - u^k\|^2 &= \|J_B^\lambda(z^{k+1}) - J_B^\lambda(z^k)\|^2 \\ &\le \|z^{k+1} - z^k\|^2 \\ &= \gamma^2\|e(z^k, \lambda)\|^2 \\ &\le \frac{\gamma}{(2-\gamma)(k+1)}\|z^0 - z^*\|^2 \end{aligned} \tag{3.4}$$

for any z^* satisfying (2.1). Since we are considering the space $\Re^n$ for (1.1), the sequence $\{u^k\}$ generated by (1.5) converges (in fact, strongly) to a solution point of (1.1), see e.g. Theorem 3.15 in [8]. The convergence of $\{u^k\}$ and the inequality (3.4) thus imply a worst-case $O(1/\sqrt{k})$ convergence rate of $\{u^k\}$ generated by the generalized Douglas–Rachford scheme (2.5).

Acknowledgments The authors are grateful to three anonymous referees for their constructive suggestions which have helped us improve the presentation of this paper substantially. In particular, one referee's valuable comments including the suggestion of using the reflection operator to address the general case where both the operators A and B are set-valued are highly appreciated.

References

1. Aubin, J.P., Frankowska, H.: Set-Valued Analysis. Birkhäuser, Boston (1990)
2. Auslender, A., Teboulle, M.: Interior projection-like methods for monotone variational inequalities. Math. Program. **104**, 39–68 (2005)
3. Bauschke, H.H., Combettes, P.L.: Resolvents of Monotone Operators, in Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, New York (2011)
4. Cai, X.J., Gu, G.Y., He, B.S.: On the O(1/t) convergence rate of the projection and contraction methods for variational inequalities with Lipschitz continuous monotone operators. Comput. Optim. Appl. **57**, 339–363 (2014)
5. Chan, T.F., Glowinski, R.: Finite Element Approximation and Iterative Solution of a Class of Mildly Non-linear Elliptic Equations. Technical report, Stanford University, Stanford (1978)
6. Combettes, P.L.: Solving monotone inclusions via compositions of nonexpansive averaged operators. Optimization **53**, 475–504 (2004)
7. Douglas, J., Rachford, H.H.: On the numerical solution of the heat conduction problem in 2 and 3 space variables. Trans. Am. Math. Soc. **82**, 421–439 (1956)
8. Eckstein, J.: Splitting methods for monotone operators with applications to parallel optimization, Doctoral dissertation, Department of Civil Engineering, Massachusetts Institute of Technology. Available as Report LIDS-TH-1877, Laboratory for Information and Decision Sciences, MIT Cambridge (1989)
9. Eckstein, J., Bertsekas, D.P.: On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. **55**, 293–318 (1992)
10. Fukushima, M.: The primal Douglas–Rachford splitting algorithm for a class of monotone mappings with application to the traffic equilibrium problem. Math. Program. **72**, 1–15 (1996)
11. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite-element approximations. Comput. Math. Appl. **2**, 17–40 (1976)
12. Glowinski, R., Marrocco, A.: Approximation par é'ements finis d'ordre un et résolution par pénalisation-dualité d'une classe de problèmes non linéaires, R.A.I.R.O., R2, pp. 41–76 (1975)
13. Glowinski, R., Le Tallec, P.: Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics. SIAM Studies in Applied Mathematics, Philadelphia (1989)
14. Güler, O.: On the convergence of the proximal point algorithm for convex minimization. SIAM J. Optim. **29**(2), 403–419 (1991)
15. He, B.S., Liao, L.-Z., Wang, S.L.: Self-adaptive operator splitting methods for monotone variational inequalities. Numer. Math. **94**, 715–737 (2003)

16. He, B.S., Yuan, X.M.: On the O(1/n) convergence rate of Douglas–Rachford alternating direction method. SIAM J. Numer. Anal. **50**, 700–709 (2012)
17. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**, 964–979 (1979)
18. Martinet, B.: Regularisation, d'inéquations variationelles par approximations succesives. Rev. Francaise d'Inform. Recherche Oper. **4**, 154–159 (1970)
19. Minty, G.J.: Monotone (nonlinear) operators in Hilbert space. Duke Math. J. **29**(2), 341–346 (1962)
20. Moreau, J.J.: Proximité et dualit 'e dans un espace Hilbertien. Bull. Soc. Math. Fr. **93**, 273–299 (1965)
21. Nemirovski, A.: Prox-method with rate of convergence $O(1/t)$ for variational inequality with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. SIAM J. Optim. **15**, 229–251 (2005)
22. Rockafellar, R.T.: Monotone operators and proximal point algorithm. SIAM J. Control Optim. **14**, 877–898 (1976)
23. Varga, R.S.: Matrix Iterative Analysis. Prentice-Hall, Englewood Cliffs (1966)