J. Aust. Math. Soc. **110** (2021), 333–370
doi:10.1017/S1446788719000570

SURVEY: SIXTY YEARS OF DOUGLAS–RACHFORD

SCOTT B. LINDSTROM and BRAILEY SIMS

(Received 19 September 2018; accepted 30 October 2019; first published online 20 February 2020)

Communicated by G. Willis

This work is dedicated to the memory of Jonathan M. Borwein our greatly missed friend, mentor and colleague. His influence on both the topic at hand, as well as his impact on the present authors personally, cannot be overstated.

Abstract

The Douglas–Rachford method is a splitting method frequently employed for finding zeros of sums of maximally monotone operators. When the operators in question are normal cone operators, the iterated process may be used to solve *feasibility* problems of the following form: Find $x \in \bigcap_{k=1}^{N} S_k$. The success of the method in the context of closed, convex, nonempty sets $S_1, \ldots, S_N$ is well known and understood from a theoretical standpoint. However, its performance in the nonconvex context is less well understood, yet it is surprisingly impressive. This was particularly compelling to Jonathan M. Borwein who, intrigued by Elser, Rankenburg and Thibault's success in applying the method to solving sudoku puzzles, began an investigation of his own. We survey the current body of literature on the subject, and we summarize its history. We especially commemorate Professor Borwein's celebrated contributions to the area.

2010 *Mathematics subject classification*: primary 46N10, 49J53, 49-02; secondary 49-03, 49Mxx.

Keywords and phrases: Douglas–Rachford, feasibility, projection algorithms, iterative methods, discrete dynamical systems.

1. Introduction

In 1996 Heinz Bauschke and Jonathan Borwein broadly classified the commonly applied projection algorithms for solving convex feasibility problems as falling into four categories. These were: best approximation theory, discrete models for image reconstruction, continuous models for image reconstruction and subgradient algorithms [17]. One such celebrated iterative process has been known by many names in many contexts and is possibly best known as the *Douglas–Rachford* method (DR).

DR is frequently used for the more general problem of finding a zero of the sum of maximally monotone operators, which itself is a generalization of the problem of minimizing a sum of convex functions. Many volumes could be written on monotone operator theory, convex optimization and splitting algorithms specifically, the definitive work being that of Bauschke and Combettes [20]; the story of DR is inextricably entwined with each of these.

More recently, the method has become famous for its surprising success in solving nonconvex feasibility problems, notwithstanding the lack of theoretical justification. The recent investigation of these methods in the nonconvex setting has been both motivated by and advanced through experimental application of the algorithms to nonconvex problems in a variety of different settings. In many cases, impressive performance has been observed despite having previously been thought of as ill-adapted to projection algorithms.

The task of choosing what to include in a condensed survey of DR is thus necessarily difficult. We therefore choose to adopt an approach which balances reasonable brevity with the goal that a reader unfamiliar with DR should be able to at least glean the following: the basic history of the method, an understanding of the various motivating contexts in which it has been 'discovered', an appreciation for the diversity of problems to which it is applied, and a sense of which research topics are currently being explored.

1.1. Outline. This paper is divided into four sections.

Section 1: In 1.2, we provide preliminaries on DR and feasibility. In 1.3, we briefly motivate its history and explain how feasibility problems are a special case of finding a zero for a sum of maximal monotone operators, and in 1.4 we explore its use for finding zeros of maximal monotone operator sums, including its connection with the *alternating direction method of multipliers (ADMM)* in 1.4.1. In 1.5, we analyze the ways in which it has been extended from two-set feasibility problems to N-set feasibility problems.

Section 2: We consider the role of DR in solving convex feasibility problems. In 2.1, we catalogue some of the convergence results, and in 2.2 we mention some of its better known applications.

Section 3: We consider the context of nonconvex feasibility. We first consider discrete problems in 3.1 and go on to discuss hypersurface problems in 3.2. In 3.3, we explore some of the possibly nonconvex convergence results which employ notions of regularity and transversality. In 3.3.3, we describe some of the recent work applying DR to nonconvex minimization problems.

Section 4: Finally, we mention two open problems and summarize the current state of research in the field.

Appendix A: This appendix provides a more detailed summary of Gabay's exposition on the connection between DR and ADMM.

1.2. Preliminaries. The method of alternating projections (AP) and the DR are frequently used to find a *feasible point* (point in the intersection) of two closed constraint sets A and B in a Hilbert space H. The *feasibility problem* is the following.

$$\text{Find} \quad x \in A \cap B. \tag{1-1}$$

The projection onto a subset C of H is defined for all $x \in H$ by

$$\mathbb{P}_C(x) := \left\{ z \in C : \|x - z\| = \inf_{z' \in C} \|x - z'\| \right\}.$$

Note that $\mathbb{P}_C$ is, generically, a set-valued map where values may be empty or contain more than one point. In the cases of interest to us, $\mathbb{P}_C$ has nonempty values (indeed, throughout, $\mathbb{P}_C$ is nonempty and so C is said to be *proximal*) and, in order to simplify both notation and implementation, we will work with a selector for $\mathbb{P}_C$, that is, a map $P_C : H \to C : x \mapsto P_C(x) \in \mathbb{P}_C(x)$, so $P_C^2 = P_C$.

When C is nonempty, closed and convex, the projection operator P_C is uniquely determined by the variational inequality

$$\langle x - P_C(x), c - P_C(x) \rangle \le 0 \quad \text{for all } c \in C,$$

and it is a *firmly nonexpansive* mapping; that is, for all $x, y \in H$,

$$\|P_C x - P_C y\|^2 + \|(I - P_C)x - (I - P_C)y\| \le \|x - y\|^2.$$

See, for example, [20, Ch. 4]. When C is a closed subspace it is also a self-adjoint linear operator [20, Corollary 3.22].

The reflection mapping through the set C is defined by

$$R_C := 2P_C - I,$$

where I is the identity map.

Definition 1.1 (Method of AP). For two closed sets A and B and an initial point $x_0 \in H$, the method of AP generates a sequence $(x_n)_{n=1}^{\infty}$ as follows.

$$x_{n+1} := P_B P_A x_n. \tag{1-2}$$

Definition 1.2 (DR method). For two closed sets A and B and an initial point $x_0 \in H$, the DR generates a sequence $(x_n)_{n=1}^{\infty}$ as follows.

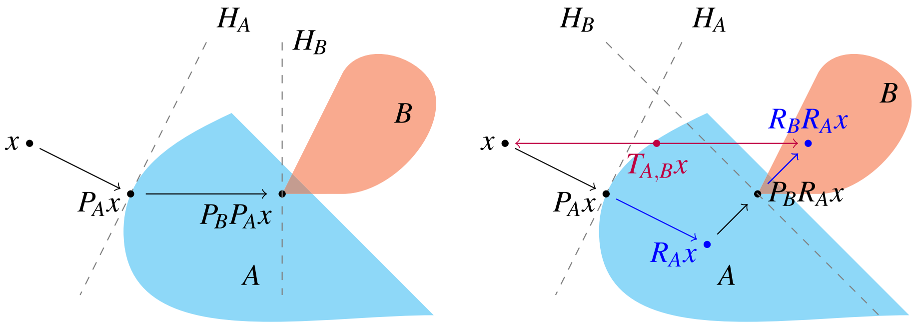
$$x_{n+1} \in T_{A,B}(x_n) \quad \text{where } T_{A,B} := \tfrac{1}{2}(I + R_B R_A). \tag{1-3}$$

DR is often referred to as *reflect-reflect-average*. Both DR and AP are special cases of averaged relaxed projection methods. We denote a *relaxed projection* by

$$R_C^{\gamma}(x) := (2 - \gamma)(P_C - \mathrm{Id}) + \mathrm{Id}, \tag{1-4}$$

for a fixed *reflection parameter* $\gamma \in [0, 2)$. Observe that, when $\gamma = 0$, the operator $R_C^{\gamma=0} = 2P_C - \mathrm{Id}$ is the standard *reflection* employed by DR, and, for $\gamma = 1$, we obtain the *projection* $R_C^{\gamma} = R_C^1 = P_C$. For $\gamma \in (1, 2)$, the operator R_C^{γ} can be called an *under-relaxed projection* following [71]. Here we are using the terminology in (1-4). However, the reader is cautioned that in some articles R_C^{γ} is written as P_C^{γ}, while in others the role of γ is reversed so that $\gamma = 2$ corresponds to a reflection and $\gamma = 0$ is the identity: $\gamma(P_C - \mathrm{Id}) + \mathrm{Id}$.

https://doi.org/10.1017/S1446788719000570 Published online by Cambridge University Press



(a) One step of alternating projections (b) One step of Douglas–Rachford method

FIGURE 1. The operator $T_{A,B}$.

In addition to using relaxed projections as in (1-4), the averaging step of the DR iteration can also be relaxed by choosing an arbitrary point on the interval between the second reflection and the initial iterate. This can be parametrized by some $\lambda \in (0, 1]$. Accordingly we define a λ-averaged relaxed sequence $\{x_n\}$ by

$$x_n := \left(T^{\lambda}_{A^{\gamma},B^{\mu}}\right)^n x_0 := \left(\lambda(R^{\mu}_B \circ R^{\gamma}_A) + (1-\lambda)\mathrm{Id}\right)^n x_0. \tag{1-5}$$

When $\lambda = \gamma = \mu = 1$, this is the sequence generated by AP (1-2), and for $\lambda = 1/2$ and $\gamma = \mu = 0$, this is the standard DR sequence (1-3). For $\gamma = \mu = 0$ and $\lambda = 1$, this is the *Peaceman–Rachford* sequence [125] (see also Lions and Mercier [117, Algorithm 1]).

We note that the framework introduced here does not cover all possible projection methods. For example, one may want to vary the parameters γ, μ and λ on every step or to consider other variations of DR operators (see [9], for example). Single steps of the AP and DR methods are illustrated in Figure 1, which originally appeared in [73].

DEFINITION 1.3. The fixed point set for a mapping $T : H \to H$ is $\mathrm{Fix}\, T = \{x \in H \mid Tx = x\}$ (in the case when T is set-valued, $\mathrm{Fix}\, T = \{x \in H \mid x \in Tx\}$.

1.3. History. Projection methods date at least as far back as 1933 when J. von Neumann considered the method of AP when A and B are affine subsets of H establishing its norm convergence to $P_{A\cap B}(x_0)$ [138]. In 1965, Bregman showed that, in the more general setting where A and B are closed convex sets, AP converges weakly to a point in $A \cap B$ [55] (see also [17]). In 2002, Hundal [110] provided an example in infinite dimensions of a hyperplane and closed cone for which AP fails to converge in norm. However, the cone constructed by Hundal is somewhat unnatural. In [50], Borwein *et al.* explored the possibility of norm convergence for sets occurring more naturally in applications.

Sixty years ago, the DR method was introduced, somewhat indirectly, in connection with nonlinear heat flow problems [74]; see [120] for a thorough treatment of the connection with the form we recognize today. The definitive statement of the weak

https://doi.org/10.1017/S1446788719000570 Published online by Cambridge University Press

convergence result was given by Lions and Mercier in the more general setting of maximal monotone operators [117]. We will first state the problem and result, and then we explain the connection. The problem is

$$\text{Find} \quad x \quad \text{such that } 0 \in (\mathbb{A} + \mathbb{B})x. \tag{1-6}$$

Let the resolvent for a set-valued mapping F be defined by $J_F^\lambda := (\mathrm{Id} + \lambda F)^{-1}$ with $\lambda > 0$. The classical result is as follows.

Theorem 1.4 (Lions and Mercier [117]). *Assume that $\mathbb{A}, \mathbb{B}$ are maximal monotone operators with $\mathbb{A} + \mathbb{B}$ also maximal monotone. Then, for*

$$T_{\mathbb{A},\mathbb{B}} : X \to X : x \mapsto J_{\mathbb{B}}^\lambda (2J_{\mathbb{A}}^\lambda - I)x + (I - J_{\mathbb{A}}^\lambda)x, \tag{1-7}$$

the sequence given by $x_{n+1} = T_{\mathbb{A},\mathbb{B}} x_n$ converges weakly to some $v \in H$ as $n \to \infty$ such that $J_{\mathbb{A}}^\lambda v$ is a zero of $\mathbb{A} + \mathbb{B}$.

The *normal cone* to a set C at $x \in C$ is $N_C(x) = \{y \in H : \langle y, c - x\rangle \le 0 \text{ for all } c \in C\}$. The normal cone operator associated to C is

$$N_C : H \to H : x \mapsto \begin{cases} N_C(x) & \text{when } x \in C, \\ \emptyset & \text{when } x \notin C. \end{cases} \tag{1-8}$$

See, for example, [20, Definition 6.37]. One may think of the feasibility problem (1-1) as a special case of the optimization problem

$$\text{Find} \quad x \in \operatorname{argmin} \{\iota_A + \iota_B\}, \tag{1-9}$$

where the indicator function ι_C for a set C is defined by

$$\iota_C : H \to \mathbb{R}^\infty \quad \text{by } \iota_C : x \mapsto \begin{cases} 0 & \text{if } x \in C, \\ \infty & \text{otherwise.} \end{cases}$$

Whenever A and B are closed and convex, ι_A and ι_B are lower semicontinuous and convex, and their subdifferential operators $\partial\iota_A = N_A$ and $\partial\iota_B = N_B$ are maximal monotone. In this case, under satisfactory constraint qualifications on A, B to guarantee the sum rule for subdifferentials $\partial(\iota_A + \iota_B) = \partial\iota_A + \partial\iota_B$ (see [20, Corollary 16.38]), the problem (1-9) reduces to

$$\text{Find} \quad x \quad \text{such that } 0 \in (\partial\iota_A + \partial\iota_B)(x) = (N_A + N_B)(x),$$

which we recognize as (1-6). Seen through this lens, two-set convex feasibility is a special case of an extremely common problem in convex optimization: that of minimizing a sum of two convex functions $f + g$, where $\mathbb{A} = \partial f$ and $\mathbb{B} = \partial g$. This illuminates its close relationship to many other proximal iteration methods, including the various augmented Lagrangian techniques with which it is often studied in tandem (see § 1.4.1).

Where $\mathbb{A} = N_A$ and $\mathbb{B} = N_B$ are the normal cone operators for closed convex sets A and B, the resolvents $J_{\mathbb{A}}^\lambda, J_{\mathbb{B}}^\lambda$ are the projection operators P_A, P_B, respectively,

https://doi.org/10.1017/S1446788719000570 Published online by Cambridge University Press

$T_{\mathbb{A},\mathbb{B}} = \frac{1}{2}R_BR_A + \frac{1}{2}\mathrm{Id}$ is what we recognize as the operator of the usual DR method, and $J^{\lambda}_{\mathbb{A}}v = P_Av \in A \cap B$ is a solution for the feasibility problem (1-1). For details, see, for example, [20, Example 23.4]. An operator $T : D \to H$ with $D \neq \emptyset$ satisfies $T = J_A$, where $A := T^{-1} - \mathrm{Id}$. Moreover, T is firmly nonexpansive if and only if A is monotone, and T is firmly nonexpansive with full domain if and only if A is maximally monotone. See [20, Proposition 23.7] for details. Rockafellar [128] and Brezis [56] (as cited in [14]) showed that the condition $\mathrm{dom}\mathbb{A} \cap \mathrm{intdom}\mathbb{B} \neq \emptyset$ is sufficient to ensure that $\mathbb{A}$ and $\mathbb{B}$ maximal monotone implies that $\mathbb{A} + \mathbb{B}$ is also maximal monotone. In 1979, Hedy Attouch showed that the weaker condition $0 \in \mathrm{int}(\mathrm{dom}\mathbb{A} - \mathrm{dom}\mathbb{B})$ is sufficient [14].

However, Attouch's condition may not be satisfied if $\mathbb{A} = N_A$ and $\mathbb{B} = N_B$, where A and B meet at a single point, since $\mathrm{dom}N_A = A$ and $\mathrm{dom}N_B = B$. In the following theorem, Bauschke, Combettes and Luke [21] showed that, in the case of the feasibility problem (1-1), the requirement that $\mathbb{A} + \mathbb{B}$ be maximal monotone may be relaxed.

Theorem 1.5 [21, Fact 5.9]. *Suppose $A, B \subseteq H$ are closed and convex with nonempty intersection. Given $x_0 \in H$, the sequence of iterates defined by $x_{n+1} := T_{A,B}x_n$ converges weakly to an $x \in \mathrm{Fix}T_{A,B}$ with $P_Ax \in A \cap B$.*

It should be noted that Zarantonello gave an example showing that, when C is not closed and affine, P_C need not be weakly continuous [140] (see also [20, ex. 4.12]). Despite the potential discontinuity of the resolvent J^{λ}_A, Svaiter later demonstrated that $J^{\lambda}_A x_n$ converges weakly to some $v \in \mathrm{zer}(\mathbb{A} + \mathbb{B})$ [134].

Theorem 1.5 relies on the firm nonexpansivity of $T_{A,B}$. This is an immediate consequence of the fact that it is a 1/2-average of R_BR_A with the identity and that P_A, P_B are themselves firmly nonexpansive so that R_A, R_B and hence R_BR_A are nonexpansive. The proof of Theorem 1.4 similarly relies on the firm nonexpansivity of $J^{\lambda}_{\mathbb{A}}$ and $J^{\lambda}_{\mathbb{B}}$; its requirement that $\mathbb{A} + \mathbb{B}$ be maximal monotone was later relaxed by Svaiter [134].

1.4. Through the lens of monotone operator sums. While our principal interest lies in the less general setting of projection operators, much of the investigation of the DR algorithm has centered on analysis of the problem (1-6). We provide a brief summary.

In 1989 [77], Eckstein and Bertsekas motivated the advantage of $T_{\mathbb{B},\mathbb{A}}$ among resolvent methods as a *splitting method*: a method which employs separate computation of resolvents for $\mathbb{A}$ and $\mathbb{B}$ in lieu of attempting to compute the resolvent of $\mathbb{A} + \mathbb{B}$ directly. They showed that, in the case where $\mathrm{zer}(\mathbb{A} + \mathbb{B}) = \emptyset$, the sequence (1-3) is unbounded, which is a useful diagnostic observation. They also demonstrated that, with exact evaluation of resolvents, the DR method is a special case of the *proximal point algorithm* [77, Theorem 6] in the sense of iterating a resolvent operator [129]: that is,

$$x_{n+1} := J_{\delta_n A} \quad \text{where } \delta_n > 0, \ \sum_{n \in \mathbb{N}} \delta_n = +\infty, \tag{1-10}$$

$$\text{and } \ A : H \to 2^H \text{ is maximally monotone with } \mathrm{zer}A \neq \emptyset.$$

https://doi.org/10.1017/S1446788719000570 Published online by Cambridge University Press

For more information on this characterization, see [20, Theorem 23.41]. In his PhD dissertation [76], Eckstein went on to show that the DR operator may, however, fail to be a *proximal mapping* [20, Theorem 27.1] in the sense of satisfying

$$x_{n+1} := \operatorname{prox}_{\delta_n f} x_n \quad \text{where } \delta_n > 0, \sum_{n\in\mathbb{N}} \delta_n = +\infty \text{ and } f \in \Gamma_0(H) \qquad (1\text{-}11)$$
$$\text{and } \operatorname{prox}_{\delta_n f} x := \operatorname*{argmin}_{y\in X} \left(\delta_n f(y) + \frac{1}{2}\|x-y\|^2\right).$$

Since $\operatorname{prox}_{\delta_n f} = J_{\partial(\delta_n f)}$ (see, for example, [20]), clearly (1-11) implies (1-10). This is also why, in the literature, *DR splitting* is frequently described in terms of prox operators as follows.

$$\textbf{Step 0.} \quad \text{Set initial point } x_0 \text{ and parameter } \eta > 0. \qquad (1\text{-}12)$$

$$\textbf{Step 1.} \quad \text{Set} \begin{cases} y_{n+1} & \in \operatorname*{argmin}_y \left\{ f(y) + \frac{1}{2\eta}\|y - x_n\|^2 \right\} = \operatorname{prox}_{\eta f}(x_n), \\ z_{n+1} & \in \operatorname*{argmin}_z \left\{ g(z) + \frac{1}{2\eta}\|2y_{n+1} - x_n - z\|^2 \right\} = \operatorname{prox}_{\eta g}(2y_{n+1} - x_n), \\ x_{n+1} & = x_n + (z_{n+1} - y_{n+1}), \end{cases}$$

which simplifies to (1-3) when $f := \iota_A$ and $g := \iota_B$ are indicator functions for convex sets (see, for example, [115, 124]).

In 2018, Bauschke *et al.* [40] investigated DR operators which fail to satisfy (1-11), demonstrating that, for linear relations which are maximally monotone, $T_{\mathbb{A},\mathbb{B}}$ generically does not satisfy (1-11).

In 2004, Combettes provided an excellent illumination of the connections between the DR method, the Peaceman–Rachford method, the backward-backward method and the forward-backward method [63]. He also established the following result on a perturbed, relaxed extension of DR, which we quote with minor notation changes.

THEOREM 1.6 (Combettes, 2004). *Let $\gamma \in]0, +\infty[$, let $(\nu_n)_{n\in\mathbb{N}}$ be a sequence in $]0, 2[$ and let $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ be sequences in H. Suppose that $0 \in \operatorname{ran}(\mathbb{A} + \mathbb{B})$, $\sum_{n\in\mathbb{N}} \nu_n(2 - \nu_n) = +\infty$ and $\sum_{n\in\mathbb{N}}(\|a_n\| + \|b_n\|) < +\infty$. Take $x_0 \in \mathcal{H}$ and set*

$$(\forall n \in \mathbb{N})\ x_{n+1} = x_n + \nu_n(J_{\gamma\mathbb{A}}(2(J_{\gamma\mathbb{B}}x_n + b_n) - x_n) + a_n - (J_{\gamma\mathbb{B}x_n} + b_n)).$$

Then $(x_n)n \in \mathbb{N}$ converges weakly to some point $x \in H$ and $J_{\gamma\mathbb{B}}x \in (\mathbb{A} + \mathbb{B})^{-1}(0)$.

At the same time, Eckstein and Svaiter conducted a similar investigation through the lens of Fejér monotonicity, allowing the proximal parameter to vary from operator to operator and iteration to iteration [78].

In 2011, He and Yuan provided a simple proof of the worst case $O(1/k)$ convergence rate in the case where the maximally monotone operators $\mathbb{A}$ and $\mathbb{B}$ are continuous on $\mathbb{R}^n$ [107].

In 2011 [18], Bauschke *et al.* analyzed the Attouch–Théra duality of the problem (1-6), providing a new characterization of $\operatorname{Fix} T_{\mathbb{B},\mathbb{A}}$. In their 2013 article [31],

https://doi.org/10.1017/S1446788719000570 Published online by Cambridge University Press

Bauschke *et al.* introduced a 'normal problem' associated with (1-6), which introduces a perturbation based on an infimal displacement vector (see equation (2-1)). In 2014, they went on to rigorously investigate the range of $T_{\mathbb{A},\mathbb{B}}$ [32].

In 2015, Combettes and Pesquet introduced a random sweeping block coordinate variant, along with an analogous variant for the forward-backward method [66]. In so doing, they furnished a thorough investigation of quasi-Fejér monotonicity.

In 2017, Bauschke *et al.* [34] provided a detailed unpacking of the connections between the original context of Douglas and Rachford [74] and the classical statement of the weak convergence provided by Lions and Mercier [117]. In addition, they provided numerous extensions of the original theory in the case where $\mathbb{A}$ and $\mathbb{B}$ are maximally monotone and affine, including results in the infinite-dimensional setting.

In the same year, Pontus Giselsson and Stephen Boyd established bounds for the rates of global linear convergence under assumptions of strong convexity of g (where $\mathbb{B} = \partial g$) and smoothness, with a relaxed averaging parameter [98]. Giselsson also provided tight global linear convergence rate bounds in the more general setting of monotone inclusions [96]: namely, when one of $\mathbb{A}$ or $\mathbb{B}$ is strongly monotone and the other cocoercive, when one of $\mathbb{A}$ or $\mathbb{B}$ is both strongly monotone and cocoercive, and when one of $\mathbb{A}$ or $\mathbb{B}$ is strongly monotone and Lipschitz continuous. In the case where one operator is strongly monotone and Lipschitz continuous, Giselsson demonstrated that the linear convergence rate bounds provided by Lions and Mercier are not tight. In his analysis, he introduced and made use of *negatively averaged operators*—T such that $-T$ is averaged—proving and exploiting the fact that averaged maps of negatively averaged operators are contractive, in order to obtain the linear convergence results.

In 2018, Moursi and Vandenberghe [121] supplemented Giselsson's work by providing linear convergence results in the case where $\mathbb{A}$ is Lipschitz continuous and $\mathbb{B}$ is strongly monotone, a result that is not symmetric in $\mathbb{A}$ and $\mathbb{B}$ except when $\mathbb{B}$ is a linear mapping.

The DR operator has also been employed as a step in the construction of a more complicated iterated method. For example, in 2015, Luis Briceño-Arias considered the problem of finding a zero for a sum of a normal cone to a closed vector subspace of H, a maximally monotone operator and a cocoercive operator. They provided weak convergence results for a method which employs a DR step applied to the normal cone operator and the maximal monotone operator [58].

Recently, Dao and Phan [68] have introduced what they call an adaptive DR splitting algorithm in the context where one operator is strongly monotone and the other weakly monotone.

Svaiter has also analyzed the semiinexact and fully inexact cases where one or both proximal subproblems are solved only approximately, within a relative error tolerance [135].

The definitive modern treatment of the above history—including the most detailed version of the exposition from [34] on the connections between the contexts of Douglas and Rachford [74] and Lions and Mercier [117]—was given by Moursi in her PhD dissertation [120].

1.4.1. Connection with method of multipliers (ADMM). We provide here an abbreviated discussion of the connection between the DR method and the so-called *method of multipliers* or the ADMM. For a more detailed exposition, see Appendix A. In 1983 [94], Gabay showed that, under appropriate constraint qualifications, the Lagrangian method of Uzawa applied to finding

$$\mathbf{p} := \inf_{v \in V} \{F(Bv) + G(v)\},$$

where B is a linear operator with adjoint B^* and F, G are convex, is equivalent to DR in the Lions and Mercier sense of iterating resolvents (1-7) applied to the problem of finding

$$\mathbf{d} := \inf_{\mu \in H} \{G^*(-B^*\mu) + F^*(\mu)\},$$

where the former is the primal value and the latter is the dual value associated through Fenchel duality (see, for example, [46, Theorem 3.3.5]). We have presented here a more specific case of his result, namely, where $B^t = B^*$; the more general version is in Appendix A.

Gabay gave this method what is now the commonly accepted name *method of multipliers*. It is also frequently referred to as the *alternating direction method of multipliers (ADMM)*. Gabay went on to also consider an analysis of the Peaceman–Rachford algorithm [125] (see also Lions and Mercier [117, Algorithm 1]). Because of this connection, DR, PR and ADMM are frequently studied together. Indeed, another name by which ADMM is known is the *DR ADM*.

Remark (On a point of apparently common confusion). In the literature, we have found it indicated that the close relationship between the ADMM and the iterative schemes in Douglas and Rachford's article [74] and in Lions and Mercier's article [117] was explained by Chan and Glowinski in 1978 [61]. However, both Glowinski and Marroco's 1975 paper [100] and Chan and Glowinski's 1978 paper [61] pre-date Lions and Mercier's 1979 paper [117], and neither of them contains any reference to Douglas' and Rachford's article [74].

Lions and Mercier made a note that DR (which they called simply *Algorithm II*) is equivalent to one of the penalty-duality methods studied in 1975 by Gabay and Mercier [95] and by Glowinski and Marocco [100]. In both of these articles, the method under consideration is simply identified as *Uzawa's algorithm*. The source of the confusion remains unclear, but the explicit explanation of the connection that we have followed is that of Gabay in 1983 [94]. In fact, clearly explaining the connection appears to have been one of his main intentions in writing his 1983 book chapter.

Reasonable brevity precludes an in-depth discussion of Lagrangian duality beyond establishing the connection of ADMM with DR. Instead, we refer the interested reader to a recent survey of Moursi and Zinchenko [122], who drew Gabay's work to the attention of the present authors. We refer the reader also to the sources mentioned in Remark 1, to the recent book by Glowinski *et al.* on splitting methods [101, Ch. 2] and to the following selected resources, which are by no means comprehensive: [44, 80, 83, 89, 99, 105, 106].

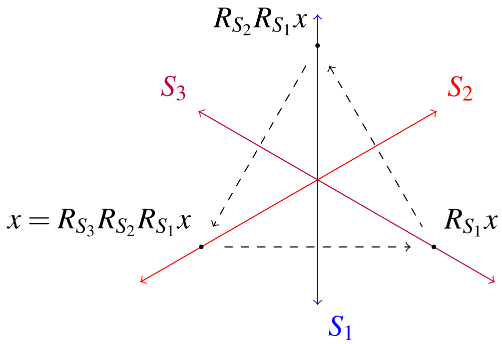


FIGURE 2. The algorithm $x_n := (\frac{1}{2}R_C R_B R_A + \frac{1}{2}\mathrm{Id})^n x_0$ may cycle.

1.5. Extensions to N sets. The method of AP, and the associated convergence results, readily extend to the feasibility problem for N sets

$$\text{Find} \quad x \in \bigcap_{k=1}^{N} S_k, \tag{1-13}$$

to yield the method of *cyclic projections* that involves iterating $T_{S_1 S_2 \cdots S_N} = P_{S_N} P_{S_{N-1}} \cdots P_{S_1}$.

However, even for three sets, the matching extension of DR,

$$x_{n+1} = \tfrac{1}{2}(I + R_{S_3}R_{S_2}R_{S_1})(x_n),$$

may cycle and so fail to solve the feasibility problem (see Figure 2, an example due to Sims that has previously appeared in [136]).

The most commonly used extension of DR from two sets to N sets is Pierra's product space method [127]. More recently, Borwein and Tam have introduced a cyclic variant [52].

1.5.1. Pierra's product space reformulation: 'divide and concur' method. To apply DR for finding $x \in \bigcap_{k=1}^{N} S_k \neq \emptyset$, we may work in the Hilbert product space $\mathbf{H} = H^N$ as follows.

$$\begin{aligned} &\text{Let } S := S_1 \times \cdots \times S_N \\ &\text{and } D := \{(x_1, \ldots, x_N) \in \mathbf{H} : x_1 = x_2 = \cdots = x_N\} \end{aligned} \tag{1-14}$$

and apply the DR method to the two sets S and D. The product space projections for $x = (x_1, \ldots, x_N) \in \mathbf{H}$ are

$$P_S(x_1, \ldots, x_N) = (P_{S_1}(x_1), \ldots, P_{S_N}(x_N)),$$

$$\text{and} \quad P_D(x_1, \ldots, x_N) = \left(\frac{1}{N}\sum_{k=1}^{N} x_k, \ldots, \frac{1}{N}\sum_{k=1}^{N} x_k\right).$$

https://doi.org/10.1017/S1446788719000570 Published online by Cambridge University Press

The method was first nicknamed *divide and concur* by Gravel and Elser [103]—the latter of whom credits the former for the name [81]—and the diagonal set D in this context is referred to as the *agreement set*. It is clear that any point $x \in S \cap D$ has the property that $x_1 = x_2 = \cdots = x_N \in \bigcap_{k=1}^N S_k$. It is also clear that D is a closed subspace of $\mathbf{H}$ (so, P_D is weakly continuous) and that, when $S_1, \ldots, S_N$ are convex, so too is S.

The form of P_D and its consequent linearity allows us to readily unpack the product space formulation to yield the iteration

$$(x_k(n+1))_{k=1}^N = (x_k(n) - a(n) + 2A(n) - P_{S_k}(x_k(n)))_{k=1}^N,$$

where $a(n) = 1/N \sum_{k=1}^N x_k(n)$ and $A(n) = 1/N \sum_{k=1}^N P_{S_k}(x_k(n))$, under which, in the convex case, the sequence of successive iterates weakly converges (by Theorem 1.5) to a limit $(x_1(\infty), x_2(\infty), \ldots, x_N(\infty))$ for which $P_{S_k}(x_k(\infty))$ is, for any $k = 1, 2, \ldots, N$, a solution to the N-set feasibility problem.

A product space schema can also be applied with AP instead of DR, to yield the method of *averaged projections*

$$x_{n+1} = \frac{1}{N}\sum_{i=1}^N P_i(x_i).$$

1.5.2. Cyclic variant: Borwein–Tam method. The cyclic version of DR, also called the *Borwein–Tam* method, is defined by

$$T_{[S_1 S_2 \ldots S_N]} := T_{S_N,S_1} T_{S_{N-1} S_N} \ldots T_{S_2,S_3} T_{S_1,S_2},$$

where each T_{S_i,S_j} is as defined in (1-3). The key convergence result is as follows.

THEOREM 1.7 (Borwein and Tam, 2014). *Let $S_1, \ldots, S_N \subset H$ be closed and convex with nonempty intersection. Let $x_0 \in H$ and set*

$$x_{n+1} := T_{[S_1 S_2 \ldots S_N]} x_n.$$

Then x_n converges weakly to x, which satisfies $P_{S_1} x = P_{S_2} x = \cdots = P_{S_N} x \in \bigcap_{k=1}^N S_k$.

For a proof, see [52, Theorem 3.1] or [136, Theorem 2.4.5], the latter of which—Matthew Tam's dissertation—is the definitive treatise on the cyclic variant.

1.5.3. Cyclically anchored variant (CADRA). Bauschke, Noll and Phan provided linear convergence results for the Borwein–Tam method in the finite-dimensional case in the presence of transversality [39]. At the same time, they introduced the *cyclically anchored DR algorithm (CADRA)* and defined closed, convex sets A (the *anchor* set) and $(B_i)_{i \in \{1,\ldots,m\}}$, where

$$A \cap \bigcap_{i \in \{1,\ldots,m\}} B_i \neq \emptyset$$

$$\text{and} \quad (\forall i \in \{1,\ldots,m\}) T_i = P_{B_i} R_A + \mathrm{Id} - P_A, \quad Z_i = \operatorname{Fix} T_i, \quad Z = \bigcap_{i \in \{1,\ldots,m\}} Z_i,$$

$$\text{where} \quad (\forall n \in \mathbb{N}) \quad x_{n+1} := T x_n \quad \text{where } T := T_m \ldots T_2 T_1. \tag{1-15}$$

When $m = 1$, CADRA becomes regular DR, which is not the case for the Borwein–Tam method. The convergence result is as follows.

https://doi.org/10.1017/S1446788719000570 Published online by Cambridge University Press

THEOREM 1.8 CADRA (Bauschke, Noll and Phan, 2015 [39, Theorem 8.5]). *The sequence $(x_n)_{n\in\mathbb{N}}$ from (1-15) converges weakly to $x \in Z$ with $P_A x \in A \cap \bigcap_{i\in\{1,\ldots,m\}} B_i$. Convergence is linear if one of the following hold.*

(1) *X is finite-dimensional and $\mathrm{ri}A \cap \bigcap_{i\in\{1,\ldots,m\}} \mathrm{ri}B_i \neq \emptyset$.*
(2) *A and each B_i is a subspace with $A + B_i$ closed and $(Z_i)_{i\in\{1,\ldots,m\}}$ is boundedly linearly regular.*

1.5.4. String-averaging and block-iterative variants. In 2016, Censor and Mansour introduced the string-averaging DR (SA-DR) and block-iterative DR (BI-DR) variants [60]. SA-DR involves separating the index set $I := \{1, \ldots, N\}$ (where N is as in (1-13)) into strings along which the two-set DR operator is applied and taking a convex combination of the strings' endpoints to be the next iterate. Formally, letting $I_t := (i_1^t, i_2^t, \ldots, i_{\gamma(t)}^t)$ be an ordered, nonempty subset of I with length $\gamma(t)$ for $t = 1, \ldots, M$ and $x_0 \in H$, set

$$x_{k+1} := \sum_{t=1}^{M} w_t \mathbb{V}_t(x_k) \quad \text{with } w_t > 0 \ \ (\forall t = 1, \ldots, M) \ \text{ and } \ \sum_{t=1}^{M} w_t = 1,$$

$$\text{where } \mathbb{V}_t(x_k) := T_{i_{\gamma(t)}^t, i_1^t} T_{i_{\gamma(t)-1}^t, i_{\gamma(t)}^t} \ldots T_{i_2^t, i_3^t} T_{i_1^t, i_2^t}(x_k),$$

where $T_{A,B}$ is the two-set DR operator. The principal result is as follows.

THEOREM 1.9 SA-DR (Censor and Mansour, 2016 [60, Theorem 18]). *Let $S_1, \ldots, S_N \subset H$ be closed and convex with $\mathrm{int} \bigcap_{i\in I} S_i \neq \emptyset$. Then, for any $x_0 \in H$, any sequence $(x_k)_{k=1}^{\infty}$ generated by the SA-DR algorithm with strings satisfying $I = I_1 \cup I_2 \cup \cdots \cup I_M$ converges strongly to a point $x^* \in \bigcap_{i\in I} S_i$.*

The BI-DR algorithm involves separating I into subsets and applying the two-set DR to each of them by choosing a block index according to the rule $t_k = k \mod M + 1$ and setting

$$x_{k+1} := \sum_{j=1}^{\gamma(t_k)} w_j^{t_k} z_j \quad \text{with } w_{t_k} > 0 \ \ (\forall j = 1, \ldots, \gamma(t_k)) \ \text{ and } \ \sum_{j=1}^{\gamma(t_k)} w_j^{t_k} = 1,$$

$$\text{where } \ z_j := T_{i_j^{t_k}, i_{j+1}^{t_k}}(x_k) \ \ (\forall j = 1, \ldots, \gamma(t_k) - 1) \ \text{ and } \ z_{\gamma(t_k)} := T_{i_{\gamma(t_k)}^{t_k}, i_1^{t_k}}(x_k).$$

The principal result is as follows.

THEOREM 1.10 BI-DR (Censor and Mansour, 2016 [60, Theorem 19]). *Let $S_1, \ldots, S_N \subset H$ be closed and convex with $\bigcap_{i\in I} S_i \neq \emptyset$. For any $x_0 \in H$, the sequence $(x_k)_{k=1}^{\infty}$ of iterates generated by the BI-DR algorithm with $I = I_1 \cup \cdots \cup I_M$, after full sweeps through all blocks, converges:*

(1) *weakly to a point x^* such that $P_{S_{i_j^t}}(x^*) \in \bigcap_{j=1}^{\gamma(t)} S_{i_j^t}$ for $j = 1, \ldots, \gamma(t)$ and $t = 1, \ldots, M$; and*
(2) *strongly to a point x^* such that $x^* \in \bigcap_{i=1}^{N} S_i$ if the additional assumption $\mathrm{int} \bigcap_{i\in I} S_i \neq \emptyset$ holds.*

https://doi.org/10.1017/S1446788719000570 Published online by Cambridge University Press

1.5.5. Cyclic r-sets DR: Aragón Artacho–Censor–Gibali method. Motivated by the intuition of the Borwein–Tam method and the example in Figure 2, Artacho, Censor and Gibali have recently introduced another method which simplifies to the classical DR method in the two-set case [11, Theorem 3.7].

For the feasibility problem of N sets $S_0, \ldots, S_{N-1}$, we denote by $S_{N,r}(d)$ the finite sequence of sets: that is,

$$S_{N,r}(d) := S_{((r-1)d-(r-1)) \bmod N}, S_{((r-1)d-(r-2)) \bmod N}, \ldots, S_{((r-1)d) \bmod N}.$$

The method is then given by

$$\begin{aligned} x_{n+1} &:= \mathbb{V}_N \mathbb{V}_{N-1} \ldots \mathbb{V}_1(x_n), \\ \text{where } \mathbb{V}_d &:= \tfrac{1}{2}(\mathrm{Id} + V_{C_{m,r}(d)}) \\ \text{and} \quad V_{C_0,C_1,\ldots,C_{r-1}} &:= R_{C_{r-1}} R_{C_{r-2}} R_{C_0}. \end{aligned}$$

Provided $\mathrm{int}(\bigcap_{i=0}^{N-1} S_i) \neq \emptyset$, the sequence $(x_n)_{n=1}^{\infty}$ converges weakly to a solution of the feasibility problem. They also provided a more general result, [11, Theorem 3.4], whose sufficiency criteria are, generically, more difficult to verify.

1.5.6. Sums of N operators: Spingarn's method. One popular method for finding a point in the zero set of a sum of N monotone operators $T_1, \ldots, T_N$ is the reduction to a two-operator problem given by

$$\begin{aligned} \mathbb{A} &:= T_1 \otimes T_2 \otimes \cdots \otimes T_N, \\ \mathbb{B} &:= N_B, \end{aligned}$$

where N_B is the normal cone operator (1-8) for B and B is the *agreement set* defined in (1-14). As $\mathbb{A}$ and $\mathbb{B}$ are maximal monotone, the weak convergence result is given by Svaiter's relaxation [134] of Theorem 1.4. The application of DR to this problem is analogous to the product space method discussed in 1.5.1. In 2007, Eckstein and Svaiter [79] described this as *Spingarn's method*, referencing Spingarn's 1983 article [132]. They also established a general projective framework for such problems which does not require reducing the problem to the case $N = 2$.

2. Convex setting

Throughout the rest of the exposition, we will take the DR operator and sequence to be as in (1-5). Where no mention is made of the parameters λ, μ, γ, it is understood that they are as in Definition 1.2. While Theorems 1.5 and 1.4 guarantee weak convergence for DR in the convex setting, only in finite dimensions is this sufficient to guarantee strong convergence. An important result of Hundal shows that AP may not converge in norm for the convex case when H is infinite-dimensional [110] (see also [19, 119]). Although no analogue of Hundal's example seems known to date, for DR in the infinite-dimensional case, norm convergence has been verified under additional assumptions on the nature of A and B.

https://doi.org/10.1017/S1446788719000570 Published online by Cambridge University Press

2.1. Convergence. Borwein *et al.* [47] attribute the first convergence rate results for DR to Hesse, Luke and Neumann who, in 2014, showed local linear convergence in the possibly nonconvex context of sparse affine feasibility problems [109]. Bauschke, Cruz, Nghia, Phan and Wang extended this work by showing that the rate of linear convergence of DR for subspaces is the cosine of the Friedrichs angle [16].

In 2014, Bauschke *et al.* [24] used the convergence rates of matrices to find optimal convergence rates of DR for subspaces with more general averaging parameters as in (1-5). In 2017, Fält and Giselsson characterized the parameters that optimize the convergence rate in this setting [87].

In 2014, Giselsson and Boyd demonstrated methods for preconditioning a particular class of problems with linear convergence rate in order to optimize a bound on the rate [97].

Motivated by the recent local linear convergence results in the possibly nonconvex setting [108, 109, 115, 126], Borwein *et al.* asked whether a global convergence rate for DR in finite dimensions might be found for a reasonable class of convex sets even when the regularity condition $\mathrm{ri}A \cap \mathrm{ri}B \neq \emptyset$ is potentially not satisfied. They provided some partial answers in the context of Hölder regularity with special attention given to convex semialgebraic sets [47].

Borwein, Sims and Tam established sufficient conditions to guarantee norm convergence in the setting where one set is the positive Hilbert cone and the other set is a closed affine subspace which has finite codimension [50].

In 2015, Bauschke *et al.* studied the setting of $\mathbb{R}^2$ where one set is the epigraph of a convex function and the other is the real axis, obtaining various convergence rate results [30]. In their follow-up article in 2016, they demonstrated finite convergence when Slater's condition holds in both the case where one set is an affine subspace and the other is a polyhedron and in the case where one set is a hyperplane and the other is an epigraph [29]. They included an analysis of the relevance of their results in the product space setting of Spingarn [132] and numerical experiments comparing the performance of DR and other methods for solving linear equations with a positivity constraint. In the same year, Bauschke *et al.* provided a characterization of the behavior of the sequence $(T^n x - T^n y)_{n\in\mathbb{N}}$ [27].

In 2015, Davis and Yin showed that DR might converge arbitrarily slowly in the infinite-dimensional setting [70].

2.1.1. Order of operators. In 2016, Bauschke and Moursi investigated the order of operators: $T_{A,B}$ versus $T_{B,A}$. In so doing, they demonstrated that $R_A : \mathrm{Fix}\, T_{A,B} \to \mathrm{Fix}\, T_{B,A}$ and $R_B : \mathrm{Fix}\, T_{B,A} \to \mathrm{Fix}\, T_{A,B}$ are bijections [36].

2.1.2. Best approximations and the possibly infeasible case. The behavior of DR in the inconsistent setting is most often studied using the minimal displacement vector

$$v := P_{\overline{\mathrm{ran}}(\mathrm{Id}-T_{A,B})}0. \tag{2-1}$$

The set of best approximation solutions relative to A is $A \cap (v + B)$; when it is nonempty, the following have also been shown.

https://doi.org/10.1017/S1446788719000570 Published online by Cambridge University Press

In 2004, Bauschke *et al.* considered the algorithm under the name *averaged alternating reflections (AAR)*. They demonstrated that, in the possibly inconsistent case, the shadow sequence $P_A x_n$ remains bounded with its weak sequential cluster points being in $A \cap (v + B)$ [23].

In 2015, Bauschke and Moursi [35] analyzed the more specific setting of two affine subspaces, showing that $P_A x_n$ will converge to a best approximation solution. In 2016, Bauschke *et al.* [28] furthered this work by considering the affine-convex setting, showing that, when one of A and B is a closed affine subspace, $P_A x_n$ will converge to a best approximation solution. They then applied their results to solving the least squares problem of minimizing $\sum_{k=1}^{M} d_{C_k}(x)^2$ with Spingarn's splitting method [132].

In 2016, Bauschke and Moursi provided a more general sufficient condition for the weak convergence [37], and in 2017 they characterized the magnitudes of minimal displacement vectors for more general compositions and convex combinations of operators.

2.1.3. Nearest feasible points (Aragón Artacho–Campoy method). In 2017, Artacho and Campoy introduced what they called the *averaged alternating modified reflections (AAMR)* method for finding the nearest feasible point for a given starting point [9]. The operator and method are defined with parameters $\alpha, \beta \in]0, 1[$ by

$$T_{A,B,\alpha,\beta} := (1 - \alpha)\mathrm{Id} + \alpha(2\beta P_B - \mathrm{Id})(2\beta P_A - \mathrm{Id})$$
$$x_n := T_{A-q,B-q,\alpha,\beta} x_n, n = 0, 1, \ldots, \tag{2-2}$$

which we recognize as DR in the case $\alpha = 1/2, \beta = 1, q = 0$. The convergence result is as follows.

THEOREM 2.1 Aragón Artacho and Campoy 2017, [9, Theorem 4.1]. *Given A, B closed and convex, $\alpha, \beta \in]0, 1[$ and $q \in H$, choose any $x_0 \in H$. Let $(x_n)_{n\in\mathbb{N}}$ be as in (2-2). Then, if $A \cap B \neq \emptyset$ and $q - P_{A\cap B}(q) \in (N_A + N_B)(P_{A\cap B}(q))$, the following hold.*

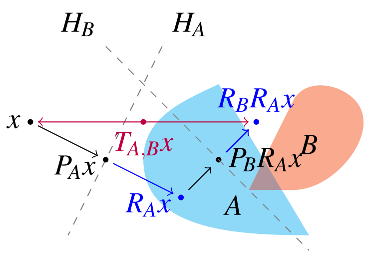
(1) *$(x_n)_{n\in\mathbb{N}}$ is weakly convergent to a point $x \in \mathrm{Fix}\, T_{A-q,B-q,\alpha,\beta}$ such that $P_A(q + x) = P_{A\cap B}(q)$.*
(2) *$(x_{n+1} - x_n)_{n\in\mathbb{N}}$ is strongly convergent to 0.*
(3) *$(P_A(q + x_n))_{n\in\mathbb{N}}$ is strongly convergent to $P_{A\cap B}(q)$.*

Otherwise, $\|x_n\| \to \infty$. Moreover, if A, B are closed affine subspaces, $A \cap B \neq \emptyset$ and $q - P_{A\cap B}(q) \in (A - A)^\perp + (B - B)^\perp$, then $(x_n)_{n\in\mathbb{N}}$ is strongly convergent to $P_{\mathrm{Fix}\, T_{A-q,B-q,\alpha,\beta}}(x_0)$.

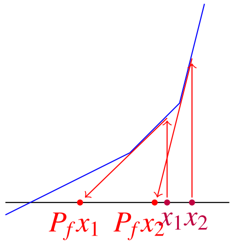
The algorithm may be thought of as another approach to the convex optimization problem of minimizing the convex function $y \mapsto \|q - y\|^2$ subject to constraints on the solution.

It is quite natural to consider the theory of the algorithm in the case where projection operators $P_A = J_{N_A}, P_B = J_{N_B}$ are replaced with more general resolvents for maximally monotone operators [8], an extension Artacho and Campoy gave in 2018. This work has already been extended by Alwadani, Bauschke, Moursi and Wang, who analyzed the asymptotic behavior and gave the algorithm the more specific name of the *Aragón Artacho–Campoy algorithm (AACA)* [1].

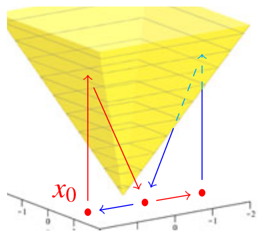
https://doi.org/10.1017/S1446788719000570 Published online by Cambridge University Press



(a) DR with cutters



(b) Subgradient projectors may not be nonexpansive



(c) A fixed point of subgradient DR

Figure 3. Cutter methods.

2.1.4. Cutter methods. Another computational approach is to replace true projections with approximate projections or *cutter* projections onto separating hyperplanes, as in Figure 3(a). Prototypical of this category are subgradient projection methods which may be used to find $x \in \bigcap_{i=1}^m \mathrm{lev}_{\le 0} f_i$ for m convex functions $f_1, \ldots, f_m$; see Figure 3(b). Such methods are not generally nonexpansive (as shown in Figure 3(b)) but may be easier to compute. When true reflection parameters are allowed, the method is no longer immune from 'bad' fixed points, as illustrated in Figure 3(c). However, with a suitable restriction on reflection parameters and under other modest assumptions, convergence may be guaranteed through Fejér monotonicity methods (see, for example, the works of Cegielski and Fukushima [59, 92]). More recently, Díaz *et al.* have provided a standalone analysis of DR with cutter projections [73].

2.2. Notable applications. While the DR operator is firmly nonexpansive in the convex setting, the volume of literature about it is indeed expansive. While reasonable brevity precludes us from furnishing an exhaustive catalogue, we provide a sample of the relevant literature.

As early as 1961, working in the original context of Douglas and Rachford, P.L.T. Brian introduced a modified version of DR for high-order accuracy solutions of heat flow problems [57].

In 1995, Fukushima applied DR to the traffic equilibrium problem, comparing its performance (and the complexity and applicability of the induced algorithms) to ADMM [93].

In 2007, Combettes and Pesquet applied a DR splitting to nonsmooth convex variational signal recovery, demonstrating their approach on image denoising problems [64].

In 2009, Setzer showed that the *alternating split Bregman algorithm* from [102] could be interpreted as a special case of DR in order to interpret its convergence properties, applying the former to an image denoising problem [131]. In the same year, Steidl and Teuber applied DR for removing multiplicative noise, analyzing its linear

https://doi.org/10.1017/S1446788719000570 Published online by Cambridge University Press

convergence in their context and providing computational examples by denoising images and signals [133].

In 2011, Combettes and Pesquet contrasted and compared various proximal point algorithms for signal processing [65].

In 2012, Demanet and Zhang applied DR to l_1 minimization problems with linear constraints, analyzing its convergence and bounding the convergence rate in the context of compressed sensing [72].

In 2012, Boţ and Hendrich proposed two algorithms based on DR splitting, which they used to solve a generalized Heron problem and to deblur images [54]. In 2014, they analyzed with Csetnek an *inertial* DR algorithm and used it to solve clustering problems [53].

In 2015, Bauschke *et al.* applied DR for a road design optimization problem in the context of minimizing a sum of proper convex lower semicontinuous functions, demonstrating its effectiveness on real-world data [33].

In 2017, Wang *et al.* applied DR with facial reduction for a set of matrices of a given rank and a linear constraint set in order to find maximum rank moment matrices [139].

3. Nonconvex setting

Investigation in the nonconvex setting has been two-pronged, with the theoretical inquiry chasing the experimental discovery. The investigation has also taken place in, broadly, two contexts: that of curves and/or hypersurfaces and that of discrete sets.

While Jonathan Borwein's exploration spanned both of the aforementioned contexts, his interest in DR appears to have been initially sparked by its surprising performance in the latter [81], specifically, the application of the method by Elser, Rankenburg and Thibault to solving a wide variety of combinatorial optimization problems, including sudoku puzzles [86]. Where the product space reformulation is applied to feasibility problems with discrete constraint sets, DR often succeeds while AP does not. Early wisdom suggested that one reason for its superior performance is that DR, unlike AP, is immune from false fixed points regardless of the presence or absence of convexity, as shown in the following proposition (see, for example, [136, Proposition 1.5.1] or [86]).

Proposition 3.1 (Fixed points of DR). *Let $A, B \subset H$ be proximal. Then $x \in \operatorname{Fix} T_{A,B}$ implies that $P_A(x) \in A \cap B$.*

Proof. Let $x \in \operatorname{Fix} T_{A,B}$. Then $x = x + P_B(2P_A(x) - x) - P_A(x)$ and so $P_B(2P_A(x) - x) - P_A(x) = 0$, so $P_A(x) \in B$. □

A typical example where $A := \{a_1, a_2\}$ is a doubleton and B a subspace (analogous to the agreement set) is illustrated in Figure 4, where DR is seen to solve the problem while AP becomes trapped by a fixed point.

If the germinal work on DR in the nonconvex setting is that of Elser *et al.* [86] (caution: the role of A and B are reversed from those here), then the seminal work is

https://doi.org/10.1017/S1446788719000570 Published online by Cambridge University Press

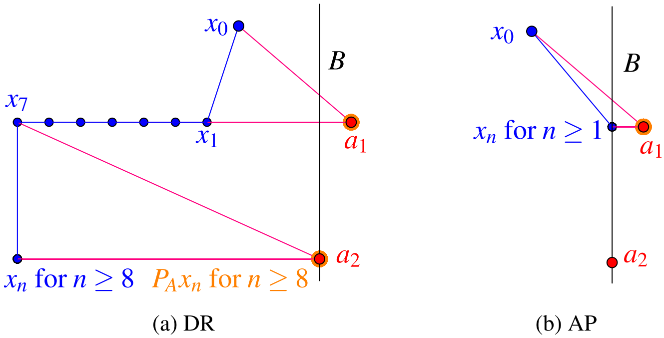


FIGURE 4. DR and AP for a doubleton and a line in $\mathbb{R}^2$.

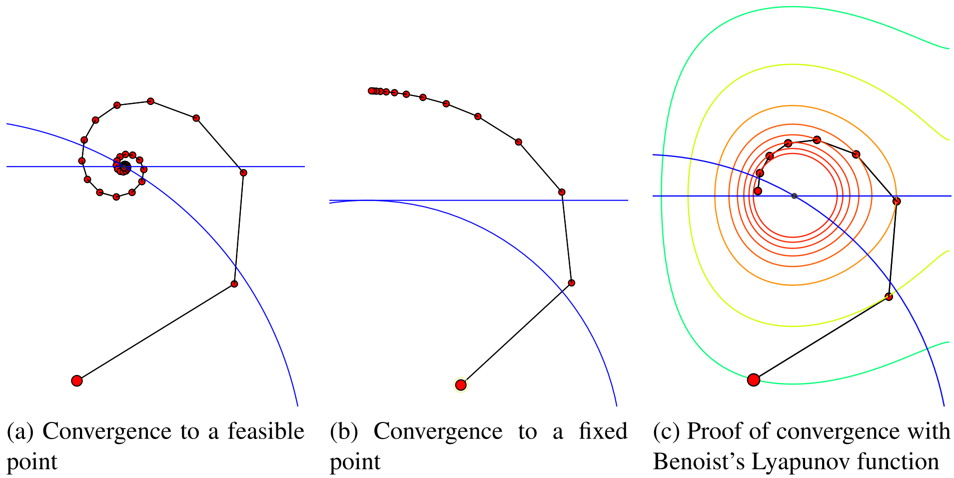


FIGURE 5. Behavior of DR where A is a circle and B is a line; (c) is discussed in § 3.2.

that of J.R. Fienup who applied the method to solve the phase retrieval problem [88]. In [86], Elser *et al.* referred to DR as *Fienup's iterated map* and the *difference map*, while Fienup himself called it the *hybrid input–output algorithm* (HIO) [88]. Elser explains that, originally, neither Fienup nor Elser *et al.* were aware of the work of Lions and Mercier [117], and so the seminal work on DR in the nonconvex setting is, surprisingly, an independent discovery of the method [81]. Fienup constructed the method by combining aspects of two other methods he considered—the *basic input–output* algorithm and the *output–output* algorithm—with the intention of obviating stagnation. Here, again, one may think of the behavior illustrated in Figure 4.

Figure 5 shows behavior of DR in the case where A is a circle and B is a line, a situation prototypical of the phase retrieval problem. For most arrangements, DR converges to a feasible point, as in Figure 5(a). However, when the line and circle meet tangentially, as in Figure 5(b), DR converges to a fixed point which is not feasible, and the sequence $P_A x_n$ converges to the true solution.

Elser notes that it is unclear whether or not Fienup understood that a fixed point of the algorithm is not necessarily feasible, as his approach was largely empirical. Elser sought to clarify this point in his follow-up article in which he augmented the study of DR for phase retrieval by replacing support constraints with object histogram and atomicity constraints for crystallographic phase retrieval [82, Section 5]. In 2001, when [82] was submitted, Elser was not yet aware of Lions' and Mercier's characterization of DR as the *DR method*; it may be recognized in [82] as a special instance of the *difference map* (which we define in (3-1)), a generalization of *Fienup's input–output map*.

In 2002, Bauschke *et al.* finally demonstrated that Fienup's basic input–output algorithm is an instance of Dykstra's algorithm and that HIO (hybrid input–output) with the support constraint alone corresponds to DR [21] (see also their 2003 follow-up article [22]). In another follow-up article [23], they showed that, with support and nonnegativity constraints, HIO corresponds to the HPR (hybrid projection reflection) algorithm, a point that Luke sought to clarify in his succinct 2017 summary of the investigation of DR in the context of phase retrieval [118].

More recently, in 2017, Elser *et al.* published a set of benchmark problems for phase retrieval [85]. They considered DR with true reflections and a relaxed averaging parameter—$\mu = \gamma = 0, \lambda \in]0, 1]$, as in (1-5)—under the name *relaxed–reflect–reflect (RRR)*. In particular, they provided experimental evidence for the exponential growth of DR's mean iteration count as a function of the autocorrelation sparsity parameter, which seems well suited for revealing behavioral trends. They also provided an important clarification of the different algorithms which have been labelled 'Fienup' algorithms in the literature, some of which are not DR.

3.1. Discrete sets. The landmark experimental work on discrete sets is that of Elser, Rankenburg and Thibault [86]. They considered the performance of what they called the *difference map* for various values of the parameter β: that is,

$$T : x \mapsto x + \beta \left(P_A \circ f_B(x) - P_B \circ f_A(x)\right), \tag{3-1}$$
$$\text{where } f_A : x \mapsto P_A(x) - (P_A(x) - x)/\beta,$$
$$\text{and } f_B : x \mapsto P_B(x) + (P_B(x) - x)/\beta.$$

When $\beta = -1$, we recover the DR operator $T_{A,B}$, and when $\beta = 1$, we obtain $T_{B,A}$.

3.1.1. Stochastic problems. Much of the surprising success of DR has been in the setting where some of the sets of interest have had the form $\{0, 1\}^p$. Elser *et al.* adopted the approach of using stochastic feasibility problems to study the performance of DR [86]. They began with the problem of solving the linear Diophantine equation $Cx = b$, where C, a stochastic $p \times q$ matrix, and $b \in \mathbb{N}^p$ are 'known.' Requiring the solution $x \in \{0, 1\}^q$ that is used to generate the problem to also be stochastic ensures uniqueness of the solution for the feasibility problem: find $x \in A \cap B$, where

$$A := \{0, 1\}^q \quad \text{and} \quad B := \{x \in \mathbb{R}^q \text{ such that } Cx = b\}.$$

https://doi.org/10.1017/S1446788719000570 Published online by Cambridge University Press

They continued by solving Latin squares of n symbols. Where $x_{ijk} = 1$ indicates that the cell in the ith row of the jth column of the square is k, the problem is stochastic and the constraint that $x_{\hat{i}\hat{j}\hat{k}} = 1$ if and only if $(\forall i \neq \hat{i})\ x_{i\hat{j}\hat{k}} = 0, (\forall j \neq \hat{j}) x_{\hat{i}j\hat{k}} = 0$ determines the set of allowable solutions. The most familiar form of a Latin square is the sudoku puzzle where $n = 9$ and we require the additional constraint that the complete square consists of a grid of nine smaller Latin squares. For more on the history of the application of projection algorithms to solving sudoku puzzles, see Schaad's master's thesis [130] in which he also applies the method to the eight queens problem.

This work of Elser *et al.* piqued the interest of Borwein who, in 2013, together with Artacho and Tam, continued the investigation of sudoku puzzles [3, 5], exploring the effect of formulation (integer feasibility versus stochastic) on performance. They also extended the approach by solving nonogram problems.

3.1.2. Matrix completion and decomposition. Another application for which DR has shown promising results is finding the remaining entries of a partially specified matrix in order to obtain a matrix of a given type. Borwein, Artacho and Tam considered the behavior of DR for such matrix completion problems [4]. They provided a discussion of the convex setting, including positive semidefinite matrices, correlation matrices and doubly stochastic matrices. They went on to provide experimental results for a number of nonconvex problems, including for rank minimization, protein reconstruction and finding Hadamard and skew-Hadamard matrices. In 2017, Artacho, Campoy, Kotsireas and Tam applied DR to constructing various classes of circulant combinatorial designs [10], reformulating them as three-set feasibility problems. Designs they studied included Hadamard matrices with two circulant cores, as well as circulant weighing matrices and D-optimal matrices.

Even more recently, Franklin used DR to find compactly supported, nonseparable wavelets with orthonormal shifts, subject to the additional constraint of regularity [90, 91]. Reformulating the search as a three-set feasibility problem in $\{\mathbb{C}^{2\times2}\}^M$ for $M = \{4, 6, 8, \dots\}$, they compared the performance of cyclic DR, product space DR, cyclic projections and the *proximal alternating linear minimization* (or *PALM*) algorithm. Impressively, product space DR solved every problem it was presented with.

In 2017, Elser applied DR—under the name RRR (short for *relaxed reflect-reflect*)—for matrix decomposition problems, making several novel observations about DR's tendency to wander, by searching in an apparently chaotic manner, until it happens upon the basin for a fixed point [83]. These observations have motivated the open question we pose in 4.1.2.

3.1.3. The study of proteins. In 2014, Borwein and Tam went on to consider protein conformation determination, reformulating such problems as matrix completion problems [51]. An excellent resource for understanding the early class problems studied by Borwein, Tam and Artacho—as well as the cyclic DR algorithm described in § 1.5—is Tam's PhD dissertation [136].

Elser *et al.* applied DR to study protein folding problems, discovering much faster performance than that of the landscape sampling methods commonly used [86].

3.1.4. Where A is a subspace and B a restriction of allowable solutions. Elser *et al.* applied DR to the study of 3-SAT problems, comparing its performance to that of another solver, Walksat [86] (see also [103]). They found that DR solved all instances without requiring random restarts. They also applied the method to the spin glass ground state problem, an integer quadratic optimization program with nonpositive objective function.

3.1.5. Graph coloring. Elser *et al.* applied DR to find colorings of the *edges* of complete graphs with the constraint that no triangle may have all its edges of the same color [86]. They compared its performance to CPLEX, and included an illustration showing the change of edge colors over time. DR solved all instances, and outperformed CPLEX in harder instances.

In 2016, Artacho and Campoy applied DR to solving graph coloring problems in the usual context of coloring *nodes* [7]. They constructed the feasibility problem by attaching one of two kinds of gadget to the graphs, and they compared performance with the two different gadget types both with and without the inclusion of maximal clique information. They also explored the performance for several other problems reformulated as graph coloring problems; these included: 3-SAT, sudoku puzzles, the eight queens problem and generalizations thereof, and Hamiltonian path problems.

More recently, Artacho, Campoy and Elser [13] have considered a reformulation of the graph coloring problem based on semidefinite programming, demonstrating its superiority through numerical experimentation.

3.1.6. Other implementations. Elser *et al.* went on to consider the case of bit retrieval, where A is a Fourier magnitude/autocorrelation constraint and B is the binary constraint set $\{\pm 1/2\}^n$ [86]. They found its performance to be superior to that of CPLEX.

Bansal used DR to solve Tetravex problems [15].

More recently, in 2018, Elser expounded further upon the performance of DR under varying degrees of complexity by studying its behavior on bit retrieval problems [84]. Of his findings of its performance he observes the following.

> These statistics are consistent with an algorithm that blindly and repeatedly reaches into an urn of M solution candidates, terminating when it has retrieved one of the 4×43 solutions. Two questions immediately come to mind. The easier of these is: How can an algorithm that is deterministic over most of its run-time behave randomly? The much harder question is: How did the $M = 2^{43}$ solution candidates get reduced, apparently, to only about $1.7 \times 10^5 \times (4 \times 43) \approx 2^{24}$?

The behavior of DR under varying complexity remains a fascinating open topic, and we provide it as one of our two open problems in 4.1.2.

https://doi.org/10.1017/S1446788719000570 Published online by Cambridge University Press

3.1.7. Theoretical analysis. One of the first global convergence results in the nonconvex setting was given by Artacho, Borwein and Tam in the setting where one set is a half space and the second set finite [6]. Bauschke, Dao and Lindstrom have since fully categorized the global behavior for the case of a hyperplane and a doubleton (a set of two points) [26]. Both problems are prototypical of discrete combinatorial feasibility problems, the latter especially, insofar as the hyperplane is analogous to the agreement set in the product space version of the method discussed in § 1.5.1, which is the most commonly employed method for problems of more than two sets.

3.2. Hypersurfaces. In 2011, Borwein and Sims made the first attempt at deconstructing the behavior of DR in the nonconvex setting of hypersurfaces [49]. In particular, they considered in detail the case of a circle A and a line B, a problem prototypical of phase retrieval. Here the dynamical geometry software Cinderella [62] first played an important role in the analysis: the authors paired Cinderella's graphical interface with accurate computational output from *Maple* in order to visualize the behavior of the dynamical system. Borwein and Sims went on to show local convergence in the feasible case where the line is not tangential to the two-sphere by using a theorem of Perron. They concluded by suggesting analysis for a generalization of the two-sphere: p-spheres.

In 2013 Artacho and Borwein revisited the case of a two-sphere and line intersecting nontangentially [2]. When x_0 lies in the subspace perpendicular to B, the sequence $(x_n)_{n=0}^{\infty}$ is contained in the subspace and exhibits chaotic behavior. For x_0 not in the aforementioned subspace—which we call the *singular set*—they provided a conditional proof of global convergence of iterates to the nearer of the two feasible points. The proof relied upon constructing and analyzing the movement of iterates through different regions. Borwein humorously remarked of the result, 'This was definitely not a proof from *the book*. It was a proof from the *anti-book*'. Benoist later provided an elegant proof of global convergence by constructing the Lyapunov function seen in Figure 5(c) [43].

In one of his later posthumous publications on the subject [45], Borwein, together with Giladi, demonstrated that the *DR* operator for a sphere and a convex set may be approximated by another operator satisfying a weak ergodic theorem.

In 2016, Borwein *et al.* undertook Borwein's suggested follow-up work in $\mathbb{R}^2$, analyzing not only the case of p-spheres more generally but also of ellipses [48]. They discovered incredible sensitivity of the global behavior to small perturbations of the sets, with some arrangements eliciting a complex and beautiful geometry characterized by *periodic points* with corresponding basins of attraction. A point x satisfying $T^n x = x$ is said to be periodic with *period* the smallest n for which this holds; Figure 6 from [48] shows 13 different DR sequences for an ellipse and line from which subsequences converge to periodic points. Borwein *et al.* combined data from *Cinderella* with parallelization techniques in order to visualize the global behavior. An artistic rendering of the basins with colors chosen based on Aboriginal Australian artwork may be seen in Figure 7; this image appears on the poster for *Mathematics of*

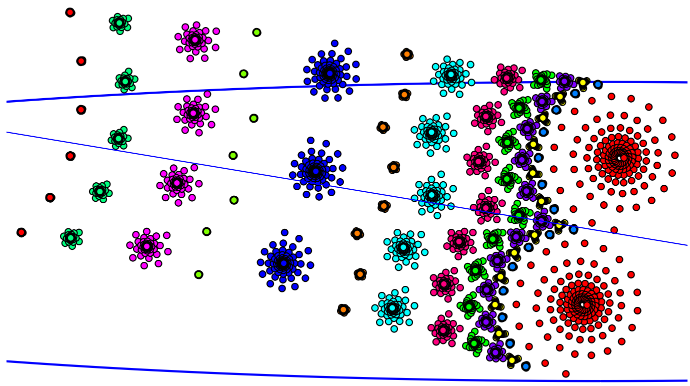


FIGURE 6. Basins of attraction for periodic points with an ellipse and line.

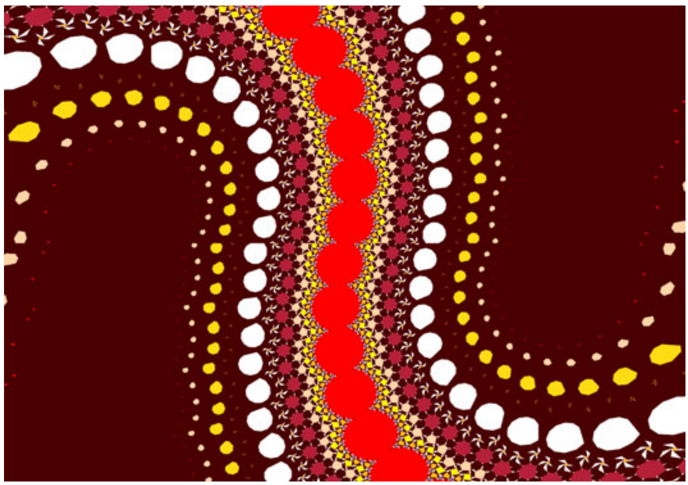


FIGURE 7. Basins of attraction for an ellipse and line with colors based on Aboriginal Australian artwork. This image appears on the poster for *MoCaO*.

Computation and Optimization (MoCaO), an Australian Mathematical Society special interest group founded by Borwein and Jérôme Droniou.

Borwein *et al.* went on to show local convergence to feasible points in the case where the ellipse and line intersect nontangentially, and they extended a best approximation result of Moursi and Bauschke [37] in the setting of boundaries of convex sets.

In order to check the potential influence of sensitivity to compounding numerical error on their discoveries, Borwein *et al.* used Schwarzian reflection to compute approximate projections as an alternative to the numerical solution of a Lagrangian problem (see, for example, [123]). This work inspired a 2017 follow-up article by

https://doi.org/10.1017/S1446788719000570 Published online by Cambridge University Press

Lindstrom *et al.* [116] that analyzed the performance of DR for finding intersections of smooth curves in $\mathbb{R}^2$ more generally, and it showed that local convergence extends to the more general case of smooth plane curves intersecting nontangentially with reasonable limits on their curvature (in Definition 3.3, we will introduce the notion of *superregularity*). Dao and Tam [69] have since adapted Benoist's Lyapunov approach to beautifully illuminate the behavior for more general curves, including showing global behavior for many curve pairs.

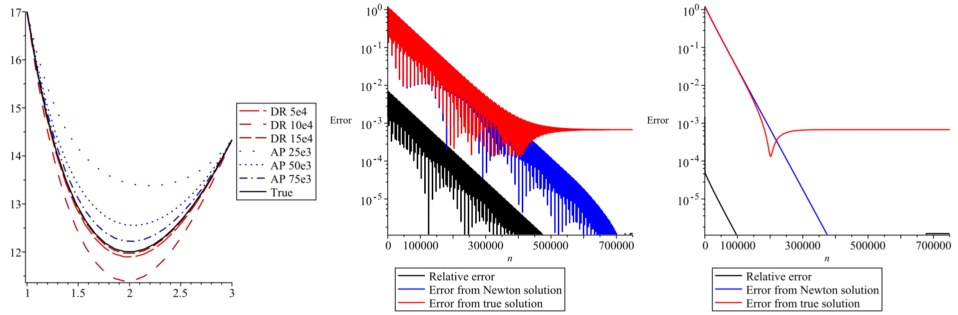
Lamichhane, Lindstrom and Sims used AP and DR to find numerical solutions for boundary-value ordinary differential equations on closed intervals in $\mathbb{R}$ by reformulating the problem of N node approximation as a feasibility problem of satisfying N equations which define possibly discontinuous hypersurfaces [116]. The approach is mostly experimental, and they compared the observed convergence with that explicitly visible in the two-set ellipse/line setting. They also compared the behavior of DR and AP on each test problem and found that AP generally performs faster.

The above studies on hypersurfaces have uncovered a general trend which distinguishes AP from DR: namely, AP is more prone to becoming trapped by extraneous fixed points but demonstrates monotonicity in convergence with an asymptotic direction of approach, while DR tends to escape from false solutions and its basins of convergence persistently feature spiralling trajectories which induce observed oscillations in plots of change and error. Some of this behavior may be seen in Figure 8 from [113] which shows the behavior, as measured for the agreement set shadow sequence $P_B x_n$, when seeking the solution to an N set feasibility problem corresponding to the numerical solution of a boundary-value problem. In Figures 8(b) and 8(c), relative error (change from iterate to iterate), error from numerical solution (obtained by applying Newton's method to the discretized problem) and error from the true solution (analytically obtained) are monotonic for AP but oscillate for DR. This monotonicity may be further observed in Figure 8(a), where approximate solutions to a boundary value problem—corresponding to various step intervals for DR and AP—may be seen along with the true solution; AP approaches the true solution from one side, while DR exhibits more exotic behavior. The authors of [113] hypothesize that the observed left-right-left wandering of $P_B x_n$ visible in Figure 5(a), which results from the spiralling of x_n, is prototypical of the numerically observed oscillation in more complicated settings like Franklin's work on wavelets.

3.3. Results on regularity, transversality and rates of convergence. Much of the convergence analysis in the nonconvex setting has focused on regularity assumptions. Throughout this section, A and B continue to be closed subsets of a finite-dimensional Euclidean space X.

Definition 3.2 (Regularity and transversality). The closed sets $\{C_i\}_{i\in I}$, $I = \{1, \ldots, m\}$ are said to be:

https://doi.org/10.1017/S1446788719000570 Published online by Cambridge University Press



(a) Approximate solutions for an ODE (b) DR error plot (c) AP error plot

FIGURE 8. Comparison of DR and AP convergence behavior.

(1) *κ-subtransversal* or *κ-linearly regular* with regularity modulus $\kappa \in]0, \infty[$ on $U \subset X$ if

$$(\forall x \in U) \quad d_C(x) \le \kappa \max_{i \in I} d_{C_i}(x) \quad \text{where } C := \bigcap_{i \in I} C_i;$$

(2) *subtransversal* around $x \in X$ or *linearly regular* at $x \in X$ if there exist δ and κ greater than zero such that $\{C_i\}_{i \in I}$ is κ-linearly regular on $\mathbb{B}(x, \delta)$;
(3) *boundedly linearly regular* if, for every bounded set $U \subset X$, there exists $\kappa_U > 0$ such that $\{C_i\}_{i \in I}$ is κ-linearly regular on U;
(4) U-regular at $x \in X$ if U is an affine subspace of X with $x \in U$ and

$$\sum_{i \in I} u_i = 0 \quad \text{and} \quad u_i \in N_{C_i}(x) \cap (U - x) \implies (\forall i \in I)\ u_i = 0;$$

(5) *transversal* or *strongly regular* at $x \in X$ if $\{C_i\}_{i \in I}$ is U-regular with $U = X$; and
(6) *affine-hull regular* at x in the two-set case $m = 2$ when $L = \operatorname{aff}(C_1 \cup C_2)$ if $N_A^L(x) \cap (-N_B^L(x)) = \{0\}$.

See, for example, [67, 108, 111, 126].

More recently, the notion of *intrinsic transversality* has been introduced which fills a theoretical gap between the regularity conditions of transversality and subtransversality [75] (see also [112]).

It may be readily seen that an ellipse and line which intersect nontangentially are transversal at the point of intersection. Indeed, the regularity framework locally describes many hypersurface feasibility problems. The notion of *superregularity* for a single set C may be thought of as a smoothness condition.

DEFINITION 3.3 (Superregularity). A closed subset $A \subset X$ is (ε, δ)-regular at x if $\varepsilon \ge 0$, $\delta > 0$ and

$$\begin{Bmatrix} y, z \in A \cap \mathbb{B}_\delta(x) \\ u \in N_A^{\text{prox}}(x) = \operatorname{cone}(P_A^{-1} x - x) \end{Bmatrix} \implies \langle u, z - y \rangle \le \varepsilon \|u\| \cdot \|z - y\|.$$

https://doi.org/10.1017/S1446788719000570 Published online by Cambridge University Press

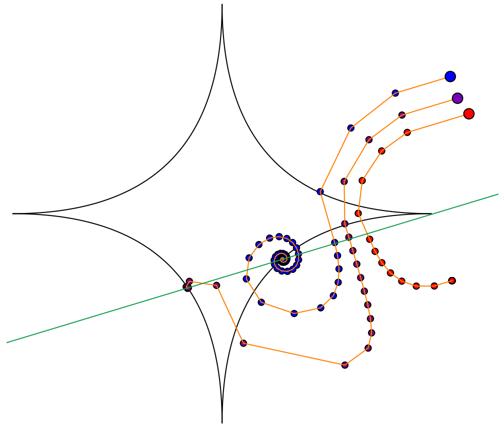


FIGURE 9. DR convergence for a $1/2$-sphere and a line.

C is said to be superregular at x if, for every $\varepsilon > 0$, there exists $\delta > 0$ such that C is (ϵ, δ)-regular at x (see, for example, [126]).

It may be seen that in the case $X = \mathbb{R}^2$ and $A = \operatorname{graph} f = \{(x_1, x_2) | f(x_1) = x_2\}$, where $f : \mathbb{R} \to \mathbb{R}$, superregularity of A at $(x, f(x))$ implies smoothness of f at x.

Figure 5(a) shows how DR may behave when regularity conditions are not satisfied at the feasible point, while the rightmost sequence in Figure 9 illustrates what may happen when two sets meet subtransversally but superregularity fails for one of them (the p-sphere). The other two sequences illustrate how the angle at which the sets meet at the feasible point determines the linear rate of convergence.

As early as 2013, Lewis, Luke and Malick analyzed the local convergence for alternating and averaged nonconvex projection methods in the presence of regularity conditions [114]. In the same year, Hesse and Luke undertook a theoretical study of DR in the presence of local regularity conditions in finite dimensions [108]. They showed that, when the sets involved are affine, strong regularity is necessary for linear convergence, in contradistinction with AP for which such conditions are sufficient but not necessary. They also established a number of linear convergence results for DR, the first of which is as follows.

THEOREM 3.4 Linear convergence of DR (Luke and Hess, 2013 [108, Theorem 3.16]). *Let the pair of closed sets $\{A, B\}$ be linearly regular at $x \in A \cap B$ on $\mathbb{B}_\delta(x)$ with regularity modulus $\kappa > 0$ for some $\delta > 0$. Suppose, further, that B is a subspace and that A is (ε, δ)-regular at x with respect to $A \cap B$. Assume that, for some $c \in]0, 1[$,*

$$\left.\begin{cases} z \in A \cap \mathbb{B}_\delta(x), & u \in N_A(z) \cap \mathbb{B}_1(0) \\ y \in B \cap \mathbb{B}_\delta(x), & v \in N_B(y) \cap \mathbb{B}_1(0) \end{cases}\right\} \implies \langle u, v \rangle \geq -c.$$

https://doi.org/10.1017/S1446788719000570 Published online by Cambridge University Press

If $x_n \in \mathbb{B}_{\delta/2}(x)$ *and* $x_{n+1} \in T_{A,B}x_n$*, then*

$$d(x_{n+1}, A \cap B) \leq \sqrt{1 + 2\varepsilon(1+\varepsilon) - \frac{1-c}{\kappa^2}}\, d(x_n, A \cap B).$$

Another of their results, [108, Theorem 3.18], has since been strengthened by Phan [126, Theorem 4.3] to the following.

THEOREM 3.5 Linear convergence of DR (Phan, 2016 [126, Theorem 4.3]). *Let the closed sets* A, B *be superregular at* $x \in A \cap B$ *and let* $\{A, B\}$ *be strongly regular at* x*. Then if* x_0 *is sufficiently close to* x*, the sequence* $x_{n+1} := T_{A,B}x_n$ *converges to a point* $\overline{x} \in A \cap B$ *with R-linear rate.*

Phan provided additional information about the rate R in [126, Remark 4.5] and gave the following second main result on affine-hull regularity.

THEOREM 3.6 Linear convergence of DR (Phan, 2016 [126, Theorem 4.7]). *Let* A, B *be closed and let* $L := \mathrm{aff}(A \cup B)$*. Further, suppose that* A, B *are superregular at* $x \in A \cap B$ *and* $\{A, B\}$ *is affine-hull regular at* x*. Then, if the the shadow sequence* $P_L x_0$ *is sufficiently close to* x*, the DR sequence* $x_{n+1} := T_{A,B}x_n$ *converges to a point* $\overline{x} \in \operatorname{Fix} T_{A,B}$ *with R-linear rate. Moreover,*

$$P_A\overline{x} \equiv P_B\overline{x} = \overline{x} - (x_0 - P_L x_0) \in A \cap B,$$

and so $P_A\overline{x} \equiv P_B\overline{x}$ *solves the feasibility problem.*

Phan also provided a more detailed description of the region of convergence, and extended the analysis into the convex setting.

In 2016 [67], Dao and Phan went on to consider the more general framework of cyclic relaxed projection methods for the feasibility problem of m sets $\{C_i\}_{i\in I}$, $I = \{1, \ldots, m\}$, where the sequence is defined in terms of l operators

$$T_{nl+j} := T_j \quad \text{and} \quad x_n := T_n x_{n-1} \quad \text{with } J := \{1, \ldots, l\}. \tag{3-2}$$

Where we have modified the notation to be consistent with (1-4), Dao and Phan considered the following cyclic generalized DR algorithm defined by (3-2) and the following. For every $j \in J$, let $\mu_j, \gamma_j \in [0, 2[$, $\lambda_j \in]0, 1[$ and $s_j, t_j \in I$ such that $s_j \neq t_j$ and

$$\begin{aligned} I =& \{s_j | j \in J\} \cup \{t_j | j \in J\}, \\ T_j :=& (1 - \lambda_j)\mathrm{Id} + \lambda_j R_{C_{t_j}}^{\mu_j} R_{C_{s_j}}^{\gamma_j}, \end{aligned}$$

where $R_{C_j}^{\mu_j}$ is defined as in (1-4). The convergence results are as follows.

THEOREM 3.7 Linear convergence of cyclic generalized DR (Dao and Phan, 2016 [67, Theorem 5.21]). *Let* $I := \{1, \ldots, m\}$ *and* $x \in \bigcap_{i\in I} C_i$*. Suppose that* $\{C_i\}_{i\in I}$ *is superregular at* x *and is linearly regular around* x *and that* $\{C_{s_j}, C_{t_j}\}$ *is strongly regular at* x *for every* $j \in J$*. Then, when started at a point* x_0 *sufficiently close to* x*, the cyclic generalized DR sequence generated by* $(T_j)_{j\in J}$ *converges R-linearly to a point* $\overline{x} \in \bigcap_{i\in I} C_i$*.*

https://doi.org/10.1017/S1446788719000570 Published online by Cambridge University Press

Theorem 3.8 Affine reduction for generalized DR sequences (Dao and Phan, 2016 [67, Theorem 5.25]). *Let A, B be closed, let $x \in A \cap B$ and let $L := \mathrm{aff}(A \cup B)$. Suppose that $\{A, B\}$ is superregular and is affine-hull regular at x. Let $(x_n)_{n\in\mathbb{N}}$ be defined by $x_{n+1} := ((1-\lambda)\mathrm{Id} + R_B^{\mu} P_A^{\gamma})x_n$, where $\mu, \gamma \in [0, 2[$ and $\lambda \in]0, 1[$. Then the following hold.*

(1) *If $\gamma = \mu = 0$, then, whenever $P_L x_0$ is sufficiently close to x, $(x_n)_{n\in\mathbb{N}}$ converges R-linearly to a point $\overline{x} \in \mathrm{Fix}\, T$ with $P_A\overline{x} = P_B\overline{x} \in A \cap B$.*

(2) *If either $\lambda > 0$ or $\mu > 0$, then, whenever $P_L x_0$ is sufficiently close to x, $(x_n)_{n\in\mathbb{N}}$ converges R-linearly to a point $\overline{x} \in A \cap B$.*

3.3.1. Other convergence results. Numerous other investigations of convergence for DR have also been undertaken. In 2014 Bauschke and Noll proved local convergence to a fixed point in the case where A and B are finite unions of convex sets [38]. In 2016, Bauschke and Dao provided various sufficient conditions for finite convergence of the DR sequence [25].

3.3.2. Further variants. If one considers the spiralling behavior characteristic of local convergence of DR, it is very natural to seek faster convergence by taking a step towards the center of the spiral. This intuition has given birth to the notion of the method of *circumcentering* [41, 42].

3.3.3. Nonconvex minimization. In 2014, Patrinos *et al.* introduced the so-called *DR envelope* whose stationary points correspond to solutions for the problem of minimizing a sum of two convex functions $f + g$ subject to linear constraints [124].

In 2015, motivated by properties of the DR envelope, Li and Pong introduced the DR merit function [115]

$$\mathfrak{D}_\eta(y, z, x) := f(y) + g(z) - \frac{1}{2\eta}\|y - z\|^2 + \frac{1}{\eta}\langle x - y, z - y\rangle.$$

Li and Pong analyzed the limiting characteristics of $\mathfrak{D}_\eta(y_n, z_n, x_n)$, where y_n, z_n, x_n are either as in (1-12) or are obtained from a modified variant, where $x_0 \in X$ and

$$\begin{cases} y_{n+1} = \dfrac{1}{1+\eta}(x_n + \eta P_A(x_n)), \\ z_{n+1} \in \operatorname*{argmin}_{z\in B}\{\|2y_{n+1} - x_n - z_n\|\}, \\ x_{n+1} = x_n + (z_{n+1} - y_{n+1}), \end{cases} \qquad (3\text{-}3)$$

which arises from applying (1-12) to the problem of minimizing $\frac{1}{2}d_A^2(x)$ subject to $x \in B$, where A is convex but B may not be. They showed the following theorem.

Theorem 3.9 Global subsequential convergence (Li and Pong, 2015 [115, Theorem 1]). *Let g be proper and closed and let f have Lipschitz continuous gradient whose Lipschitz continuity modulus is bounded by L. Choose $\nu \in \mathbb{R}$ so that $f + (\nu/2)\|\cdot\|^2$ is convex. Suppose that η is chosen so that $(1+\eta L)^2 + 5\eta\nu/2 - (3/2) < 0$. Let y_n, z_n, x_n*

https://doi.org/10.1017/S1446788719000570 Published online by Cambridge University Press

be as in 1-12. Then $\{\mathfrak{D}_\eta(y_n, z_n, x_n)\}_{n\geq 1}$ *is nonincreasing. Moreover, if a cluster point of* (y_n, z_n, x_n) *exists, then*

$$\lim_{n\to\infty} \|x_{n+1} - x_n\| = \lim_{n\to\infty} \|z_{n+1} - y_n\| = 0, \tag{3-4}$$

and, for any cluster point $(\overline{y}, \overline{z}, \overline{x})$*, we have* $\overline{z} = \overline{y}$ *and* $0 \in \nabla f(\overline{z}) + \partial g(\overline{z})$*.*

THEOREM 3.10 Global convergence of the whole sequence (Li and Pong, 2015 [115, Theorem 2]). *Let* $f, g, l, L, x_n, y_n, z_n, \eta$ *be as in Theorem 3.9. Additionally, suppose that* f, g *are algebraic and that* $\{(y_n, z_n, x_n)\}$ *has a cluster point* $(\overline{y}, \overline{z}, \overline{x})$*. Then the sequence* $\{(y_n, z_n, x_n)\}$ *is convergent.*

THEOREM 3.11 Convergence of DR splitting method for nonconvex feasibility problems involving two sets (Li and Pong, 2015 [115, Theorem 5]). *Let A be a nonempty, closed, convex set, and let B be a nonempty closed set with either A or B compact. Suppose, in addition, that* $0 < \eta < \sqrt{3/2} - 1$*. Then the sequence* $\{(y_n, z_n, x_n)\}$*, where* y_n, z_n, x_n *are as in (3-3), is bounded. Moreover, any cluster point* $(\overline{y}, \overline{z}, \overline{x})$ *of the sequence satisfies* $\overline{z} = \overline{y}$ *and* $\overline{z}$ *is a stationary point of (3-3). Additionally, (3-4) holds.*

Li and Pong also provided detailed results on the convergence rates. Andreas Themelis and Panos Patrinos have since published a follow-up article [137] in which they relax some of the restrictions on the step size η, as well as providing a discussion of the connections with ADMM.

In 2017, Grussler and Giselsson [104] analyzed the specific case of minimizing $f + g$ with both forward-backward splitting and the DR operator $T_{\partial f, \partial g}$, where g is convex and

$$f : M \mapsto k(\|M\|) + \iota_{\mathrm{rank}(M)\leq r}(M)$$

is nonconvex, where $k(\cdot)$ is increasing and convex, $\|\cdot\|$ is a unitarily invariant norm and $\iota_{\mathrm{rank}(M)\leq r}$ is the indicator function for matrices that have at most rank r. They provided conditions under which prox_f and $\mathrm{prox}_{f^{**}}$ coincide, constructing a framework under which they showed local convergence when solutions to the convex problem of minimizing $f^{**} + g$ coincide with solutions to the nonconvex problem of minimizing $f + g$.

4. Summary

The goal of this survey has been to illuminate the history, motivations and robustness of DR in each of the broad settings wherein it has been considered. Much more could be said, and certainly much more will be. As noted by Glowinski *et al.* in the preface of their new book on the subject, new applications of splitting methods are being introduced almost daily [101].

https://doi.org/10.1017/S1446788719000570 Published online by Cambridge University Press

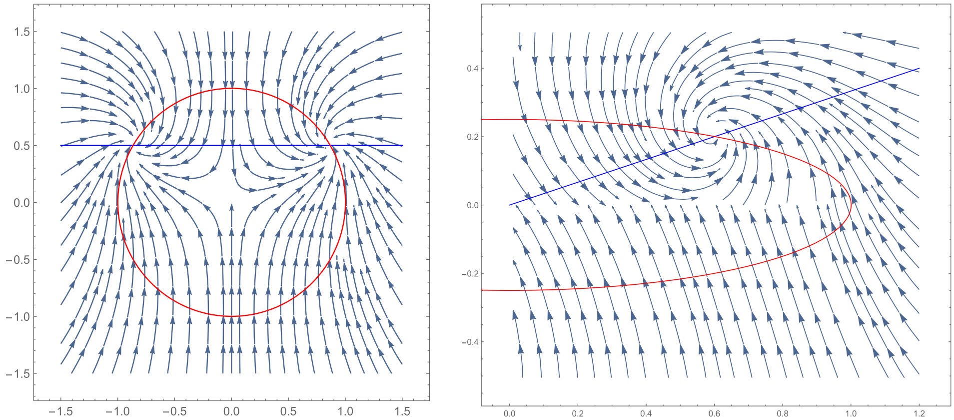


Figure 10. The flowfield (4.1.1) with a circle/line (left) and ellipse/line (right). Images courtesy of Veit Elser.

4.1. Future avenues of inquiry. These directions include the continued analysis of the Artacho–Campoy method in the convex setting, wavelet discovery in the nonconvex setting, nonconvex minimization through the framework of Li and Pong and the analysis of convergence rates under general parameters in all of these. We choose to state here two problems in the nonconvex setting—both suggested by Veit Elser—which have received little attention despite their particularly intriguing nature.

4.1.1. Continuous time variant. For the case of a circle and line, Borwein and Sims [49] considered the 'continuous time' version of the algorithm whose flow field is at left in Figure 10 and corresponds to the solution of the differential equation

$$\frac{dx}{dt} = T(x) \quad \text{when } \lambda \to 0^+.$$

Elser has suggested analyzing the continuous time variant in the more general setting of ellipses and plane curves. Elser provided flow field images for a curve and integer lattice in [83], and he has generously furnished the images in Figure 10.

4.1.2. Complexity theory. Elser hypothesizes that, for Latin square problems, higher dimensionality is associated with greater robustness for the algorithm. The idea is that, as the complexity of the problem grows, the singular regions—of chaotic or periodic behavior—account for a smaller share of the total space. For most starting points, then, the iterates tend to explore the space without becoming stuck, as in Figure 4(a), until eventually they fall into the basin of attraction for a feasibility point. Evidence abounds, as in [83–85]. Can the behavior of DR and similar methods under complexity be rigorously catalogued through experimental analysis?

4.2. Conclusion. The role of DR in the convex setting is both well known and celebrated. More novel and striking is its success in the nonconvex setting. Borwein

https://doi.org/10.1017/S1446788719000570 Published online by Cambridge University Press

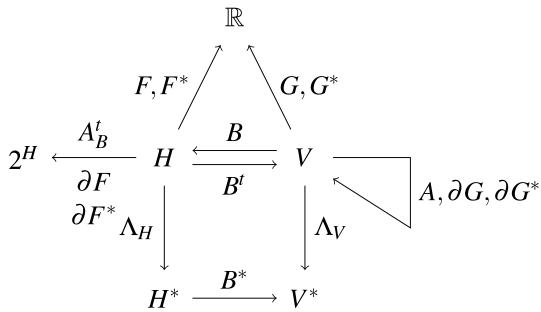


FIGURE 11. Function diagram for Gabay's exposition.

described DR as an 'out-of-the-box solver,' whose robustness for a given nonconvex problem cannot be simply explained by its having been originally designed with that specific problem in mind. While the exact formulation for an embedding of a problem in $\mathbb{R}^n$—for example, the stochastic representation of a sudoku puzzle or the number of gadgets used in [7]—may affect performance, DR fundamentally requires very little: if one can compute the projections, one can use the solver. Perhaps this is why its performance consistently surprises those who study or use it. One thing is certain: the complexity of the behavior is astounding, and much of the space remains to be explored.

Acknowledgements

We are particularly grateful to Heinz Bauschke and Veit Elser, whose expertise on DR has been instrumental in our reconstruction of its history. We also extend our thanks to Pontus Giselsson, whose careful reading and helpful suggestions helped to improve the manuscript.

Appendix A. ADMM and DR

Throughout this section, the function diagram in Figure 11 is a useful reference. In particular, it should be noted that Gabay defined the conjugates $F^* : H \to \mathbb{R}$ and $G^* : V \to \mathbb{R}$ on the primal spaces.

In 1983 [94], Gabay considered the application of (1-7) with $\mathbb{B} := \partial F^* = (\partial F)^{-1}$ for $F : H \to]0, \infty]$ a proper convex lower semicontinuous function and $\mathbb{A} := A^t_B : H \to 2^H$ by

$$A^t_B(\mu) = \{q \in H | \exists v \in V \text{ such that } q = -Bv,\ -B^t\mu \in A(v)\},$$

for a maximally monotone operator A, and where $B : V \to H$ is a continuous linear operator with adjoint $B^* : H^* \to V^*$

$$\text{where } \begin{cases} \langle \Lambda_V u, v\rangle_{V^*\times V} = \langle u, v\rangle_V \ \forall\, u, v \in V & \text{with } \Lambda_V u \in V^*, \\ \langle \Lambda_H p, q\rangle_{H^*\times H} = \langle p, q\rangle_H \ \forall\, p, q \in H & \text{with } \Lambda_H p \in H^* \end{cases}$$

$$\text{and} \quad B^t : H \to V \text{ by } B^t := \Lambda_V^{-1} \circ B^* \circ \Lambda_H.$$

The motivating variational inequality problem is to find $u \in V$ such that

$$\exists w \in A(u) \ \text{ where } \ (\forall v \in V) \ \langle w, v-u\rangle_V + F(Bv) - F(Bu) \geq 0. \tag{A-1}$$

When $A = \partial G$ for $G : V \to]-\infty, \infty]$ a convex, proper, lower semicontinuous function, the variational inequality (A-1) is just

$$\mathbf{p} := \inf_{v\in V}\{F(Bv) + G(v)\}. \tag{A-2}$$

When A is coercive or B^tB is an isomorphism of V, then

$$J^{\lambda}_{A^t_B}(y) = y + \lambda B(A + \lambda B^tB)^{-1}(-B^ty).$$

Gabay showed that (1-7) then becomes the following.

Step 0 Choose ω_0 to be an approximate solution of the problem: Find ω such that $0 \in (A^t_B + \partial F^*)(\omega)$.

Step 1 Choose x_0, p_0 such that $p_0 \in \partial F^*(\omega_0)$, $t_0 = \omega_0 + \lambda p_0$. (This ensures that $\omega_0 = J^r_{\partial F^*}(x_0)$.)

Step 2
$$\begin{cases} u_{n+1} := (A + \lambda B^tB)^{-1}(\lambda B^t p_n - B^t\omega_n), \\ p_{n+1} := (\partial F + \lambda \mathrm{Id})^{-1}(\omega_n + \lambda B u_{n+1}), \\ \omega_{n+1} := \omega_n + \lambda(Bu_{n+1} - p_{n+1}), \\ x_{n+1} := \omega_n + \lambda B u_{n+1}. \end{cases} \tag{A-3}$$

In this case, $(\omega_n)_{n\in\mathbb{N}}$ is the sequence of multipliers, and $\omega_n := J^{\lambda}_{\partial F^*}(x_n)$ is the shadow sequence iterate corresponding to the nth iterate of the DR sequence $(x_n)_{n\in\mathbb{N}}$. In terms of Figure 1(b), if we take $\lambda = 1$, $B = \mathrm{Id}$, $F^* = N_A$ and $A^t_B = N_B$, then, in (A-3), $x_n = x_n$, $\omega_n = P_A x_n$, $p_n = (x_n - P_A x_n)$ and $u_{n+1} = (P_B R_A x_n - R_A x_n)$.

Gabay rewrites (A-3) as in terms of the sequences u_n, p_n, ω_n as follows.

Step 0 Find $u_{n+1} \in V$ satisfying the variational inequality: $\exists w_{n+1} \in A(u_{n+1})$ such that $(\forall v \in V)\langle w_{n+1}, v\rangle_V + \langle \omega_n - \lambda p_n + \lambda B u_{n+1}, Bv\rangle_H = 0$.

Step 1 Find p_{n+1} which solves the minimization problem:

$$F(p_{n+1}) - F(q) - \langle \omega_n, p_{n+1} - q\rangle_H + \frac{\lambda}{2}\|Bu_{n+1} - p_{n+1}\|^2_H - \frac{\lambda}{2}\|Bu_{n+1} - q\|^2_H \leq 0.$$

Step 2 Update multiplier by $\omega_{n+1} \leftarrow \omega_n + \lambda(Bu_{n+1} - p_{n+1})$.

Gabay highlights that this is a variant of Uzawa's algorithm [12] for the augmented Lagrangian

$$\mathfrak{L}_r(v, q, u) = F(q) + G(v) + \langle \mu, Bv - q\rangle_H + \frac{\lambda}{2}\|Bv - q\|^2_H$$

for solving the optimization problem (A-2). When $A = \partial G$, under qualification conditions, $A^t_B = \partial(G^* \circ (-B^t))$ and so

$$\mathbf{d} := \inf_{\mu\in H}\{G^*(-B^t\mu) + F^*(\mu)\} \tag{A-4}$$

is the dual value associated with the primal value (A-2) (see, for example, [46, Theorem 3.3.5]). Thus the Lagrangian method of Uzawa applied to finding $\mathbf{p}$ (A-2) is equivalent to DR applied to finding $\mathbf{d}$ (A-4).

https://doi.org/10.1017/S1446788719000570 Published online by Cambridge University Press

References

[1] S. Alwadani, H. H. Bauschke, W. M. Moursi and X. Wang, 'On the asymptotic behaviour of the Aragón Artacho–Campoy algorithm', *Oper. Res. Let.* **46**(6) (2018), 585–587.

[2] F. J. A. Artacho and J. M. Borwein, 'Global convergence of a non-convex Douglas–Rachford iteration', *J. Global Optim.* **57**(3) (2013), 753–769.

[3] F. J. A. Artacho, J. M. Borwein and M. K. Tam, 'Recent results on Douglas–Rachford methods', *Serdica Math. J.* **39** (2013), 313–330.

[4] F. J. A. Artacho, J. M. Borwein and M. K. Tam, 'Douglas–Rachford feasibility methods for matrix completion problems', *ANZIAM J.* **55**(4) (2014), 299–326.

[5] F. J. A. Artacho, J. M. Borwein and M. K. Tam, 'Recent results on Douglas–Rachford methods for combinatorial optimization problems', *J. Optim. Theory Appl.* **163**(1) (2014), 1–30.

[6] F. J. A. Artacho, J. M. Borwein and M. K. Tam, 'Global behavior of the Douglas–Rachford method for a nonconvex feasibility problem', *J. Global Optim.* **65**(2) (2016), 309–327.

[7] F. J. A Artacho and R. Campoy, 'Solving graph coloring problems with the Douglas–Rachford algorithm', *Set-Valued Var. Anal.* **26**(2) (2018), 277–304.

[8] F. J. A. Artacho and R. Campoy, 'Computing the resolvent of the sum of maximally monotone operators with the averaged alternating modified reflections algorithm', *J. Optim. Theory Appl.* **181**(3) (2019), 709–726.

[9] F. J. A. Artacho and R. Campoy, 'A new projection method for finding the closest point in the intersection of convex sets', *Comput. Optim. Appl.* **69**(1) (2018), 99–132.

[10] F. J. A. Artacho, R. Campoy, I. Kotsireas and M. K. Tam, 'A feasibility approach for constructing combinatorial designs of circulant type', *J. Combin. Optim.* **35**(4) (2018), 1061–1085.

[11] F. J. A. Artacho, Y. Censor and A. Gibali, 'The cyclic Douglas–Rachford algorithm with r-sets-Douglas–Rachford operators', *Optim. Methods Softw.* **34**(4) (2019), 875–889.

[12] K. Arrow, L. Hurwicz and H. Uzawa, 'Studies in nonlinear programming', (Cambridge University Press, 1958).

[13] F. J. Artacho, R. Campoy and V. Elser, 'An enhanced formulation for solving graph coloring problems with the Douglas–Rachford algorithm', Preprint, (2018), arXiv:1808.01022.

[14] H. Attouch, 'On the maximality of the sum of two maximal monotone operators', Technical report, Wisconsin University, Madison Mathematics Research Center, (1979).

[15] P. Bansal, 'Code for solving Tetravex using Douglas–Rachford algorithm', (2010), available at https://people.ok.ubc.ca/bauschke/Pulkit/pulkitreport.pdf.

[16] H. H. Bauschke, J. Y. B. Cruz, T. T. A. Nghia, H. M. Phan and X. Wang, 'The rate of linear convergence of the Douglas–Rachford algorithm for subspaces is the cosine of the Friedrichs angle', *J. Approx. Theory* **185** (2014), 63–79.

[17] H. H. Bauschke and J. M. Borwein, 'On projection algorithms for solving convex feasibility problems', *SIAM Rev.* **38**(3) (1996), 367–426.

[18] H. H. Bauschke, R. I. Boţ, W. L. Hare and W. M. Moursi, 'Attouch–Théra duality revisited: paramonotonicity and operator splitting', *J. Approx. Theory* **164**(8) (2012), 1065–1084.

[19] H. H. Bauschke, J. Burke, F. Deutsch, H. Hundal and J. Vanderwerff, 'A new proximal point iteration that converges weakly but not in norm', *Proc. Amer. Math. Soc.* **133**(6) (2005), 1829–1835.

[20] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 1st edn, CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC (Springer, Cham, 2011).

[21] H. H. Bauschke, P. L. Combettes and D. R. Luke, 'Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization', *J. Opt. Soc. Amer. A* **19**(7) (2002), 1334–1345.

[22] H. H. Bauschke, P. L. Combettes and D. R. Luke, 'Hybrid projection–reflection method for phase retrieval', *JOSA A* **20**(6) (2003), 1025–1034.

https://doi.org/10.1017/S1446788719000570 Published online by Cambridge University Press

[23] H. H. Bauschke, P. L. Combettes and D. R. Luke, 'Finding best approximation pairs relative to two closed convex sets in Hilbert spaces', *J. Approx. Theory* **127**(2) (2004), 178–192.
[24] H. H. Bauschke, J. Y. B. Cruz, T. T. A. Nghia, H. M. Pha and X. Wang, 'Optimal rates of linear convergence of relaxed alternating projections and generalized Douglas–Rachford methods for two subspaces', *Numer. Algorithms* **73**(1) (2016), 33–76.
[25] H. H. Bauschke and M. N. Dao, 'On the finite convergence of the Douglas–Rachford algorithm for solving (not necessarily convex) feasibility problems in Euclidean spaces', *SIAM J. Optim.* **27**(1) (2017), 507–537.
[26] H. H. Bauschke, M. N. Dao and S. B. Lindstrom, 'The Douglas–Rachford algorithm for a hyperplane and a doubleton', *J. Glob. Optim.* **74**(1) (2019), 79–93.
[27] H. H. Bauschke, M. N. Dao and W. M. Moursi, 'On Fejér monotone sequences and nonexpansive mappings', Preprint, (2015), arXiv:1507.05585.
[28] H. H. Bauschke, M. N. Dao and W. M. Moursi, 'The Douglas–Rachford algorithm in the affine-convex case', *Oper. Res. Lett.* **44**(3) (2016), 379–382.
[29] H. H. Bauschke, M. N. Dao, D. Noll and H. M. Phan, 'On Slaters condition and finite convergence of the Douglas–Rachford algorithm for solving convex feasibility problems in Euclidean spaces', *J. Global Optim.* **65**(2) (2016), 329–349.
[30] H. H. Bauschke, M. N. Dao, D. Noll and H. M. Phan, 'Proximal point algorithm, Douglas–Rachford algorithm and alternating projections: a case study', *J. Convex Anal.* **23**(1) (2016), 237–261.
[31] H. H. Bauschke, W. L. Hare and W. M. Moursi, 'Generalized solutions for the sum of two maximally monotone operators', *SIAM J. Control Optim.* **52**(2) (2014), 1034–1047.
[32] H. H. Bauschke, W. L. Hare and W. M. Moursi, 'On the range of the Douglas–Rachford operator', *Math. Oper. Res.* **41**(3) (2016), 884–897.
[33] H. H. Bauschke, V. R. Koch and H. M. Phan, 'Stadium norm and Douglas–Rachford splitting: a new approach to road design optimization', *Oper. Res.* **64**(1) (2015), 201–218.
[34] H. H. Bauschke, B. Lukens and W. M. Moursi, 'Affine nonexpansive operators, Attouch–Théra duality and the Douglas–Rachford algorithm', *Set-Valued Var. Anal.* **25**(3) (2017), 481–505.
[35] H. H. Bauschke and W. M. Moursi, 'The Douglas–Rachford algorithm for two (not necessarily intersecting) affine subspaces', *SIAM J. Optim.* **26**(2) (2016), 968–985.
[36] H. H. Bauschke and W. M. Moursi, 'On the order of the operators in the Douglas–Rachford algorithm', *Opt. Lett.* **10**(3) (2016), 447–455.
[37] H. H. Bauschke and W. M. Moursi, 'On the Douglas–Rachford algorithm', *Math. Program.* **164**(1-2) (2017), 263–284.
[38] H. H. Bauschke and D. Noll, 'On the local convergence of the Douglas–Rachford algorithm', *Arch. Math. (Basel)* **102**(6) (2014), 589–600.
[39] H. H. Bauschke, D. Noll and H. M. Phan, 'Linear and strong convergence of algorithms involving averaged nonexpansive operators', *J. Math. Anal. Appl.* **421**(1) (2015), 1–20.
[40] H. H. Bauschke, J. Schaad and X. Wang, 'On Douglas–Rachford operators that fail to be proximal mappings', *Math. Program.* **168**(1-2) (2018), 55–61.
[41] R. Behling, J. Y. Bello-Cruz and L.-R. Santos, 'On the linear convergence of the circumcentered-reflection method', *Oper. Res. Lett.* **46**(2) (2018), 159–162.
[42] R. Behling, J. Y. Bello-Cruz and L.-R. Santos, 'Circumcentering the Douglas–Rachford method', *Numer. Algorithms* **78** (2018), 759–776.
[43] J. Benoist, 'The Douglas–Rachford algorithm for the case of the sphere and the line', *J. Glob. Optim.* **63** (2015), 363–380.
[44] D. P. Bertsekas, *Convex Optimization Algorithms* (Athena Scientific, Belmont, MA, 2015).
[45] J. M. Borwein and O. Giladi, 'Ergodic behaviour of a Douglas–Rachford operator away from the origin', Preprint, (2017), arXiv:1708.09068.
[46] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization: Theory and Examples*, 2nd edition (Springer, New York, 2006).
[47] J. M. Borwein, G. Li and M. K. Tam, 'Convergence rate analysis for averaged fixed point iterations in common fixed point problems', *SIAM J. Optim.* **27**(1) (2017), 1–33.

[48] J. M. Borwein, S. B. Lindstrom, B. Sims, M. Skerritt and A. Schneider, 'Dynamics of the Douglas–Rachford method for ellipses and p-spheres', *Set-Valued Anal.* **26**(2) (2018), 385–403.

[49] J. M. Borwein and B. Sims, 'The Douglas–Rachford algorithm in the absence of convexity', in: *Fixed Point Algorithms for Inverse Problems in Science and Engineering*, Springer Optimization and Its Applications, 49 (eds. H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke and H. Wolkowicz) (Springer, Science and Business Media, 2011), 93–109.

[50] J. M. Borwein, B. Sims and M. K. Tam, 'Norm convergence of realistic projection and reflection methods', *Optimization* **64**(1) (2015), 161–178.

[51] J. M. Borwein and M. K. Tam, 'Reflection methods for inverse problems with applications to protein conformation determination', in: *Springer Volume on the CIMPA School Generalized Nash Equilibrium Problems, Bilevel Programming and MPEC, New Delhi, India* (Springer, Singapore, 2017), 83–100.

[52] J. M. Borwein and M. K. Tam, 'A cyclic Douglas–Rachford iteration scheme', *J. Optim. Theory Appl.* **160** (2014), 1–29.

[53] R. I. Boţ, E. R. Csetnek and C. Hendrich, 'Inertial Douglas–Rachford splitting for monotone inclusion problems', *Appl. Math. Comput.* **256** (2015), 472–487.

[54] R. I. Boţ and C. Hendrich, 'A Douglas–Rachford type primal-dual method for solving inclusions with mixtures of composite and parallel-sum type monotone operators', *SIAM J. Optim.* **23**(4) (2013), 2541–2565.

[55] L. M. Bregman, 'The method of successive projection for finding a common point of convex sets', *Sov. Math. Dok.* **162**(3) (1965), 688–692.

[56] H. Brezis, *Operateurs Maximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert*, North-Holland Mathematics Studies, 5 (Elsevier, 1973).

[57] P. L. T. Brian, 'A finite-difference method of high-order accuracy for the solution of three-dimensional transient heat conduction problems', *AIChE J.* **7**(3) (1961), 367–370.

[58] L. M. Briceño-Arias, 'Forward-Douglas–Rachford splitting and forward-partial inverse method for solving monotone inclusions', *Optimization* **64**(5) (2015), 1239–1261.

[59] A. Cegielski, *Iterative Methods for Fixed Point Problems in Hilbert Spaces*, Lecture Notes in Mathematics, 2057 (Springer, Heidelberg, 2012).

[60] Y. Censor and R. Mansour, 'New Douglas–Rachford algorithmic structures and their convergence analyses', *SIAM J. Optim.* **26**(1) (2016), 474–487.

[61] T. F. C. Chan and R. Glowinski, *Finite Element Approximation and Iterative Solution of a Class of Mildly Non-Linear Elliptic Equations* (Computer Science Department, Stanford University, Stanford, 1978).

[62] Cinderella (software). Available at https://cinderella.de/tiki-index.php, (2016).

[63] P. L. Combettes, 'Solving monotone inclusions via compositions of nonexpansive averaged operators', *Optimization* **53**(5-6) (2004), 475–504.

[64] P. L. Combettes and J.-C. Pesquet, 'A Douglas–Rachford splitting approach to nonsmooth convex variational signal recovery', *IEEE J. Sel. Top. Signal Process.* **1**(4) (2007), 564–574.

[65] P. L. Combettes and J.-C. Pesquet, 'Proximal splitting methods in signal processing', in: *Fixed-Point Algorithms for Inverse Problems in Science and Engineering* (Springer, New York, 2011), 185–212.

[66] P. L. Combettes and J.-C. Pesquet, 'Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping', *SIAM J. Optim.* **25**(2) (2015), 1221–1248.

[67] M. N. Dao and H. M. Phan, 'Linear convergence of projection algorithms', *Math. Oper. Res.* **44**(2) (2019), 715–738.

[68] M. N. Dao and H. M. Phan, 'Adaptive Douglas–Rachford splitting algorithm for the sum of two operators', *SIAM J. Optim.* **29**(4) (2019), 2697–2724.

[69] M. N. Dao and Matthew .K. Tam, 'A Lyapunov-type approach to convergence of the Douglas–Rachford algorithm', *J. Global Optim.* **73**(1) (2019), 83–112.

[70] D. Davis and W. Yin, 'Convergence rate analysis of several splitting schemes', in: *Splitting Methods in Communication, Imaging, Science, and Engineering* (Springer, Cham, 2016), 115–163.

https://doi.org/10.1017/S1446788719000570 Published online by Cambridge University Press

[71] A. R. D. Pierro, 'From parallel to sequential projection methods and vice versa in convex feasibility: results and conjectures', in: *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications (Haifa, 2000)*, Studies in Computational Mathematics, 8 (North-Holland, Amsterdam, 2001), 187–201.
[72] L. Demanet and X. Zhang, 'Eventual linear convergence of the Douglas–Rachford iteration for basis pursuit', *Math. Comput.* **85**(297) (2016), 209–238.
[73] R. Díaz Millán, Scott B. Lindstrom and Vera Roshchina, 'Comparing averaged relaxed cutters and projection methods: theory and examples', *From Analysis to Visualization: A Celebration of the Life and Legacy of Jonathan M. Borwein, Callaghan, Australia, September 2017*, Springer Proceedings in Mathematics and Statistics (2020), to appear.
[74] J. Douglas Jr. and H. H. Rachford Jr, 'On the numerical solution of heat conduction problems in two and three space variables', *Trans. Amer. Math. Soc.* **82** (1956), 421–439.
[75] D. Drusvyatskiy, A. D. Ioffe and A. S. Lewis, 'Alternating projections and coupling slope', Preprint, (2014), arXiv:1401.7569, 1–17.
[76] J. Eckstein, 'Splitting methods for monotone operators with applications to parallel optimization', PhD Thesis, Massachusetts Institute of Technology, (1989).
[77] J. Eckstein and D. P. Bertsekas, 'On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators', *Math. Prog.* **55**(3) (1992), 293–318.
[78] J. Eckstein and B. F. Svaiter, 'A family of projective splitting methods for the sum of two maximal monotone operators', *Math. Program.* **111**(1-2) (2008), 173–199.
[79] J. Eckstein and B. F. Svaiter, 'General projective splitting methods for sums of maximal monotone operators', *SIAM J. Control Optim.* **48**(2) (2009), 787–811.
[80] J. Eckstein and W. Yao, 'Understanding the convergence of the alternating direction method of multipliers: theoretical and computational perspectives', *Pacific J. Optim.* **11**(4) (2015), 619–644.
[81] V. Elser, Private communication.
[82] V. Elser, 'Phase retrieval by iterated projections', *JOSA A* **20**(1) (2003), 40–55.
[83] V. Elser, 'Matrix product constraints by projection methods', *J. Global Optim.* **68**(2) (2017), 329–355.
[84] V. Elser, 'The complexity of bit retrieval', *IEEE Trans. Inform. Theory* **64**(1) (2018), 412–428.
[85] V. Elser, T.-Y. Lan and T. Bendory, 'Benchmark problems for phase retrieval', *SIAM J. Imag. Sci.* **11**(4) (2018), 2429–2455.
[86] V. Elser, I. Rankenburg and P. Thibault, 'Searching with iterated maps', *Proc. Natl. Acad. Sci. USA* **104**(2) (2007), 418–423.
[87] M. Fält and P. Giselsson, 'Optimal convergence rates for generalized alternating projections', in: *2017 IEEE 56th Annual Conference on Decision and Control (CDC)* (IEEE, 2017), 2268–2274.
[88] J. R. Fienup, 'Phase retrieval algorithms: a comparison', *Appl. Opt.* **21**(15) (1982), 2758–2769.
[89] M. Fortin and R. Glowinski, *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, Vol. 15 (Elsevier, 2000).
[90] D. J. Franklin, Private communication.
[91] D. J. Franklin, 'Projection algorithms for non-separable wavelets and Clifford Fourier analysis', PhD Thesis, University of Newcastle, (2018).
[92] M. Fukushima, 'A relaxed projection method for variational inequalities', *Math. Program.* **35**(1) (1986), 58–70.
[93] M. Fukushima, 'The primal Douglas–Rachford splitting algorithm for a class of monotone mappings with application to the traffic equilibrium problem', *Math. Program.* **72**(1) (1996), 1–15.
[94] D. Gabay, 'Applications of the method of multipliers to variational inequalities', in: *Studies in Mathematics and its Applications,* Vol. 15, Ch. ix (Elsevier, 1983), 299–331.
[95] D. Gabay and B. Mercier, 'A dual algorithm for the solution of nonlinear variational problems via finite element approximation', *Comput. Math. Appl.* **2**(1) (1976), 17–40.

[96] P. Giselsson, 'Tight global linear convergence rate bounds for Douglas–Rachford splitting', *J. Fixed Point Theory Appl.* **19**(4) (2017), 2241–2270.

[97] P. Giselsson and S. Boyd, 'Diagonal scaling in Douglas–Rachford splitting and ADMM', in: *53rd IEEE Conference on Decision and Control, Los Angeles, CA, 15–17 December, 2014* (IEEE, 2014), 5033–5039.

[98] P. Giselsson and S. Boyd, 'Linear convergence and metric selection for Douglas–Rachford splitting and ADMM', *IEEE Trans. Automat. Control* **62**(2) (2017), 532–544.

[99] R. Glowinski, 'On alternating direction methods of multipliers: a historical perspective', in: *Modeling, Simulation and Optimization for Science and Technology* (Springer, Dordrecht, 2014), 59–82.

[100] R. Glowinski and A. Marroco, 'Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires', *Revue française d'automatique, informatique, recherche opérationnelle. Analyse numérique* **9**(R2) (1975), 41–76.

[101] R. Glowinski, S. J. Osher and W. Yin, *Splitting Methods in Communication, Imaging, Science, and Engineering* (Springer, 2017).

[102] T. Goldstein and S. Osher, 'The split Bregman method for L1-regularized problems', *SIAM J. Imag. Sci.* **2**(2) (2009), 323–343.

[103] S. Gravel and V. Elser, 'Divide and concur: a general approach to constraint satisfaction', *Phys. Rev. E* **78**(3) (2008), 036706.

[104] C. Grussler and P. Giselsson, 'Local convergence of proximal splitting methods for rank constrained problems', in: *56th IEEE Conference on Decision and Control, Melboune, 12–15 December, 2017* (IEEE, 2017), 702–708.

[105] B. He and X. Yuan, 'On the O(1/n) convergence rate of the Douglas–Rachford alternating direction method', *SIAM J. Numer. Anal.* **50**(2) (2012), 700–709.

[106] B. He and X. Yuan, 'On non-ergodic convergence rate of Douglas–Rachford alternating direction method of multipliers', *Numer. Math.* **130**(3) (2015), 567–577.

[107] B. He and X. Yuan, 'On the convergence rate of Douglas–Rachford operator splitting method', *Math. Program.* **153**(2) (2015), 715–722.

[108] R. Hesse and D. R. Luke, 'Nonconvex notions of regularity and convergence of fundamental algorithms for feasibility problems', *SIAM J. Optim.* **23**(4) (2013), 2397–2419.

[109] R. Hesse, D. R. Luke and P. Neumann, 'Alternating projections and Douglas–Rachford for sparse affine feasibility', *IEEE Trans. Signal Process.* **62**(18) (2014), 4868–4881.

[110] H. S. Hundal, 'An alternating projection that does not converge in norm', *Nonlinear Anal.: Theory Methods Appl.* **57**(1) (2004), 35–61.

[111] A. Y. Kruger, 'About regularity of collections of sets', *Set-Valued Anal.* **14**(2) (2006), 187–206.

[112] A. Y. Kruger, 'About intrinsic transversality of pairs of sets', *Set-Valued Var. Anal.* **26**(1) (2018), 111–142.

[113] B. P. Lamichhane, S. B. Lindstrom and B. Sims, 'Application of projection algorithms to differential equations: boundary value problems', *ANZIAM J.* **61**(1) (2019), 23–46.

[114] A. S. Lewis, D. R. Luke and J. Malick, 'Local linear convergence for alternating and averaged nonconvex projections', *Found. Comput. Math.* **9**(4) (2009), 485–513.

[115] G. Li and T. K. Pong, 'Douglas–Rachford splitting for nonconvex optimization with application to nonconvex feasibility problems', *Math. Program.* **159**(1-2, Ser. A) (2016), 371–401.

[116] S. B. Lindstrom, B. Sims and M. P. Skerritt, 'Computing intersections of implicitly specified plane curves', *Nonlinear Convex. Anal.* **18**(3) (2017), 347–359.

[117] P.-L. Lions and B. Mercier, 'Splitting algorithms for the sum of two nonlinear operators', *SIAM J. Numer. Anal.* **16**(6) (1979), 964–979.

[118] D. R. Luke, 'Phase retrieval, what's new', *SIAG/OPT Views and News* **25**(1) (2017), 1–5.

[119] E. Matoušková and S. Reich, 'The Hundal example revisited', *J. Nonlinear Convex Anal.* **4**(3) (2003), 411–427.

[120] W. M. Moursi, 'The Douglas–Rachford operator in the possibly inconsistent case: static properties and dynamic behaviour', PhD Thesis, University of British Columbia, (2016).

https://doi.org/10.1017/S1446788719000570 Published online by Cambridge University Press

[121] W. M. Moursi and L. Vandenberghe, ‘Douglas–Rachford splitting for a Lipschitz continuous and a strongly monotone operator’, Preprint, (2018), arXiv:1805.09396.
[122] W. M. Moursi and Y. Zinchenko, ‘A note on the equivalence of operator splitting methods’, in: *Splitting Algorithms, Modern Operator Theory, and Applications* (Springer, 2019), 331–349.
[123] T. Needham, *Visual Complex Analysis* (Oxford University Press, 1997).
[124] P. Patrinos, L. Stella and A. Bemporad, ‘Douglas–Rachford splitting: Complexity estimates and accelerated variants’, in: *53rd IEEE Conference on Decision and Control, Los Angeles, CA, 15–17 December, 2014* (IEEE, 2014), 4234–4239.
[125] D. W. Peaceman and H. H. Rachford Jr, ‘The numerical solution of parabolic and elliptic differential equations’, *J. Soc. Ind. Appl. Math.* **3**(1) (1955), 28–41.
[126] H. M. Phan, ‘Linear convergence of the Douglas–Rachford method for two closed sets’, *Optimization* **65**(2) (2016), 369–385.
[127] G. Pierra, ‘Decomposition through formalization in a product space’, *Math. Program.* **28**(1) (1984), 96–115.
[128] R. T. Rockafellar, *Conjugate Duality and Optimization*, Regional Conference Series in Applied Mathematics, 16 (SIAM, 1976).
[129] R. T. Rockafellar, ‘Monotone operators and the proximal point algorithm’, *SIAM J. Control Optim.* **14**(5) (1976), 877–898.
[130] J. Schaad, ‘Modeling the 8-queens problem and sudoku using an algorithm based on projections onto nonconvex sets’, PhD Thesis, University of British Columbia, (2010).
[131] S. Setzer, ‘Split Bregman algorithm, Douglas–Rachford splitting and frame shrinkage’, in: *International Conference on Scale Space and Variational Methods in Computer Vision* (Springer, Berlin–Heidelberg, 2009), 464–476.
[132] J. E. Spingarn, ‘Partial inverse of a monotone operator’, *Appl. Math. Optim.* **10**(1) (1983), 247–265.
[133] G. Steidl and T. Teuber, ‘Removing multiplicative noise by Douglas–Rachford splitting methods’, *J. Math. Imag. Vis.* **36**(2) (2010), 168–184.
[134] B. F. Svaiter, ‘On weak convergence of the Douglas–Rachford method’, *SIAM J. Control Opt.* **49**(1) (2011), 280–287.
[135] B. F. Svaiter, ‘A weakly convergent fully inexact Douglas–Rachford method with relative error tolerance’, *ESAIM: COCV* **25** (2019), 57 pages.
[136] M. K. Tam, ‘Iterative projection and reflection methods: theory and practice’, PhD Thesis, University of Newcastle, (2016).
[137] A. Themelis and P. Patrinos, ‘Douglas–Rachford splitting and ADMM for nonconvex optimization: tight convergence results’, (2018).
[138] J. von Neumann, *Functional Operators*, Vol. II (Oxford University Press, 1950), a reprint of mimeographed notes first distributed in 1933.
[139] F. Wang, G. Reid and H. Wolkowicz, ‘Finding maximum rank moment matrices by facial reduction on primal form and Douglas–Rachford iteration’, *ACM Commun. Comput. Algebra* **51**(1) (2017), 35–37.
[140] E. H. Zarantonello, ‘Projections on convex sets in Hilbert space and spectral theory: Part I. Projections on convex sets: Part II. Spectral theory.’, in: *Contributions to Nonlinear Functional Analysis* (Academic Press, 1971), 237–424.

SCOTT B. LINDSTROM, CARMA, University of Newcastle, Australia
e-mail: scott.b.lindstrom@polyu.edu.hk

BRAILEY SIMS, CARMA, University of Newcastle, Australia