The Douglas–Rachford Algorithm for Convex and Nonconvex Feasibility Problems

Francisco J. Aragón Artacho[*]   Rubén Campoy[†]   Matthew K. Tam[‡]

April 22, 2019

Abstract

The Douglas–Rachford method, a projection algorithm designed to solve continuous optimization problems, forms the basis of a useful heuristic for solving combinatorial optimization problems. In order to successfully use the method, it is necessary to formulate the problem at hand as a feasibility problem with constraint sets having efficiently computable nearest points. In this self-contained tutorial, we develop the convergence theory of projection algorithms within the framework of fixed point iterations, explain how to devise useful feasibility problem formulations, and demonstrate the application of the Douglas–Rachford method to said formulations. The paradigm is then illustrated on two concrete problems: a generalization of the "eight queens puzzle" known as the "(m,n)-queens problem", and the problem of constructing a probability distribution with prescribed moments.

Keywords Projection methods · Douglas–Rachford · Feasibility problem · Eight queens problem

MSC2010 65K05 · 90C27 · 90-01 · 65-01

1   Introduction

The so-called *feasibility problem* asks for a point contained in the intersection of a finite collection of constraint sets. Precisely, given a family of sets $C_1, C_2, \dots, C_r$ contained in a Euclidean space E, the corresponding feasibility problem takes the form

$$\text{Find } x \in C := \bigcap_{i=1}^{r} C_i. \tag{1}$$

A feasibility problem is said to be *consistent* when it has solution (i.e. when $\cap_{i=1}^{r} C_i \neq \emptyset$), otherwise it is said to be *inconsistent*. This seemingly simple problem provides a modeling framework with great flexibility and power. For the purpose of numerical schemes however, devising computationally tractable formulations is often a nontrivial task and some creativity is required.

In many situations of practical interest, it is difficult to find a point in C directly. On the other hand, the individual constraint sets, C_i, can often be chosen to have relatively simple structure. *Projection methods* are a family of iterative algorithms for solving (1) which aim to exploit this observation with a solution to the problem being obtained in the limit. In this context, "simple" is understood in the sense of easy-to-compute *projection mappings*.

Definition 1 (Projection mapping). *Given a nonempty subset $C \subseteq E$, the* projection mapping *(or projector) onto C is the possibly set-valued operator, $P_C : E \rightrightarrows C$, defined at each $x \in E$ by*

$$P_C(x) := \left\{ p \in C : \|x - p\| = d_C(x) := \inf_{c \in C} \|c - x\| \right\}.$$

---

[*]Department of Mathematics, University of Alicante, Spain. Email: francisco.aragon@ua.es
[†]Department of Mathematics, University of Alicante, Spain. Email: ruben.campoy@ua.es
[‡]Inst. for Num. and Appl. Math., University of Göttingen, Germany. Email: m.tam@math.uni-goettingen.de

It is a straightforward exercise to verify that the projection mapping P_C has nonempty values (i.e. $P_C(x) \neq \emptyset$ for all $x \in E$) when C is nonempty and closed. If $P_C(x)$ is a singleton set for all $x \in E$, then C is said to be *Chebyshev*. In this case, the projector P_C is a single-valued mapping that sends each point $x \in E$ to its unique nearest point in C. Amongst other names, the projector is sometimes also called the *projection operator*, the *metric projection*, *nearest point mapping* or the *best approximation operator*.

Exercise 1. *Let $C \subseteq E$ be a nonempty, closed set. Show that $P_C(x) \neq \emptyset$ for all $x \in E$.*

The simplest projection algorithm for solving (1) is the *method of cyclic projections* which iterates by successively applying the projectors onto each of the constraint sets. More precisely, given an initial point $x_0 \in E$, it generates a sequence $(x_k)_{k=0}^{\infty}$ according to

$$x_{k+1} \in \left(P_{C_r} P_{C_{r-1}} \cdots P_{C_1} \right)(x_k), \quad \text{for } k = 0, 1, 2, \ldots, \tag{2}$$

where we note that, for an operator T and a set C, $T(C)$ is defined by $T(C) := \cup_{x \in C} T(x)$.

In the case that the sets $C_1, \ldots, C_r$ are Chebyshev, (2) can be written with equality. However, in general, the inclusion is required due to potential set-valuedness of the projection mapping. In such cases, we will be interested in any sequence $(x_k)_{k=0}^{\infty}$ satisfying (2). Two illustrations of the method, where the limit of $(x_k)_{k=0}^{\infty}$ is contained in the intersection, are provided in Figure 1.
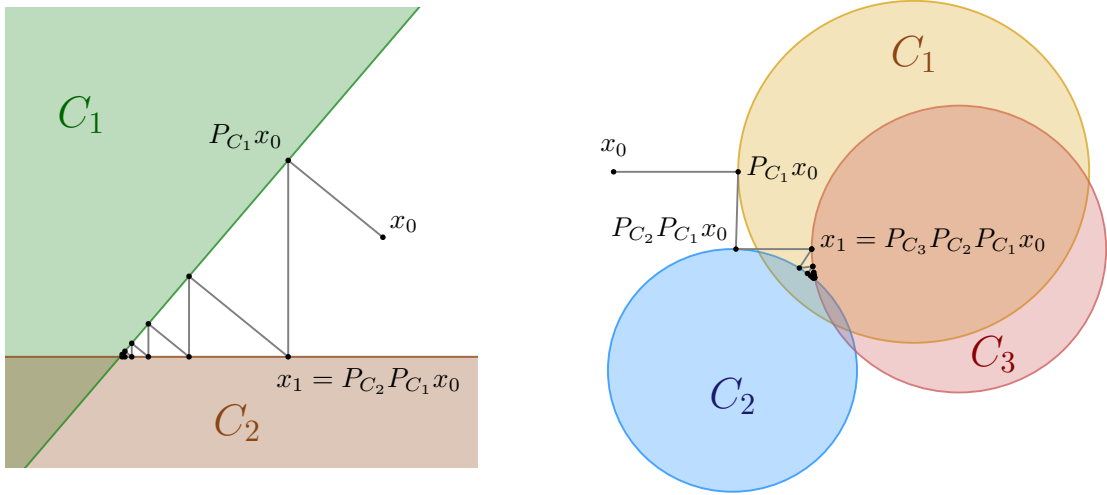
(a) The method of cyclic projections for two halfspaces   (b) The method of cyclic projections for three balls

Figure 1: The method of cyclic projections applied to families of closed and convex sets with nonempty intersections.

The origins of the method of cyclic projections can be traced back at least to the work of von Neumann (1950), who proved convergence of (2) when $r = 2$ and the sets are subspaces. Halperin (1962) subsequently showed that the result remains true for $r > 2$. Bregman (1965) established convergence of the method when the constraints are closed and convex sets.

The following familiar example illustrates the potential advantages of projection methods and the feasibility problem paradigm.

Example 1 (Linear systems). *Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map and $b \in \mathbb{R}^m$. Consider the problem of solving the linear equation $Ax = b$. By letting a_i^T denote the ith row of the matrix A, this can be expressed as*

$$\begin{pmatrix} a_1^T x \\ \vdots \\ a_m^T x \end{pmatrix} = \begin{pmatrix} a_1^T \\ \vdots \\ a_m^T \end{pmatrix} x = Ax = b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

In other words, $Ax = b$ if and only if

$$x \in \bigcap_{i=1}^{m} C_i \quad where \quad C_i := \left\{ x \in \mathbb{R}^n : a_i^T x = b_i \right\}. \tag{3}$$

The projector onto the hyperplane C_i is given by

$$P_{C_i}(x) = x + \left(b_i - a_i^T x\right) \frac{a_i}{\|a_i\|^2}, \tag{4}$$

which can be easily computed using only vector arithmetic. The well-known Kaczmarz's method *(Kaczmarz (1937)) is precisely the method of cyclic projections applied to* (3).

One utility of viewing this problem as a feasibility problem is that additional constraints can be easily incorporated without changing the algorithm used. For instance, a nonnegative solution to the linear system $Ax = b$ can be found by augmenting (3) *with the constraint $C_{m+1} := \{x \in \mathbb{R}^n : x \geq 0\}$, whose projection is given by $P_{C_{m+1}}(x) = \max\{0, x\}$ (understood in the pointwise sense).*

Another possible variant of the problem involves considering the linear inequality system $Ax \leq b$. This too can be cast in terms of the feasibility problem (3) *with constraint sets $C_i := \{x \in \mathbb{R}^n : a_i^T x \leq b_i\}$. The projection onto these sets are given by*

$$P_{C_i}(x) = x + \min\left\{0, b_i - a_i^T x\right\} \frac{a_i}{\|a_i\|^2}, \tag{5}$$

which also only requires vector arithmetic.   ◇

This manuscript is intended to be a condensed tutorial on the Douglas–Rachford algorithm for solving convex and nonconvex feasibility problems for the unacquainted. For general reviews on projection algorithms, we refer the reader to Bauschke and Combettes (2017); Cegielski (2012); Censor (1984); Censor and Cegielski (2015); Deutsch (2001). The remainder of this tutorial is organized as follows. In Section 2, we discuss mathematical preliminaries including general properties of nonexpansive operators, fixed point iterations and projection mappings. In Section 3, we use these preliminaries to analyze the convergence behavior of numerous projection methods including the Douglas–Rachford algorithm. In Section 4, we apply the Douglas–Rachford method to solve a combinatorial problem known as the (m, n)-*queens puzzle*. For simplicity, this tutorial mainly focuses on the finite-dimensional setting, although appropriately modified, most of the results presented remain true in infinite dimensions. Indeed, the technical difficulties associated with the infinite-dimensional setting are explained in Section 5, where the problem of constructing a probability distribution with prescribed moments is also discussed. We finish with some conclusions and open questions in Section 6.

2   Preliminaries

Throughout this paper, E denotes a *Euclidean space* equipped with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\| \cdot \|$. Given a non-empty set $D \subseteq E$, $T : D \rightrightarrows E$ denotes a *set-valued operator* that maps a point in D to a subset of E (i.e. $T(x) \subseteq E$ for all $x \in D$). In the case when $T(x) = \{u\}$ for all $x \in D$, T is said to be a *single-valued mapping*, which is denoted as $T : D \rightarrow E$. In an abuse of notation, we write $T(x) = u$ whenever $T(x) = \{u\}$. The set of *fixed points* of an operator T, denoted by $\operatorname{Fix} T$, is given by

$$\operatorname{Fix} T := \{x \in E \mid x \in T(x)\}.$$

The *identity operator* is the mapping $\operatorname{Id} : E \rightarrow E$ that maps every point to itself.

2.1   Convergence of fixed point algorithms

The mathematical analysis of projection algorithms can be performed within the framework of *fixed-point theory*. In this section, we establish general results concerning convergence of *Banach–Picard iterations*, which we will later specialize to projection algorithms.

We take the following abstract representation of an iterative algorithm: Given a mapping $T : E \rightrightarrows E$ and a point $x_0 \in E$, we consider the scheme defined by

$$x_{k+1} \in T(x_k), \quad \text{for } k = 0, 1, 2, \ldots. \tag{6}$$

We refer to (6) as the *fixed point iteration* or *Banach–Picard iteration* defined by T. Within this setting, the operator T is an abstract representation of the algorithm and its fixed points are assumed to provide solutions to the problem at hand. A classical result due to Banach (1922), known as *Banach's contraction principle*, is perhaps the most well-known criteria for ensuring convergence of (6).

Theorem 1 (Banach's contraction principle). *Let $T \colon E \to E$ be a contraction mapping, i.e., there exists $\kappa \in [0, 1[$ such that*

$$\|T(x) - T(y)\| \le \kappa \|x - y\| \quad \forall x, y \in E. \tag{7}$$

Then T has exactly one fixed point $x^\star \in E$. Furthermore, for any $x_0 \in X$ the sequence $x_{k+1} = T(x_k)$ converges to $x^\star$ with the linear rate κ; i.e., $\|x_k - x^\star\| \le \kappa^k \|x_0 - x^\star\|$ for all $k = 0, 1, 2, \ldots$.

For the purposes of projection algorithms for feasibility problem, Banach's contraction principle is too restricted. We will rarely be fortunate enough for the operator T to be contractive and it will usually be the case that our problem has more than one solution (i.e., $\operatorname{Fix} T$ is not singleton). On the other hand, nonexpansivity of T (i.e., (7) with $\kappa = 1$) is also not sufficient to ensure convergence of its Banach–Picard iteration. For a simple example, consider the mapping $T = -\operatorname{Id}$ and any $x_0 \ne 0$. Therefore, notions stronger than nonexpansivity but weaker than contractivity are required. In what follows, we explore such notions.

Definition 2 (Notions of nonexpansiveness). *Let D be a nonempty subset of E and let $T : D \to E$. The operator T is said to be*

 (i) nonexpansive *if*

$$\|T(x) - T(y)\| \le \|x - y\|, \quad \forall x, y \in D;$$

 (ii) firmly nonexpansive *if*

$$\|T(x) - T(y)\|^2 + \|(\operatorname{Id} - T)(x) - (\operatorname{Id} - T)(y)\|^2 \le \|x - y\|^2, \quad \forall x, y \in D,$$

 or, equivalently,

$$\langle x - y, T(x) - T(y) \rangle \ge \|T(x) - T(y)\|^2, \quad \forall x, y \in D; \tag{8}$$

 (iii) α-averaged *for $\alpha \in \,]0, 1[$ if there exists a nonexpansive operator $R : D \to E$ such that*

$$T = (1 - \alpha)\operatorname{Id} + \alpha R.$$

It follows immediately from the respective definitions that both firm nonexpansiveness and α-averagedness imply nonexpansiveness, while the converse implications are not true in general. Moreover, it also follows that if T is α-averaged, then it is also α'-averaged for all $\alpha' \in [\alpha, 1[$. Further relationships between the notions are shown in the following proposition. For more, see, e.g., (Bauschke and Combettes, 2017, Chapter 4).

Proposition 1. *Let $D \subseteq E$ be nonempty and let $T : D \to E$. The following hold:*

 (i) If T is α-averaged for $\alpha \in \,]0, 1[$, then

$$\|T(x) - T(y)\|^2 \le \|x - y\|^2 - \frac{1 - \alpha}{\alpha}\|(\operatorname{Id} - T)(x) - (\operatorname{Id} - T)(y)\|^2, \quad \forall x, y, \in D.$$

 (ii) T is firmly nonexpansive if and only if $2T - \operatorname{Id}$ is nonexpansive.

 (iii) If T is α-averaged with $\alpha \in \,]0, \frac{1}{2}]$, then T is firmly nonexpansive.

Proof. (i) By definition, there exists a nonexpansive operator $R : D \to E$ such that $T = (1 - \alpha)\operatorname{Id} + \alpha R$. We therefore have that

$$\|T(x) - T(y)\|^2 = (1 - \alpha)^2\|x - y\|^2 + 2\alpha(1 - \alpha)\langle x - y, R(x) - R(y) \rangle + \alpha^2\|R(x) - R(y)\|^2,$$

and that

$$\frac{1}{\alpha}\|(\operatorname{Id} - T)(x) - (\operatorname{Id} - T)(y)\|^2 = \alpha\|(\operatorname{Id} - R)(x) - (\operatorname{Id} - R)(y)\|^2$$
$$= \alpha\left(\|x - y\|^2 - 2\langle x - y, R(x) - R(y) \rangle + \|R(x) - R(y)\|^2\right).$$

By summing these two equalities together, we obtain

$$\|T(x) - T(y)\|^2 + \frac{1 - \alpha}{\alpha}\|(\operatorname{Id} - T)(x) - (\operatorname{Id} - T)(y)\|^2 = (1 - \alpha)\|x - y\|^2 + \alpha\|R(x) - R(y)\|^2.$$

The result then follows by using the nonexpansivity of R. (ii) Left as an exercise for the reader. (iii) Follows immediately from (i). $\qquad\square$

Exercise 2. *Prove Proposition 1(ii).*

Our next result provides a useful criteria for convergence of the Banach–Picard iteration in (6); namely, it suffices to assume that T is averaged and possesses a fixed point. This result is actually a special case of *Opial's Theorem* (Opial (1967)), which remains true in Banach spaces satisfying *Opial's property*.

Theorem 2 (Convergence of averaged iterations). *Let D be a nonempty, closed and convex subset of E and let $T : D \to D$ be an α-averaged operator with $\operatorname{Fix} T \neq \emptyset$. Given any $x_0 \in D$, set*

$$x_{k+1} = T(x_k), \quad for \ k = 0, 1, 2 \ldots .$$

Then $(x_k)_{k=0}^{\infty}$ converges to a point $x^\star \in \operatorname{Fix} T$.

Proof. Let $x \in \operatorname{Fix} T$. Since T is α-averaged, by Proposition 1(i), we have

$$\|x_{k+1} - x\|^2 + \frac{1 - \alpha}{\alpha} \|x_k - x_{k+1}\|^2 \leq \|x_k - x\|^2 \quad \text{for all } k = 0, 1, 2, \ldots .$$

It then follows that $(\|x_k - x\|)_{k=1}^{\infty}$ is non-increasing, $(x_k)_{k=0}^{\infty}$ is bounded and $x_k - x_{k+1} \to 0$. Now, as a bounded sequence, $(x_k)_{k=0}^{\infty}$ has a cluster point $x^\star$. Let $(x_{k_n})_{n=0}^{\infty}$ be a subsequence such that $x_{k_n} \to x^\star$. We therefore have that

$$(\operatorname{Id} - T)(x^\star) = \lim_{n \to \infty} (\operatorname{Id} - T)(x_{k_n}) = \lim_{n \to \infty} (x_{k_n} - x_{k_n + 1}) = 0,$$

which shows that $x^\star \in \operatorname{Fix} T$. Since $(\|x_k - x^\star\|)_{k=1}^{\infty}$ is non-increasing and $\|x_{k_n} - x^\star\| \to 0$, it follows that $\|x_k - x^\star\| \to 0$, which completes the proof. $\qquad\square$

Clearly, a Banach–Picard iteration (6) can only converge when the operator T possesses at least one fixed point. Nevertheless, when $\operatorname{Fix} T = \emptyset$, it is still possible to say something about the asymptotic behavior when T is averaged. In this tutorial, we prove a simplified version of the asymptotic behavior known from (Pazy (1971); Baillon et al. (1978)). In order to do so, we first recall the following *summability method*.

Lemma 1 (Cesáro summation). *Let $(x_k) \subseteq E$ be a convergent sequence with limit $x^\star \in E$. Then*

$$\frac{1}{k} \sum_{n=1}^{k} x_n \to x^\star \quad as \ \ k \to \infty.$$

Proof. Left as an exercise for the reader. $\qquad\square$

Exercise 3. *Prove Lemma 1 and give an example of a non-convergent sequence $(x_k) \subseteq E$ which is* Cesáro *summable, i.e., $\frac{1}{k} \sum_{n=1}^{k} x_n \to x^\star$ for some $x^\star \in E$. Is a Cesáro summable sequence necessarily bounded?*

The following lemma will help to streamline the proof of Theorem 3

Lemma 2. *Suppose $T : E \to E$ is an α-averaged operator for some $\alpha \in]0, 1[$. Let $x_0, y_0 \in E$ and denote $x_k = T^k(x_0)$ and $y_k = T^k(y_0)$ for $k = 1, 2, \ldots$. Then*

$$\frac{1 - \alpha}{\alpha} \sum_{k=0}^{n} \|(\operatorname{Id} - T)(x_k) - (\operatorname{Id} - T)(y_k)\|^2 \leq \|x_0 - y_0\|^2 - \|x_{n+1} - y_{n+1}\|^2, \quad for \ all \ n = 0, 1, 2, \ldots . \quad (9)$$

In particular, $(\operatorname{Id} - T)(x_k) - (\operatorname{Id} - T)(y_k) \to 0$ as $k \to \infty$.

Proof. Since T is α-averaged, Proposition 1(i) implies that

$$\|x_{k+1} - y_{k+1}\|^2 + \frac{1 - \alpha}{\alpha} \|(\operatorname{Id} - T)(x_k) - (\operatorname{Id} - T)(y_k)\|^2 \leq \|x_k - y_k\|^2, \quad \text{for all } k = 0, 1, 2, \ldots,$$

which telescopes to yield

$$\|x_{n+1} - y_{n+1}\|^2 + \frac{1 - \alpha}{\alpha} \sum_{k=0}^{n} \|(\operatorname{Id} - T)(x_k) - (\operatorname{Id} - T)(y_k)\|^2 \leq \|x_0 - y_0\|^2, \quad \text{for all } n = 0, 1, 2, \ldots .$$

The series in the previous equation therefore converges as $n \to \infty$, from which the result follows. $\qquad\square$

Theorem 3 (Asymptotic behavior of averaged iterations). *Let $T : E \to E$ be an α-averaged operator for some $\alpha \in]0, 1[$. Then*

$$\operatorname{Fix} T = \emptyset \quad \Longleftrightarrow \quad \|T^k(x)\| \to \infty \ \text{ for any } x \in E.$$

Proof. To prove the result, we establish its contrapositive: $\operatorname{Fix} T \neq \emptyset$ if and only if there exists an $x \in E$ such that $\|T^k(x)\| \not\to \infty$. To this end, first take any $x \in E$ and suppose that $\operatorname{Fix} T \neq \emptyset$. Then the sequence $T^k(x)$ is convergent by Theorem 2 and hence, in particular, also bounded.

To prove the reverse implication, suppose there exists an $x \in E$ such that $x_k = T^k(x)$ contains a bounded subsequence. The sequence (x_k) then possesses a convergent subsequence, say $x_{k_n} \to x^\star$. Since T is α-averaged, applying Lemma 2 with $x_0 = x$ and $y_0 = x^\star$ yields

$$(\operatorname{Id} -T)(x_k) - (\operatorname{Id} -T)T^k(x^\star) \to 0 \implies (\operatorname{Id} -T)T^{k_n}(x^\star) \to (\operatorname{Id} -T)(x^\star).$$

Also, applying Lemma 2 with $x_0 = x^\star$ and $y_0 = T(x^\star)$ yields

$$\frac{1-\alpha}{\alpha} \sum_{j=1}^{k} \|(\operatorname{Id} -T)T^{j-1}(x^\star) - (\operatorname{Id} -T)T^j(x^\star)\|^2 \leq \|(\operatorname{Id} -T)(x^\star)\|^2 - \|(\operatorname{Id} -T)T^k(x^\star)\|^2.$$

Taking the limit along the subsequence (k_n) gives

$$\frac{1-\alpha}{\alpha} \sum_{j=1}^{\infty} \|(\operatorname{Id} -T)T^{j-1}(x^\star) - (\operatorname{Id} -T)T^j(x^\star)\|^2 \leq \|(\operatorname{Id} -T)(x^\star)\|^2 - \|(\operatorname{Id} -T)(x^\star)\|^2 = 0,$$

which shows that all terms in the summation are identically zero. Consequently, we have

$$(\operatorname{Id} -T)T^j(x^\star) = (\operatorname{Id} -T)(x^\star) \implies (\operatorname{Id} -T)(x_k) \to (\operatorname{Id} -T)(x^\star).$$

Denoting $x_0 := x$, we observe that

$$x_0 = \sum_{n=0}^{k-1} (\operatorname{Id} -T)(x_n) + x_k \implies \frac{1}{k} x_0 = \frac{1}{k} \sum_{n=0}^{k-1} (\operatorname{Id} -T)(x_n) + \frac{1}{k} x_k.$$

Taking the limit along the subsequence x_{k_n} and using Lemma 1 gives

$$0 \cdot x_0 = (\operatorname{Id} -T)(x^*) + 0 \cdot x^\star \implies x^\star = T(x^\star).$$

In order words, $x^\star \in \operatorname{Fix} T$ which shows that $\operatorname{Fix} T \neq \emptyset$. The proof is now complete. $\qquad \square$

2.2 The projection mapping

In this section, we examine properties of the projector onto closed and convex sets.

Proposition 2. *Let $C \subseteq E$ be nonempty, closed and convex. Then C is a Chebyshev set and, for every $x \in E$, we have*

$$p = P_C(x) \quad \Longleftrightarrow \quad p \in C \ \text{ and } \ \langle c - p, x - p \rangle \leq 0 \ \text{ for all } c \in C. \tag{10}$$

Proof. Let $p, q \in P_C(x)$, which is nonempty because C is closed. By the parallelogram law and the definition of the projector, we have

$$
\begin{aligned}
\|p - q\|^2 &= \|(p - x) - (q - x)\|^2 \\
&= 2\|p - x\|^2 + 2\|q - x\|^2 - \|p + q - 2x\|^2 \\
&= 4d_C^2(x) - 4 \left\| \frac{p+q}{2} - x \right\|^2 \leq 0,
\end{aligned}
$$

where the last inequality follows because convexity of C yields $(p + q)/2 \in C$. This implies $p = q$, so we conclude that the projection must be unique.

To prove the forward implication in (10), let $x \in E$ and $p = P_C(x)$. Pick any $c \in C$, and define $z_\lambda := (1 - \lambda)p + \lambda c \in C$, for all $\lambda \in {]0, 1[}$. Then,

$$\|x - p\|^2 \leq \|x - z_\lambda\|^2 = \|x - p - \lambda(c - p)\|^2$$
$$= \|x - p\|^2 + \lambda^2\|c - p\|^2 - 2\lambda\langle x - p, c - p\rangle.$$

Since $\lambda > 0$, we deduce $\langle x - p, c - p\rangle \leq \frac{\lambda}{2}\|c - p\|^2$, from which $\langle x - p, c - p\rangle \leq 0$ follows by letting $\lambda \to 0$. The backward implication in (10) is left as an exercise for the reader. $\qquad\square$

Exercise 4. *Complete the proof of Proposition 2. That is, prove the backward implication of (10).*

Exercise 5 (Affine projectors)**.** *A non-empty set $C \subseteq E$ is an* affine subspace *if $\lambda c_1 + (1 - \lambda)c_2 \in C$ for all $\lambda \in \mathbb{R}$ and $c_1, c_2 \in C$. Show that the projection onto an affine subspace satisfies (10) with equality. Use this equality to verify the formula for the projection onto the hyperplane given in Example 1.*

In the subsequent sections, it will sometimes be easier to work with the following operator, which is defined in terms of the projector.

Definition 3 (Reflection mapping)**.** *Given a nonempty subset $C \subseteq E$, the* reflection mapping *(or* reflector*) with respect to C is the operator $R_C : E \rightrightarrows E$ given by*

$$R_C := 2P_C - \mathrm{Id}\,.$$

Each element $r \in R_C(x)$ is called a reflection *of x with respect to C.*

It is immediate from the definition that the reflector is single-valued precisely when the corresponding projector is. Illustrations of reflection mappings are shown in Figure 2.

(a) A closed and convex set C (Chebyshev). The projection and reflection of the point x are unique

(b) A closed but nonconvex set C. The projector and reflector are multivalued at x, with $P_C(x) = \{p_1, p_2\}$ and $R_C(x) = \{r_1, r_2\}$
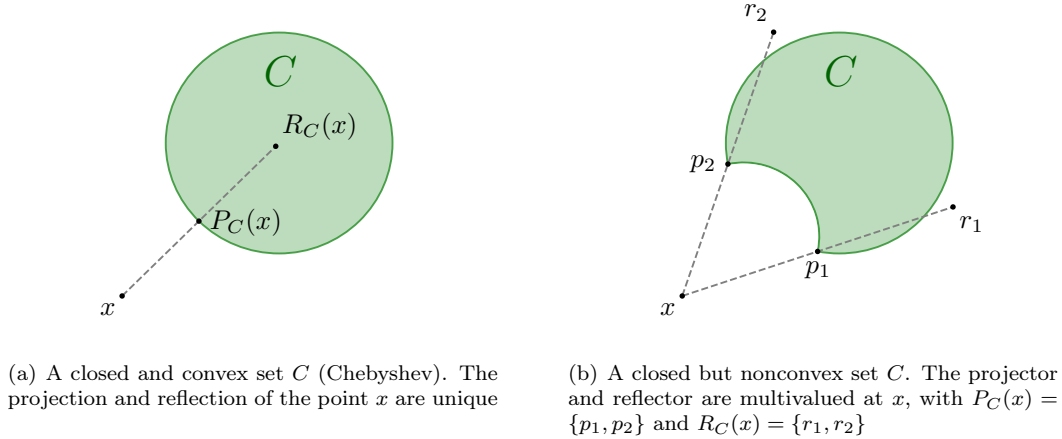
Figure 2: Examples of projectors and reflectors onto convex and nonconvex sets.

We conclude this section with the following fundamental result, which is responsible for the good behavior of projection algorithms in the convex setting.

Proposition 3. *Let $C \subseteq E$ be nonempty, closed and convex. Then the projector, P_C, is firmly nonexpansive and the reflector, R_C, is nonexpansive.*

Proof. Let $x, y \in E$. By applying Proposition 2 to $P_C(x)$ and $P_C(y)$, respectively, we obtain

$$\langle P_C(y) - P_C(x), x - P_C(x)\rangle \leq 0 \quad \text{and} \quad \langle P_C(x) - P_C(y), y - P_C(y)\rangle \leq 0.$$

The addition of these inequalities gives

$$\langle P_C(x) - P_C(y), P_C(x) - P_C(y) - (x - y)\rangle \leq 0,$$

which, by (8), implies that P_C is firmly nonexpansive. The fact that R_C is nonexpansive now follows from Proposition 1(ii). $\qquad\square$

Although projectors onto convex sets are firmly nonexpansive by Proposition 3, in general, their compositions need not be.

Example 2. *Let $E = \mathbb{R}^2$. Consider the constraint sets $C_1 := \mathbb{R} \times \{0\}$ and $C_2 := \{(\lambda, \lambda) : \lambda \in \mathbb{R}\}$. Then $P_{C_2} P_{C_1}(x, y) = P_{C_2}(x, 0) = (x, x)/2$ and hence*

$$\langle (1, -2) - (0, 0), P_{C_2} P_{C_1}(1, -2) - P_{C_2} P_{C_1}(0, 0) \rangle = -1/2.$$

Consequently, the composition $P_{C_2} P_{C_1}$ is not firmly nonexpansive, according to (8). $\Diamond$

Exercise 6 (Further examples of projectors). *By appealing to Proposition 2, show that each of the following projector formulae holds.*

(i) *(Closed balls) Suppose $r > 0$ and denote $\mathbb{B}_r = \{x \in E : \|x\| \leq r\}$. Then*

$$P_{\mathbb{B}_r}(x) = \frac{r}{\max\{\|x\|, r\}} x.$$

(ii) *(Nonnegative orthant) Suppose $E = \mathbb{R}^n$ and denote $\mathbb{R}^n_+ = \{x \in E : x_i \geq 0 \text{ for } i = 1, \ldots, n\}$. Then*

$$p = P_{\mathbb{R}^n_+}(x) \quad \text{where} \quad p_i = \max\{0, x_i\} \text{ for all } i \in \{1, \ldots, n\}.$$

(iii) *(Linear systems) Consider a matrix $A \in \mathbb{R}^{m \times n}$ such that $A^T A$ is invertible and suppose $b \in \text{range}(A) \subseteq \mathbb{R}^m$. Denote $C = \{x \in \mathbb{R}^n : Ax = b\}$, which is nonempty by assumption. Then*

$$P_C(x) = x - (A^T A)^{-1} A^T (Ax - b). \tag{11}$$

Compare the computational advantages and disadvantages of using (11) *with the feasibility problem formulation (for the same problem) outlined in Example 1.*

Exercise 7. *Let $C \subseteq E$ be a nonempty, closed set. Prove that the following formulae hold.*

(i) *(Translation formula) $P_{y+C}(x) = y + P_C(x - y)$ for all $x, y \in E$.*

(ii) *(Dilution formula) $P_{\alpha C}(x) = \alpha P_C(x/\alpha)$ for all $x \in E$ and $\alpha \in \mathbb{R} \setminus \{0\}$.*

3 Projection algorithms for solving feasibility problems

3.1 The method of cyclic projections

In this section, we prove convergence of the method of cyclic projections, discussed in Section 2, for convex feasibility problems. Given nonempty, closed and convex sets $C_1, \ldots, C_r \subseteq E$, we recall that this algorithm is the fixed point iteration defined by

$$x_{k+1} = (P_{C_r} P_{C_{r-1}} \ldots P_{C_1})(x_k), \quad \text{for } k = 0, 1, 2, \ldots.$$

In light of Theorem 2, the general recipe for proving convergence is as follows: (i) establish that the underlying fixed operator is averaged, (ii) characterize its fixed points in a *meaningful* way, and (iii) apply Theorem 2 to deduce convergence of the sequence. To this end, we begin with the following result.

Proposition 4. *Let $C_1, \ldots, C_r \subseteq E$ be nonempty, closed and convex. Then the operator $P_{C_r} P_{C_{r-1}} \cdots P_{C_1}$ is α-averaged with $\alpha = 1 - 2^{-r}$.*

Proof. For each $i \in \{1, \ldots, r\}$, the projector P_{C_i} is 1/2-averaged by Proposition 3. To prove the result it therefore suffices to establish the following: if T is 1/2-averaged and Q is $(1 - 2^{-k})$-averaged for some $k \in \{1, 2, \ldots\}$, then TQ is $(1 - 2^{-(k+1)})$-averaged. To this end, we first write

$$T = \frac{1}{2} \text{Id} + \frac{1}{2} R \quad \text{and} \quad Q = \frac{1}{2^k} \text{Id} + \left(1 - \frac{1}{2^k}\right) S,$$

for nonexpansive mappings R and S. Next, observe that

$$
\begin{aligned}
TQ = \frac{1}{2}Q + \frac{1}{2}RQ &= \frac{1}{2^{k+1}}\operatorname{Id} + \frac{2^k - 1}{2^{k+1}}S + \frac{2^k}{2^{k+1}}RQ \\
&= \frac{1}{2^{k+1}}\operatorname{Id} + \frac{2^{k+1} - 1}{2^{k+1}}\left(\frac{2^k - 1}{2^{k+1} - 1}S + \frac{2^k}{2^{k+1} - 1}RQ\right) \\
&= \frac{1}{2^{k+1}}\operatorname{Id} + \left(1 - \frac{1}{2^{k+1}}\right)\left(\frac{2^k - 1}{2^{k+1} - 1}S + \frac{2^k}{2^{k+1} - 1}RQ\right).
\end{aligned}
$$

Since S and RQ are nonexpansive, so is their convex combination. It then follows that TQ is $(1 - 2^{-(k+1)})$-averaged. $\square$

Proposition 5. *Let $C_1, C_2, \ldots, C_r \subseteq E$ be closed and convex sets with nonempty intersection. Then*

$$
\operatorname{Fix}\left(P_{C_r}P_{C_{r-1}} \cdots P_{C_1}\right) = C_1 \cap \ldots \cap C_r.
$$

Proof. The inclusion $\operatorname{Fix}\left(P_{C_r}P_{C_{r-1}} \cdots P_{C_1}\right) \supseteq \cap_{i=1}^r C_i$ is immediate, so we need only consider the other inclusion. To this end, let $x \in \operatorname{Fix}\left(P_{C_r}P_{C_{r-1}} \cdots P_{C_1}\right)$ and let $c \in \cap_{i=1}^r C_i$. Since P_{C_i} is firmly nonexpansive by Proposition 3, we have

$$
\begin{aligned}
\|x - c\|^2 &\geq \|(\operatorname{Id} - P_{C_1})x\|^2 + \|P_{C_1}(x) - c\|^2 \\
&\geq \left(\|(\operatorname{Id} - P_{C_1})x\|^2 + \|(\operatorname{Id} - P_{C_2})P_{C_1}(x)\|^2\right) + \|P_{C_2}P_{C_1}(x) - c\|^2 \\
&\geq \ddots \\
&\geq \left(\|(\operatorname{Id} - P_{C_1})x\|^2 + \cdots + \|(\operatorname{Id} - P_{C_r})(P_{C_{r-1}} \cdots P_{C_1})(x)\|^2\right) + \|(P_{C_r}P_{C_{r-1}} \cdots P_{C_1})(x) - c\|^2.
\end{aligned}
$$

Since $(P_{C_r}P_{C_{r-1}} \cdots P_{C_1})(x) = x$ by assumption, it follows that

$$
\|(\operatorname{Id} - P_{C_1})x\| = \cdots = \|(\operatorname{Id} - P_{C_r})(P_{C_{r-1}} \cdots P_{C_1})(x)\| = 0,
$$

which implies that $x \in \cap_{i=1}^r C_i$, as claimed. $\square$

Theorem 4 (Method of cyclic projections). *Let $C_1, C_2, \ldots, C_r \subseteq E$ be closed and convex sets with nonempty intersection. Given an initial point $x_0 \in E$, consider the iteration generated by*

$$
x_{k+1} = \left(P_{C_r}P_{C_{r-1}} \cdots P_{C_1}\right)(x_k), \quad \text{for } k = 0, 1, 2, \ldots.
$$

Then $(x_k)_{k=0}^\infty$ converges to a point $x^\star \in \cap_{i=1}^r C_i$.

Proof. The operator $T = P_{C_r}P_{C_{r-1}} \cdots P_{C_1}$ is $(1 - 2^{-r})$-averaged by Proposition 4 with $\operatorname{Fix} T = \cap_{i=1}^r C_i \neq \emptyset$ by Proposition 5. The result immediately follows by applying Theorem 2. $\square$

3.2 The Douglas–Rachford algorithm

In this section, we introduce another projection algorithm known as the *Douglas–Rachford (DR) algorithm*. It was originally proposed by Douglas and Rachford (1956) for solving a system of linear equations arising in heat conduction problems, and later extended to solving feasibility problems with arbitrary closed and convex sets by Lions and Mercier (1979). For further details regarding the connection between the original algorithm and the extension of Lions and Mercier, the reader is referred to (Bauschke et al., 2017, Appendix) as well as Eckstein and Bertsekas (1992); Svaiter (2011). A recent survey on Douglas–Rachford can be found in Lindstrom and Sims (2018).

Given two nonempty, closed and convex sets $A, B \subseteq E$, the DR algorithm is the fixed point iteration generated by the *Douglas–Rachford operator*, $T_{A,B}$, defined by

$$
x_{k+1} = T_{A,B}(x_k), \quad \text{with } T_{A,B} := \frac{\operatorname{Id} + R_B R_A}{2}. \tag{12}
$$

The algorithm iterates by computing an average between the current point and the composition of two reflectors (see Figure 3). Due to this geometric interpretation, the DR algorithm is sometimes called the *averaged alternating reflections (AAR) method*. In the imaging community, it is also known as the *difference-map algorithm* (Bauschke et al. (2002)).

As in the previous section, we follow the same general recipe with the aim of applying Theorem 2. The next result analyzes the nonexpansiveness properties of the DR operator for closed and convex sets.
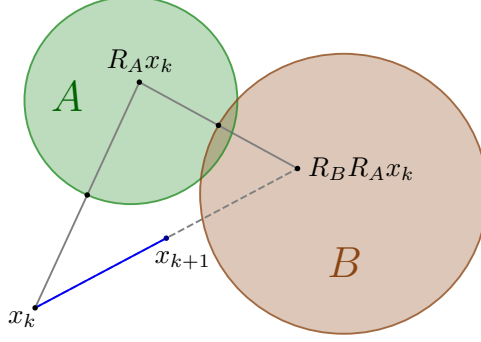
Figure 3: Geometric interpretation of the Douglas–Rachford iteration.

Proposition 6 (Nonexpansiveness properties of the DR operator)**.** *Let A and B be two nonempty, closed and convex subsets of E. Then the DR operator $T_{A,B}$ in* (12) *is $\frac{1}{2}$-averaged and, thus, firmly nonexpansive.*

Proof. Since A and B are nonempty, closed and convex, the reflectors R_A and R_B are nonexpansive by Proposition 3 and, consequently, so too is their composition. We therefore have that $T_{A,B} = \left(1 - \frac{1}{2}\right)\operatorname{Id} + \frac{1}{2}R_B R_A$ with $R_B R_A$ nonexpansive; i.e., $T_{A,B}$ is $\frac{1}{2}$-averaged. The remaining assertion follows from Proposition 1(iii). □

Therefore, according to the previous result and taking into account Theorem 2, the convergence of the DR algorithm only depends on the nonemptiness of the set of fixed points of the operator. Furthermore, when they exist, these fixed points need to be useful for solving the feasibility problem. Let us show that this is indeed the case.

Proposition 7 (Fixed points of the DR operator)**.** *Let $A, B \subseteq E$ be nonempty, closed and convex, and let $T_{A,B}$ be the DR operator. Then*

$$\operatorname{Fix} T_{A,B} \neq \emptyset \quad \Longleftrightarrow \quad A \cap B \neq \emptyset, \tag{13}$$

and, moreover, $P_A(x) \in A \cap B$ for all $x \in \operatorname{Fix} T$.

Proof. Observe that $A \cap B \subseteq \operatorname{Fix}(R_B R_A) = \operatorname{Fix} T_{A,B}$. Moreover, $x \in \operatorname{Fix}(R_B R_A)$ if and only if

$$x = (2P_B - \operatorname{Id})(2P_A - \operatorname{Id})(x) = 2P_B(2P_A(x) - x) - 2P_A(x) + x;$$

that is, $x \in \operatorname{Fix} T_{A,B}$ if and only if $P_B(2P_A(x) - x) = P_A(x)$. Consequently, we have $P_A(\operatorname{Fix} T_{A,B}) = A \cap B$ and the result follows. □

We are now in position to derive the following theorem which establishes the convergence of the DR algorithm for convex feasibility problems.

Theorem 5 (Douglas–Rachford method)**.** *Let $A, B \subseteq E$ be nonempty, closed and convex sets, and consider the DR operator $T_{A,B}$. Given any initial point $x_0 \in E$, consider the iteration generated by*

$$x_{k+1} = T_{A,B}(x_k), \quad for\ k = 0, 1, 2, \dots. \tag{14}$$

Exactly one of the following alternatives holds.

(i) $A \cap B \neq \emptyset$ and $(x_k)_{k=0}^{\infty}$ converges to a point $x^\star \in \operatorname{Fix} T_{A,B}$ with $P_A(x^\star) \in A \cap B$.

(ii) $A \cap B = \emptyset$ and $\|x_k\| \to +\infty$.

Proof. The operator $T_{A,B}$ is $\frac{1}{2}$-averaged by Proposition 6. Suppose first that $A \cap B \neq \emptyset$. Then, according to Proposition 7, we have $\operatorname{Fix} T_{A,B} \neq \emptyset$ and therefore we may apply Theorem 2 to deduce that

$$x_k \to x^\star \in \operatorname{Fix} T_{A,B}.$$

Furthermore, Proposition 7 yields that $P_A(x^\star) \in A \cap B$ and thus (i) is proved. Consider now the case when $A \cap B = \emptyset$. Again, by applying Proposition 7, we deduce that $\operatorname{Fix} T_{A,B} = \emptyset$ and hence (ii) immediately follows from Theorem 3. □

Remark 1 (Shadow sequence). *We emphasize that, in general, the limit point $x^\star$ in Theorem 5(i) need not be contained the intersection. Indeed, this is a consequence of the fact that $A \cap B$ can be a strict subset of $\operatorname{Fix} T_{A,B}$. To generate a solution to the feasibility problem using Theorem 5 is therefore necessary to compute $P_A(x^\star)$ (see Figures 4(a) and 4(b)). In other words, the sequence of interest is not the DR sequence $(x_k)_{k=0}^\infty$ itself but rather $(P_A(x_k))_{k=0}^\infty$, which is known as the* shadow sequence. *Also note that, as the projector is a continuous mapping, it immediately follows that*

$$P_A(x_k) \to a^\star = P_A(x^\star) \in A \cap B.$$

Remark 2 (Displacement vector). *In the inconsistent case, one typically looks for appropriate surrogate solutions. In the two set case with $A \cap B = \emptyset$, the notion of a* best approximation pair *provides one possibility. A pair of points $(a,b) \in A \times B$ is said to be a* best approximation pair *for (A,B) if*

$$\|a - b\| = d(A,B) := \inf \left\{ \|x - y\| : (x,y) \in A \times B \right\}.$$

The finer properties of the DR operator in the potentially inconsistent case have been studied by Bauschke et al. (2004). Their analysis relies on the notion of displacement vector, *which is defined as $v := P_{\overline{A-B}}(0)$ (i.e., v is the element of minimum norm in the closure of $A - B$). One can easily check that v measures the gap between the sets, since $\|v\| = d(A,B)$. In addition, $v \in A - B$ if and only if the distance $d(A,B)$ is attained, which means that there exists a best approximation pair. In this case, any inconsistent feasibility problem can be turned into a consistent one described as $A \cap (v + B)$ and the sequence $(x_k)_{k=0}^\infty$ generated by the DR algorithm (14) satisfies $x_{k+1} - x_k \to v$. Furthermore, Bauschke and Moursi (2017) showed that when $v \in A - B$, we have*

$$P_A(x_k) \to a^\star \in A \cap (v + B).$$

Finally, recall that, according to Theorem 5(ii), the DR sequence is unbounded when the sets do not intersect. Despite this, the latter result asserts that the shadow sequence remains convergent to a point $a^\star$ so long as the distance between the sets, $d(A,B)$, is attained. In this case, $(a^\star, P_B(a^\star))$ would constitute a best approximation pair (see Figure 4(c)).

(a) $A \cap B \neq \emptyset$ and $x^\star \in A \cap B$ (b) $A \cap B \neq \emptyset$ but $x^\star \notin A \cap B$ (c) $A \cap B = \emptyset$
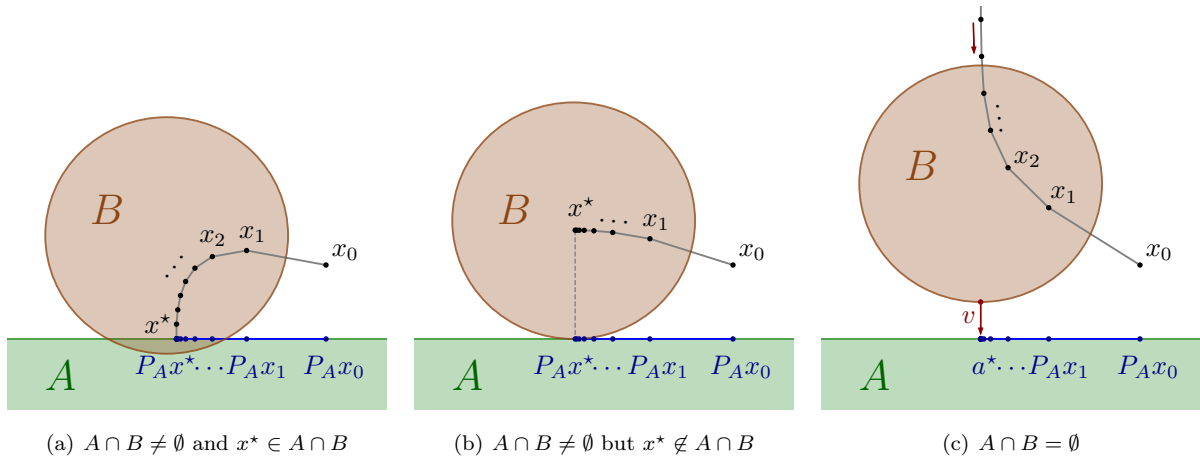
Figure 4: Behavior of the Douglas–Rachford algorithm in three possible scenarios.

3.3   Product space reformulations and Douglas–Rachford variants

Of the algorithms presented in the previous two sections, only the method of cyclic projections can be applied to the feasibility problem (1) (without modification) when $r \geq 3$. A naive approach to extending the DR algorithm to the case of three sets $A, B, C \subseteq E$ would be to consider the fixed point iteration generated by

$$T_{A,B,C} := \frac{\operatorname{Id} + R_C R_B R_A}{2}. \tag{15}$$

The corresponding iteration is convergent, since the operator $T_{A,B,C}$ in (15) is firmly nonexpansive with a nonempty set of fixed points, provided that the three sets intersect (see Exercise 8). Unfortunately, the operator's fixed points can not always be used to produce an intersection point, as is shown in Figure 5.
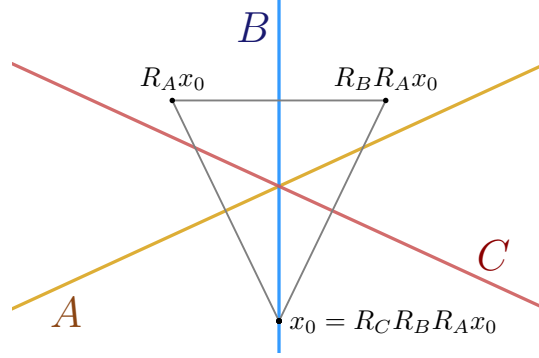
Figure 5: Failure of the 3-sets Douglas–Rachford iteration.

Exercise 8. *Let $A, B, C \subseteq E$ be nonempty, closed and convex sets. Show that the operator $T_{A,B,C}$, defined in* (15)*, is firmly nonexpansive with $A \cap B \cap C \subseteq \operatorname{Fix} T_{A,B,C}$. Deduce that its fixed point iteration converges whenever $A \cap B \cap C \neq \emptyset$.*

The classical way to apply the original two-set DR algorithm to problems having $r \geq 3$ sets is to use the following *product space reformulation*, due to Pierra (1984). To this end, consider the Euclidean product space

$$\mathbf{E} := E^r = E \times \overset{(r)}{\cdots} \times E, \tag{16}$$

endowed with the inner product

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \sum_{i=1}^{r} \langle x_i, y_i \rangle, \quad \text{for all } \boldsymbol{x} = (x_i)_{i=1}^r, \boldsymbol{y} = (y_i)_{i=1}^r \in \mathbf{E}.$$

Consider the two-set feasibility problem

$$\text{Find } \mathbf{x} = (x_1, x_2, \ldots, x_r) \in \mathbf{C} \cap \mathbf{D}, \tag{17}$$

where the constraint sets $\mathbf{C}$ and $\mathbf{D}$ are given by

$$\boldsymbol{C} := C_1 \times C_2 \times \cdots \times C_r \quad \text{and} \quad \boldsymbol{D} := \{(x, x, \ldots, x) \in \mathbf{E} : x \in E\}. \tag{18}$$

Although the set $\boldsymbol{D}$, sometimes called the *diagonal* subspace, is always a closed subspace, the properties of $\boldsymbol{C}$ are largely inherited from the individual constraint sets. For instance, $\boldsymbol{C}$ is nonempty, closed and convex precisely when $C_1, \ldots, C_r$ are. Note that any solution $\mathbf{x}$ to (17) satisfies $x_1 = x_2 = \cdots = x_r$.

It is easily seen that the original r-set feasibility problem (1) is equivalent to the two-set problem (17) in the following sense

$$x \in \bigcap_{i=1}^{r} C_i \subseteq E \quad \Longleftrightarrow \quad \boldsymbol{x} = (x, x, \ldots, x) \in \boldsymbol{C} \cap \boldsymbol{D} \subseteq \mathbf{E}. \tag{19}$$

In other words, an r-set feasibility problem in E can always be formulated as a two-set feasibility problem in $\mathbf{E}$. It is worth mentioning that when the number of constraints, r, is large, this formulation can become numerically inefficient because it requires computations to be performed in a space with much higher dimension. In some special cases, this can be avoided (see Exercise 10).

In order to apply projection methods to the reformulated problem, it is necessary to have access to the projectors onto the sets $\boldsymbol{C}$ and $\boldsymbol{D}$. As we summarize in the following proposition, this is indeed the case whenever the projectors onto the underlying constraint sets in the original problem, $C_1, \ldots, C_r$, are available.

Proposition 8 (Product-space projectors)**.** *Let $\boldsymbol{x} = (x_1, \ldots, x_r) \in \boldsymbol{E}$. The projectors onto the sets $\boldsymbol{C}$ and $\boldsymbol{D}$ in* (18) *at $\boldsymbol{x}$ are given by*

$$P_{\boldsymbol{C}}(\boldsymbol{x}) = P_{C_1}(x_1) \times P_{C_2}(x_2) \times \cdots \times P_{C_r}(x_r) \quad \text{and} \quad P_{\boldsymbol{D}}(\boldsymbol{x}) = \left(\frac{1}{r} \sum_{i=1}^{r} x_i, \frac{1}{r} \sum_{i=1}^{r} x_i, \ldots, \frac{1}{r} \sum_{i=1}^{r} x_i \right).$$

Proof. The proof is left as an exercise. □

Exercise 9. *Prove Proposition 8. Hint: To prove the formula for P_C, use the definition directly. To prove the formula for P_D, note that D is a subspace and use Proposition 2.*

Exercise 10 (The method of averaged projections)**.** *Suppose $C_1, C_2, \ldots, C_r \subseteq E$ are closed and convex sets with nonempty intersection. Given a point $x_0 \in E$, the* method of averaged projections *is the fixed point iteration governed by*

$$x_{k+1} = T(x_k), \quad where \quad T = \frac{P_{C_r} + \cdots + P_{C_1}}{r}.$$

By applying the method of cyclic projections to (17), show that $(x_k)_{k=0}^{\infty}$ converges to a point in $\cap_{i=1}^{r} C_i$.

The product space reformulation explained above leads to the following DR scheme that can be applied to r closed and convex sets $C_1, C_2, \ldots, C_r \subseteq E$.

Douglas–Rachford in the product space. Given r arbitrary starting points $x_{1,0}, x_{2,0}, \ldots, x_{r,0} \in E$, the iterative scheme can be expressed as

$$\text{for } k = 0, 1, 2, \ldots :$$
$$\left\lfloor \begin{array}{l} p_k = \frac{1}{r} \sum_{i=1}^{r} x_{i,k}, \\ \text{for } i = 1, 2, \ldots, r : \\ \quad \lfloor \; x_{i,k+1} = \frac{1}{2} x_{i,k} + \frac{1}{2} R_{C_i} (2 p_k - x_{i,k}). \end{array} \right. \tag{20}$$

Observe that (20) is just the iteration

$$\boldsymbol{x}_{k+1} = \frac{1}{2} \boldsymbol{x}_k + \frac{1}{2} R_{\boldsymbol{C}} R_{\boldsymbol{D}}(\boldsymbol{x}_k), \quad \text{for } k = 0, 1, 2, \ldots,$$

where $\boldsymbol{C}$ and $\boldsymbol{D}$ are the product and the diagonal sets defined in (18), and $\boldsymbol{x}_k := (x_{1,k}, x_{2,k}, \ldots, x_{r,k})$. Consequently, Theorem 5 guarantees convergence of the sequences generated by (20).

We remark that the choice to compute the reflection with respect to the diagonal $\boldsymbol{D}$ first is deliberate, so that the shadow sequence, $(P_{\boldsymbol{D}}(x_k))_{k=0}^{\infty} \subseteq \mathbf{E}$, can be unambiguously identified with the sequence $(p_k)_{k=0}^{\infty}$ in E (rather than $\mathbf{E}$). The latter will converge to a common point of the sets, whenever such a point exists (see Figure 6).
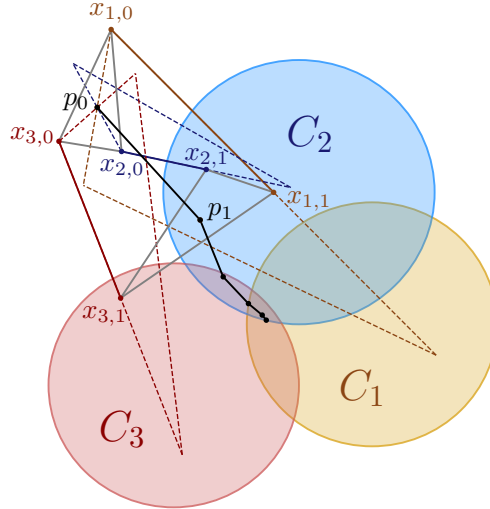

Figure 6: Illustration of the Douglas–Rachford iteration in the product space.

Note that the iteration in (20) has the advantage that each of the individual projectors, P_{C_i}, can be computed in parallel. On the other hand, as many points as constraint sets need to be stored at each step to compute the next iteration. This makes the algorithm computationally inefficient when the number of constraints is large.

One possible modification of the DR algorithm which avoids using a product reformulation is based on cyclic application of DR operators.

3.4 Cyclic Douglas–Rachford methods

Instead of the operator in (15), Borwein and Tam (2014) introduced the *cyclic DR operator* defined as

$$T_{[C_1,C_2,\ldots,C_r]} := T_{C_r,C_1} T_{C_{r-1},C_r} \cdots T_{C_2,C_3} T_{C_1,C_2}. \tag{21}$$

The fixed point iteration generated by this operator, known as the *cyclic DR method*, can be applied to convex feasibility problems defined by an arbitrary number of sets, without recourse to the product space reformulation (see Borwein and Tam, 2014, Theorem 3.2). Note the operator (21) can be viewed as cyclically applying the classical DR method to pairs of sets (see Figure 7(a)). The analysis of the method in the inconsistent case was developed in Borwein and Tam (2015), and a further extension of the algorithm has recently been proposed by Aragón et al. (2018c), which incorporates k-sets-DR operators (of type (15) when $k = 3$). Observe that

$$T_{[A,B]} = T_{B,A} T_{A,B} \neq T_{A,B},$$

that is, (21) does not coincide with the classical DR operator for $r = 2$. Alternatively, Bauschke et al. (2015) proposed the *cyclically anchored DR method*, which is defined by the operator

$$T_{C_1,[C_2,\ldots,C_r]} := T_{C_1,C_r} T_{C_1,C_{r-1}} \cdots T_{C_1,C_2}. \tag{22}$$

Here the set C_1 is called the *anchor set*. In contrast to the cyclic DR method, the latter scheme does reduce to the original DR scheme when dealing with only two sets. Also, when $C_1 = E$, observe that the cyclically anchored DR method coincides with the method of cyclic projections, since $T_{E,C_i} = P_{C_i}$.

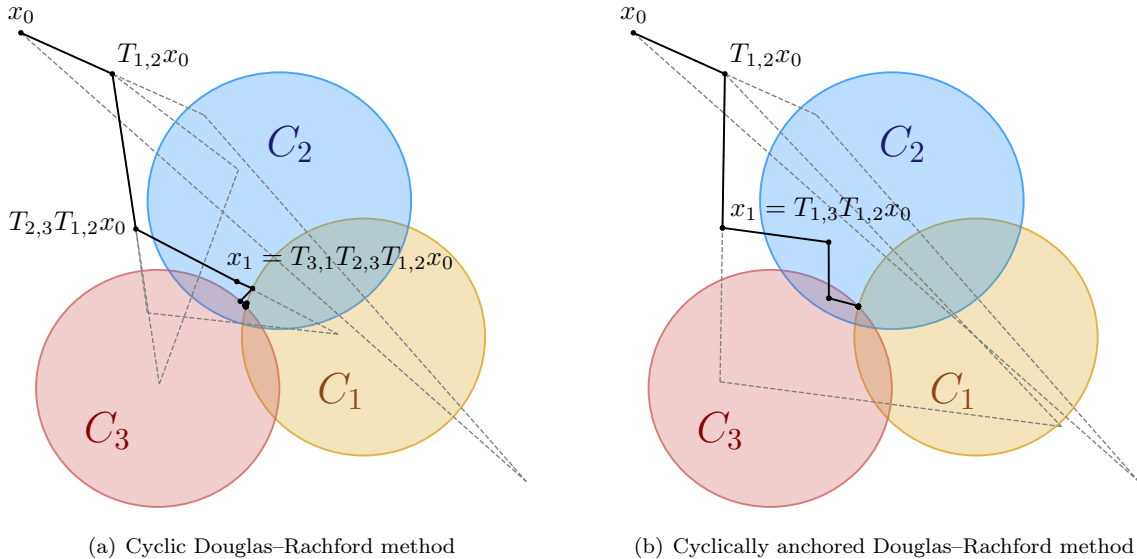(a) Cyclic Douglas–Rachford method   (b) Cyclically anchored Douglas–Rachford method

Figure 7: Illustration of two cyclic versions of Douglas–Rachford for finitely many sets.

3.5 Some modified and relaxed versions

To conclude this section, we briefly mention some other modifications and relaxations of the classical DR algorithm in the literature. Such variants are typically constructed by relaxing or modifying some of the involved operators, so that parameters can be incorporated into the scheme. The introduction of these parameters, which can be tuned, allows the possibility of accelerating the method or even completely changing the dynamics of the algorithm.

In what follows, we present four such variants. Each of their iterations are illustrated in Figure 9.

Generalized Douglas–Rachford. The *generalized Douglas–Rachford (GDR) method* is the most evident generalization, which was in fact the method studied by Eckstein and Bertsekas (1992). The algorithm relies on iteratively applying a general α-averaged version of the DR operator, that is,

$$x_{k+1} = T_{A,B,\alpha}(x_k) := (1-\alpha)\, x_k + \alpha R_B R_A(x_k), \quad \text{with } \alpha \in \,]0,1[. \tag{23}$$

Note that for $\alpha = \frac{1}{2}$ it becomes the classical DR algorithm.

Relaxed Averaged Alternating Reflections. Luke (2008) proposed the *relaxed averaged alternating reflections (RAAR) method*, whose iteration is defined as an average between the classical DR and the projection onto the first set, that is,

$$x_{k+1} = (1 - \beta) P_A(x_k) + \beta T_{A,B}(x_k), \quad \text{with } \beta \in]0, 1[.$$

Unlike the classical DR method, this scheme converges even for inconsistent feasibility problems (without turning to the shadow sequence) as long as the distance between the sets is attained. In this case, the method provides a best approximation pair. It is worth mentioning that a more general relaxed version of DR has been recently proposed by Thao (2018).

Circumcentered Douglas–Rachford. Motivated by the "spiraling dynamics" exhibited by DR when applied to subspaces (see Figure 8(a)), Behling et al. (2018) proposed a modification of the algorithm termed the *circumcentered Douglas–Rachford (CDR) method*. In this method, the next iterate is computed as the circumcenter of the triangle implicitly generated by the classical DR iteration; that is,

$$x_{k+1} = C_T[x_k, R_A(x_k), R_B R_A(x_k)], \tag{24}$$

where $C_T[a, b, c]$ denotes the circumcenter of the triangle of vertices a, b and c. This method requires the sets to be closed subspaces since, otherwise the circumcenter operator is not necessarily well-defined. In fact, even when it is, the method may still fail to converge. An example with two intersecting balls in $\mathbb{R}^2$ is shown in Figure 10. For one-dimensional subspaces, however, the algorithm always converges in one iteration.

Averaged Alternating Modified Reflections. In the particular case when the DR algorithm is applied to closed (affine) subspaces, it does not yield just any point in the intersection but rather the intersection nearest to the initial one, i.e., $P_{A \cap B}(x_0)$ (see, e.g., Bauschke et al., 2014, Corollaries 4.4 and 4.5). In this context, the method can be used to solve best approximation problems. This is not the case for arbitrary convex sets, as shown in Figure 8. Aragón and Campoy (2018a) proposed the *averaged alternating modified reflections (AAMR) method* to solve best approximation problems described by arbitrary closed and convex sets. This algorithm can be seen as a variant of DR where the reflection steps are slightly modified. For a given point $q \in E$, the iteration is given by

$$x_{k+1} = (1 - \alpha) x_k + \alpha (2\beta P_{B-q} - \text{Id})(2\beta P_{A-q} - \text{Id})(x_k), \quad \text{with } \alpha \in]0, 1], \beta \in]0, 1[. \tag{25}$$

Under a constraint qualification on the sets at the point q, the scheme converges to a point $x^\star$ such that $P_A(x^\star + q) = P_{A \cap B}(q)$. The convergence of the sequence $(x_k)_{k=0}^\infty$ for the case $\alpha = 1$ was proved by Alwadani et al. (2018). Observe that the DR algorithm can be obtained as the limit case $\beta = 1$ in (25) with $\alpha = \frac{1}{2}$ and $q = 0$.
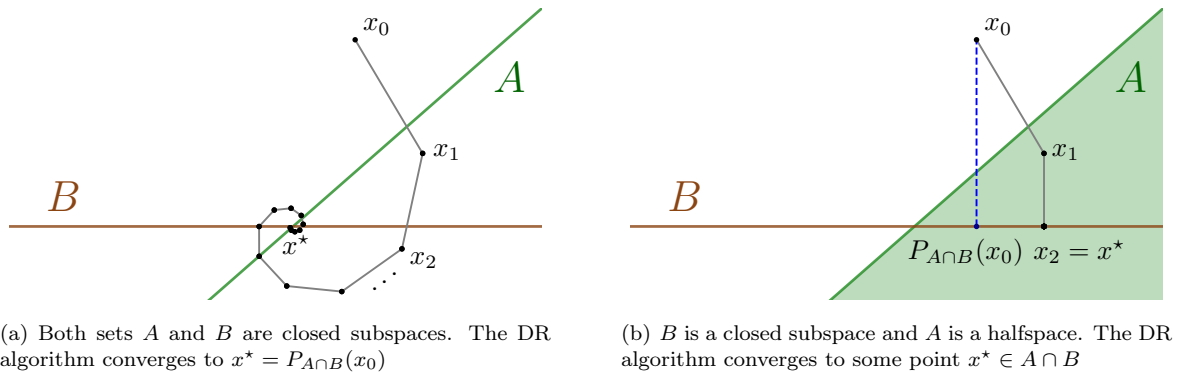

(a) Both sets A and B are closed subspaces. The DR algorithm converges to $x^\star = P_{A \cap B}(x_0)$

(b) B is a closed subspace and A is a halfspace. The DR algorithm converges to some point $x^\star \in A \cap B$

Figure 8: Failure of the DR method for solving the best approximation problem for arbitrary convex sets.

(a) GDR with $\alpha = 0.8$

(b) RAAR with $\beta = 0.4$

(c) CDR

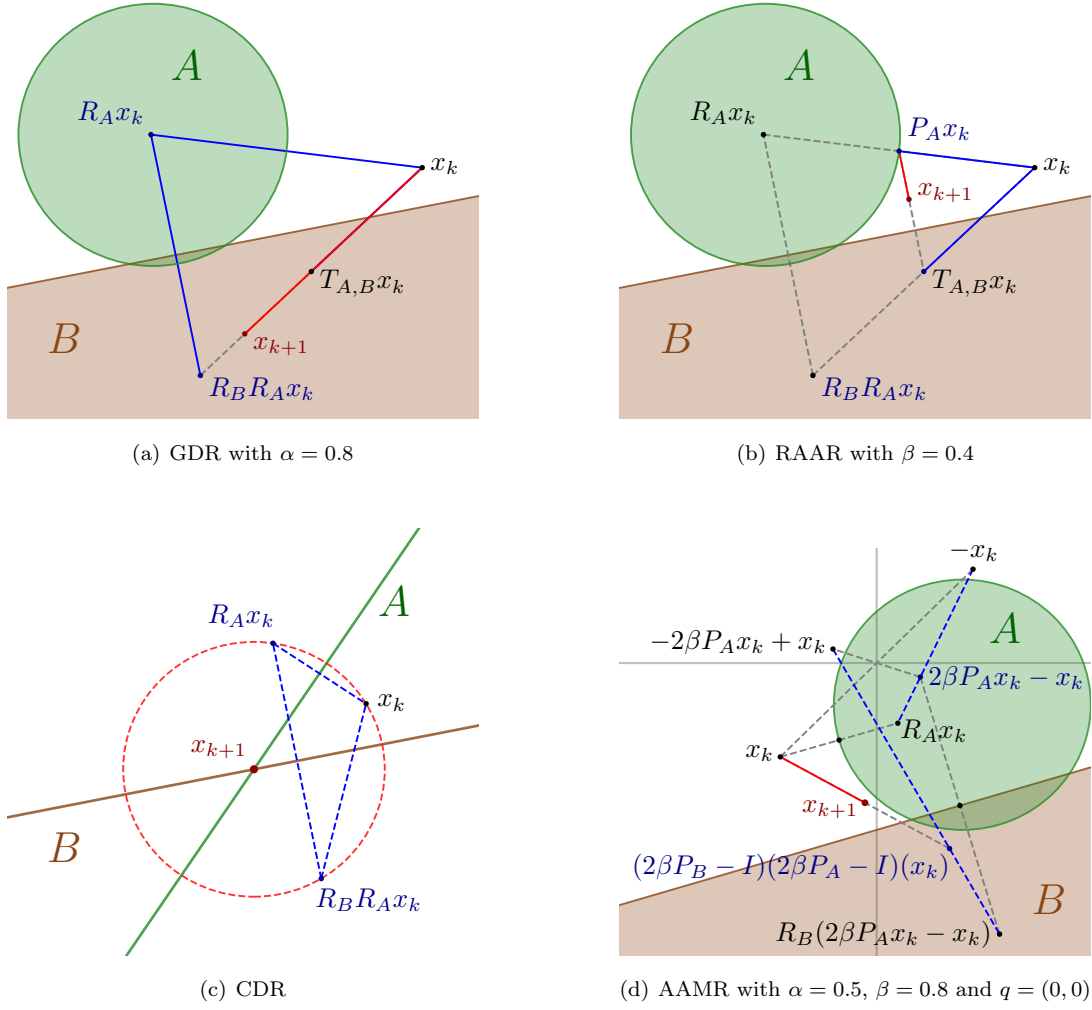(d) AAMR with $\alpha = 0.5$, $\beta = 0.8$ and $q = (0,0)$

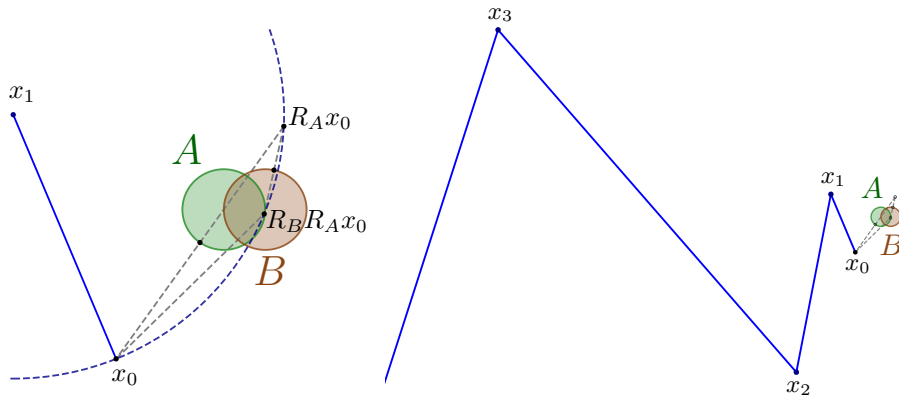Figure 9: One iteration of GDR, RAAR, CDR and AAMR methods.

Figure 10: Failure of the circumcentered Douglas–Rachford method when applied to two balls of the same radius $A, B \in \mathbb{R}^2$: the sequence $(x_k)_{k=0}^{\infty}$ diverges. The left figure shows construction of the first iteration from the right figure in more detail.

4 A successful heuristic for nonconvex problems

The DR algorithm has recently gained significant popularity, in part, thanks to its good behavior when applied in nonconvex settings. In this context, we note that the DR operator may be multivalued due to the fact that the projection onto a nonconvex set is not necessarily unique. Consequently, the equality in (14) must be replaced by an inclusion, and the iteration thus takes the form

$$x_{k+1} \in T_{A,B}(x_k) := \{x_k + b_k - a_k \in E : a_k \in P_A(x_k), b_k \in P_B(2a_k - x_k)\}. \tag{26}$$

The fixed point set of this operator remains useful for solving the feasibility problem without assuming convexity of the sets A and B. This is explained in the following straightforward exercise which generalizes Proposition 7.

Exercise 11 (Fixed points of the DR operator). *Let $A, B \subseteq E$ be nonempty and closed, and let $T_{A,B}$ be the DR operator given by (26). Prove that*

$$\operatorname{Fix} T_{A,B} \neq \emptyset \quad \Longleftrightarrow \quad A \cap B \neq \emptyset,$$

and that, for any $x \in \operatorname{Fix} T_{A,B}$, there exists a point $p \in P_A(x)$ such that $p \in A \cap B$.

Despite the fact that convergence of the algorithm is generally only ensured when applied to convex sets, the method has nevertheless been successfully employed for solving a variety of nonconvex optimization problems, especially those of combinatorial nature. Examples of such applications include matrix completion (Aragón et al., 2014a), protein conformation determination (Borwein and Tam, 2017), phase and bit retrieval (Bauschke et al., 2002; Elser, 2003, 2018), differential equations (Lamichhane et al., 2017), graph coloring (Aragón and Campoy, 2018b; Aragón et al., 2018a), combinatorial designs (Aragón et al., 2018b), and a wide variety of NP-hard problems such as Sudoku or 3-Satisfiability (Aragón et al., 2014b; Elser et al., 2007; Schaad, 2010).

The nonconvex theory for the DR algorithm is much less developed; there are very few results explaining why the algorithm works, and even less justifying its good global performance. The first nonconvex scenario was considered by Borwein and Sims (2011), who established local convergence of the method near each of the intersection points of a line and a sphere in a Euclidean space. An explicit description of the regions of convergence was later provided by Aragón and Borwein (2013). It was finally Benoist (2015) who, via the construction of a Lyapunov function, established the convergence of the algorithm for every starting point not lying on the hyperplane of symmetry. Lyapunov functions are a powerful tool from difference inclusions whose existence guarantees the convergence of the iteration. By using this approach, Dao and Tam (2019a) proved global convergence of the DR algorithm for finding a zero of a function, with applications to several nonconvex feasibility problems.

From a different perspective, Aragón et al. (2016) proved global convergence for the case of a halfspace and a finite set. We make use of this scenario to illustrate in Figure 11 the difference between the behavior of DR and cyclic projections when addressing combinatorial problems. While the method of cyclic projections usually gets stuck in those points which are close to be solutions (as it finds a local best approximation pair), DR is satisfactorily capable to escape from them (thanks to Theorem 5(ii)). Even so, the DR algorithm does not break free from getting caught by cycles in other nonconvex settings. Although this is something that does not seem to happen very often, it may be hard to detect. It is worth mentioning that the cycling of the algorithm for a simple inconsistent nonconvex feasibility problem, specifically, a hyperplane and a doubleton, was recently analyzed by Bauschke et al. (2019).

Local convergence of the algorithm (i.e., convergence for starting points sufficiently close to a solution) in nonconvex settings has been established, for instance, for the case of a line and an ellipse or a p-sphere (Borwein et al., 2018), and for union of convex sets (Bauschke and Noll, 2014; Tam, 2018; Dao and Tam, 2019b). Other results regarding local convergence are usually obtained by requiring regularity properties of the sets and/or of their intersection, see e.g. Bauschke and Dao (2017); Hesse and Luke (2013); Hesse et al. (2014); Phan (2016).

4.1 Application to a generalized "8-queens problem"

In this section, we use the DR algorithm to solve a generalization of the so-called 8-*queens problem*, which asks for the placement of 8 mutually non-attacking queens on an 8×8 chessboard. Recall that, in chess,

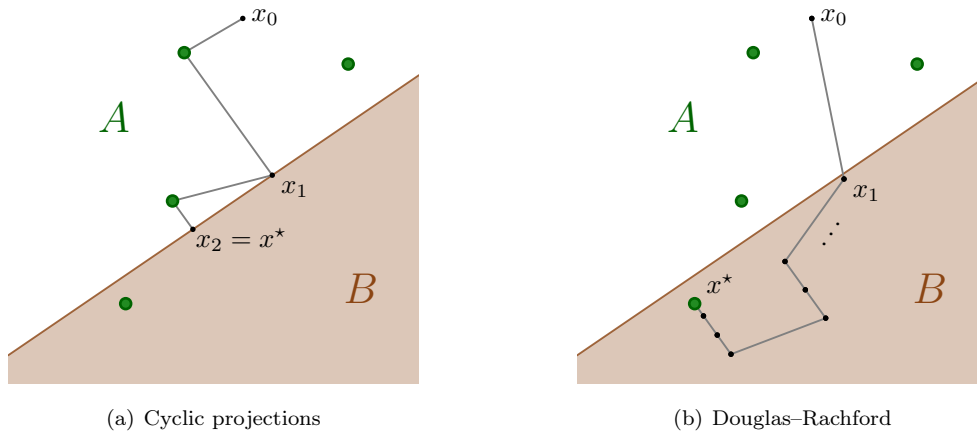(a) Cyclic projections   (b) Douglas–Rachford

Figure 11: The method of cyclic projections and DR applied to a finite set (the set of green points) and a halfspace.

queens can move any number of squares along a vertical, horizontal or diagonal of the board. For this reason, two or more queens are said to be *attacking* if they lie in the same vertical, horizontal or diagonal.

The 8-queens puzzle therefore consists in placing 8 queens so that each row and column contain precisely one queen, and each diagonal contains at most one. The problem we consider here generalizes this puzzle in two ways. Firstly, we consider $n \times n$ chessboards for an arbitrary positive integer n, and secondly, we consider placing mn queens ($1 \leq m \leq n$) so that there are precisely m queens in each row and column, and each diagonal contains at most m. We refer to the general problem as the (m, n)-*queens problem*. For the sake of clarity, we shall restrict our exposition to the $m = 2$ case, however the results easily generalize to the $m > 2$ case. An example of a solution to the $(2, 8)$-queens problem is shown in Figure 12.

For the case when $m = 1$, the $(1, n)$-queens problem is known as the n-*queens problem*. In his master's thesis, Schaad (2010) successfully demonstrated that the DR method can be used to solve the n-queens problem.

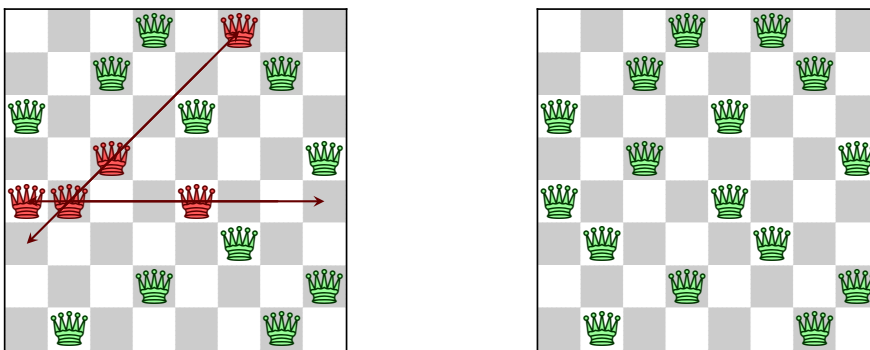Figure 12: Two configurations of 16 queens on an 8×8 chessboard. The left board is not a solution to the $(2, 8)$-queens problem as one row and one backward diagonal both have three queens (red). In contrast, the right board is a solution.

To model queens on a chessboard as a feasibility problem, consider a matrix $X \in \{0, 1\}^{n \times n}$. We interpret the entries of this matrix as squares on a chessboard, with non-zero entries corresponding to queens. The $(2, n)$-queens problem can thus be described in terms of the following constraints:

1. Exactly two entries in each row of X are non-zero.

2. Exactly two entries in each column of X are non-zero.

3. At most two entries in each forward diagonal of X are non-zero.

4. At most two entries in each backward diagonal of X are non-zero.

We now express the problem mathematically using these constraints. To this end, observe that a matrix $X \in \mathbb{R}^{n \times n}$ is a solution to the $(2, n)$-queens problem if and only if its entries are contained in $\{0, 1\}$ and it belongs to the following constraint sets

$$C_1 := \left\{ X \in \mathbb{R}^{n \times n} : \sum_{i=1}^{n} x_{ij} = 2, \text{ for all } j = 1, \ldots, n. \right\}, \tag{27a}$$

$$C_2 := \left\{ X \in \mathbb{R}^{n \times n} : \sum_{j=1}^{n} x_{ij} = 2, \text{ for all } i = 1, \ldots, n. \right\}, \tag{27b}$$

$$C_3 := \left\{ X \in \mathbb{R}^{n \times n} : \sum_{i=1}^{n-k} x_{i,i+k} \leq 2 \text{ and } \sum_{j=1}^{n-k} x_{j+k,j} \leq 2, \text{ for all } k = 0, 1, \ldots, n-3. \right\}, \tag{27c}$$

$$C_4 := \left\{ X \in \mathbb{R}^{n \times n} : \sum_{j=1}^{n-k} x_{n-j-k,j} \leq 2 \text{ and } \sum_{i=1}^{n-k} x_{n-i,i+k} \leq 2, \text{ for all } k = 0, 1, \ldots, n-3. \right\}. \tag{27d}$$

Now, even with these constraints sets fixed, there are still many possible formulations which model the desired feasible set. Let us denote $\widehat{C}_i := C_i \cap \{0, 1\}^{n \times n}$ for each $i = 1, 2, 3, 4$. We shall concern ourselves with the following four formulations:

$$\text{Find } X \in C_1 \cap C_2 \cap C_3 \cap C_4 \cap \{0, 1\}^{n \times n} \qquad \text{(Formulation 1)}$$
$$= C_1 \cap C_2 \cap \widehat{C}_3 \cap \widehat{C}_4 \qquad \text{(Formulation 2)}$$
$$= \widehat{C}_1 \cap \widehat{C}_2 \cap C_3 \cap C_4 \qquad \text{(Formulation 3)}$$
$$= \widehat{C}_1 \cap \widehat{C}_2 \cap \widehat{C}_3 \cap \widehat{C}_4. \qquad \text{(Formulation 4)}$$

In order to apply any of the algorithms previously presented, we need to be able to compute the projections onto each of these sets. Note that a projection of any $X \in \mathbb{R}^{n \times n}$ onto $\{0, 1\}^{n \times n}$, $\pi_{\{0,1\}^{n \times n}}(X) \in P_{\{0,1\}^{n \times n}}(X)$, can be easily computed component-wise as

$$\left(\pi_{\{0,1\}^{n \times n}}(X) \right)_{ij} = \begin{cases} 1, & \text{if } x_{ij} > 0.5, \\ 0, & \text{otherwise,} \end{cases} \quad \text{for } i, j = 1, 2, \ldots, n.$$

In order to derive explicit formulae to compute projections onto the remaining sets, consider, for each $p = 1, 2, \ldots, n$, the sets

$$S_p := \left\{ x \in \mathbb{R}^p : \sum_{i=1}^{p} x_i = 2 \right\} \quad \text{and} \quad H_p := \left\{ x \in \mathbb{R}^p : \sum_{i=1}^{p} x_i \leq 2 \right\},$$

where we use the notation $x = (x_1, x_2, \ldots, x_p)$. Similarly as before, let us denote $\widehat{S}_p := S_p \cap \{0, 1\}^p$ and $\widehat{H}_p := H_p \cap \{0, 1\}^p$. Although S_p and H_p are, respectively, a hyperplane and a halfspace and thus, convex, the sets $\widehat{S}_p$ and $\widehat{B}_p$ are discrete. The constraint sets in (27), as well as their discrete counterparts, can be represented in terms of these four sets. The following propositions provide specific formulae for computing a projection onto each of these sets.

Proposition 9 (Projectors onto S_p and H_p). *Let $x = (x_1, x_2, \ldots, x_p) \in \mathbb{R}^p$. Then the projectors onto the sets S_p and H_p are given component-wise by*

$$\left(P_{S_p}(x) \right)_i = x_i + \frac{1}{p} \left(2 - \sum_{j=1}^{p} x_j \right), \quad \text{for } i = 1, 2, \ldots, p;$$

$$\left(P_{H_p}(x) \right)_i = x_i + \frac{1}{p} \min \left\{ 0, 2 - \sum_{j=1}^{p} x_j \right\}, \quad \text{for } i = 1, 2, \ldots, p.$$

Proof. Observe that $S_p = \{x \in \mathbb{R}^p : e^T x = 2\}$ and $H_p = \{x \in \mathbb{R}^p : e^T x \leq 2\}$ with $e := (1, 1, \ldots, 1) \in \mathbb{R}^p$. Thus, the result follows directly from (4) and (5). $\qquad\square$

Proposition 10 (Projections onto $\widehat{S}_p$ and $\widehat{H}_p$). *Let $x = (x_1, x_2, \ldots, x_p) \in \mathbb{R}^p$. Then a projection onto the sets $\widehat{S}_p$ and $\widehat{H}_p$, $\pi_{\widehat{S}_p}(x) \in P_{\widehat{S}_p}(x)$ and $\pi_{\widehat{H}_p}(x) \in P_{\widehat{H}_p}(x)$, can be computed component-wise as*

$$\left(\pi_{\widehat{S}_p}(x)\right)_i = \begin{cases} 1, & \text{if } i \in Q_2(x), \\ 0, & \text{otherwise,} \end{cases} \quad \text{for } i = 1, 2, \ldots, p, \tag{29}$$

$$\left(\pi_{\widehat{H}_p}(x)\right)_i = \begin{cases} 1, & \text{if } i \in Q_2(x) \text{ and } x_i > 0.5, \\ 0, & \text{otherwise,} \end{cases} \quad \text{for } i = 1, 2, \ldots, p; \tag{30}$$

where $Q_2(x)$ is the set of indices in $\{1, 2, \ldots, p\}$ corresponding to the two largest values in $\{x_1, x_2, \ldots, x_p\}$ (largest index is chosen in case of tie).

Proof. Let $e_1, \ldots, e_p$ denote the unit vectors of the standard basis of $\mathbb{R}^p$ and note that

$$\widehat{S}_p = \{e_i + e_j : i, j = 1, 2, \ldots, p, \; i \neq j\},$$
$$\widehat{H}_p = \{0_p\} \cup \{e_i : i = 1, 2, \ldots, p\} \cup \{e_i + e_j : i, j = 1, 2, \ldots, p, \; i \neq j\}.$$

Since the sets $\widehat{S}_p$ and $\widehat{H}_p$ contain a finite number of elements, they are clearly closed nonempty sets. To characterize the projector onto $\widehat{S}_p$ observe that

$$e_{i_0} + e_{j_0} \in P_{\widehat{S}_p}(x) \quad \Longleftrightarrow \quad \|x - e_{i_0} - e_{j_0}\| \leq \|x - e_i - e_j\|, \quad \text{for all } i, j \in \{1, 2, \ldots, p\} \text{ with } i \neq j,$$

$$\Longleftrightarrow \quad \|x - e_{i_0} - e_{j_0}\|^2 \leq \|x - e_i - e_j\|^2, \quad \text{for all } i, j \in \{1, 2, \ldots, p\} \text{ with } i \neq j,$$

$$\Longleftrightarrow \quad \|x\|^2 + 2 - 2x^T e_{i_0} - 2x^T e_{j_0} \leq \|x\|^2 + 2 - 2x^T e_i - 2x^T e_j,$$
$$\text{for all } i, j \in \{1, 2, \ldots, n\}, i \neq j,$$

$$\Longleftrightarrow \quad x_i + x_j \leq x_{i_0} + x_{j_0}, \quad \text{for all } i, j \in \{1, 2, \ldots, n\}, i \neq j,$$

$$\Longleftrightarrow \quad x_i \leq x_{i_0} \text{ and } x_i \leq x_{j_0}, \quad \text{for all } i \in \{1, 2, \ldots, n\}. \tag{31}$$

By using a case distinction and following analogous reasoning, we can easily obtain that

$$0_p \in P_{\widehat{H}_p}(x) \quad \Longleftrightarrow \quad x_i \leq \frac{1}{2}, \text{ for all } i \in \{1, 2, \ldots, p\}, \tag{32a}$$

$$e_{i_0} \in P_{\widehat{H}_p}(x) \quad \Longleftrightarrow \quad x_j \leq \frac{1}{2} \leq x_{i_0}, \text{ for all } j \in \{1, 2, \ldots, p\} \setminus \{i_0\}; \tag{32b}$$

$$e_{i_0} + e_{j_0} \in P_{\widehat{H}_p}(x) \quad \Longleftrightarrow \quad \begin{array}{c} \text{with } \frac{1}{2} \leq x_{i_0}, \quad \frac{1}{2} \leq x_{j_0}, \\ x_i \leq x_{i_0}, \; x_i \leq x_{j_0}, \text{ for all } i \in \{1, 2, \ldots, p\} \setminus \{i_0, j_0\}. \end{array} \tag{32c}$$

Finally, observe that (29) and (30) are directly derived from (31) and (32), respectively. $\qquad\square$

Remark 3 (On the multivaluedness of the projectors). *In contrast to Proposition 9, where the projectors onto S_p and H_p were given, we only use the qualifier "a projection" in Proposition 10. Since the sets $\widehat{S}_p$ and $\widehat{H}_p$ are nonconvex, the projections may not be unique. Multivaluedness occurs when either there is a tie in the largest components of the vector to be projected, or when some of them are exactly 0.5. Nonetheless, none of these situations will happen very often in practice, due to the finite precision of the computations, specially if the algorithm is run from a randomly generated starting point.*

Observe that the constraint sets C_1 and C_2 (resp. $\widehat{C}_1$ and $\widehat{C}_2$) determine that each row and each column of X, respectively, lays in the set S_n (resp. $\widehat{S}_n$). On the other hand, C_3 and C_4 (resp. $\widehat{C}_3$ and $\widehat{C}_4$) determine that each forward and each backward diagonal of X, respectively, belongs to H_p (resp. $\widehat{H}_p$) for some $p \in \{1, 2, \ldots, n\}$. Thus, we can compute projections onto all these sets by using Proposition 9 (resp. Proposition 10).

Remark 4 (Fixed points and solutions). *As demonstrated in Figure 11, the method of cyclic projections often fails to be useful when applied to combinatorial problems as there exist fixed points of the operator which do not provide a solution. Consider, for instance, one iteration of this method, for the $(2,3)$-queens problem generated with Formulation 3 as*

$$X_0 := \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \xrightarrow{P_{\widehat{C}_1}} \begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \xrightarrow{P_{\widehat{C}_2}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \xrightarrow{P_{C_3}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \xrightarrow{P_{C_4}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} = X_0.$$

Here we remark that projections onto $\widehat{C}_1$ and $\widehat{C}_2$ are computed using Proposition 10, so in case of tie between the two largest components of a row or a column, we choose those ones of largest indexes. The point X_0 is therefore a fixed point in the sense that

$$X_0 \in P_{C_4} P_{C_3} P_{\widehat{C}_2} P_{\widehat{C}_1}(X_0).$$

However, X_0 is not a solution to the problem since the second row contains three queens.

The same problem occurs when we use a cyclic variant of the DR algorithm. For instance, we have

$$Y_0 \in T_{[\widehat{C}_1, \widehat{C}_2, C_3, C_4]}(Y_0) \quad and \quad Y_0 \in T_{\widehat{C}_1, [\widehat{C}_2, C_3, C_4]}(Y_0), \quad for \ Y_0 := \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix},$$

where $T_{[\widehat{C}_1, \widehat{C}_2, C_3, C_4]}$ and $T_{\widehat{C}_1, [\widehat{C}_2, C_3, C_4]}$ are the cyclic and the cyclically anchored DR operators defined in (21) and (22), respectively. When we use a cyclic DR method, the point that solves the problem is not the fixed point itself, but its projections onto any of the constraint sets. Observe that in this case

$$P_{\widehat{C}_1}(Y_0) = P_{C_3}(Y_0) = P_{C_4}(Y_0) = \{Y_0\} \quad and \quad P_{\widehat{C}_2}(Y_0) \cap \widehat{C}_1 = \emptyset,$$

so a solution cannot be obtained from Y_0. The existence of this undesired type of fixed points cannot happen to the DR algorithm in the product space, thanks to Exercise 11.

Numerical experiment. We performed a numerical experiment to compare the behavior of the DR algorithm on the four formulations presented for the $(2,n)$-queens problem. Note that all the feasibility problems considered above are described by more than two sets, so we have to turn to the product space. We shall not present the results of running the method of cyclic projections and the other versions of the DR algorithm presented in Section 3.3, because all of them are highly unsuccessful in finding a solution, mainly due to Remark 4. Therefore, for each of the four formulations defined by r sets, we use the following implementation:

- *Initialization*: Generate a random vector $y \in \{0,1\}^{n \times n}$ and set $x_{1,0} := y, x_{2,0} := y, \ldots, x_{r,0} := y$.

- *Iteration*: For $k = 0, 1, 2, \ldots$, compute p_k and update $x_{1,k+1}, x_{2,k+1}, \ldots, x_{r,k+1}$ according to (20).

- *Termination*: The algorithm is terminated if either a solution is found or a maximum time of 300 seconds has been reached. In order to test if the current iterate provides a solution, we pointwise round the current shadow point, p_k, to the nearest integer and check if this integer point satisfies all constraints.

All implementations were coded in Python 2.7 and run on an Intel Core i7-4770 CPU 3.40 GHz with 16 GB RAM running Windows 10 (64-bit).

Our experiments were performed with chessboards of size n for $n \in \{10, 20, \ldots, 100\}$. For each size and each formulation, the DR algorithm was run from 20 random starting points. The results of the experiment are summarized in Figure 13. We observe that Formulations 1 and 2 both performed poorly, only being able to solve small sized boards. Formulation 4 performed well, however Formulation 3 was the best with at most one failed instance for all sizes. Again, we wish to emphasize that it is quite remarkable that the DR method works here at all, let alone so well, given the absence of any sound theoretical justification for this behavior.
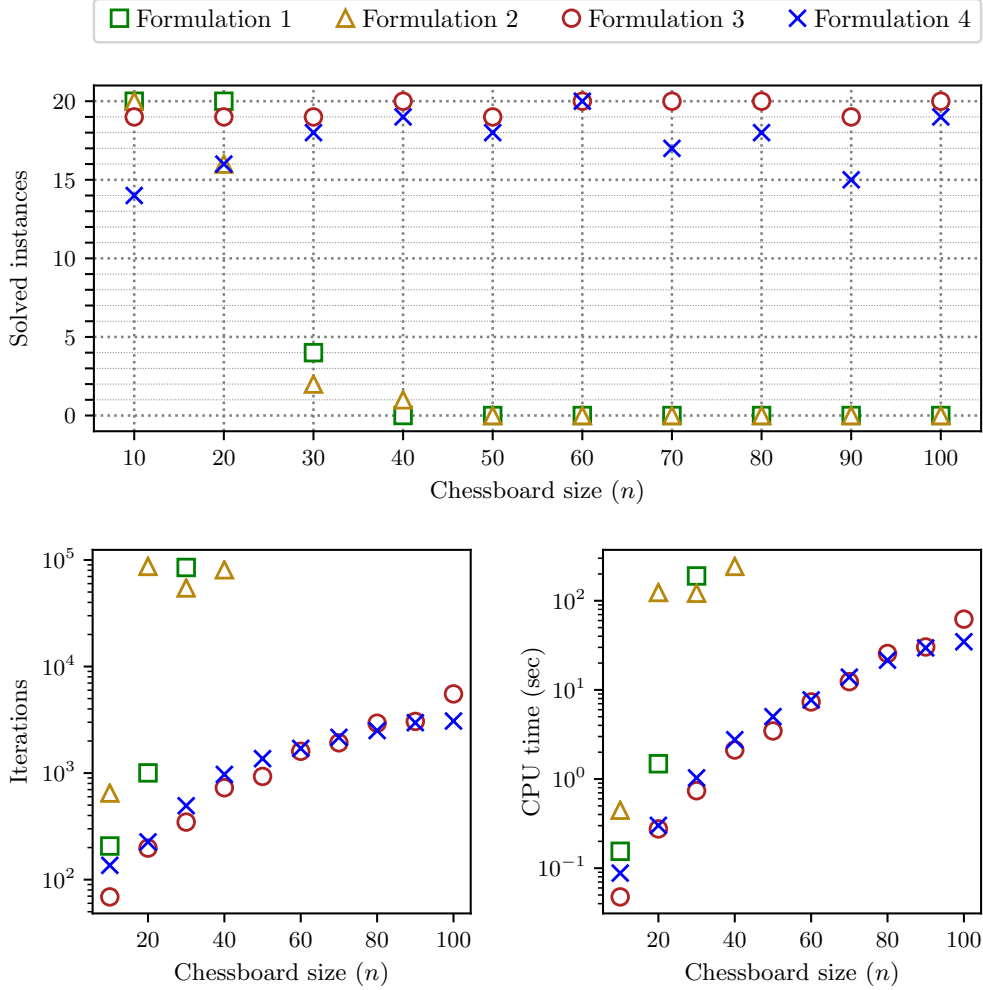
Figure 13: Results from the $(2, n)$-queens experiment. The top figure shows the number of instances (out of 20) solved by DR in less than 300 seconds. The bottom figures plots the average number of iterations (left) and average CPU time (right) over the solved instances for each formulation and each board size.

5 The infinite-dimensional case

In order to avoid the topological technicalities associated with working in an infinite dimensional space, we have so far restricted our setting to a Euclidean space (i.e., a finite dimensional Hilbert space). Nevertheless, most of the presented theory on fixed-point iterations and projection algorithms remain valid in a potentially infinite-dimensional Hilbert space, $\mathcal{H}$, with *strong convergence* of sequences replaced by *weak convergence*.

Definition 4 (Weak/strong convergence)**.** *Consider a sequence $(x_k)_{k=0}^{\infty} \subset \mathcal{H}$ and a point $x^\star \in \mathcal{H}$. Then $(x_k)_{k=0}^{\infty} \subset \mathcal{H}$ is said to:*

  (i) converge strongly to $x^\star$, denoted $x_k \to x^\star$, if $\|x_k - x^*\| \to 0$ as $k \to \infty$.

  (ii) converge weakly to $x^\star$, denoted $x_k \rightharpoonup x^\star$, if

$$\langle x_k, z \rangle \to \langle x^\star, z \rangle \quad \text{for all } z \in \mathcal{H}. \tag{33}$$

In general, strong convergence implies weak convergence, whereas the opposite is false. In fact, the *Kadec–Klee property* (Exercise 12) gives a precise contain between the two notions. Moreover, when $\mathcal{H}$ is finite dimensional, both notions coincide, as can be seen by applying (33) with z equal to each element of an orthonormal basis for the space.

Exercise 12 (Kadec–Klee property)**.** *Consider a sequence $(x_k) \subseteq \mathcal{H}$ and a point $x^\star \in \mathcal{H}$. Show that $x_k \to x^\star$ if and only if $x_k \rightharpoonup x^\star$ and $\|x_k\| \to \|x^\star\|$.*

The main technical difficulty arising from weak convergence is that projectors onto closed and convex sets are generally not weakly (sequentially) continuous, as the following example shows. Note there is no ambiguity here in using the term "closed" without qualification for convex sets, since a convex set is weakly (sequentially) closed if and only if it is strongly (sequentially) closed (Bauschke and Combettes, 2017, Theorem 3.34).

Example 3 (Projectors are not weakly continuous (Bauschke and Combettes, 2017, Example 4.20))**.** *Consider the infinite dimensional Hilbert space of square summable real sequences*

$$\mathcal{H} := \ell_2(\mathbb{N}) = \left\{ x = (x_i)_{i=1}^\infty \subset \mathbb{R} : \|x\|^2 = \sum_{i=1}^\infty x_i^2 < +\infty \right\},$$

endowed with the inner product $\langle x, y \rangle = \sum_{i=1}^\infty x_i y_i$. Denote its closed unit ball by $\mathbb{B} := \{x \in \mathcal{H} : \|x\| \leq 1\}$. Consider the bounded sequence $(e^k)_{k=2}^\infty \subset \mathcal{H}$ where, for all $k \geq 1$, the sequence $e^k = (e_i^k)_{i=1}^\infty \subset \mathbb{R}$ is given by

$$e_i^k := \begin{cases} 1 & \text{if } i \in \{1, k\}, \\ 0 & \text{otherwise.} \end{cases}$$

Then $e^k \rightharpoonup e^1$ as $k \to \infty$ and $P_\mathbb{B}(e^1) = e^1$. However, $P_\mathbb{B}(e^k) = \frac{1}{\sqrt{2}} e^k \rightharpoonup \frac{1}{\sqrt{2}} e^1 \neq P_\mathbb{B}(e^1)$. ◊

Exercise 13. *Verify the details of Example 3. That is, show that $e^k \rightharpoonup e^1$ but $P_\mathbb{B}(e^k) \not\rightharpoonup P_\mathbb{B}(e^1)$.*

As pointed out in Remark 1, when applying the DR algorithm, the sequence of interest is the shadow sequence $(P_A(x_k))_{k=0}^\infty$. For the reasons given in Example 3, weak convergence of this sequence cannot be directly derived from weak converge of $(x_k)_{k=0}^\infty$, since the projector P_A may not be weakly sequentially continuous. Nevertheless, Svaiter (2011) proved that the shadow sequence is indeed weakly convergent to $P_A(x^\star)$, where $x^\star$ denotes the weak limit of $(x_k)_{k=0}^\infty$. A simpler proof of this result was later provided by Bauschke (2013) by means of a *demiclosedness principle*.

5.1 A non-negative moment problem

Consider the infinite dimensional Hilbert space of square-integrable real-valued functions on the interval $[a, b]$ given by

$$\mathcal{H} := L^2([a, b]) = \left\{ x : [a, b] \to \mathbb{R} : \int_a^b x(t)^2 dt < \infty \right\},$$

endowed with the inner product $\langle x, y \rangle = \int_a^b x(t) y(t) dt$. Suppose that we are interested in finding a probability density function $x \in \mathcal{H}$ with mean μ and variance σ^2. Hence, we aim to find a non-negative function $x \in \mathcal{H}$ that satisfies the underdetermined linear system

$$\langle x, 1 \rangle = 1, \quad \langle x, t \rangle = \mu \quad \text{and} \quad \langle x, t^2 \rangle = \sigma^2 + \mu^2.$$

Defining $C_i := \{x \in \mathcal{H} : \langle x, t^{i-1} \rangle = c_i\}$ for $i = 1, 2, 3$ with $c_1 := 1$, $c_2 := \mu$ and $c_3 := \sigma^2 + \mu^2$, and $C_4 := \{x \in \mathcal{H} : x \geq 0\}$, we observe that the problem is an infinite dimensional modification of Example 1 (where we have added the nonnegativity constraint). The projectors onto C_i with $i = 1, 2, 3$ are given by

$$P_{C_i}(x)(t) = x(t) + \left(c_i - \langle t^{i-1}, x \rangle\right) \frac{t^{i-1}}{\|t^{i-1}\|^2} = x(t) + \left(c_i - \int_a^b t^{i-1} x(t) dt\right) \frac{(2i-1)t^{i-1}}{b^{2i-1} - a^{2i-1}},$$

whereas

$$P_{C_4}(x)(t) = \max\{0, x(t)\}.$$

In Figures 14(a) to 14(d), we show the result of applying four of the projection methods presented in Section 3.3 to the problem of finding a probability density function defined on $[0, 1]$, with mean $\mu = \frac{1}{2}$ and variance $\sigma^2 = \frac{1}{20}$, using $x_0 = 1$ as starting point for all the algorithms. These computations were performed symbolically, without discretizing the problem.
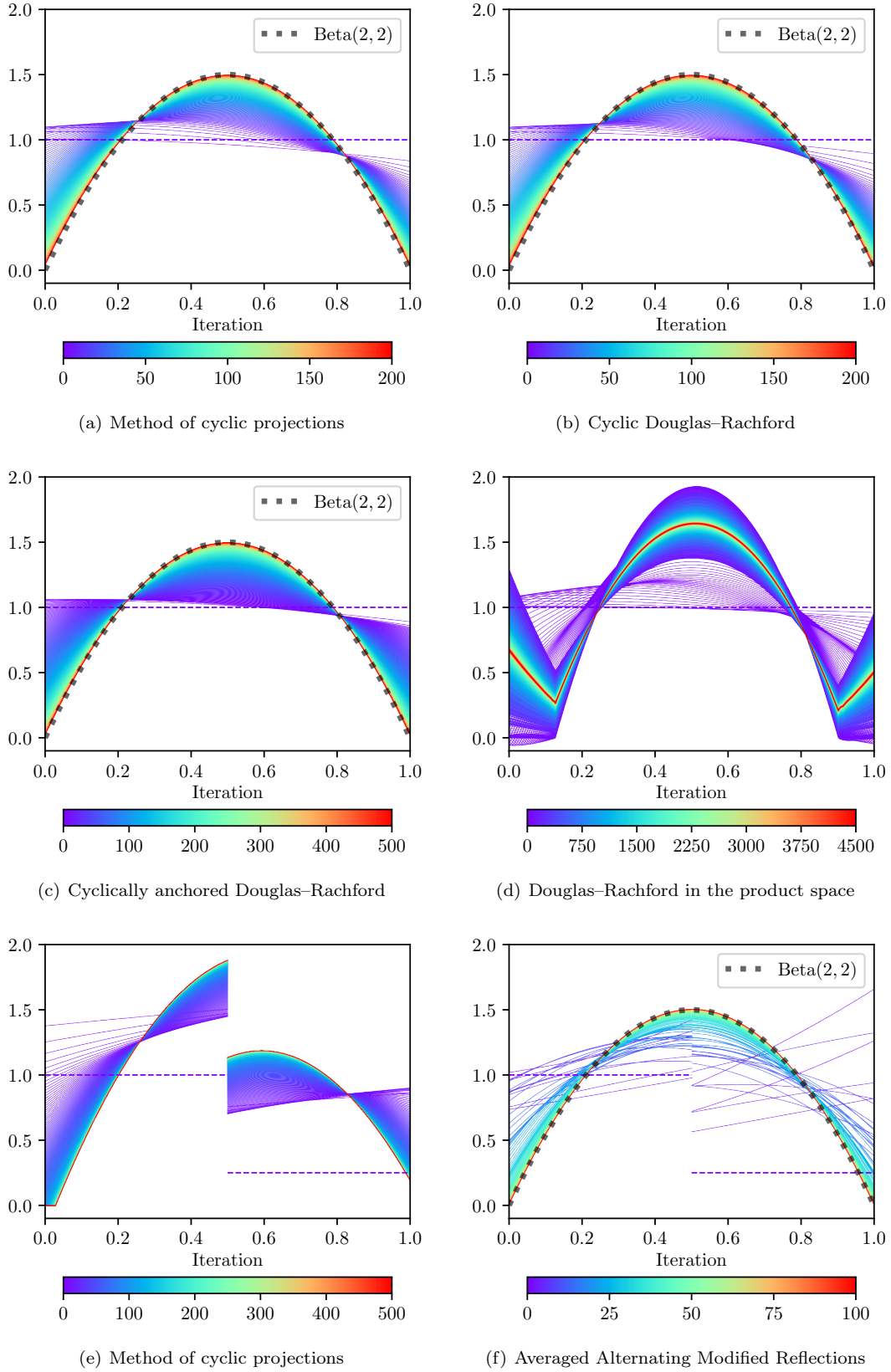
Figure 14: Sequences generated by various projection algorithms for the non-negative moment problem with $a = 0$, $b = 1$, $\mu = 1/2$ and $\sigma^2 = 1/20$. The starting point (function) is represented with a dashed line.

On one hand, we clearly observe that the method of cyclic projections, the cyclic DR algorithm, and the cyclically anchored DR algorithm all converge to a Beta distribution with parameters $(2, 2)$. Of these three methods, the cyclically anchored DR algorithm was the slowest. On the other hand, we realize that the behavior of the sequence generated by the DR algorithm in the product space is completely different to that of the other methods: the algorithm converges to a different (nonsmooth) solution to the problem. As it was also (by far) the slowest algorithm among the four, the DR algorithm in the product space seems to be the worst choice for this problem. Note none of these four algorithms guarantee strong convergence in this infinite dimensional setting.

In Fig. 14(e), we show the result of running the method of cyclic projections from a different starting point, given by

$$\widetilde{x}_0(t) := \begin{cases} 1 & \text{if } 0 \leq t \leq \frac{1}{2}, \\ \frac{1}{4} & \text{if } \frac{1}{2} < t \leq 1. \end{cases}$$

This time, the method converges to a different (discontinuous) solution. In Fig. 14(f), we present the result of running the averaged alternating modified reflections method in the product space, with $q = 0$ and parameters $\alpha = \beta = 0.95$, using $\widetilde{x}_0$ as a starting point. The sequence converges in much less iterations than the ones generated by the other methods, even when the continuous starting point $x_0 = 1$ was used. Moreover, in contrast with the other methods, the shadow sequence is now guaranteed to be strongly convergent to a fixed point of the operator whenever the set of fixed points is nonempty (according to Aragón and Campoy, 2018a, Theorem 5.1). If a fixed point exists, its projection onto the diagonal $\boldsymbol{D}$ would solve the best approximation problem consisting in finding the probability density function on $[0, 1]$ of minimum norm with $(\mu, \sigma^2) = \left(\frac{1}{2}, \frac{1}{20}\right)$.

We observe in Fig. 14(f) that the shadow sequence converges to the probability density function of a Beta$(2, 2)$. In Fig. 15 we present the sequences in the product space resulting from running again the averaged alternating modified reflections method from the continuous starting point $x_0 = 1$.

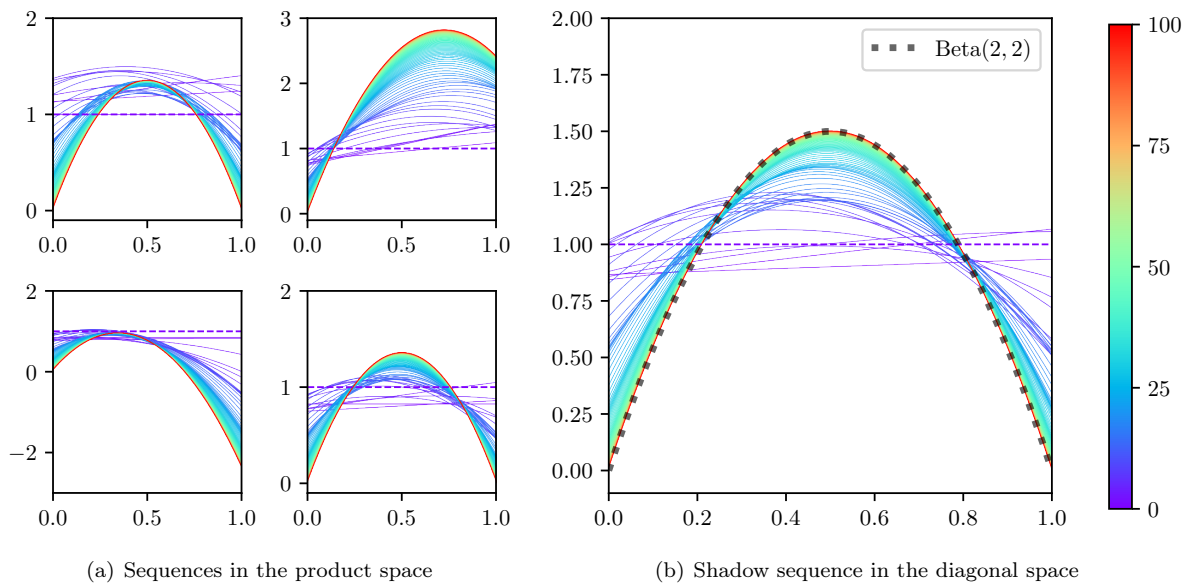(a) Sequences in the product space   (b) Shadow sequence in the diagonal space

Figure 15: Sequences generated by the averaged alternating modified reflections method in the product space for the non-negative moment problem with $a = 0$, $b = 1$, $\mu = 1/2$ and $\sigma^2 = 1/20$. The starting point (function) is represented with a dashed line.

It is apparent that each of the four sequences generated is convergent to some function in $\mathcal{H}$. Looking at the expression of the sequences, it can be observed that these limiting functions are

$$\overline{x}_1(t) := \frac{27}{5}t(1-t), \quad \overline{x}_2(t) := \frac{3}{5}t(13-9t), \quad \overline{x}_3(t) := \frac{3}{5}t(9-13t), \quad \overline{x}_4(t) := \frac{27}{5}t(1-t). \quad (34)$$

It can be checked that $\overline{\boldsymbol{x}} := (\overline{x}_1, \overline{x}_2, \overline{x}_3, \overline{x}_4) \in \boldsymbol{\mathcal{H}} = \mathcal{H}^4$ is a fixed point of the averaged alternating modified

reflections operator $T_{\boldsymbol{D},\boldsymbol{C},0.95,0.95}$ given by (25). Therefore, since

$$P_{\boldsymbol{D}}(\overline{\boldsymbol{x}})(t) = \frac{\overline{x}_1 + \overline{x}_2 + \overline{x}_3 + \overline{x}_4}{4} = 6t(1-t),$$

which coincides with the probability density function of a Beta$(2,2)$, we can conclude that Beta$(2,2)$ is the probability density function of minimum norm in $L^2([0,1])$ with $(\mu,\sigma^2) = \left(\frac{1}{2}, \frac{1}{20}\right)$.

Exercise 14. *Verify that $\overline{\boldsymbol{x}}$ given by* (34) *is a fixed point of $T_{\boldsymbol{D},\boldsymbol{C},0.95,0.95}$.*

6 Conclusions and open questions

In this self-contained tutorial, we introduced the feasibility problem paradigm and developed the convergence theory for projection algorithms within the framework of nonexpansive fixed point theory. The level of abstraction provided by this convergence framework allowed us to easily analyze several well-known projection algorithms in the literature. Particular attention was given to one member from the projection algorithm family, the Douglas–Rachford method, due to its known ability to solve certain combinatorial problems. By way of computational experiments on the $(2,n)$-queens problem, we demonstrated that choosing a good feasibility formulation is absolutely essential for this observed success. Although the product space formulation is a crucial ingredient of the DR method for feasibility problems with more than two constraint sets, we observed that it can actually make the method very slow, even in the convex setting with only four constraints as was considered in Section 5. We remark that the Douglas–Rachford projection algorithm, as was presented here, can be viewed as special case of an algorithm (which bears the same name) for finding a zero in the sum of two *maximal monotone operators*. This was, in fact, the algorithm developed by Lions and Mercier (1979). For a modern treatment, see (Bauschke and Combettes, 2017, §25.2).

To conclude this tutorial, we propose three open problems suitable as starting points for open-ended research projects, organized in order of decreasing difficulty.

- *An explicit counterexample for weak convergence* – As discussed in Section 5, in the infinite dimensional setting, the DR method is only known to converge weakly when applied to feasibility problems with convex sets. Despite this fact, to best of the authors' knowledge, there is no explicitly known example of two sets and an initial point for which the DR method converges weakly but not strongly. For the method of cyclic projections, a construction due to Hundal (2004) provides such a counterexample. The goal of this research project is to formulate an explicit counterexample for strong converge of the DR method.

- *Global convergence of DR for combinatorial problems* – As stated in Section 4.1, it is quite remarkable that the DR method applied to the (m,n)-queens problem works so well, especially given the absence of any applicable theory for the method in such settings. The goal of this research project is to find sufficient criteria for global convergence of the DR method involving more realistic nonconvex sets than the ones existing in the literature (e.g., involving the product space reformulation). Knowledge of such conditions would provide insights into general properties of useful feasibility formulations for combinatorial problems.

- *AAMR for combinatorial problems* – As demonstrated in Figure 11 and Remark 4, most projection algorithms possess fixed points which are not necessary related to the feasibility problem at hand. The DR method however does not suffer this potential shortcoming (Exercise 11) and nor does the AAMR discussed in Section 3.5. The latter suggests the AAMR method as a potentially useful heuristic. The goal of this project is to numerically compare the performance of the AAMR method to the DR method and GDR method on non-convex and combinatorial problems, such as the (m,n)-queens problem. A number of additional feasibility formulations for numerical testing can be found in Aragón et al. (2014a,b). Our initial numerical results for the $(2,n)$-queens problem suggest that the AAMR performs quite well (see Figure 16).
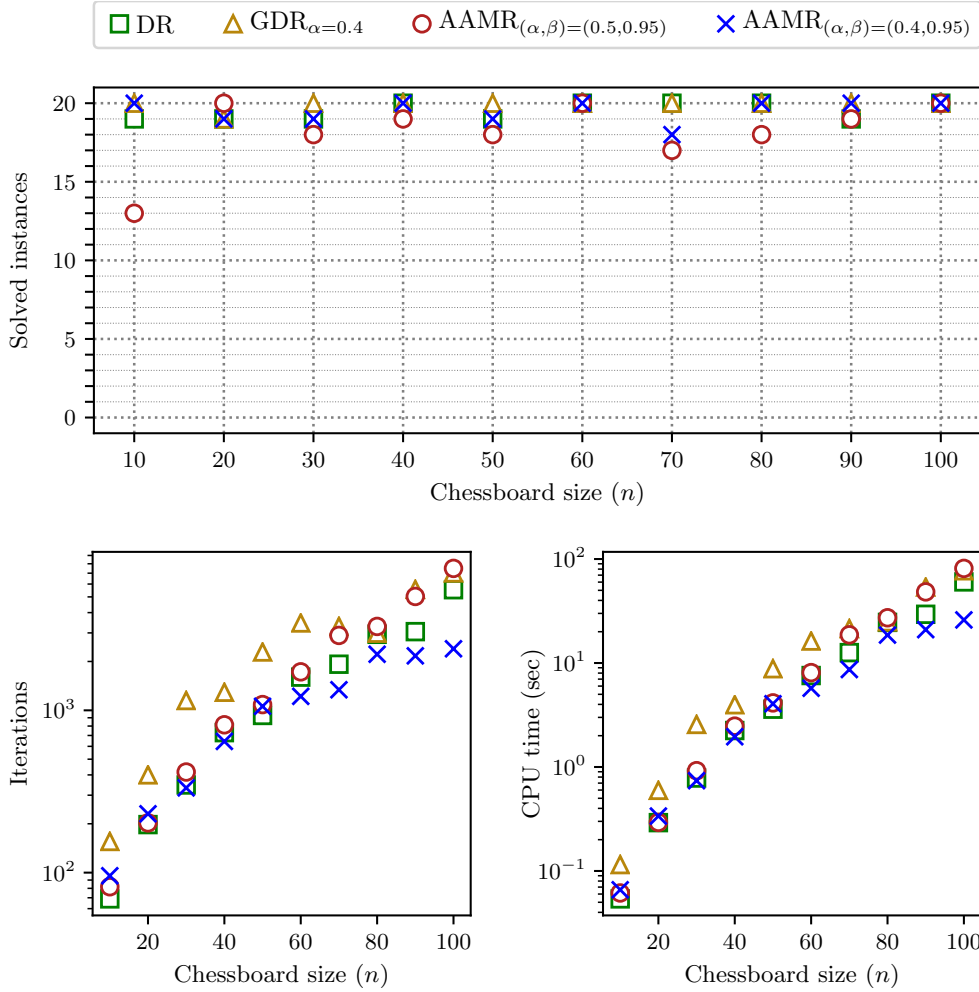
Figure 16: Results comparing DR, GDR and AAMR for the $(2, n)$-queens problem with Formulation 3. The top figure shows the number of instances (out of 20) solved in less than 300 seconds. The bottom figures plots the average number of iterations (left) and average CPU time (right) over the solved instances for each formulation and each board size.

References

Alwadani S, Bauschke HH, Moursi WM, Wang X (2018) On the asymptotic behaviour of the Aragón Artacho–Campoy algorithm. Oper Res Lett 46(6):585–587

Aragón Artacho FJ, Borwein JM (2013) Global convergence of a non-convex Douglas–Rachford iteration. J Glob Optim 57(3):753–769

Aragón Artacho FJ, Campoy R (2018a) A new projection method for finding the closest point in the intersection of convex sets. Comput Optim Appl 69(1):99–132

Aragón Artacho FJ, Campoy R (2018b) Solving graph coloring problems with the Douglas–Rachford algorithm. Set-Valued Var Anal 26(2):277–304

Aragón Artacho FJ, Borwein JM, Tam MK (2014a) Douglas–Rachford feasibility methods for matrix completion problems. ANZIAM J 55(4):299–326

Aragón Artacho FJ, Borwein JM, Tam MK (2014b) Recent results on Douglas–Rachford methods for combinatorial optimization problem. J Optim Theory Appl 163(1):1–30

Aragón Artacho FJ, Borwein JM, Tam MK (2016) Global behavior of the Douglas–Rachford method for a nonconvex feasibility problem. J Glob Optim 65(2):309–327

Aragón Artacho FJ, Campoy R, Elser V (2018a) An enhanced formulation for successfully solving graph coloring problems with the Douglas–Rachford algorithm. arXiv e-prints https://arxiv.org/abs/1808.01022

Aragón Artacho FJ, Campoy R, Kotsireas IS, Tam MK (2018b) A feasibility approach for constructing combinatorial designs of circulant type. J Comb Optim 35(4):1061–1085

Aragón Artacho FJ, Censor Y, Gibali A (2018c) The cyclic Douglas–Rachford algorithm with r-sets-Douglas–Rachford operators. Optim Methods Softw DOI: 10.1080/10556788.2018.1504049

Baillon JB, Bruck RE, Reich S (1978) On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces. Houston J Math 4(1):1–9

Banach S (1922) Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales. Fund math 3(1):133–181

Bauschke HH (2013) New demiclosedness principles for (firmly) nonexpansive operators. In: Computational and Analytical Mathematics, Springer, pp 19–28

Bauschke HH, Combettes PL (2017) Convex Analysis and Monotone Operator Theory in Hilbert Spaces, 2nd edn. Springer

Bauschke HH, Dao MN (2017) On the finite convergence of the Douglas–Rachford algorithm for solving (not necessarily convex) feasibility problems in Euclidean spaces. SIAM J Optim 27(1):507–537

Bauschke HH, Moursi WM (2017) On the Douglas–Rachford algorithm. Math Program, Ser A 164(1–2):263–284

Bauschke HH, Noll D (2014) On the local convergence of the Douglas–Rachford algorithm. Arch Math 102(6):589–600

Bauschke HH, Combettes PL, Luke DR (2002) Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization. J Opt Soc Am A 19(7):1334–1345

Bauschke HH, Combettes PL, Luke DR (2004) Finding best approximation pairs relative to two closed convex sets in Hilbert spaces. J Approx Theory 127(2):178–192

Bauschke HH, Bello Cruz JY, Nghia TT, Phan HM, Wang X (2014) The rate of linear convergence of the Douglas–Rachford algorithm for subspaces is the cosine of the Friedrichs angle. J Approx Theory 185:63–79

Bauschke HH, Noll D, Phan HM (2015) Linear and strong convergence of algorithms involving averaged nonexpansive operators. J Math Anal Appl 421(1):1–20

Bauschke HH, Lukens B, Moursi WM (2017) Affine nonexpansive operators, Attouch–Théra duality and the Douglas–Rachford algorithm. Set-Valued Var Anal 25(3):481–505

Bauschke HH, Dao MN, Lindstrom SB (2019) The Douglas–Rachford algorithm for a hyperplane and a doubleton. J Glob Optim pp 1–15, DOI: 10.1007/s10898-019-00744-7

Behling R, Bello Cruz JY, Santos L (2018) Circumcentering the Douglas–Rachford method. Numer Algor 78(3):759–776

Benoist J (2015) The Douglas–Rachford algorithm for the case of the sphere and the line. J Global Optim 63(2):363–380

Borwein JM, Sims B (2011) The Douglas–Rachford algorithm in the absence of convexity. In: Bauschke H, Burachik R, Combettes P, Elser V, Luke D, Wolkowicz H (eds) Fixed-Point Algorithms for Inverse Problems in Science and Engineering, Springer Optimization and Its Applications, vol 49, Springer, New York, pp 93–109

Borwein JM, Tam MK (2014) A cyclic Douglas–Rachford iteration scheme. J Optim Theory Appl 160(1):1–29

Borwein JM, Tam MK (2015) The cyclic Douglas–Rachford method for inconsistent feasibility problems. J Nonlinear Convex Anal 16(4):573–584

Borwein JM, Tam MK (2017) Reflection methods for inverse problems with applications to protein conformation determination. In: Aussel D, Lalitha C (eds) Generalized Nash Equilibrium Problems, Bilevel Programming and MPEC. Forum for Interdisciplinary Mathematics, Springer Singapore, Singapore, pp 83–100

Borwein JM, Lindstrom SB, Sims B, Schneider A, Skerritt MP (2018) Dynamics of the Douglas–Rachford method for ellipses and p-spheres. Set-Valued Var Anal 26(2):385–403

Bregman LM (1965) The method of successive projection for finding a common point of convex sets. Soviet Math Dokl 162(3):688–692

Cegielski A (2012) Iterative Methods for Fixed Point Problems in Hilbert Spaces, Lecture Notes in Mathematics, vol 2057. Springer

Censor Y (1984) Iterative methods for convex feasibility problems. Ann Discrete Math 20:83–91

Censor Y, Cegielski A (2015) Projection methods: an annotated bibliography of books and reviews. Optimization 64(11):2343–2358

Dao MN, Tam MK (2019a) A Lyapunov-type approach to convergence of the Douglas–Rachford algorithm for a nonconvex setting. J Glob Optim 73(1):83–112

Dao MN, Tam MK (2019b) Union averaged operators with applications to proximal algorithms for min-convex functions. J Optim Theory and Appl pp 1–34

Deutsch F (2001) Best Approximation in Inner Product Spaces, CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, vol 7. Springer-Verlag

Douglas J, Rachford HH (1956) On the numerical solution of heat conduction problems in two and three space variables. Trans Amer Math Soc 82:421–439

Eckstein J, Bertsekas DP (1992) On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math Program 55(1):293–318

Elser V (2003) Phase retrieval by iterated projections. J Opt Soc Am A 20(1):40–55

Elser V (2018) The complexity of bit retrieval. IEEE Transactions on Information Theory 64(1):412–428

Elser V, Rankenburg I, Thibault P (2007) Searching with iterated maps. Proc Natl Acad Sci 104(2):418–423

Halperin I (1962) The product of projection operators. Acta Sci Math 23:96–99

Hesse R, Luke DR (2013) Nonconvex notions of regularity and convergence of fundamental algorithms for feasibility problems. SIAM J Optim 23(4):2397–2419

Hesse R, Luke DR, Neumann P (2014) Alternating projections and Douglas–Rachford for sparse affine feasibility. IEEE Transactions on Signal Processing 62(18):4868–4881

Hundal HS (2004) An alternating projection that does not converge in norm. Nonlin Anal: Theory, Methods & Appl 57(1):35–61

Kaczmarz S (1937) Angenäherte Auflösung von Systemen linearer Gleichungen. Bull Int Acad Sci Pologne, A 35:355–357

Lamichhane BP, Lindstrom SB, Sims B (2017) Application of projection algorithms to differential equations: boundary value problems. arXiv e-prints https://arxiv.org/abs/1705.11032

Lindstrom SB, Sims B (2018) Survey: Sixty years of Douglas–Rachford. arXiv e-prints https://arxiv.org/abs/1809.07181

Lions PL, Mercier B (1979) Splitting algorithms for the sum of two nonlinear operators. SIAM J Numer Anal 16(6):964–979

Luke DR (2008) Finding best approximation pairs relative to a convex and a prox-regular set in a Hilbert space. SIAM J Optim 19(2):714–739

von Neumann J (1950) Functional Operators II: The Geometry of Orthogonal Spaces. Princeton University Press

Opial Z (1967) Weak convergence of the sequence of successive approximations for nonexpansive mappings. Bulletin of the American Mathematical Society 73(4):591–597

Pazy A (1971) Asymptotic behavior of contractions in Hilbert space. Israel J Math 9:235–240

Phan HM (2016) Linear convergence of the Douglas–Rachford method for two closed sets. Optim 65(2):369–385

Pierra G (1984) Decomposition through formalization in a product space. Math Program 28:96–115

Schaad J (2010) Modeling the 8-queens problem and Sudoku using an algorithm based on projections onto non-convex sets. Master's thesis, University of British Columbia

Svaiter BF (2011) On weak convergence of the Douglas–Rachford method. SIAM J Control Optim 49(1):280–287

Tam MK (2018) Algorithms based on unions of nonexpansive maps. Optim Letters 12(5):1019–1027

Thao NH (2018) A convergent relaxation of the Douglas–Rachford algorithm. Comput Optim Appl 70(3):841–863