On convergence rate of the Douglas-Rachford operator splitting method

Bingsheng He[1] and Xiaoming Yuan[2]

Abstract. This note provides a simple proof on a $O(1/k)$ convergence rate for the Douglas-Rachford operator splitting method where k denotes the iteration counter.

Keywords. Douglas-Rachford operator splitting method, convergence rate.

1 Introduction

Finding a root of a monotone set-valued operator is a fundamental problem, and it serves as the uniformed model of many important problems such as fixed point problems, partial differential equations, convex programming problems and variational inequalities. Let us consider finding a root of the sum of two maximal monotone set-valued operators in an finite-dimensional space

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

where A and B are continuous monotone mappings from $\Re^n$ to itself. Throughout we assume the solution set of (1.1), denoted by S^*, to be nonempty.

A fundamental method for solving (1.1) is the proximal point algorithm (PPA) which was proposed originally in [17]. Applying the general PPA for solving (1.1) amounts to an exact estimate of the resolvent operator of $A+B$. Here, the resolvent of a monotone set-valued mapping (say T) denoted by J_T^λ is defined as $J_T^\lambda := (I + \lambda T)^{-1}$, where λ is a positive scalar, see e.g. [22]. Recall that the resolvent operator of a monotone operator is single-valued, see Corollary 2.2 in [4] or [22]. An exact estimate of the resolvent operator of $A + B$, 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 original problem to estimating the resolvent operators of A and B individually, see [4, 9, 16] to mention a few of earlier articles.

This short note only discusses the Douglas-Rachford operator splitting method in [16], which appears to have the most general convergence properties as mentioned in [4]. As elaborated in [16], starting from an arbitrary iterate u^0 in the domain of B, choosing $b^0 \in B(u^0)$ and setting $w^0 = u^0 + \lambda b^0$ in such a way that $u^0 = J_B^\lambda w^0$ (the existence of such a pair of (u^0, w^0) is unique by the Representation Lemma), the iterative scheme of the Douglas-Rachford operator splitting method in [16] generates the sequence $\{w^k\}$ by induction

$$w^{k+1} = J_A^\lambda (2J_B^\lambda - I)w^k + (I - J_B^\lambda)w^k. \tag{1.2}$$

In the case where both A and B in (1.1) are singled-valued, the scheme (1.2) reduces to the original Douglas-Rachford scheme in [3] for heat conduction problems. It turns out that the scheme (1.2) is

[1]Department of Mathematics, Nanjing University, Nanjing, 210093, China. This author was supported by the NSFC grants 10971095 and 91130007, and the grant of MOE of China 20110091110004. Email: hebma@nju.edu.cn
[2]Corresponding author. Department of Mathematics, Hong Kong Baptist University, Hong Kong, China. This author was supported by a grant from Hong Kong General Research Fund. Email: xmyuan@hkbu.edu.hk

the root of a number of celebrated methods such as the alternating direction method of multipliers (ADMM) in [6, 8] (as analyzed in [2]) and a primal Douglas-Rachford scheme in [5].

We focus on the convergence rate of the Douglas-Rachford scheme in [16]. An important result is [4], where the Douglas-Rachford operator splitting method in [16] is shown to be a special application of the original PPA in [17]. Thus, analysis for the generic PPA in [18] ensures a worst-case $O(1/k)$ convergence rate of the Douglas-Rachford scheme (1.2), but only for some special cases of (1.1) such as the variational inequality problem (VIP) where the mapping B in (1.1) is the normal cone of a nonempty closed convex set in $\Re^n$. We emphasize that a worst-case $O(1/k)$ convergence rate means the accuracy (measured by a certain criterion) to a solution is of the order $O(1/k)$ after k iterations of an iterative scheme, or equivalently, it requires at most $O(1/\epsilon)$ iterations to achieve an approximate solution with ϵ accuracy. We also refer to some recent results on the $O(1/k)$ convergence rate of Douglas-Rachford based methods for some special cases of (1.1), such as the ADMM [13] for convex programming, projection and contraction methods [11] and interior projection methods [1] for VIPs.

The purpose of this note is to provide a novel and very simple proof to derive a worst-case $O(1/k)$ convergence rate for the Douglas-Rachford operator splitting method in [16]. Note we discuss the generic case of (1.1) without any assumption on the structure of A and B. Thus, as mentioned in [18], a $O(1/k)$ rate is an optimal worst-case convergence result we can expect.

2 Preliminaries

We first recall some well known properties and then establish some preliminary results for further analysis. For simplicity of notation and clearer exposition of our idea, we present our analysis under the assumption that one mapping in (1.1), say B, is single-valued, and its extension to the case where B is also set-valued is not hard.

For any $\lambda > 0$, it is clear that u^* is a solution of (1.1) if and only if

$$u^* = J_A^\lambda(u^* - \lambda B(u^*)). \tag{2.1}$$

This fact immediately shows that we can measure the accuracy of a vector $u \in \Re^n$ to a solution of (1.1) by $\|e(u,\lambda)\|^2$ where

$$e(u,\lambda) := u - J_A^\lambda(u - \lambda B(u)). \tag{2.2}$$

By using $w = (I + \lambda B)u$ and $J_B^\lambda(w) = u$, the Douglas-Rachford scheme (1.2) in the form of u is

$$u^{k+1} + \lambda B(u^{k+1}) = J_A^\lambda(u^k - \lambda B(u^k)) + \lambda B(u^k).$$

Moreover, with the notation (2.2), it can be rewritten as

$$u^{k+1} + \lambda B(u^{k+1}) = u^k + \lambda B(u^k) - e(u^k,\lambda). \tag{2.3}$$

Inspired by [24] (Page 240) and [12] for VIP cases of (1.1), our analysis will be conducted for

$$u^{k+1} + \lambda B(u^{k+1}) = u^k + \lambda B(u^k) - \gamma e(u^k,\lambda), \tag{2.4}$$

which is a general version of (2.3) with $\gamma \in (0,2)$. Obviously, the original Douglas-Rachford scheme (2.3) is recovered when $\gamma = 1$ in (2.4), and we refer to [12] for the numerical efficiency contributed by the parameter γ. 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 (2.4). For concrete applications of (1.1) with specific structures of A and B, the subproblem (2.4) could be easy enough to have a closed-form solution; while for generic cases without any structure, Newton-like methods for solving smooth equations in the literature can be applied to solve (2.4).

We now provide some elementary results which will be used later. Proof of Lemma 2.1 could be found easily in many literatures such as [4, 22], and we omit it.

Lemma 2.1 *The resolvent operator of a monotone operator T defined on $\Re^n$ is firmly nonexpansive. That is,*

$$(u-v)^T(J_T^\lambda(u)-J_T^\lambda(v)) \geq \|J_T^\lambda(u)-J_T^\lambda(v)\|^2 \quad \forall\, u,v\in\Re^n. \tag{2.5}$$

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

Lemma 2.2 *For any $u,v\in\Re^n$ and $\lambda>0$, we have*

$$\{(u-v)+\lambda(B(u)-B(v))\}^T\big(e(u,\lambda)-e(v,\lambda)\big) \geq \|e(u,\lambda)-e(v,\lambda)\|^2. \tag{2.6}$$

Proof. Since A is monotone, by setting $u:=u-\lambda B(u)$ and $v:=v-\lambda B(v)$ in (2.5), we have

$$\begin{aligned}&\{(v-\lambda B(v))-(u-\lambda B(u))\}^T\{J_A^\lambda(v-\lambda B(v))-J_A^\lambda(u-\lambda B(u))\}\\ \geq\;& \|J_A^\lambda(v-\lambda B(v))-J_A^\lambda(u-\lambda B(u))\|^2.\end{aligned} \tag{2.7}$$

Adding the following term

$$2(u-v)^T\{J_A^\lambda(v-\lambda B(v))-J_A^\lambda(u-\lambda B(u))\}$$

to the both sides of (2.7) and by a simple manipulation, it follows that

$$\begin{aligned}&\{(u-v)+\lambda(B(u)-B(v))\}^T\{J_A^\lambda(v-\lambda B(v))-J_A^\lambda(u-\lambda B(u))\}\\ \geq\;& \|(u-v)+[J_A^\lambda(v-\lambda B(v))-J_A^\lambda(u-\lambda B(u))]\|^2-\|u-v\|^2.\end{aligned} \tag{2.8}$$

Because B is monotone, we have

$$\{(u-v)+\lambda(B(u)-B(v))\}^T(u-v) \geq \|u-v\|^2. \tag{2.9}$$

Adding (2.8) and (2.9) and using the notation of $e(u,\lambda)$ in (2.2), the assertion (2.6) follows directly and the lemma is proved. □

Corollary 2.1 *For any $u\in\Re^n$, we have*

$$\{(u-u^*)+\lambda(B(u)-B(u^*))\}^T e(u,\lambda) \geq \|e(u,\lambda)\|^2, \quad \forall u^*\in S^*. \tag{2.10}$$

Proof. By setting $v=u^*$ and using $e(u^*,\lambda)=0$, (2.10) is derived from (2.6) immediately. □

3 Main results

We establish a $O(1/k)$ convergence rate for the generalized Douglas-Rachford scheme (2.4) in this section. First of all, we show that both of the sequences $\{\|(u^k-u^*)+\lambda(B(u^k)-B(u^*))\|\}$ and $\{\|e(u^k,\lambda)\|\}$ generated by the scheme (2.4) are monotonically decreasing whenever a solution of (1.1) is not found (i.e, $\|e(u^k,\lambda)\|\neq 0$).

Lemma 3.1 *Let $\{u^k\}$ be the sequence generated by (2.4). Then we have*

$$\begin{aligned} &\|(u^{k+1}-u^*)+\lambda(B(u^{k+1})-B(u^*))\|^2 \\ \le\ & \|(u^k-u^*)+\lambda(B(u^k)-B(u^*))\|^2-\gamma(2-\gamma)\|e(u^k,\lambda)\|^2 \quad \forall\, u^*\in S^*. \end{aligned} \tag{3.1}$$

Proof. Using (2.4) by a simple computation, we get

$$\begin{aligned} &\|(u^{k+1}-u^*)+\lambda(B(u^{k+1})-B(u^*))\|^2 \\ =\ & \|(u^k-u^*)+\lambda(B(u^k)-B(u^*))-\gamma e(u^k,\lambda)\|^2 \\ =\ & \|(u^k-u^*)+\lambda(B(u^k)-B(u^*))\|^2 \\ & -2\gamma\{(u^k-u^*)+\lambda(B(u^k)-B(u^*))\}^Te(u^k,\lambda)+\gamma^2\|e(u^k,\lambda)\|^2. \end{aligned} \tag{3.2}$$

Substituting (2.10) in the right-hand side of (3.2), we obtain (3.1) and the lemma is proved. □

Lemma 3.2 *Let $\{u^k\}$ be the sequence generated by (2.4). Then we have*

$$\|e(u^{k+1},\lambda)\|^2 \le \|e(u^k,\lambda)\|^2 - \frac{2-\gamma}{\gamma}\|e(u^k,\lambda)-e(u^{k+1},\lambda)\|^2. \tag{3.3}$$

Proof. Setting $u=u^k$ and $v=u^{k+1}$ in (2.6), we get

$$\{(u^k-u^{k+1})+\lambda(B(u^k)-B(u^{k+1}))\}^T\{e(u^k,\lambda)-e(u^{k+1},\lambda)\} \ge \|e(u^k,\lambda)-e(u^{k+1},\lambda)\|^2. \tag{3.4}$$

Note that (see (2.4))

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

Hence, it follows from the last equation and (3.4) that

$$e(u^k,\lambda)^T\{e(u^k,\lambda)-e(u^{k+1},\lambda)\} \ge \frac{1}{\gamma}\|e(u^k,\lambda)-e(u^{k+1},\lambda)\|^2. \tag{3.5}$$

Substituting it in the identity

$$\|e(u^{k+1},\lambda)\|^2 = \|e(u^k,\lambda)\|^2 - 2e(u^k,\lambda)^T\{e(u^k,\lambda)-e(u^{k+1},\lambda)\} + \|e(u^k,\lambda)-e(u^{k+1},\lambda)\|^2,$$

we obtain (3.3) and thus the lemma is proved. □

Now, we are ready to show a worst-case $O(1/k)$ convergence rate for the generalized Douglas-Rachford scheme (2.4).

Theorem 3.1 *Let $\{u^k\}$ be the sequence generated by (2.4). For any integer number $k>0$, we have*

$$\|e(u^k,\lambda)\|^2 \le \frac{1}{\gamma(2-\gamma)(k+1)}\|(u^0-u^*)+\lambda(B(u^0)-B(u^*))\|^2, \quad \forall u^*\in S^*. \tag{3.6}$$

Proof. First, it follows from (3.1) that

$$\sum_{i=0}^{\infty}\gamma(2-\gamma)\|e(u^i,\lambda)\|^2 \le \|(u^0-u^*)+\lambda(B(u^0)-B(u^*))\|^2, \quad \forall\, u^*\in S^*. \tag{3.7}$$

In addition, it follows from (3.3) that the sequence $\{\|e(u^i,\lambda)\|^2\}$ is monotonically non-increasing. Therefore, we have

$$(k+1)\|e(u^k,\lambda)\|^2 \le \sum_{i=0}^{k}\|e(u^i,\lambda)\|^2. \tag{3.8}$$

The assertion (3.6) follows from (3.7) and (3.8) directly. □

Thus, a convergence rate $O(1/k)$ of the recursion (2.4) in the worst case is established based on (3.6), where k denotes the iteration counter and the accuracy to a solution of (1.1) is measured by $\|e(u^k,\lambda)\|^2$.

4 Applications to monotone variational inequalities

We take the variational inequality problem as an example to show that the presented approach can recover the $O(1/k)$ convergence rate for some classical methods such as the proximal point algorithm (PPA) in [17] and its relaxed version in [7]. Analysis of the presented approach is much easier than existing approaches in the literature such as [18].

The monotone variational inequality problem (VIP) in $\Re^n$ is to find $u^* \in \Omega$ such that

$$\text{VI}(\Omega, F) \qquad\qquad (u-u^*)^T F(u^*) \geq 0, \quad \forall\, u \in \Omega, \tag{4.1}$$

where Ω is a nonempty closed convex set in $\Re^n$ and F is a monotone mapping from $\Re^n$ to itself. We denote by Ω^* the solution set of (4.1).

The VIP (4.1) can be reformulated as finding u^* such that

$$0 \in N_\Omega(u^*) + F(u^*), \tag{4.2}$$

where $N_\Omega(\cdot)$ is the normal cone operator to Ω, i.e.,

$$N_\Omega(u) := \begin{cases} \{w \mid (v-u)^T w \leq 0, \ \ \forall v \in \Omega\}, & \text{if } u \in \Omega, \\ \emptyset, & \text{otherwise.} \end{cases}$$

Therefore, (4.1) is a special case of (1.1) with $A = N_\Omega$ and $B = F$. Let $P_\Omega(v)$ denote the orthogonal projection of v on Ω, *i.e.*,

$$P_\Omega(v) = \arg\min\{\|u-v\| \mid u \in \Omega\}.$$

Then, by noticing the obvious facts $N_\Omega = \lambda N_\Omega$ for any $\lambda > 0$ and $J_A^\lambda = (I+\lambda N_\Omega)^{-1} = P_\Omega$, the vector $e(u,\lambda)$ defined in (2.2) can be specified into

$$e(u,\lambda) = u - P_\Omega[u - \lambda F(u)] \tag{4.3}$$

for (4.1), and the Douglas-Rachford scheme (2.4) is accordingly simplified to

$$u^{k+1} + \lambda F(u^{k+1}) = u^k + \lambda F(u^k) - \gamma e(u^k, \lambda). \tag{4.4}$$

We refer to [10] for similar discussions for general monotone variational inequalities. Based on this analysis, the conclusion in Theorem 3.1 can be specified as: The sequence $\{u^k\}$ generated by the recursion (4.4) satisfies

$$\|e(u^k,\lambda)\|^2 \leq \frac{1}{\gamma(2-\gamma)(k+1)} \|(u^0-u^*) + \lambda(F(u^0)-F(u^*))\|^2, \ \ \forall u^* \in \Omega^*.$$

Note that the Douglas-Rachford scheme (2.4) converts the VIP (4.1) to a series of systems of nonlinear equations (4.4), which are in general easier than (4.1).

For solving (4.1), a fundamental method is the PPA proposed originally in [17] which turns out to be the root of a number of celebrated methods such as the projected gradient method [20], the augmented Lagrangian method [14, 21], the extragradient method [15], the accelerated projected gradient method [19, 23], and so on. We consider the relaxed PPA in [7] whose iterative scheme for (4.1) can be written as

$$\begin{cases} \tilde{u}^k \in \Omega, \ \ (u-\tilde{u}^k)^T\{\lambda F(\tilde{u}^k) + (\tilde{u}^k - u^k)\} \geq 0, \ \ \forall\, u \in \Omega, \\ u^{k+1} = u^k - \gamma(u^k - \tilde{u}^k), \ \gamma \in (0,2), \end{cases} \tag{4.5}$$

where the parameter λ serves as the proximal parameter. Note that the original PPA in [17] is a special case of (4.5) where $\gamma = 1$. In [18], a $O(1/k)$ convergence rate of the original PPA was shown. Now, we show that the scheme (4.5) is also a special case of the Douglas-Rachford recursion (2.4) and thus our general result in Theorem 3.1 immediately indicates the same $O(1/k)$ convergence rate for the relaxed version of PPA (4.5).

Let $A := N_\Omega + F$ and $B := 0$. For given $\lambda > 0$ and $u \in R^n$, we define $\tilde{u} := J_A^\lambda(u)$ which means

$$u \in \tilde{u} + \lambda F(\tilde{u}) + \lambda N_\Omega(\tilde{u}).$$

Note that J_A^λ defined above is firmly nonexpansive. In fact, since $J_A^\lambda(v) \in \Omega$, it follows from the definition of J_A^λ that

$$(J_A^\lambda(v) - J_A^\lambda(u))^T \{\lambda F(J_A^\lambda(u)) + (J_A^\lambda(u) - u)\} \geq 0.$$

Analogously, it follows that

$$(J_A^\lambda(u) - J_A^\lambda(v))^T \{\lambda F(J_A^\lambda(v)) + (J_A^\lambda(v) - v)\} \geq 0.$$

Adding the above two inequalities, using the monotonicity of F and by a manipulation, we have

$$(u - v)^T (J_A^\lambda(u) - J_A^\lambda(v)) \geq \|J_A^\lambda(u) - J_A^\lambda(v)\|^2, \quad \forall\, u, v \in R^n,$$

which implies the nonexpensiveness of J_A^λ.

In addition, when $B = 0$, it follows from (2.2) that

$$e(u, \lambda) = u - J_A^\lambda(u) = u - \tilde{u}.$$

Thus, the Douglas-Rachford scheme (2.4) is simplified to

$$u^{k+1} = u^k - \gamma(u^k - J_A^\lambda(u^k)),$$

or equivalently,

$$u^{k+1} = u^k - \gamma(u^k - \tilde{u}^k),$$

which is exactly the scheme of relaxed PPA (4.5) (since the iterate $\tilde{u}^k$ generated by (4.5) can be expressed as $\tilde{u}^k := J_A^\lambda(u^k)$). Therefore, Theorem 3.1 implies that the sequence $\{u^k\}$ generated by (4.5) satisfies

$$\|e(u^k, \lambda)\|^2 = \|u^k - \tilde{u}^k\|^2 \leq \frac{1}{\gamma(2-\gamma)(k+1)} \|u^0 - u^*\|^2, \quad \forall u^* \in \Omega^*.$$

A simple proof for the $O(1/k)$ convergence rate of the relaxed PPA in [7] for solving the VIP (4.1) is thus established.

References

[1] A. Auslender and M. Teboulle, *Interior projection-like methods for monotone variational inequalities*, Math. Program., 104 (2005), pp. 39-68.

[2] T. F. Chan and R. Glowinski, *Finite element approximation and iterative solution of a class of mildly non-linear elliptic equations*, technical report, Stanford University, 1978.

[3] J. Douglas and H.H. Rachford, *On the numerical solution of the heat conduction problem in 2 and 3 space variables*, Trans. Amer. Math. Soc., 82 (1956), pp. 421–439.

[4] J. Eckstein and D.P. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Program., 55 (1992), pp. 293-318.

[5] M. Fukushima, *The primal Douglas-Rachford splitting algorithm for a class of monotone mappings with application to the traffic equilibrium problem*, Math. Program., 72 (1996), pp. 1-15.

[6] D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite-element approximations*, Comput. Math. Appli., 2 (1976), pp. 17-40.

[7] E. G. Gol'shtein and N. V. Tret'yakov, *Modified Lagrangian in convex programming and their generalizations*, Math. Program. Studies, 10 (1979), pp. 86-97.

[8] R. Glowinski and A. Marrocco, *Approximation par éléments finis d'ordre un et résolution par pénalisation-dualité d'une classe de problèmes non linéaires*, R.A.I.R.O., R2 (1975), pp. 41-76.

[9] R. Glowinski and P. Le Tallec, Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics, SIAM Studies in Applied Mathematics, Philadelphia, PA, 1989.

[10] B. S. He, *Inexact implicit methods for monotone general variational inequalities*, Math. Program., 86 (1999), pp. 199-217.

[11] B. S. He, *On the O(1/t) convergence rate of the projection and contraction methods for variational inequalities with Lipschitz continuous monotone operators*, manuscript, 2011, available at http://www.optimization-online.org/DB_HTML/2011/07/3097.html.

[12] B. S. He, L-Z Liao and S. L. Wang, *Self-adaptive operator splitting methods for monotone variational inequalities*, Numer. Math., 94 (2003), pp. 715-737.

[13] B. S. He and X. M. Yuan, *On the $O(1/t)$ convergence rate of Douglas-Rachford alternating direction method*, SIAM. J. Numer. Anal., under minor revision.

[14] M. R. Hestenes, *Multiplier and gradient methods*, J. Optim. Theory Appli., 4 (1969), pp. 303-320.

[15] G.M. Korpelevich, *The extragradient method for finding saddle points and other problems*, Ekonomika i Matematchskie Metody, 12 (1976), pp. 747-756.

[16] P. L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal., 16 (1979), pp. 964-979.

[17] B. Martinet, *Regularisation, d'inéquations variationelles par approximations succesives*, Rev. Francaise d'Inform. Recherche Oper., 4 (1970), pp. 154-159.

[18] A. Nemirovski, *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 (2005), pp. 229-251.

[19] Y. E. Nesterov, *A method for solving the convex programming problem with convergence rate $O(1/k^2)$*, Dokl. Akad. Nauk SSSR, 269 (1983), pp. 543-547.

[20] B. T. Polyak, *Introduction to Optimization, Translations Series in Mathematics and Engineering, Optimization Software*, Publications Division, New York, 1987.

[21] M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, In Optimization edited by R. Fletcher, pp. 283-298, Academic Press, New York, 1969.

[22] R. T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, NJ, 1970.

[23] P. Tseng, *On accelerated proximal gradient methods for convex-concave optimization*, manuscript, 2008.

[24] R. S. Varga, Matrix Iterative Analysis. Prentice-Hall, Englewood Cliffs, New Jersey, 1966.