On Douglas–Rachford operators that fail to be proximal mappings

Heinz H. Bauschke[*], Jason Schaad[†], and Xianfu Wang[‡]

Dedicated to Terry Rockafellar on the occasion of his 80th birthday

August 3, 2016

Abstract

The problem of finding a zero of the sum of two maximally monotone operators is of central importance in optimization. One successful method to find such a zero is the Douglas–Rachford algorithm which iterates a firmly nonexpansive operator constructed from the resolvents of the given monotone operators.

In the context of finding minimizers of convex functions, the resolvents are actually proximal mappings. Interestingly, as pointed out by Eckstein in 1989, the Douglas–Rachford operator itself may fail to be a proximal mapping. We consider the class of symmetric linear relations that are maximally monotone and prove the striking result that the Douglas–Rachford operator is generically not a proximal mapping.

2010 Mathematics Subject Classification: Primary 47H09, Secondary 47H05, 90C25.

Keywords: Douglas–Rachford algorithm, firmly nonexpansive mapping, maximally monotone operator, nowhere dense set, proximal mapping, resolvent.

[*]Mathematics, University of British Columbia, Kelowna, B.C. V1V 1V7, Canada. E-mail: heinz.bauschke@ubc.ca.

[†]Department of Mathematics and Statistics, Okanagan College, 1000 K.L.O. Road, Kelowna, B.C. V1Y 4X8, Canada. E-mail: JSchaad@okanagan.bc.ca.

[‡]Mathematics, University of British Columbia, Kelowna, B.C. V1V 1V7, Canada. E-mail: shawn.wang@ubc.ca.

1 Introduction

Throughout this paper, we work in the standard Euclidean space

$$X = \mathbb{R}^n, \tag{1}$$

equipped with the standard inner product $\langle \cdot, \cdot \rangle$ and induced Euclidean norm $\|\cdot\|$. Recall that a set-valued operator

$$A\colon X \rightrightarrows X \tag{2}$$

is *monotone* if $\langle x - y, x^* - y^* \rangle \geq 0$ whenever (x, x^*) and (y, y^*) belong to gra A, the graph of A; A is *maximally monotone* if any proper enlargement of A fails to be monotone. Maximally monotone operators are of importance in modern optimization (see [1], [2], [5], [6], [7], [8], [9], [12], [23]) as they cover subdifferential operators of functions that are convex lower semicontinuous and proper as well as matrices whose symmetric part is positive semidefinite. A central problem is to

$$\text{find } x \in X \text{ such that } 0 \in Ax + Bx, \tag{3}$$

where A and B are maximally monotone on X. For instance, if $A = \partial f$ and $B = \partial g$, where f and g belong to $\Gamma_0(X)$, the set of functions that are convex, lower semicontinuous and proper on X, then the sum problem (3) is tied to the problem of finding a minimizer of $f + g$. A popular iterative method, dating back to Lions and Mercier's seminal work [15], to solve (3) is the *Douglas–Rachford algorithm* whose governing sequence $(x_n)_{n\in\mathbb{N}}$ is given by

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = T_{A,B} x_n, \tag{4}$$

where $T_{A,B} = \operatorname{Id} - J_A + J_B R_A = (\operatorname{Id} + R_B R_A)/2$ is the Douglas–Rachford splitting operator, $J_A = (\operatorname{Id} + A)^{-1}$ is the *resolvent* of A and and $R_A = 2J_A - \operatorname{Id}$ is the *reflected resolvent*. If Z, the set of solutions to (3), is nonempty, then $(x_n)_{n\in\mathbb{N}}$ converges to a fixed point of $T_{A,B}$ and $(J_A x_n)_{n\in\mathbb{N}}$ converges to a point in Z. In fact, as pointed in [15], one has $T_{A,B} = J_C$ for some maximally monotone operator C depending on (A, B). That is, (4) is actually the iteration of a resolvent — the resulting method was carefully studied by Rockafellar [20]. If the operator C is actually a subdifferential operator, i.e., $C = \partial h$, where $h \in \Gamma_0(X)$; or equivalently if J_C is a *proximal map* (also known as proximity operator) [17], then stronger statements are available concerning the resolvent iteration [14]. This prompts interest in the question whether $C = \partial h$. Unfortunately, in general, $T_{A,B} = J_C$ is only a resolvent, not a proximal map as demonstrated by Eckstein [11]; the following simpler example is from Schaad's thesis [22]. Suppose that $X = \mathbb{R}^2$, and that A and B are the normal cone operators of the subspaces $\mathbb{R}(1, 0)$ and $\mathbb{R}(1, 1)$. Then the associated maximally monotone operator is given by the matrix

$$C = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \tag{5}$$

which is *not* symmetric and hence C is not a subdifferential operator. The corresponding Douglas–Rachford operator

$$J_C = T_{A,B} = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \tag{6}$$

which is also *not* symmetric, is therefore only a resolvent but *not* a proximal mapping. This is surprising because (3) corresponds in this case to the convex feasibility problem asking to find a point in $\mathbb{R}(1,0) \cap \mathbb{R}(1,1)$ (which is $\{(0,0)\}$).

In this paper, we show that, in the context of linear relations, it is generically *the case that the Douglas–Rachford operator is only a resolvent and* not *a proximal mapping.*

The rest of the paper is organized as follows. In Section 2, we develop auxiliary results on matrices, proximal mappings and convergence. Section 3 contains our main result.

Finally, notation and notions not explicitly defined may be found in, e.g., [2], [16], [19], or [21].

2 Auxiliary results

2.1 Matrices

Unless stated otherwise, we view $\mathbb{R}^{n\times n}$, the set of real $n \times n$ matrices, as a Banach space, with norm $\|R\| := \sup_{\|x\|\leq 1} \|Rx\|$, which is the square root of the largest eigenvalue of $R^{\mathsf{T}}R$. We denote by $\mathbb{S}^n$ the subspace of symmetric $n \times n$ matrices. A matrix R is *nonexpansive* if $\|R\| \leq 1$, i.e., R belongs to the unit ball of $\mathbb{R}^{n\times n}$. This set is convex, closed, and has 0 in its interior. The set of nonexpansive symmetric matrices is likewise in $\mathbb{S}^n$.

Lemma 2.1. *Let R_0, S_0, R_1, S_1 be matrices in $\mathbb{R}^{n\times n}$. Suppose that R_0 commutes with S_0, but that R_1 does not commute with S_1. For each $\lambda \in \,]0,1[$, set $R_\lambda = (1-\lambda)R_0 + \lambda R_1$ and set $S_\lambda = (1-\lambda)S_0 + \lambda S_1$. Then $\{\lambda \in \,]0,1[\mid R_\lambda \text{ commutes with } S_\lambda\}$ is either empty or a singleton.*

Proof. For $\lambda \in [0,1]$, consider the matrix

$$M_\lambda = R_\lambda S_\lambda - S_\lambda R_\lambda. \tag{7}$$

By hypothesis, $M_0 = 0$ but $M_1 \neq 0$. Since $M_1 \neq 0$, there exist $(i,j) \in \{1,\ldots,n\}^2$ such that the (i,j) entry of M_1 is not 0. Denote by $q(\lambda)$ the (i,j) entry of M_λ. Then $q(\lambda)$ is a polynomial in λ of degree at most 2, with $q(0) = 0$ and $q(1) \neq 0$. On $]0,1[$, q has at most one root. Therefore, with the possible exception of one value $\lambda \in \,]0,1[$, $M_\lambda \neq 0$. ■

Example 2.2. *Suppose that $n \geq 2$ and define matrices in $\mathbb{S}^n$ by*

$$R_0 = R_1 = \left[\begin{array}{cc|c} 1 & 0 & \multirow{2}{*}{0} \\ 0 & -1 & \\ \hline \multicolumn{2}{c|}{0} & 0 \end{array}\right], \quad S_0 = \left[\begin{array}{cc|c} -1 & 0 & \multirow{2}{*}{0} \\ 0 & 1 & \\ \hline \multicolumn{2}{c|}{0} & 0 \end{array}\right], \quad S_1 = \left[\begin{array}{cc|c} 0 & 1 & \multirow{2}{*}{0} \\ 1 & 0 & \\ \hline \multicolumn{2}{c|}{0} & 0 \end{array}\right]. \tag{8}$$

For each $\lambda \in \,]0,1[$, set $R_\lambda = (1-\lambda)R_0 + \lambda R_1$ and set $S_\lambda = (1-\lambda)S_0 + \lambda S_1$. Let $\lambda \in [0,1]$. Then

$$\|R_\lambda\| = 1, \ \|S_\lambda\| = \sqrt{(1-\lambda)^2 + \lambda^2} \in \left[1/\sqrt{2}, 1\right] \tag{9}$$

and

$$R_\lambda S_\lambda = \left[\begin{array}{cc|c} \lambda-1 & \lambda & \multirow{2}{*}{0} \\ -\lambda & \lambda-1 & \\ \hline \multicolumn{2}{c|}{0} & 0 \end{array}\right], \ S_\lambda R_\lambda = \left[\begin{array}{cc|c} \lambda-1 & -\lambda & \multirow{2}{*}{0} \\ \lambda & \lambda-1 & \\ \hline \multicolumn{2}{c|}{0} & 0 \end{array}\right]. \tag{10}$$

Consequently, R_λ commutes with S_λ if and only if $\lambda = 0$ and therefore the set $\{\lambda \in \,]0,1[\mid R_\lambda \text{ commutes with } S_\lambda\}$ appearing in the conclusion of Lemma 2.1 is empty.

We thank a referee for suggesting the following variant of Example 2.2.

Example 2.3. *Let R_1, S_0, S_1 be as in Example 2.2 but set $R_0 = S_0$. Because $R_{1/2} = 0$, then $R_{1/2}$ commutes with $S_{1/2}$. Therefore, the set $\{\lambda \in \,]0,1[\mid R_\lambda \text{ commutes with } S_\lambda\}$ appearing in the conclusion of Lemma 2.1 is the* singleton $\{1/2\}$.

2.2 Proximal mappings

We now characterize proximal mappings within the set of resolvents.

Lemma 2.4. *[22, Lemma 4.36] Let $T \in \mathbb{R}^{n\times n}$ be a proximal mapping. Then $T = T^{\mathsf{T}}$.*

Proof. Set $q\colon x \mapsto \frac{1}{2}\|x\|^2$. Then $\mathrm{Id} = \nabla q$. By hypothesis, T is a proximal mapping, so there exists a convex f such that

$$T = (\mathrm{Id} + \partial f)^{-1} = (\partial(q+f))^{-1} = \partial(q+f)^* = \nabla(q+f)^*. \tag{11}$$

Since T is linear, it follows that $T = \nabla T = \nabla^2(q+f)^*$ is symmetric. ■

It turns out that the converse of the previous result also holds.

Lemma 2.5. *Let $T \in \mathbb{R}^{n\times n}$ be firmly nonexpansive*[1] *and such that $T = T^{\mathsf{T}}$. Then T is a proximal mapping.*

Proof. Set $f \colon X \to \mathbb{R} \colon x \mapsto \frac{1}{2}\langle x, Tx\rangle$. Since T is symmetric, we have $\nabla f = T$. Since T is firmly nonexpansive, it is maximally monotone and thus f is convex. By the (extended form of the) Baillon–Haddad theorem (see [2, Theorem 18.15]), $\nabla f = T$ is a proximal map. ■

We thus obtain the following useful characterization of proximal mappings.

Corollary 2.6. *Let $T \in \mathbb{R}^{n\times n}$. Then T is a proximal mapping if and only if T is both firmly nonexpansive and symmetric.*

Remark 2.7. *Corollary 2.6 is closely related to a characterization due to Moreau: T is a proximal mapping if and only if T is symmetric, nonexpansive, and monotone (see [17, last paragraph in Section 3] and also [10, Example 2.13]).*

2.3 Convergence

From now on, we denote the set of maximally monotone operators on X by $\mathcal{M}$, the subset of maximally monotone linear relations[2] by $\mathcal{L}$, and the subdifferential operators of functions in $\Gamma_0(X)$ by $\mathcal{S}$.

Let $(A_k)_{k\in\mathbb{N}}$ be a sequence in $\mathcal{M}$ and let $A \in \mathcal{M}$. Then $(A_k)_{k\in\mathbb{N}}$ converges to A *graphically*, in symbols $A_k \overset{\mathsf{g}}{\to} A$ if and only if the resolvents converge pointwise, in symbols, $J_{A_k} \overset{\mathsf{p}}{\to} J_A$. This induces a metric topology on $\mathcal{M}$ (see [21] for details). Note that $\mathcal{L}$ is a closed topological subspace of $\mathcal{M}$ and that pointwise convergence by resolvents can in that setting be replaced by convergence in operator norm since X is finite-dimensional.

The following result is now easily verified.

Proposition 2.8. *Let $(A_k)_{k\in\mathbb{N}}$ be a sequence in $\mathcal{L}$, and let $A \in \mathcal{L}$. Then we have the equivalences*

$$A_k \overset{\mathsf{g}}{\to} A \Leftrightarrow J_{A_k} \overset{\mathsf{p}}{\to} J_A \Leftrightarrow R_{A_k} \overset{\mathsf{p}}{\to} R_A \Leftrightarrow J_{A_k} \to J_A \Leftrightarrow R_{A_k} \to R_A, \tag{12}$$

where the last two convergences are with respect to the operator norm.

[1] For further information on firmly nonexpansive mappings, see [2, Section 5.1] and [13, Section 1.11].

[2] A linear relation on X is set-valued map from X to X such that its graph is a linear subspace of $X \times X$. In relationship to the present paper, we refer the reader to [4] for more on maximally monotone linear relations. Furthermore, a resolvent J_A is linear if and only if $A \in \mathcal{L}$ by [3, Theorem 2.1(xviii)].

We thus are able to define a metric on $\mathcal{L}$ by

$$(A_1, A_2) \mapsto \|J_{A_1} - J_{A_2}\| \tag{13}$$

and a metric on $\mathcal{L} \times \mathcal{L}$ by

$$\big((A_1, B_1), (A_2, B_2)\big) \mapsto \|J_{A_1} - J_{A_2}\| + \|J_{B_1} - J_{B_2}\|. \tag{14}$$

Note that in view of the pointwise characterization of Proposition 2.8, both $\mathcal{L}$ and $\mathcal{L} \times \mathcal{L}$ are *complete* and so are $\mathcal{L} \cap \mathcal{S}$ and $(\mathcal{L} \cap \mathcal{S}) \times (\mathcal{L} \cap \mathcal{S})$.

These topological notions are used in the next section which contains our main result.

3 Main result

Recall that for A and B in $\mathcal{M}$, the Douglas–Rachford operator is defined by

$$T(A, B) = T_{(A,B)} = \tfrac{1}{2}\big(\operatorname{Id} + R_B R_A\big). \tag{15}$$

Note that $T_{(A,B)}$ is firmly nonexpansive and the resolvent of some maximally monotone operator $M(A, B) \in \mathcal{M}$ but it may be the case that $M(A, B) \notin \mathcal{S}$ even when A and B belong to $\mathcal{S}$ (see Section 1).

We are ready for our main result.

Theorem 3.1. *Suppose that $n \geq 2$. Then generically, the Douglas–Rachford operators for symmetric linear relations are not proximal mappings; in fact, the set*

$$D := \big\{(A, B) \in (\mathcal{L} \cap \mathcal{S})^2 \;\big|\; T_{(A,B)} \text{ is a proximal map}\big\} \tag{16}$$

is a closed subset of $(\mathcal{L} \cap \mathcal{S})^2$ that is nowhere dense.

Proof. We start by verifying that D is closed. To this end, let $(A_k, B_k)_{k\in\mathbb{N}}$ be a sequence in D converging to $(A, B) \in (\mathcal{L} \cap \mathcal{S})^2$. By definition, $T(A_k, B_k)$ is a proximal mapping for every $k \in \mathbb{N}$. By Corollary 2.6, $(\forall k \in \mathbb{N})$ $T(A_k, B_k)^\mathsf{T} = T(A_k, B_k)$. Hence $(\forall k \in \mathbb{N})$ $R_{A_k} R_{B_k} = R_{B_k} R_{A_k}$. In view of Proposition 2.8, we take the limit and obtain $R_A R_B = R_B R_A$. Thus $T(A, B) = T(A, B)^\mathsf{T}$. Using Corollary 2.6 again, we deduce that $T(A, B)$ is a proximal map.

We now show that D is nowhere dense. Let (A_0, B_0) be in D. Then $R_{A_0} R_{B_0} = R_{B_0} R_{A_0}$. Next, set $A_1 = ((R_1 + \operatorname{Id})/2)^{-1} - \operatorname{Id}$ and $B_1 = ((S_1 + \operatorname{Id})/2)^{-1} - \operatorname{Id}$, where R_1 and S_1 are as in Example 2.2. Then $R_{A_1} = R_1$ and $R_{B_1} = S_1$ do not commute and hence $(A_1, B_1) \notin D$. Now set

$$(\forall \lambda \in\,]0,1[\,) \quad A_\lambda = \Big(\tfrac{1}{2}\operatorname{Id} + \tfrac{1}{2}\big((1-\lambda)R_{A_0} + \lambda R_{A_1}\big)\Big)^{-1} - \operatorname{Id} \tag{17a}$$

and

$$(\forall \lambda \in\,]0,1[\,) \quad B_\lambda = \Big(\tfrac{1}{2}\operatorname{Id} + \tfrac{1}{2}\big((1-\lambda)R_{B_0} + \lambda R_{B_1}\big)\Big)^{-1} - \operatorname{Id}. \tag{17b}$$

Then, as $\lambda \to 0^+$,

$$R_{A_\lambda} = (1-\lambda)R_{A_0} + \lambda R_{A_1} \to R_{A_0} \text{ and } R_{B_\lambda} = (1-\lambda)R_{B_0} + \lambda R_{B_1} \to R_{B_0}. \tag{18}$$

It follows from Lemma 2.1, Proposition 2.8, and Corollary 2.6 that there exists $\mu \in\,]0,1]$ such that $(\forall \lambda \in\,]0,\mu])$ $(A_\lambda, B_\lambda) \notin D$. Therefore, (A_0, B_0) does not belong to the interior of D. ■

Remark 3.2. *The assumption that $n \geq 2$ in Theorem 3.1 is important: indeed, when $n = 1$, it is well known that every maximally monotone operator is actually a subdifferential operator (see, e.g., [2, Corollary 22.19]) and therefore every Douglas–Rachford operator is a proximal mapping in this case.*

Remark 3.3 (open problems)**.** *The following questions appear to be of interest:*

(i) *Does Theorem 3.1 admit an extension from symmetric linear relations to general subdifferential operators?*

(ii) *The set D in Theorem 3.1 is closed and nowhere dense. Is it also a porous[3] set?*

Acknowledgements

The authors thank Patrick Combettes and two anonymous referees for their comments. HHB was partially supported by the Natural Sciences and Engineering Research Council of Canada and by the Canada Research Chair Program. XW was partially supported by the Natural Sciences and Engineering Research Council of Canada.

References

[1] A. Auslender and M. Teboulle, *Asymptotic Cones and Functions in Optimization and Variational Inequalities*, Springer, 2003.

[3]See [18] for further information on porous sets.

[2] H.H. Bauschke and P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, 2011.

[3] H.H. Bauschke, S.M. Moffat, and X. Wang, Firmly nonexpansive mappings and maximally monotone operators: correspondence and duality, *Set-Valued and Variational Analysis* 20 (2012), 131–153.

[4] H.H. Bauschke, X. Wang, and L. Yao, On Borwein-Wiersma decompositions of monotone linear relations, *SIAM Journal on Optimization* 20 (2010), 2636–2652.

[5] J.M. Borwein, Fifty years of maximal monotonicity, *Optimization Letters* 4 (2010), 473–490.

[6] J.M. Borwein and J.D. Vanderwerff, *Convex Functions: Constructions, Characterizations and Counterexamples*, Cambridge University Press, 2010.

[7] J.M. Borwein and Q.J. Zhu, *Techniques of Variational Analysis*, Springer, 2005.

[8] R.I. Boţ, *Conjugate Duality in Convex Optimization*, Springer, 2010.

[9] R.S. Burachik and A.N. Iusem, *Set-Valued Mappings and Enlargements of Monotone Operators*, Springer, 2008.

[10] P.L. Combettes and V.R. Wajs, Signal recovery by proximal forward-backward splitting, *Multiscale Modeling and Simulation* 4 (2005), 1168–1200.

[11] J. Eckstein, *Splitting Methods for Monotone Operators with Applications to Parallel Optimization*, PhD thesis, Massachusetts Institute of Technology, 1989.

[12] J. Eckstein and D.P. Bertsekas, On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators, *Mathematical Programming* 55 (1992), 293–318.

[13] K. Goebel and S. Reich, *Uniform Convexity, Hyperbolic Geometry, and Nonexpansive Mappings*, Marcel Dekker, 1984.

[14] O. Güler, On the convergence of the proximal point algorithm for convex minimization, *SIAM Journal on Control and Optimization* 29 (1991), 403–419.

[15] P.-L. Lions and B. Mercier, Splitting algorithms for the sum of two nonlinear operators, *SIAM Journal on Numerical Analysis* 16 (1979), 9643–979.

[16] B.S. Mordukhovich and N.M. Nam, *An Easy Path to Convex Analysis and Applications*, Morgan & Claypool Publishers, 2014.

[17] J.-J. Moreau, Proximité et dualité dans un espace hilbertien, *Bulletin de la Société Mathématique de France* 93 (1965), 273–299.

[18] S. Reich and A.J. Zaslavski, *Genericity in Nonlinear Analysis*, Springer, 2014.

[19] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.

[20] R.T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM Journal on Control and Optimization* 14 (1976), 877–898.

[21] R.T. Rockafellar and R.J-B. Wets, *Variational Analysis*, Springer, corrected 3rd printing, 2009.

[22] J. Schaad, *Modeling the 8-Queens Problem and Sudoku using an Algorithm based on Projections onto Nonconvex Sets*, Master's thesis, The University of British Columbia, 2010.

[23] S. Simons, *From Hahn-Banach to Monotonicity*, Springer, 2008.