# The Fun is Finite: Douglas–Rachford and Sudoku Puzzle — Finite Termination and Local Linear Convergence

Robert Tovey  Jingwei Liang*  and  Carola-Bibiane Schönlieb

DAMTP, University of Cambridge, Cambridge CB3 0WA, UK

**Abstract.** In recent years, the Douglas–Rachford splitting method has been shown to be effective at solving many non-convex optimization problems. In this paper we present a local convergence analysis for non-convex feasibility problems and show that both finite termination and local linear convergence are obtained. For a generalization of the Sudoku puzzle, we prove that the local linear rate of convergence of Douglas–Rachford is exactly $\frac{\sqrt{5}}{5}$ and independent of puzzle size. For the $s$-queens problem we prove that Douglas–Rachford converges after a finite number of iterations. Numerical results on solving Sudoku puzzles and $s$-queens puzzles are provided to support our theoretical findings.

**Key words.** Douglas–Rachford · Feasibility problem · Sudoku Puzzle · Finite Termination Local Linear Convergence

**AMS subject classifications.** 49J52 · 65K05 · 65K10 · 90C25

## 1 Introduction

Given two non-empty sets $C$ and $S$ whose intersection is also non-empty, the feasibility problem aims to find a common point in the intersection $C \cap S$. In the literature, popular numerical schemes for solving feasibility problems are developed based on projection, among them alternating projection is the fundamental one. The method of alternating projection was first introduced by von Neumann for the case of two linear subspaces [29], then was extended to closed convex sets by Bregman [11]. Relaxation is a standard approach to speed up alternating projection and related work can be found in [12, 22].

Proximal splitting methods, such as Forward–Backward/Backward–Backward splitting [13, 21] and Peaceman–Rachford/Douglas–Rachford splitting [14, 25], can also be applied to solve feasibility problem either directly or up to reformulation. Moreover, equivalence between projection based methods and proximal splitting methods can be established, such as alternating projection is equivalent to Backward–Backward splitting while relaxed alternating relaxed projection covers Peaceman–Rachford/Douglas–Rachford splitting as special cases [13].

Our focus in this paper is the Douglas-Rachford splitting method, which has shown to be effective for solving feasibility problem, particularly in the non-convex setting [7]. However, the convergence property is rather less understood than its convex counter part. One reason for this is that Douglas–Rachford splitting method is not symmetric and non-descent, when compared to (proximal) gradient descent whose non-convex case is much better studied [4]. Research on non-convex Douglas–Rachford either focuses on specific cases or imposing stronger assumptions (*e.g.* smoothness) and proposes modifications to the original iteration. For instance [1] considers Douglas–Rachford splitting for solving feasibility problem of a line intersecting with a circle, and conditions for convergence are provided. In [19], the authors proposed a damped Douglas–Rachford splitting method for general non-convex optimization problem under the condition that one function has a Lipschitz continuous gradient.

The study of this paper is motivated by applying Douglas–Rachford splitting to solve Sudoku puzzle[1], for which three different convergence behaviors are observed

- Globally, the method converges sub-linearly.
- Locally, two convergence regimes occur: finite termination and local linear convergence.

Finite termination and local linear convergence are reported in the literature [7, 10], however, conditions in respective work either are designed for convex setting or cannot be satisfied by Sudoku puzzle. Therefore, a new analysis is needed for Douglas–Rachford splitting which is the aim of this paper:

*E-mail: jl993@cam.ac.uk.

[1] https://en.wikipedia.org/wiki/Sudoku

1. **Finite termination** Under a non-degeneracy condition, see (4.1), we show in Section 4 that one sequence generated by Douglas–Rachford splitting has the finite termination property. All sequences terminate in a finite number of iterations if the problem satisfies certain assumptions (*e.g.* polyhedrality, see Assumptions (**A.1**)-(**A.3**)).
2. **Local linear convergence** We also provide a precise characterization for the local linear convergence of Douglas–Rachford splitting method. Particularly, for Sudoku puzzle, we prove that locally the linear rate of convergence of Douglas–Rachford splitting method is precisely $\frac{\sqrt{5}}{5}$. Moreover, such a rate is independent of puzzle size. For the damped Douglas–Rachford splitting method, we also provide an exact estimation of the local linear rate which depends on the damping coefficient.

**Relation to Prior Work** There are several existing work studying the finite termination property of the standard Douglas–Rachford splitting method. In [10], the authors established finite convergence of Douglas–Rachford in the presence of Slater's condition, for solving convex feasibility problems where one set is an affine subspace and the other is a polyhedron, or one set is an epigraph and the other one is a hyperplane. The result was extended to general convex optimization problems in [20] under the notion of partial smoothness [18]. In [23], finite termination is proved for finding a point which is guaranteed to be in the interior of one set whose interior is assumed to be non-empty. The result of [10] was later extended to the non-convex case in [7], where one of the two sets can be finite.

For local linear convergence, results can be found in for instance [26] where linear convergence of Douglas–Rachford splitting method is established under a regularity condition. Similar results can be found in [15, 16]. Under a constraint qualification condition, [19] also discussed the local linear convergence property of the damped Douglas–Rachford splitting method.

**Paper Organization** The rest of the paper is organized as follows. Some preliminaries are collected in Section 2. Section 3 states our main assumptions on problem (3.1) and introduces the standard and damped Douglas–Rachford algorithms, global convergence is also discussed. Our main result on local convergence of Douglas–Rachford is presented in Section 4. In Section 5, we report numerical experiments on Sudoku puzzle and $s$-queens puzzle to support our theoretical findings.

## 2 Preliminaries

Throughout the paper, $\mathbb{N}$ is the set of nonnegative integers, $\mathbb{R}^n$ is a $s$-dimensional real Euclidean space equipped with scalar product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$. Id denotes the identity operator on $\mathbb{R}^n$. For a matrix $M \in \mathbb{R}^{n\times n}$, we denote $\rho(M)$ its spectral radius.

**Projection and reflection** Below we collect necessary concepts related to sets.

**Definition 2.1 (Distance and indicator function).** Let $C \subset \mathbb{R}^n$ be non-empty and $x \in \mathbb{R}^n$. The *distance function* of $x$ to $C$ is defined by

$$\mathrm{dist}(x, C) \stackrel{\text{def}}{=} \inf_{y\in C} \|x - y\|.$$

The *indicator function* of $C$ is defined by

$$\iota_C(x) = \begin{cases} 0 : x \in C, \\ +\infty : x \notin C. \end{cases}$$

**Definition 2.2 (Projection and reflection).** Let $C \subset \mathbb{R}^n$ be non-empty and $x \in \mathbb{R}^n$. The projection of $x$ onto $C$, denoted by $\mathscr{P}_C(x)$, is a set defined by

$$\mathscr{P}_C(x) \stackrel{\text{def}}{=} \big\{y \in C : \|x - y\| = \mathrm{dist}(x, C)\big\}.$$

The mapping $\mathscr{P}_C : \mathbb{R}^n \rightrightarrows C$ is called the *projection operator*. The *relaxed projection* $\mathscr{P}_C^\lambda$ is defined via

$$\mathscr{P}_C^\lambda(x) = \lambda\mathscr{P}_C(x) + (1 - \lambda)x,$$

where $\lambda \in ]0, 2]$ is the relaxation parameter. When $\lambda = 2$, the corresponding mapping is called *reflection* and denoted by $\mathscr{R}_C(x) = 2\mathscr{P}_C(x) - x$.

**Definition 2.3 (Prox-regularity).** A non-empty closed set $C \subset \mathbb{R}^n$ is *prox-regular* at $x \in C$ for $v$ if $x = \mathscr{P}_C(x + v)$. If $\mathscr{P}_C$ is single-valued in an open neighborhood of $x \in C$, $C$ is called *prox-regular* at $x$.

**Definition 2.4 (Normal vector).** Given $C \subset \mathbb{R}^n$ and $x \in C$, the *proximal normal cone* $\mathscr{N}_C^p(x)$ of $C$ at $x$ is defined by

$$\mathscr{N}_C^p(x) = \mathrm{cone}\big(\mathscr{P}_C^{-1}(x) - x\big).$$

The *limiting normal cone* $\mathscr{N}_C(x)$ is defined as any vector that can be written as the limit of proximal normals: $v \in \mathscr{N}_C(x)$ if and only if there exists sequences $\{x_k\}_{k\in\mathbb{N}} \in C$ and $\{v_k\}_{k\in\mathbb{N}}$ in $\mathscr{N}_C(x_k)$ such that $x_k \to x$ and $v_k \to v$.

Let $C_1, C_2 \subset \mathbb{R}^n$ be two sets with non-empty intersection. The feasibility problem of $C_1, C_2$ is to find a common point in the intersection, *i.e.*

$$\text{find} \quad x \in \mathbb{R}^n \quad \text{s.t.} \quad x \in C_1 \bigcap C_2.$$

A fundamental algorithm to solve the problem is the alternating projection method which, as indicated by the name, represents the procedure: from a given point $x_0$, apply projection onto each set alternatively

$$x_{k+1} = \mathscr{P}_{C_2}\mathscr{P}_{C_1}(x_k). \tag{2.1}$$

One can also consider relaxation for each projection operator and the whole iteration, which results in the following iteration

$$x_{k+1} = x_k + \lambda\big(\mathscr{P}_{C_2}^{\lambda_2}\mathscr{P}_{C_1}^{\lambda_2}(x_k) - x_k\big),$$

where $\lambda, \lambda_1, \lambda_2$ are relaxation parameters. The above scheme becomes Peaceman–Rachford splitting (alternating reflection) for $(\lambda, \lambda_1, \lambda_2) = (1, 2, 2)$ and Douglas–Rachford splitting for $(\lambda, \lambda_1, \lambda_2) = (1/2, 2, 2)$. We refer to [6] for a survey on the alternating projection method.

**Convergent Matrices** To discuss the local linear convergence, we need the following preliminary results on convergent matrices which are taken from [24].

**Definition 2.5 (Convergent matrices).** A matrix $M \in \mathbb{R}^{n\times n}$ is *convergent* to $\mathrm{M}^\infty \in \mathbb{R}^{n\times n}$ if, and only if,

$$\lim_{k\to+\infty} \|M^k - \mathrm{M}^\infty\| = 0.$$

$M$ is said to be *linearly convergent* if there exists $\eta \in [0, 1[$ and $K \in \mathbb{N}$ such that for all $k \geq K$, there holds $\|M^k - \mathrm{M}^\infty\| = O(\eta^k)$. If $M$ does not converge at any rate $\eta' \in [0, \eta[$ then $\eta$ is called the *optimum convergence rate*.

**Definition 2.6 (Semi-simple eigenvalue).** For $M \in \mathbb{R}^{n\times n}$, an eigenvalue $\eta$ is called *semi-simple* if and only if $\mathrm{rank}(M - \eta\mathrm{Id}) = \mathrm{rank}((M - \eta\mathrm{Id})^2)$.

**Theorem 2.7 (Limits of powers).** *For $M \in \mathbb{R}^{n\times n}$, the power of $M$ converges to $\mathrm{M}^\infty$ if and only if $\rho(M) < 1$ or $\rho(M) = 1$ with $1$ being the only eigenvalue on the complex unit circle and semi-simple.*

Whenever $M$ is convergent, it converges linearly to $\mathrm{M}^\infty$, and we have the following lemma.

**Lemma 2.8 (Convergence rate).** *Suppose $M \in \mathbb{R}^{n\times n}$ is convergent to some $\mathrm{M}^\infty \in \mathbb{R}^{n\times n}$, then*

(i) *for any $k \in \mathbb{N}$,*

$$M^k - \mathrm{M}^\infty = (M - \mathrm{M}^\infty)^k \quad \text{and} \quad \|M^k - \mathrm{M}^\infty\| \leq \|M - \mathrm{M}^\infty\|^k.$$

*The equality holds only when $M$ is normal.*

(ii) *We have $\rho(M - \mathrm{M}^\infty) < 1$, and $M$ is linearly convergent for any $\eta \in ]\rho(M - \mathrm{M}^\infty), 1[$.*

(iii) *$\rho(M - \mathrm{M}^\infty)$ is the optimal convergence rate if one of the following holds*

  (a) *$M$ is normal.*

  (b) *All the eigenvalues $\eta \in \Theta_M$ such that $|\eta| = \rho(M - \mathrm{M}^\infty)$ are semi-simple.*

**Proof.** See Theorems 2.12, 2.13, 2.15 and 2.16 of [9]. □

**Angles between Subspaces** To precisely characterize the local linear rate of convergence, we need the following concepts regarding the angles between subspaces. Let $T_1$ and $T_2$ be two linear subspaces with dimension $p \overset{\text{def}}{=} \dim(T_1)$ and $q \overset{\text{def}}{=} \dim(T_2)$, and without loss of generality, suppose that $1 \leq p \leq q \leq n-1$.

**Definition 2.9 (Principal angles).** The *principal angles* $\theta_k \in [0, \frac{\pi}{2}]$, $k = 1, \ldots, p$ between linear subspaces $T_1$ and $T_2$ are defined by, with $u_0 = v_0 \stackrel{\text{def}}{=} 0$ and inductively

$$\begin{aligned}\cos(\theta_k) \stackrel{\text{def}}{=} \langle u_k,\, v_k\rangle = \max\big\{\langle u,\, v\rangle\ \text{ s.t. }\ & u \in T_1, v \in T_2, \|u\| = 1, \|v\| = 1, \\ & \langle u,\, u_i\rangle = \langle v,\, v_i\rangle = 0,\, i = 0, \cdots, k-1\big\}.\end{aligned}$$

The principal angles $\theta_k$ are unique with $0 \leq \theta_1 \leq \theta_2 \leq \cdots \leq \theta_p \leq \pi/2$.

**Definition 2.10 (Friedrichs angle).** The *Friedrichs angle* $\theta_F \in [0, \frac{\pi}{2}]$ between $T_1$ and $T_2$ is

$$\cos\big(\theta_F(T_1, T_2)\big) \stackrel{\text{def}}{=} \max\langle u,\, v\rangle\ \text{ s.t. }\ u \in T_1 \cap (T_1 \cap T_2)^{\perp}, \|u\| = 1,\ \ v \in T_2 \cap (T_1 \cap T_2)^{\perp}, \|v\| = 1.$$

The following lemma shows the relation between the Friedrichs and principal angles.

**Lemma 2.11 ([9, Proposition 3.3]).** *We have* $\theta_F(T_1, T_2) = \theta_{d+1} > 0$ *where* $d \stackrel{\text{def}}{=} \dim(T_1 \cap T_2)$.

**Remark 2.12.** Singular value decomposition (SVD) can serve an efficient approach to obtain the principal angles. For instance, let $X \in \mathbb{R}^{n\times p}$ and $Y \in \mathbb{R}^{n\times q}$ form orthonormal bases for the subspaces $T_1$ and $T_2$ respectively. Let $U\Sigma V^T$ be the SVD of the matrix $X^TY \in \mathbb{R}^{p\times q}$, then $\cos(\theta_k) = \sigma_k$, $k = 1, 2, \ldots, p$ and $\sigma_k$ corresponds to the $k$'th largest singular value in $\Sigma$.

# 3 Problem and algorithm

The formal statement of the feasibility problem is written below

$$\text{find} \quad x \in \mathbb{R}^n \quad \text{s.t.} \quad x \in C \bigcap S, \tag{3.1}$$

where the following assumptions are imposed

**(A.1)** $C \subset \mathbb{R}^n$ is a closed set;
**(A.2)** $S \subset \mathbb{R}^n$ is an affine subspace;
**(A.3)** $C \bigcap S \neq \emptyset$, *i.e.* the intersection is non-empty.

Note that the problem (3.1) is not necessarily convex as we suppose $C$ is only non-empty and closed. Examples of (3.1) are provided in Section 4, including the Sudoku puzzle and $s$-queens puzzle.

## 3.1 Douglas–Rachford splitting method

The development of Douglas–Rachford (DR) splitting method [14] dates back to 1950s for solving numerical PDEs. In recently years, Douglas–Rachford splitting method has also been shown to be effective for non-convex feasibility problem [2, 17]. Details of the method for solving (3.1) is described in Algorithm 1.

**Algorithm 1:** Standard Douglas–Rachford splitting (DR)

**Initial**: $z_0 \in \mathbb{R}^n$;
**repeat**

$$\begin{aligned} x_{k+1} &= \mathscr{P}_S(z_k), \\ u_{k+1} &\in \mathscr{P}_C(2x_{k+1} - z_k), \\ z_{k+1} &= z_k + u_{k+1} - x_{k+1}, \end{aligned} \tag{3.2}$$

**until** *convergence*;

The above iteration can be written as the fixed-point iteration of variable $z_k$. Denote the fixed-point operator

$$\mathscr{F}_{\text{DR}} \stackrel{\text{def}}{=} \tfrac{1}{2}\big((2\mathscr{P}_C - \text{Id})(2\mathscr{P}_S - \text{Id}) + \text{Id}\big), \tag{3.3}$$

then we have $z_{k+1} = \mathscr{F}_{\text{DR}}(z_k)$. The other two variables $u_k, x_k$ are called the *shadow sequences* [8].

Determining the convergence properties of Douglas–Rachford splitting method for the non-convex setting is a challenging problem, the non-descent property of the method makes it much harder to obtain convergence result than the descent-type methods which includes (proximal) gradient descent [4].

Moreover, since the method has three different sequences $u_k, x_k$ and $z_k$, many different convergence behaviors may occur. We refer to [7] for a more detailed discussion but Example 3.2 presents the case of a circle intersecting with a line where:

- The shadow sequences $\{u_k\}_{k\in\mathbb{N}}$, $\{x_k\}_{k\in\mathbb{N}}$ are converge to $u^\star$ and $x^\star$ respectively.
- The fixed-point sequence $\{z_k\}_{k\in\mathbb{N}}$ diverges.
- The limiting points of $\{u_k\}_{k\in\mathbb{N}}$ and $\{x_k\}_{k\in\mathbb{N}}$ are not the same, $u^\star \neq x^\star$.

As our main interest in this paper is to study the local convergence behavior, for the rest of the paper, we suppose that the standard DR is globally convergent:

(**A.4**) The standard Douglas–Rachford splitting method for solving (3.1) is globally convergent.

Consequently, one has

$$z_k \to z^\star \in \mathrm{Fix}(\mathscr{F}_{\mathrm{DR}}) \stackrel{\mathrm{def}}{=} \big\{z \in \mathbb{R}^n : z = \mathscr{F}_{\mathrm{DR}}(z)\big\} \quad \text{and} \quad u_k, x_k \to x^\star \in \mathscr{P}_S(z^\star).$$

To avoid assumption (**A.4**), people either turn to specific cases [3] or imposing stronger assumptions such as smoothness [28]. Modifications to the original Douglas–Rachford splitting method are also considered in the literature. Below we describe a damped version of Douglas–Rachford proposed in [19].

Solving the feasibility problem (3.1) is equivalent to the following constrained smooth optimization.

$$\min_{x\in\mathbb{R}^n} \tfrac{1}{2}\mathrm{dist}^2(x, S) \quad \text{s.t.} \quad x \in C. \tag{3.4}$$

In (3.2), the update of $x_{k+1}$ is equivalent to solve the optimization problem $\min_{x\in\mathbb{R}^n} \iota_S(x) + \frac{1}{2\gamma}\|x - z_k\|^2$. Replacing the indicator function with the distance function,

$$\min_{x\in\mathbb{R}^n} \tfrac{1}{2}\mathrm{dist}^2(x, S) + \tfrac{1}{2\gamma}\|x - z_k\|^2,$$

we then get

$$x_{k+1} = \tfrac{1}{1+\gamma}\big(z_k + \gamma\mathscr{P}_S(z_k)\big) = z_k + \tfrac{\gamma}{1+\gamma}\big(\mathscr{P}_S(z_k) - z_k\big) = \mathscr{P}_S^{\frac{\gamma}{1+\gamma}}(z_k).$$

As a result, we obtain the algorithm proposed in [19].

**Algorithm 2:** A damped Douglas–Rachford splitting (dDR)

**Initial**: $\gamma > 0$, $z_0 \in \mathbb{R}^n$;
**repeat**

$$\begin{aligned} x_{k+1} &= \mathscr{P}_S^{\frac{\gamma}{1+\gamma}}(z_k), \\ u_{k+1} &\in \mathscr{P}_C(2x_{k+1} - z_k), \\ z_{k+1} &= z_k + u_{k+1} - x_{k+1}, \end{aligned} \tag{3.5}$$

**until** *convergence*;

We refer to the original work [19] for a more detailed discussion of Algorithm 2. When $\gamma = +\infty$, Algorithm 2 recovers the standard Douglas–Rachford splitting method (3.2). The fixed-point operator of dDR reads

$$\mathscr{F}_{\mathrm{dDR}} \stackrel{\mathrm{def}}{=} \tfrac{1}{2}\big((2\mathscr{P}_C - \mathrm{Id})(2\mathscr{P}_S^{\frac{\gamma}{1+\gamma}} - \mathrm{Id}) + \mathrm{Id}\big). \tag{3.6}$$

We have the following convergence result of dDR from [19].

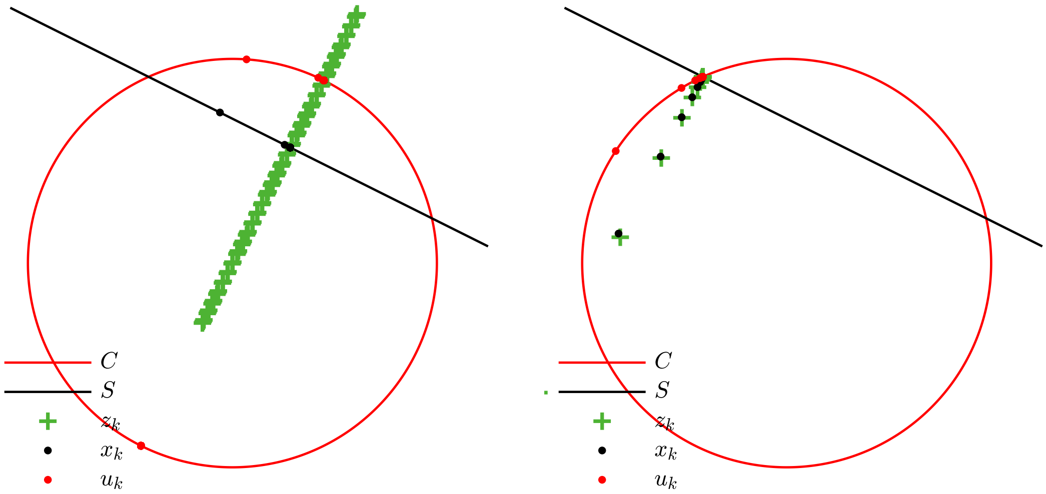
**Lemma 3.1 (Global convergence of dDR [19, Theorem 5]).** *For the non-convex feasibility problem (3.1), suppose Assumptions (**A.1**)-(**A.3**) hold and moreover $C$ is compact. Choose $\gamma \in ]0, \sqrt{3/2} - 1[$ for the Douglas–Rachford splitting method (3.5), then the sequence $\{u_k, x_k, z_k\}_{k\in\mathbb{N}}$ is bounded, and given any cluster point $(u^\star, x^\star, z^\star)$ of the sequence, there holds $\|z_k - z_{k-1}\| \to 0$, $u^\star = x^\star$ and $x^\star$ is a stationary point of the problem (3.4).*

In the example below, we demonstrate a case where DR fails to solve the problem while dDR succeeds.

**Example 3.2 (A circle intersects with a line).** Let $C = \{x \in \mathbb{R}^2 : \|x\| = 1\}$ be the unit circle and $S = \{x \in \mathbb{R}^2 : \langle x, (\frac{1}{2})\rangle = \sqrt{2}\}$ be a line that intersects with $C$ at two different points. For both methods, same initial point $z_0 = (-10, -8)$ is chosen. In Figure 1 we observe:

- For the standard DR (left): $z_k$ is not convergent, $u_k$ and $x_k$ converge to two different points and the method fails to find a feasible point.
- For the damped DR (right): all three sequences converge to the same feasible point.

We refer to [1] for a detailed discussion of the convergence properties of the standard DR for solving this feasibility problem.




(a) Standard Douglas–Rachford (b) Damped Douglas–Rachford

Figure 1: Convergence behaviour of the standard and damped Douglas–Rachford for solving the problem of a line intersecting with a circle.

Recall Example 3.2, the convergence behavior of dDR is shown in the right figure of Figure 1. Same initial point is chosen for dDR, and we observe that the method solve the problem successfully. However, as remarked in the original paper, dDR may also converges to some stationary point of (3.4) which is not a solution. In fact, as we shall see in the numerical experiments, for the Sudoku puzzle and $s$-queens puzzle, dDR fails all the tests while DR achieves very good performance; see Table 1.

### 3.2 Problems with more than two sets

Up to now, we have been dealing with the feasibility problem of two sets, while in various scenarios we need to deal with the case of finding common points of more than two sets. In what follows, we briefly show that, by a *product space trick*, we can reformulate the problem into the form of (3.1).

Let $m \geq 2$ be an integer, $C_i$ a non-empty closed set for each $i \in \{1, \dots, m\}$. Consider the following feasibility problem

$$\text{find} \quad x \in \mathbb{R}^n \quad \text{s.t.} \quad x \in \textstyle\bigcap_{i=1}^m C_i. \tag{3.7}$$

Let $\boldsymbol{\mathcal{H}} = \underbrace{\mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{m \text{ times}}$ be the product space endowed with the scalar inner-product and norm

$$\forall \boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{\mathcal{H}},\ \langle\!\langle \boldsymbol{x}, \boldsymbol{y} \rangle\!\rangle = \textstyle\sum_{i=1}^m \langle x_i,\, y_i \rangle,\ \|\boldsymbol{x}\| = \big(\sum_{i=1}^m \|x_i\|^2\big)^{1/2}.$$

Let $\boldsymbol{\mathcal{C}} \stackrel{\text{def}}{=} C_1 \times \cdots \times C_m$, then $\boldsymbol{\mathcal{C}} \subset \boldsymbol{\mathcal{H}}$, and denote the subspace $\boldsymbol{\mathcal{S}} \stackrel{\text{def}}{=} \{\boldsymbol{x} = (x_i)_i \in \boldsymbol{\mathcal{H}} : x_1 = \cdots = x_m\}$. The feasibility problem (3.7) can be reformulated into the following form

$$\text{find} \quad \boldsymbol{x} \in \boldsymbol{\mathcal{H}} \quad \text{s.t.} \quad \boldsymbol{x} \in \boldsymbol{\mathcal{C}} \textstyle\bigcap \boldsymbol{\mathcal{S}}. \tag{3.8}$$

The projection operator of $\boldsymbol{\mathcal{C}}$ is component-wise for each set $C_i, i = 1, \dots, m$, *i.e.*

$$\mathscr{P}_{\boldsymbol{\mathcal{C}}} \boldsymbol{x} = (\mathscr{P}_{C_1} x_1, \cdots, \mathscr{P}_{C_m} x_m).$$

Define the canonical isometry $\boldsymbol{\mathcal{K}} : \mathbb{R}^n \to \boldsymbol{\mathcal{S}},\ x \mapsto (x, \cdots, x)$, then we have $\mathscr{P}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) = \boldsymbol{\mathcal{K}}(\frac{1}{m} \sum_{i=1}^m x_i)$.

Adapting the standard Douglas–Rachford to the case of (3.7), we obtain the following iteration:

$$\begin{aligned} & x_{k+1} = \tfrac{1}{m} \textstyle\sum_{i=1}^m z_{i,k}, \\ & \text{For } i = 1, \dots, m: \\ & \left\lfloor \begin{aligned} & u_{i,k+1} \in \mathscr{P}_{C_i}(2x_{k+1} - z_{i,k}), \\ & z_{i,k+1} = z_{i,k} + u_{i,k+1} - x_{k+1}. \end{aligned} \right. \end{aligned} \tag{3.9}$$

Note that for the standard DR, there is no need to store $\boldsymbol{x}$ and simply $x_{k+1} = \frac{1}{m}\sum_{i=1}^m z_{i,k}$ is sufficient. Correspondingly, we also have the following iteration for the damped Douglas–Rachford splitting method:

$$\text{For } i = 1, \ldots, m: \quad \left\lfloor \begin{aligned} & x_{i,k+1} = \tfrac{1}{1+\gamma}\big(z_{i,k} + \gamma \tfrac{1}{m}\textstyle\sum_{j=1}^m z_{j,k}\big), \\ & u_{i,k+1} \in \mathscr{P}_{C_i}(2x_{i,k+1} - z_{i,k}), \\ & z_{i,k+1} = z_{i,k} + u_{i,k+1} - x_{i,k+1}. \end{aligned} \right. \tag{3.10}$$

# 4 Local convergence of Douglas–Rachford splitting

In this section we present our main result, the local convergence analysis of Douglas–Rachford splitting method. We first present the result in a general setting and then specialize to the case of Sudoku and $s$-queens puzzles.

**Non-degeneracy condition** To deliver the result, a non-degeneracy condition is needed for set $C$. Assume Assumption (**A.4**) holds for standard DR and that dDR is ran under the condition of Lemma 3.1, then at convergence for both methods we have $z_k \to z^\star$ and $u_k, x_k \to x^\star$. We assume that $C$ is prox-regular at $x^\star$ for $x^\star - z^\star$ and the following condition holds

$$x^\star - z^\star \in \operatorname{int}\big(\mathscr{N}_C(x^\star)\big) \tag{4.1}$$

where $\operatorname{int}(\cdot)$ stands for the interior of the set.

**Remark 4.1.** The non-degeneracy condition (4.1) requires $\mathscr{N}_C(u^\star)$ has a non-empty interior, which means that $x^\star$ is a vertex of the set $C$. A graphical illustration of the non-degeneracy condition (4.1) is provided in Figure 2 below.

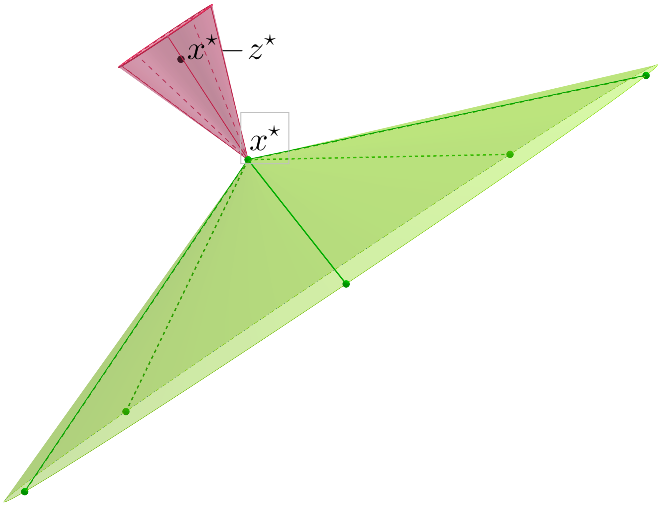



Figure 2: Normal cone (red) at a point $x^*$ in the polytope $C$ (green).

## 4.1 Local convergence of Douglas–Rachford splitting

We start with the standard Douglas–Rachford splitting method and then the damped iteration. Relation with some existing work in the literature is also discussed.

### 4.1.1 The standard Douglas–Rachford splitting

For standard Douglas–Rachford splitting method, for what follows we impose the global convergence as an assumption, *i.e.* (**A.4**) holds.

**Theorem 4.2 (Finite termination of DR).** *For the feasibility problem* (3.1) *and the Douglas–Rachford iteration* (3.2)*, suppose Assumptions* (**A.1**)-(**A.4**) *hold. Then* $\{u_k, x_k, z_k\}_{k\in\mathbb{N}}$ *converges to* $(x^\star, x^\star, z^\star)$ *with* $z^\star \in \mathrm{Fix}(\mathscr{F}_{\mathrm{DR}})$ *being a fixed point and* $x^\star = \mathscr{P}_S(z^\star)$*. If, moreover, the non-degeneracy condition* (4.1) *holds, then* $\{u_k, x_k, z_k\}_{k\in\mathbb{N}}$ *converges to* $(x^\star, x^\star, z^\star)$ *in a finite number of iterations.*

**Remark 4.3.** It is worth noting that Theorem 4.2 also holds true for the convex setting. In [10] the authors study DR for solving convex affine-polyhedral feasibility problem, and impose the following condition for finite convergence

$$S \bigcap \mathrm{int}(C) \neq \emptyset, \tag{4.2}$$

which does not hold for the non-convex case as the interior of $C$ in (3.1) can be empty; See also Section 4.2 the puzzles for which (4.2) fails. In [7], when the non-convex set is finite, finite termination is proved given that the other set is an affine subspace or a half-space. In comparison, our result here does not need the set to be finite and provides an extension to that of [10], as we characterize the situation where finite convergence happens but condition (4.2) fails.

**Proof.** The imposed global convergence of (3.2) means

$$z_k \to z^\star \in \mathrm{Fix}(\mathscr{F}_{\mathrm{DR}}) \quad \text{and} \quad u_k, x_k \to x^\star = \mathscr{P}_S(z^\star). \tag{4.3}$$

The prox-regularity of $C$ at $x^\star$ for $x^\star - z^\star$ and the non-degeneracy condition (4.1) imply that there exists an open set $\mathcal{B}$ such that

$$2x^\star - z^\star \in \mathcal{B} \subset \mathscr{N}_C(x^\star) + x^\star \quad \text{and} \quad \mathscr{P}_C(\mathcal{B}) = \{x^\star\}.$$

By the definition of convergence, there must therefore exist $K \in \mathbb{N}$ such that $2x_{k+1} - z_{k+1} \in \mathcal{B}$ for all $k \geq K$. Consequently, by the update step of $u_{k+1}$ in (3.5),

$$u_{k+1} = \mathscr{P}_C(2x_{k+1} - z_k) = x^\star$$

which is the finite convergence of $u_{k+1}$.

For the update of $x_k$ in (3.2), this time we have directly

$$x_{k+1} - x^\star = \mathscr{P}_S(z_k - z^\star).$$

For $z_{k+1}$, let $K > 0$ be such that $u_k = x^\star$ for all $k \geq K$, we have

$$\begin{aligned} z_{k+1} - z^\star = (z_k - z^\star) + (u_{k+1} - x^\star) - (x_{k+1} - x^\star) &= (z_k - z^\star) - (x_{k+1} - x^\star) \\ &= (\mathrm{Id} - \mathscr{P}_S)(z_k - z^\star) \\ &= (\mathrm{Id} - \mathscr{P}_S)^{k+1-K}(z_K - z^\star). \end{aligned}$$

Since $z_k \to z^\star$ and $(\mathrm{Id} - \mathscr{P}_S)^{k+1-K} = \mathrm{Id} - \mathscr{P}_S$, we have

$$0 = \lim_{k\to+\infty} z_{k+1} - z^\star = \lim_{k\to+\infty} (\mathrm{Id} - \mathscr{P}_S)^{k+1-K}(z_K - z^\star) = (\mathrm{Id} - \mathscr{P}_S)(z_K - z^\star) = z_{k+1} - z^\star,$$

which means $z_k = z^\star$ for all $k > K$, hence finite termination of $z_k$. The finite convergence of $x_k$ follows naturally that of $z_k$, and we conclude the proof. □

**Different order of update** In (3.2), the order of the projection operators can be switched which results in the following iterate

$$\begin{aligned} x_{k+1} &\in \mathscr{P}_C(z_k), \\ u_{k+1} &= \mathscr{P}_S(2x_{k+1} - z_k), \\ z_{k+1} &= z_k + u_{k+1} - x_{k+1}. \end{aligned} \tag{4.4}$$

The corollary below shows that the finite termination holds for the altered update.

**Corollary 4.4.** *For the feasibility problem (3.1) and the Douglas–Rachford iteration (4.4), suppose Assumptions (**A.1**)-(**A.4**) hold. Then $\{u_k, x_k, z_k\}_{k\in\mathbb{N}}$ converges to $(x^\star, x^\star, z^\star)$ with $z^\star \in \mathrm{Fix}(\mathscr{F}_{\mathrm{DR}})$ being a fixed point and $x^\star \in \mathscr{P}_C(z^\star)$. If, moreover, $C$ is prox-regular at $x^\star$ for $z^\star - x^\star$ and the following non-degeneracy condition holds,*

$$-(x^\star - z^\star) \in \mathrm{int}\big(\mathscr{N}_C(x^\star)\big), \tag{4.5}$$

*then $\{u_k, x_k, z_k\}_{k\in\mathbb{N}}$ converges to $(x^\star, x^\star, z^\star)$ in a finite number of iterations.*

**Proof.** Following the argument of the proof of Theorem 4.2, we can easily derive the finite termination of $x_k$ under the new non-degeneracy condition (4.5). In turn, for $k$ large enough, we have for $u_{k+1}$ that

$$\begin{aligned} u_{k+1} - x^\star = \mathscr{P}_S(2x_{k+1} - z_k) - \mathscr{P}_S(2x^\star - z^\star) &= 2\mathscr{P}_S(x_{k+1} - x^\star) - \mathscr{P}_S(z_k - z^\star) \\ &= -\mathscr{P}_S(z_k - z^\star). \end{aligned}$$

As a result for $z_k$,

$$\begin{aligned} z_{k+1} - z^\star = (z_k - z^\star) + (u_{k+1} - x^\star) - (x_{k+1} - x^\star) &= (z_k - z^\star) + (u_{k+1} - x^\star) \\ &= (\mathrm{Id} - \mathscr{P}_S)(z_k - z^\star), \end{aligned}$$

which is the same as the last part of proof of Theorem 4.2, and we conclude the proof. □

### 4.1.2 The damped Douglas–Rachford splitting

We now turn to the local convergence analysis of the damped Douglas–Rachford splitting method (3.5), for which we have the following theorem.

**Theorem 4.5 (Local convergence of dDR).** *For the feasibility problem (3.1) and the damped Douglas–Rachford iteration (3.5), suppose Assumptions (**A.1**)-(**A.3**) hold and (3.5) is run under the conditions of Theorem 3.1, then $\{u_k, x_k, z_k\}_{k\in\mathbb{N}} \to (x^\star, x^\star, z^\star)$ with $z^\star \in \mathrm{Fix}(\mathscr{F}_{\mathrm{dDR}})$ being a fixed point and $x^\star$ be a stationary point of (3.4). If, moreover, the non-degeneracy condition (4.1) hold, then*

(i) *$u_k$ converges in finite number of iterations, i.e. for all $k$ large enough there holds $u_k = x^\star$.*

(ii) *Let $\eta = \frac{\gamma}{1+\gamma}$, it holds $\|z_k - z^\star\| = O(\eta^k)$.*

**Proof.** The finite convergence of $u_k$ follows the argument of the proof of Theorem 4.2. For the update of $x_k$ in (3.5), since $S$ is a subspace, $\mathscr{P}_S$ is linear and we have

$$\begin{aligned} x_{k+1} - x^\star = \frac{1}{1+\gamma}\big(z_k + \gamma\mathscr{P}_S(z_k)\big) - x^\star &= \frac{1}{1+\gamma}\big(z_k + \gamma\mathscr{P}_S(z_k)\big) - \frac{1}{1+\gamma}\big(x^\star + \gamma\mathscr{P}_S(z^\star)\big) \\ &= \frac{1}{1+\gamma}(z_k - x^\star) + \frac{\gamma}{1+\gamma}\mathscr{P}_S(z_k - z^\star). \end{aligned}$$

Now for $z_{k+1}$, let $K > 0$ be such that $u_k = x^\star$ for all $k \geq K$, we have

$$\begin{aligned} z_{k+1} - z^\star = (z_k - z^\star) + (u_{k+1} - u^\star) - (x_{k+1} - x^\star) &= (z_k - z^\star) - (x_{k+1} - x^\star) \\ &= (z_k - z^\star) - \frac{1}{1+\gamma}(z_k - x^\star) - \frac{\gamma}{1+\gamma}\mathscr{P}_S(z_k - z^\star) \\ &= \frac{\gamma}{1+\gamma}(\mathrm{Id} - \mathscr{P}_S)(z_k - z^\star). \end{aligned}$$

Note that the spectral radius of the matrix appears above is

$$\rho\big(\tfrac{\gamma}{1+\gamma}(\mathrm{Id} - \mathscr{P}_S)\big) = \frac{\gamma}{1+\gamma}.$$

Combined with the fact the matrix is symmetric and normal, owing to Lemma 2.8 we conclude $\frac{\gamma}{1+\gamma}$ is the local linear convergence rate of $\|z_k - z^\star\|$. □

**Remark 4.6.** In [19], the authors also discuss the local linear convergence of damped DR under the following constraint qualification condition

$$\mathscr{N}_S\big(\mathscr{P}_S(x^\star)\big) \bigcap -\mathscr{N}_C(x^\star) = 0. \tag{4.6}$$

As shown in [19, Proposition 2], such a condition allows to show $x^\star \in C \cap S$ and $z^\star = x^\star$; See Example 3.2 which satisfies the above condition. The update of $x_{k+1}$ in (3.5) yields

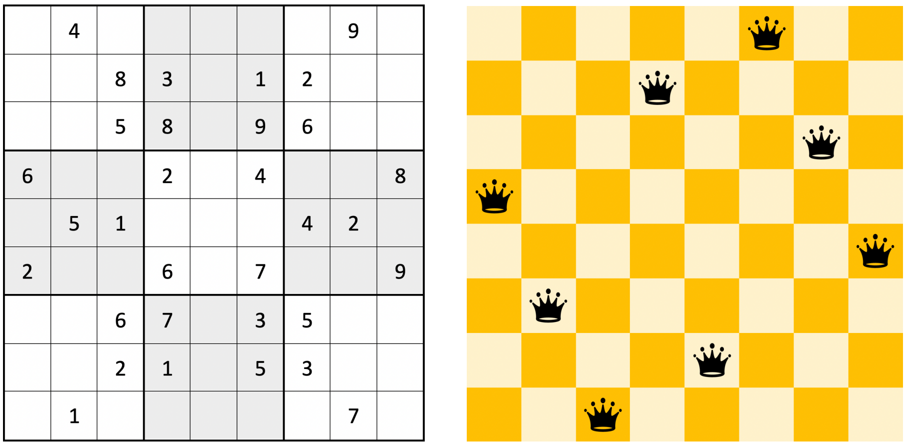
$$\begin{aligned} \frac{1+\gamma}{\gamma}(z^\star - x^\star) = z^\star - \mathscr{P}_S(z^\star) &\in \mathscr{N}_S\big(\mathscr{P}_S(x^\star)\big), \\ z^\star - x^\star &\in -\mathscr{N}_C(x^\star). \end{aligned}$$

This implies that only the fixed-points $z^\star$ such that $z^\star = x^\star$ satisfy the qualification condition (4.6). In comparison, our non-degeneracy condition is more general than (4.6) in the sense that we only focus on $\mathscr{N}_C(x^\star)$ and does not need the intersection of $\mathscr{N}_S(\mathscr{P}_S(x^\star)) \bigcap -\mathscr{N}_C(x^\star)$ to be 0, and our result holds for all fixed-points of $\mathrm{Fix}(\mathscr{F}_{\mathrm{dDR}})$.

**Remark 4.7.** When $S$, instead of being an affine subspace, has locally smooth curvature around $x^\star$, then according to the result of [20], one can show that for any $\eta \in ]\frac{\gamma}{1+\gamma}, 1[$ there holds $\|z_k - z^\star\| = O(\eta^k)$.

## 4.2 Sudoku and $s$-queens puzzles

In this part, we specialize the above result to Sudoku and $s$-queens puzzles. Examples of these two puzzles are provided in Figure 3 below.




(a) Sudoku puzzle (b) Eight queens puzzle

Figure 3: Classical examples of Sudoku and eight queens. The goal of Sudoku is to complete the grid such that each row, column, and $3 \times 3$ square contains all the digits from 1 to 9. The goal of eight queens is to place eight chess queens on an $8 \times 8$ board such that no two queens share the same row, column, or diagonal.

### 4.2.1 Sudoku puzzle

A standard Sudoku puzzle is shown in Figure 3 (a), which we generalize to grids of size $s \times s$ with the basic setting and rules:

- A partially complete $s \times s$ grid is provided
- Each *column*, each *row* and each of the $s$ *sub-grids of size* $\sqrt{s} \times \sqrt{s}$ that compose the grid contain all of the digits from 1 to $s$.

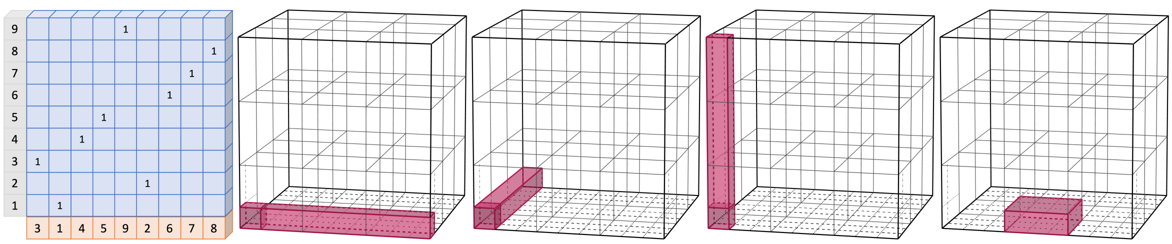
Based on the rules, we can easily formulate the Sudoku puzzle as feasibility problem. Here we consider the formulation proposed in [27], which formulates Sudoku as binary feasibility problem. We also refer to [2] for studies on Sudoku puzzle and Douglas–Rachford splitting method.

Each digit from 1 to $s$ is lifted to the set $[0,1]^s$, making the full puzzle an $s \times s \times s$ binary cube. Figure 4 (a) shows a feasible row of the lifted problem represented as a binary $s \times s$ square. Equivalently, we can say that any digit from 1 to $s$ is a permutation of unit vector $e = \{1, 0, \dots, 0\}$. This leads to four Sudoku feasibility constraints:

- Each *row* of the cube, *i.e.* $C_1(:, j, k)$, $j, k \in \{1, \dots, s\}$, is the permutation of $e$; See Figure 4 (b).
- Each *column* of the cube, *i.e.* $C_2(i, :, k)$, $i, k \in \{1, \dots, s\}$, is the permutation of $e$; See Figure 4 (c).
- Each *pillar* of the cube, *i.e.* $C_3(i, j, :)$, $i, j \in \{1, \dots, s\}$, is the permutation of $e$; See Figure 4 (d).
- For each $k \in \{1, \dots, s\}$, each of the $s$ *sub-grids* is the permutation of $e$; See Figure 4 (e). *i.e.* $C_4(\sqrt{s}(i-1)+1 : \sqrt{s}i, \sqrt{s}(j-1)+1 : \sqrt{s}j, k)$, $i, j \in \{1, \dots, \sqrt{s}\}$,

The partially completed grid forms the last constraint set

- $C_5$ is the constraint of the provided numbers.



(a) Lifted row (b) $C_1$ (c) $C_2$ (d) $C_3$ (e) $C_4$

Figure 4: Lifted Sudoku problem. (a) shows the lifted representation of a row of numbers. (b)-(e) show what is meant by a *lifted row/column/pillar/sub-grid* respectively.

At this point, solving the Sudoku puzzle is equivalent to solve the following feasibility problem of the five constraint sets

$$\text{find} \quad x \in \mathbb{R}^{s\times s\times s} \ \text{ s.t. } \ x \in C_1 \cap C_2 \cap C_3 \cap C_4 \cap C_5. \tag{4.7}$$

To obtain the product space formulation of the Sudoku puzzle we let $\boldsymbol{\mathcal{H}} = \underbrace{\mathbb{R}^{s\times s\times s} \times \cdots \times \mathbb{R}^{s\times s\times s}}_{5 \text{ times}}$, $\boldsymbol{\mathcal{C}} \stackrel{\text{def}}{=} C_1 \times \cdots \times C_5$, and $\boldsymbol{\mathcal{S}} = \{\boldsymbol{x} = (x_i)_i \in \boldsymbol{\mathcal{H}} : x_1 = \cdots = x_5\}$.

**Proposition 4.8 (Local convergence of DR).** *For the Sudoku puzzle (4.7) and Douglas–Rachford splitting method (3.9), suppose Assumptions (**A.1**)-(**A.4**) hold. Then $\{\boldsymbol{u}_k, x_k, \boldsymbol{z}_k\}_{k\in\mathbb{N}}$ converges to $(\boldsymbol{\mathcal{K}}(x^\star), x^\star, \boldsymbol{z}^\star)$ with $\boldsymbol{z}^\star \in \mathrm{Fix}(\mathscr{F}_{\mathrm{DR}})$ being a fixed point and $x^\star = \frac{1}{5}\sum_{i=1}^5 z_i^\star$. If, moreover, for $i = 1, \ldots, 4$, $C_i$ is prox-regular at $x^\star$ for $x^\star - z_i^\star$ and the following non-degeneracy condition holds*

$$x^\star - z_i^\star \in \mathrm{int}\big(\mathscr{N}_{C_i}(x^\star)\big). \tag{4.8}$$

*Then for all $k$ large enough, there holds*

- *$u_{i,k} = x^\star$ for $i = 1, \ldots, 4$,*
- *$\|\boldsymbol{z}_k - \boldsymbol{z}^\star\| = O(\eta^k)$ with $\eta = \frac{\sqrt{5}}{5}$.*

**Proof.** Denote $\boldsymbol{x}_{k+1} = \boldsymbol{\mathcal{K}}(x_{k+1})$, from the updates of $\boldsymbol{x}_k$, we have that

$$\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star = \mathscr{P}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{z}_k - \boldsymbol{z}^\star)$$

and that

$$\mathscr{P}_{\boldsymbol{\mathcal{S}}} = \tfrac{1}{5}\mathbf{1}_{5\times 5} \otimes \mathrm{Id}_{s^3\times s^3},$$

where $\mathbf{1}_{5\times5}$ stands for matrix of all 1 and $\otimes$ for Kronecker product.

The separability of $\mathscr{P}_{\boldsymbol{\mathcal{C}}}$ and the definition of projection operator lead to, for each $i = 1, \ldots, 5$

$$u_{i,k+1} = \mathscr{P}_{C_i}(2x_{k+1} - z_{i,k}) \quad \text{and} \quad u_{i,k+1} - u_i^\star = \mathscr{P}_{C_i}(2x_{k+1} - z_{i,k}) - \mathscr{P}_{C_i}(2x^\star - z_i^\star).$$

Under the non-degeneracy condition (4.8), apply the argument of Theorem 4.2 to obtain the finite convergence of $u_{i,k}$ for $i = 1, \ldots, 4$. For $C_5$, since its projection operator is linear, we have

$$\begin{aligned} u_{5,k+1} - u_5^\star &= \mathscr{P}_{C_5}(2x_{k+1} - z_{5,k}) - \mathscr{P}_{C_5}(2x^\star - z_5^\star) \\ &= 2\mathscr{P}_{C_5}(x_{k+1} - x^\star) - \mathscr{P}_{C_5}(z_{5,k} - z_5^\star). \end{aligned}$$

As a result, for $k$ large enough there holds

$$\boldsymbol{u}_{k+1} - \boldsymbol{u}^\star = 2\begin{bmatrix} \mathbf{0}_{4s^3\times4s^3} & \\ & \mathscr{P}_{C_5} \end{bmatrix} \mathscr{P}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{z}_k - \boldsymbol{z}^\star) - \begin{bmatrix} \mathbf{0}_{4s^3\times4s^3} & \\ & \mathscr{P}_{C_5} \end{bmatrix} (\boldsymbol{z}_k - \boldsymbol{z}^\star).$$

Let $\mathscr{P}_{\boldsymbol{\mathcal{C}}} \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{0}_{4s^3\times4s^3} & \\ & \mathscr{P}_{C_5} \end{bmatrix}$ and back to $\boldsymbol{z}_{k+1} - \boldsymbol{z}^\star$, we get

$$\begin{aligned} \boldsymbol{z}_{k+1} - \boldsymbol{z}^\star &= (\boldsymbol{z}_k - \boldsymbol{z}^\star) + (\boldsymbol{u}_{k+1} - \boldsymbol{u}^\star) - (\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star) \\ &= \big(\mathbf{Id} + 2\mathscr{P}_{\boldsymbol{\mathcal{C}}}\mathscr{P}_{\boldsymbol{\mathcal{S}}} - \mathscr{P}_{\boldsymbol{\mathcal{C}}} - \mathscr{P}_{\boldsymbol{\mathcal{S}}}\big)(\boldsymbol{z}_k - \boldsymbol{z}^\star), \end{aligned}$$

Since $\mathscr{P}_{C_5}$ is the projection operator onto a subspace, so is $\mathscr{P}_{\boldsymbol{\mathcal{C}}}$. As a result, the linear convergence rate is the cosine of the Friedrichs angle $\theta_F$ between the subspace of $\mathscr{P}_{\boldsymbol{\mathcal{C}}}$ and that of $\mathscr{P}_{\boldsymbol{\mathcal{S}}}$. According to Remark 2.12, we now need to analyze the singular values of $\mathscr{P}_{\boldsymbol{\mathcal{C}}}\mathscr{P}_{\boldsymbol{\mathcal{S}}}$, which essentially is the SVD of $\mathscr{P}_{C_5}\mathscr{P}_S$ where

$$\mathscr{P}_S = \tfrac{1}{5}\mathbf{1}_{1\times 5} \otimes \mathrm{Id}_{s^3\times s^3}.$$

We have

- $\mathscr{P}_{C_5}$ is diagonal matrix with only 0 and 1.
- $\mathscr{P}_S$ has a unique singular value which is $\frac{\sqrt{5}}{5}$.

As a result, $\mathscr{P}_{C_5}\mathscr{P}_S$ has only two singular values which are 0 and $\frac{\sqrt{5}}{5}$. Hence we conclude the proof. $\square$

Next we present the result for the damped Douglas–Rachford splitting method (3.5).

**Proposition 4.9 (Local convergence of dDR).** *For the Sudoku puzzle (4.7) and the damped Douglas–Rachford splitting method (3.10), suppose Assumptions (**A.1**)-(**A.3**) hold and (3.10) is run under the conditions of Theorem 3.1, then $(\boldsymbol{u}_k, \boldsymbol{x}_k, \boldsymbol{z}_k)$ converges to $(\boldsymbol{x}^\star, \boldsymbol{x}^\star, \boldsymbol{z}^\star)$ with $\boldsymbol{z}^\star$ being a fixed point of the iteration and $\boldsymbol{x}^\star$ a stationary point of $\min_{\boldsymbol{x}} \big\{\mathrm{dist}^2(\boldsymbol{x}, \boldsymbol{\mathcal{S}})$ s.t. $\boldsymbol{x} \in \boldsymbol{\mathcal{C}}\big\}$. If, moreover, the non-degeneracy condition (4.8) holds for $C_{1,\ldots,4}$, then for all $k$ large enough, it holds*

- $u_{i,k} = x^\star$ *for* $i = 1, \ldots, 4$,
- $\|\boldsymbol{z}_k - \boldsymbol{z}^\star\| = O(\eta^k)$ *with* $\eta = \frac{2\gamma+5+\sqrt{25-16\gamma^2}}{10(1+\gamma)}$.

**Proof.** From the updates of $\boldsymbol{x}_{k+1}$, we have that

$$\begin{aligned}\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star &= \frac{1}{1+\gamma}\big(\boldsymbol{z}_k + \gamma \mathscr{P}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{z}_k)\big) - \frac{1}{1+\gamma}\big(\boldsymbol{z}^\star + \gamma \mathscr{P}_{bcS}(\boldsymbol{z}^\star)\big) \\ &= \frac{1}{1+\gamma}(\boldsymbol{z}_k - \boldsymbol{z}^\star) + \frac{\gamma}{1+\gamma}\mathscr{P}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{z}_k - \boldsymbol{z}^\star)\end{aligned}$$

with $\mathscr{P}_{\boldsymbol{\mathcal{S}}} = \frac{1}{5}\mathbf{1}_{5\times 5} \otimes \mathrm{Id}_{s^3 \times s^3}$. For $\boldsymbol{u}_k$, the finite termination of $u_{i,k}$, $i = 1, \ldots, 4$ follows from the proof of Proposition 4.8. For $C_5$, again we have

$$u_{5,k+1} - u_5^\star = 2\mathscr{P}_{C_5}(x_{5,k+1} - x_5^\star) - \mathscr{P}_{C_5}(z_{5,k} - z_5^\star).$$

Let $\mathscr{P}_{\boldsymbol{\mathcal{C}}} \stackrel{\text{def}}{=} \begin{bmatrix}\mathbf{0}_{4s^3\times 4s^3} & \\ & \mathscr{P}_{C_5}\end{bmatrix}$, then

$$\begin{aligned}\boldsymbol{u}_{k+1} - \boldsymbol{u}^\star &= 2\mathscr{P}_{\boldsymbol{\mathcal{C}}}(\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star) - \mathscr{P}_{\boldsymbol{\mathcal{C}}}(\boldsymbol{z}_k - \boldsymbol{z}^\star) \\ &= \frac{2}{1+\gamma}\mathscr{P}_{\boldsymbol{\mathcal{C}}}(\boldsymbol{z}_k - \boldsymbol{z}^\star) + \frac{2\gamma}{1+\gamma}\mathscr{P}_{\boldsymbol{\mathcal{C}}}\mathscr{P}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{z}_k - \boldsymbol{z}^\star) - \mathscr{P}_{\boldsymbol{\mathcal{C}}}(\boldsymbol{z}_k - \boldsymbol{z}^\star) \\ &= \frac{1-\gamma}{1+\gamma}\mathscr{P}_{\boldsymbol{\mathcal{C}}}(\boldsymbol{z}_k - \boldsymbol{z}^\star) + \frac{2\gamma}{1+\gamma}\mathscr{P}_{\boldsymbol{\mathcal{C}}}\mathscr{P}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{z}_k - \boldsymbol{z}^\star).\end{aligned}$$

Back to $\boldsymbol{z}_{k+1} - \boldsymbol{z}^\star$, we get

$$\begin{aligned}\boldsymbol{z}_{k+1} - \boldsymbol{z}^\star &= (\boldsymbol{z}_k - \boldsymbol{z}^\star) + (\boldsymbol{u}_{k+1} - \boldsymbol{u}^\star) - (\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star) \\ &= (\boldsymbol{z}_k - \boldsymbol{z}^\star) + \frac{1-\gamma}{1+\gamma}\mathscr{P}_{\boldsymbol{\mathcal{C}}}(\boldsymbol{z}_k - \boldsymbol{z}^\star) + \frac{2\gamma}{1+\gamma}\mathscr{P}_{\boldsymbol{\mathcal{C}}}\mathscr{P}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{z}_k - \boldsymbol{z}^\star) - \frac{1}{1+\gamma}(\boldsymbol{z}_k - \boldsymbol{z}^\star) - \frac{\gamma}{1+\gamma}\mathscr{P}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{z}_k - \boldsymbol{z}^\star) \\ &= \frac{1}{1+\gamma}\big(\gamma\mathbf{Id} + 2\gamma\mathscr{P}_{\boldsymbol{\mathcal{C}}}\mathscr{P}_{\boldsymbol{\mathcal{S}}} + (1-\gamma)\mathscr{P}_{\boldsymbol{\mathcal{C}}} - \gamma\mathscr{P}_{\boldsymbol{\mathcal{S}}}\big)(\boldsymbol{z}_k - \boldsymbol{z}^\star) \\ &= \frac{1}{1+\gamma}\big(\gamma(\mathbf{Id} + 2\mathscr{P}_{\boldsymbol{\mathcal{C}}}\mathscr{P}_{\boldsymbol{\mathcal{S}}} - \mathscr{P}_{\boldsymbol{\mathcal{C}}} - \mathscr{P}_{\boldsymbol{\mathcal{S}}}) + \mathscr{P}_{\boldsymbol{\mathcal{C}}}\big)(\boldsymbol{z}_k - \boldsymbol{z}^\star).\end{aligned}$$

Denote $M_\gamma = \frac{1}{1+\gamma}(\gamma(\mathbf{Id} + 2\mathscr{P}_{\boldsymbol{\mathcal{C}}}\mathscr{P}_{\boldsymbol{\mathcal{S}}} - \mathscr{P}_{\boldsymbol{\mathcal{C}}} - \mathscr{P}_{\boldsymbol{\mathcal{S}}}) + \mathscr{P}_{\boldsymbol{\mathcal{C}}})$. Let $p, q$ be the rank of $\mathscr{P}_{\boldsymbol{\mathcal{C}}}$ and $\mathscr{P}_{\boldsymbol{\mathcal{S}}}$ respectively, also assume $p \le q$ (For the case $p \ge q$, similar discussion can be obtained). According to [5], there exists an orthogonal matrix $U$ such that

$$\mathscr{P}_{\boldsymbol{\mathcal{C}}} = U\left[\begin{array}{cc|cc}\mathrm{Id}_p & 0 & 0 & 0 \\ 0 & 0_p & 0 & 0 \\ \hline 0 & 0 & 0_{q-p} & 0 \\ 0 & 0 & 0 & 0_{n-p-q}\end{array}\right]U^* \quad \text{and} \quad \mathscr{P}_{\boldsymbol{\mathcal{S}}} = U\left[\begin{array}{cc|cc}\alpha^2 & \alpha\beta & 0 & 0 \\ \alpha\beta & \beta^2 & 0 & 0 \\ \hline 0 & 0 & \mathrm{Id}_{q-p} & 0 \\ 0 & 0 & 0 & 0_{n-p-q}\end{array}\right]U^*,$$

where $\alpha = \mathrm{diag}(\cos(\theta_1), \ldots, \cos(\theta_p))$ and $\beta = \mathrm{diag}(\sin(\theta_1), \ldots, \sin(\theta_p))$ with $\theta_{i,i=1,..p}$ being the principal angles between the subspaces of $\mathscr{P}_{\boldsymbol{\mathcal{C}}}$ and $\mathscr{P}_{\boldsymbol{\mathcal{S}}}$. Consequently,

$$\mathbf{Id} + 2\mathscr{P}_{\boldsymbol{\mathcal{C}}}\mathscr{P}_{\boldsymbol{\mathcal{S}}} - \mathscr{P}_{\boldsymbol{\mathcal{C}}} - \mathscr{P}_{\boldsymbol{\mathcal{S}}} = U\left[\begin{array}{cc|cc}\alpha^2 & \alpha\beta & 0 & 0 \\ -\alpha\beta & \alpha^2 & 0 & 0 \\ \hline 0 & 0 & 0_{q-p} & 0 \\ 0 & 0 & 0 & \mathrm{Id}_{n-p-q}\end{array}\right]U^*.$$

Therefore, we have

$$M_\gamma = \frac{1}{1+\gamma}U\left[\begin{array}{cc|cc}\gamma\alpha^2 + \mathrm{Id}_p & \gamma\alpha\beta & 0 & 0 \\ -\gamma\alpha\beta & \gamma\alpha^2 & 0 & 0 \\ \hline 0 & 0 & 0_{q-p} & 0 \\ 0 & 0 & 0 & \gamma\mathrm{Id}_{n-p-q}\end{array}\right]U^*$$

Clearly, 0 and $\frac{\gamma}{1+\gamma}$ are two eigenvalues of the matrix. For the top left block of the above matrix, as it is block diagonal, we have the following characteristic polynomial

$$0 = \prod_{i=1}^{p}\Big(\big(\tfrac{\gamma\alpha_i^2+1}{1+\gamma} - \lambda\big)\big(\tfrac{\gamma\alpha_i^2}{1+\gamma} - \lambda\big) + \tfrac{\gamma^2\alpha_i^2\beta_i^2}{(1+\gamma)^2}\Big).$$

Solving the quadratic equation for each $i$ we get

$$\lambda_i = \frac{2\gamma\alpha_i^2 + 1 \pm \sqrt{1 - 4\gamma^2\alpha_i^2\beta_i^2}}{2(1+\gamma)}.$$

As in the proof of Proposition 4.8, we have that $\alpha_i = \frac{\sqrt{5}}{5}$ for all $i = 1, \ldots, p$, therefore $M_\gamma$ has only 4 distinct eigenvalues which are

$$0, \quad \frac{2\gamma + 5 - \sqrt{25 - 16\gamma^2}}{10(1+\gamma)}, \quad \frac{\gamma}{1+\gamma} \quad \text{and} \quad \frac{2\gamma + 5 + \sqrt{25 - 16\gamma^2}}{10(1+\gamma)}$$

in the *ascending* order. To show $\eta = \frac{2\gamma+5+\sqrt{25-16\gamma^2}}{10(1+\gamma)}$ is the convergence rate, we need to show $\eta$ is semi-simple. Let $M_p = \begin{bmatrix} \gamma\alpha^2 + \mathrm{Id}_p & \gamma\alpha\beta \\ -\gamma\alpha\beta & \gamma\alpha^2 \end{bmatrix}$, since $\alpha = \frac{\sqrt{5}}{5}$ is a $p$'th order root, we can simplify $M_p$ as:

$$M_p = \tfrac{1}{5} \begin{bmatrix} (\gamma+5)\mathrm{Id}_p & 2\gamma\mathrm{Id}_p \\ -2\gamma\mathrm{Id}_p & \gamma\mathrm{Id}_p \end{bmatrix}$$

As a result, we have

$$\begin{aligned} \mathrm{rank}(M_p - \eta\mathrm{Id}_{2p}) &= \mathrm{rank}\left( \begin{bmatrix} (\gamma+5-5\eta)\mathrm{Id}_p & 2\gamma\mathrm{Id}_p \\ -2\gamma\mathrm{Id}_p & (\gamma-5\eta)\mathrm{Id}_p \end{bmatrix} \right) = p \\ \text{and} \quad \mathrm{rank}\big((M_p - \eta\mathrm{Id}_{2p})^2\big) &= \mathrm{rank}\left( \begin{bmatrix} ((\gamma+5-5\eta)^2 - 4\gamma^2)\mathrm{Id}_p & 2\gamma(\gamma+5-10\eta)\mathrm{Id}_p \\ -2\gamma(\gamma+5-10\eta)\mathrm{Id}_p & ((\gamma-5\eta)^2 - 4\gamma^2)\mathrm{Id}_p \end{bmatrix} \right) = p, \end{aligned}$$

which means $\eta$ is semi-simple by Definition 2.6, and we conclude the linear rate of convergence. □

**Remark 4.10.** The proofs of the two propositions above is dimension independent, which means the results hold true for all puzzle sizes of perfect squares $s$ with $s \geq 4$. See Section 5 for numerical illustrations.

### 4.3 $s$-queens puzzle

The rule of eight queens puzzle is rather simple: placing eight chess queens on an $8 \times 8$ chessboard so that no two queens threaten each other. The size of puzzle can be generalized to any size $s \times s$ with $s \geq 4$, while there is no solution for $s = 2, 3$ and a trivial solution for $s = 1$ which is obvious[2].

We follow the setting of [27]. On the chessboard, as there are four directions (horizontal, vertical and two diagonal directions) for the queen to move, we have four constraint sets for the problem:

- $C_1$: each *row* has only one queen.
- $C_2$: each *column* has only one queen.
- $C_3$: each diagonal direction *southeast-northwest*, there is at most one queen.
- $C_4$: each diagonal direction *southwest-northeast*, there is at most one queen.

Now we can formulate the $s$-queens puzzle as the following feasibility problem of four sets

$$\text{find} \quad x \in \mathbb{R}^{s\times s} \ \text{ s.t. } \ x \in C_1 \cap C_2 \cap C_3 \cap C_4. \tag{4.9}$$

Since all the sets above are binary, so is the set $\boldsymbol{\mathcal{C}} \stackrel{\text{def}}{=} C_1 \times \cdots \times C_4$, as a result finite convergence can be obtained under the conditions of Theorems 4.2 and 4.5, for the standard Douglas–Rachford splitting and the damped one, respectively.

# 5 Numerical results

We now provide numerical results on Sudoku and $s$-queens puzzles to support our theoretical findings. Throughout the experiments, for the damped DR, we choose $\gamma = \frac{1}{5}$ which is smaller than $\sqrt{3/2} - 1$.

Before analyzing the convergence rates, we first compare the performance of the standard Douglas–Rachford splitting method (3.2) and the damped one (3.5), on how successful are they when applied to solve these two puzzles. *i.e.* how often each method finds a feasible point.

This comparison is shown in Table 1. For both methods, the iteration is terminated if either a stopping criterion is met or $10^4$ steps of iteration are reached, then we verify the output of each method. For a given puzzle type, each method is repeated $10^5$ times with different initialization for each running.

Recall Example 3.2 the problem of a circle intersecting with a line, damped DR solves the problem while the standard DR fails under chosen initial point. However for these two puzzles, it is surprising that the damped DR simply fails all tests, while the standard DR achieves very good performance: 100% success rate for Sudoku and more than 95% for the eight queens puzzle.

[2] https://en.wikipedia.org/wiki/Eight_queens_puzzle

Table 1: Comparison of success rate of standard DR and damped DR for solving Sudoku and eight queens puzzles over 100,000 random initializations.

| | Sudoku puzzle | Eight queens puzzle |
|---|---|---|
| Standard DR (3.2) | 1 | 0.95701 |
| Damped DR (3.5) | 0 | 0 |

### 5.1 Sudoku puzzle

We consider three different puzzle sizes for Sudoku to verify out results: 4, 9 and 16, which are shown in Figure 5 (a)-(c). In each size, we have 4, 32, and 128 coefficients provided respectively. The convergence behavior of standard Douglas–Rachford splitting method can be seen in the second and third rows of Figure 5, from which we observe that for all puzzles,

- *Finite termination* of $u_{i,k}, i = 1, \ldots, 4$: in the second row of Figure 5, we provide the $\ell_0$ pseudo-norms of $\|u_{i,k} - u_i^\star\|_0, i = 1, \ldots, 4$ to show the mismatch between $u_{i,k}$ and $u_i^\star$. We observed that, for each $i \in \{1, 2, 3, 4\}$, $\|u_{i,k} - u_i^\star\|_0$ reaches 0 in finite steps, which means the finite termination.
- *Local linear convergence* In the last row of Figure 5, we provide the convergence behaviors of $\|\boldsymbol{u}_k - \boldsymbol{u}^\star\|$ (which actually reduces to $\|u_{5,k} - u_5^\star\|$), $\|x_k - x^\star\|$ and $\|\boldsymbol{z}_k - \boldsymbol{z}^\star\|$. Take $\|\boldsymbol{z}_k - \boldsymbol{z}^\star\|$ for example, its convergence has two different regimes: *sublinear* rate from the beginning, and *linear* rate locally. The *magenta* dashed line is our theoretical estimation of the linear convergence rate and the slope of the line is $\frac{\sqrt{5}}{5}$.

For all three different puzzle sizes, the local linear convergence rate is $\frac{\sqrt{5}}{5} \approx 0.45$, which confirms that the rate is independent of puzzle size.

### 5.2 $s$-queens puzzle

For the $s$-queens puzzle, we also consider three different puzzle sizes: $s = 8, 16$ and 25, which are shown in Figure 6 (a)-(c). The convergence behaviors of the Douglas–Rachford splitting method are shown in the second row of Figure 6. Since all the constraint sets are binary, we observe finite convergence for the algorithm which complies with our theoretical results.

### 5.3 The damped Douglas–Rachford splitting

We conclude our numerical experiments by showing the local linear convergence the damped Douglas–Rachford splitting method. The results on Sudoku puzzle of size $9 \times 9$ and eight queens puzzle of size $8 \times 8$ are shown below in Figure 7. For both plots, the *magenta* line is our theoretical estimation of the local linear rate:

- For Sudoku puzzle, the slope of the magenta line is $\frac{2\gamma+5+\sqrt{25-16\gamma^2}}{10(1+\gamma)} \approx 0.86$.
- For eight queens puzzle, the slope of the magenta line is $\frac{\gamma}{1+\gamma} \approx 0.17$.

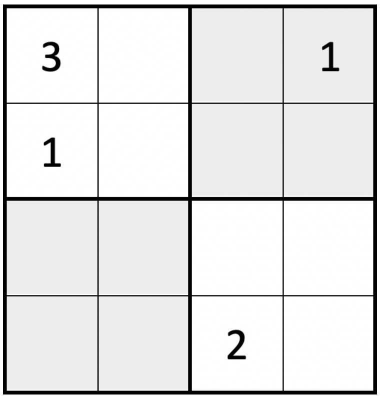
Again, our theoretical estimations are tight.

## 6 Conclusions

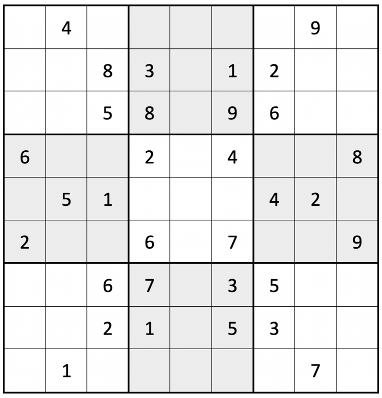
In this paper, we studied local convergence properties of Douglas–Rachford splitting method when applied to solve non-convex feasibility problems. Under a proper non-degeneracy condition, both finite convergence and local linear convergence are proved for the standard Douglas–Rachford splitting and a damped version of the method. Understanding when the methods fail, especially for the damped Douglas–Rachford splitting, require further study on the property of the methods.

### Acknowledgement

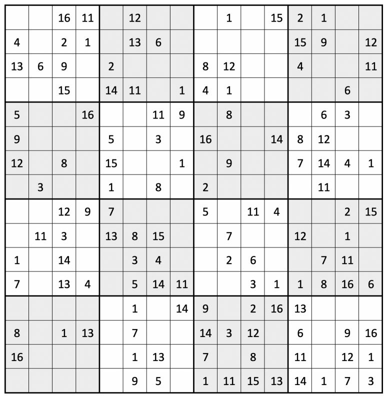
J.L. would like to thank Guoyin Li for helpful discussions on the convergence of Douglas–Rachford splitting method for solving non-convex problems. J.L. was partly supported by Leverhulme trust, Newton trust and the EPSRC centre "EP/N014588/1". R.T. acknowledges funding from EPSRC Grant No. "EP/L016516/1" for the Cambridge Centre for Analysis. C.-B.S. acknowledges support from the Leverhulme Trust project "Breaking the non-convexity barrier", EPSRC Grant No. EP/M00483X/1, EPSRC centre Grant No. "EP/N014588/1", and from CHiPS (Horizon 2020 RISE project grant). All authors were supported by the Cantab Capital Institute for Mathematics of Information.



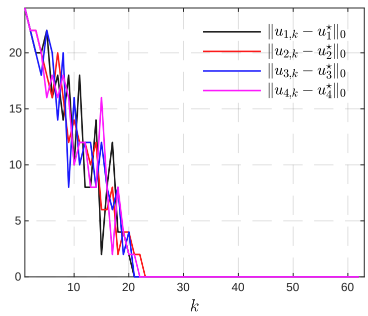

(a) Size $4 \times 4$



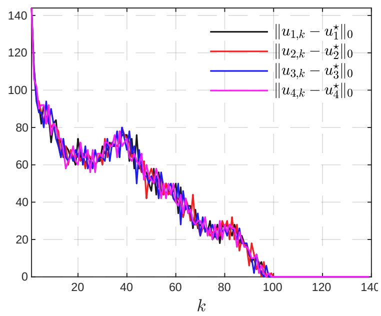

(b) Size $9 \times 9$



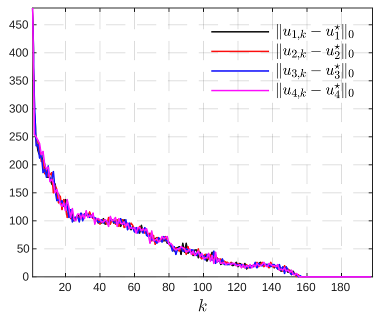
(c) Size $16 \times 16$



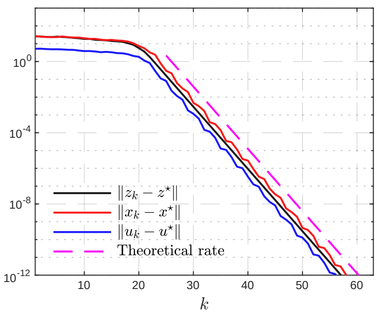

(d) Size 4: convergence of $u_{i,k}$



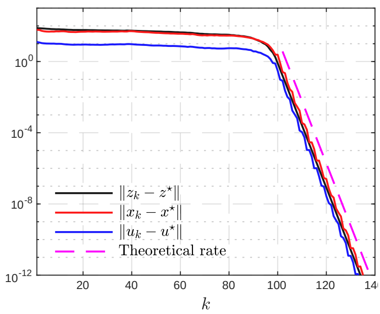

(e) Size 9: convergence of $u_{i,k}$



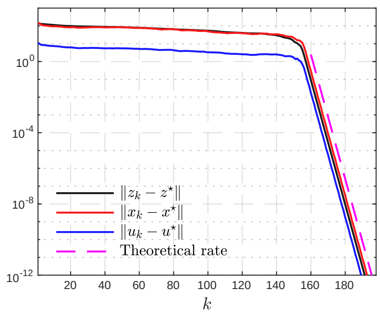

(f) Size 16: convergence of $u_{i,k}$




(g) Size 4: convergence of $(\boldsymbol{u}_k, x_k, \boldsymbol{z}_k)$




(h) Size 9: convergence of $(\boldsymbol{u}_k, x_k, \boldsymbol{z}_k)$




(i) Size 16: convergence of $(\boldsymbol{u}_k, x_k, \boldsymbol{z}_k)$

Figure 5: Different sizes of Sudoku puzzles and convergence observations.

# References

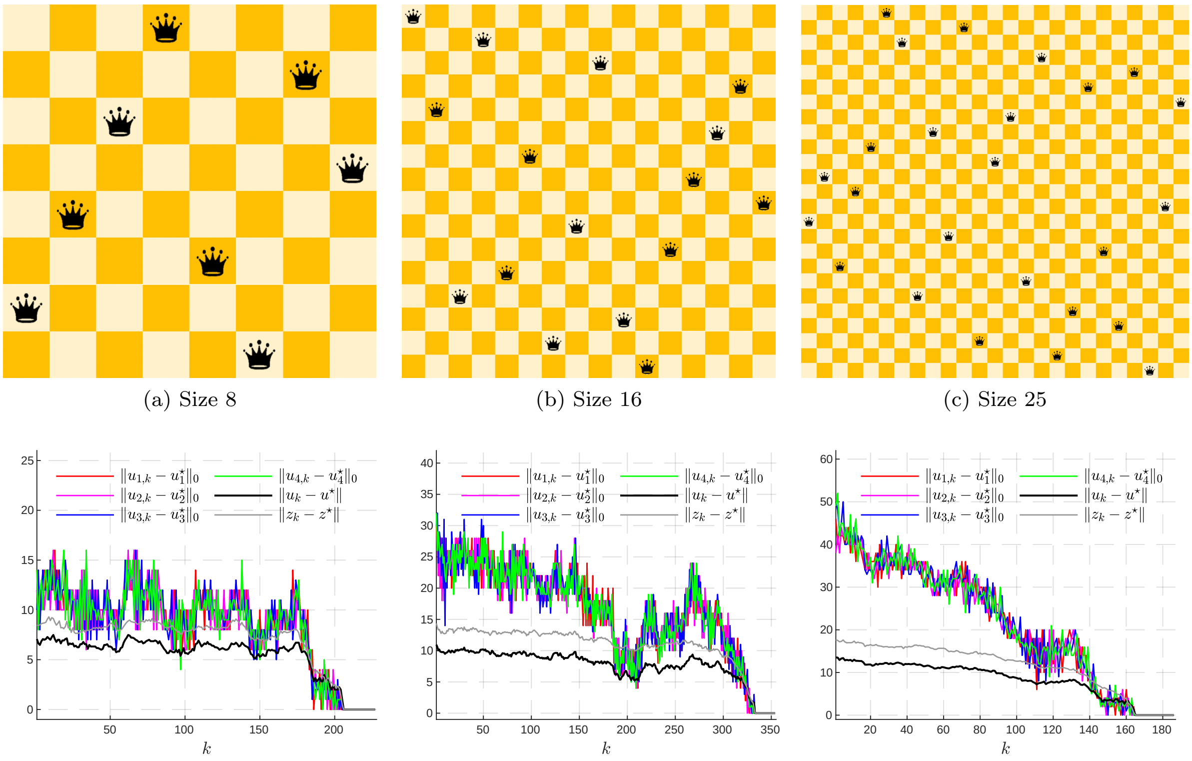

[1] F. J. A. Artacho and J. M. Borwein. Global convergence of a non-convex douglas–rachford iteration. *Journal of Global Optimization*, 57(3):753–769, 2013.

[2] F. J. A. Artacho, J. M. Borwein, and M. K. Tam. Recent results on douglas–rachford methods for combinatorial optimization problems. *Journal of Optimization Theory and Applications*, 163(1):1–30, 2014.

[3] F. J. A. Artacho, J. M. Borwein, and M. K. Tam. Global behavior of the douglas–rachford method for a nonconvex feasibility problem. *Journal of Global Optimization*, 65(2):309–327, 2016.

[4] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality. *Mathematics of Operations Research*, 35(2):438–457, 2010.

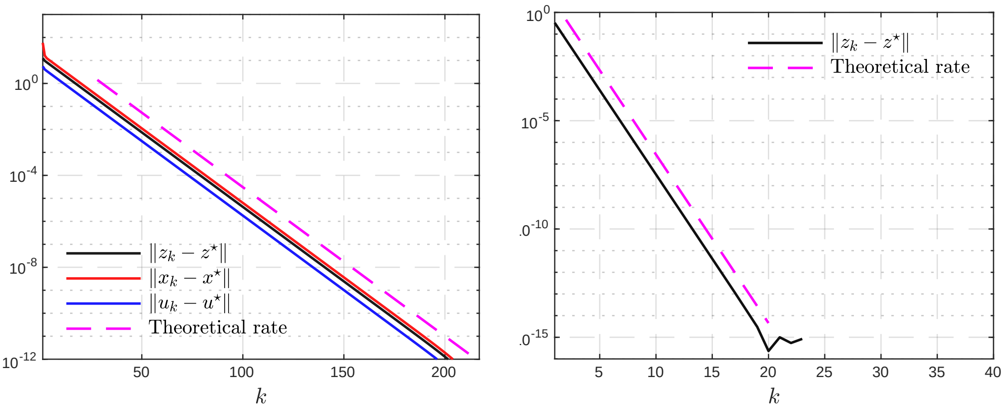
[5] H. H. Bauschke, J. Y. Bello Cruz, T. T. A. Nghia, H. M. Pha, and X. Wang. Optimal rates of linear con-




(a) Size 8 (b) Size 16 (c) Size 25

(d) Size 8 (e) Size 16 (f) Size 25

Figure 6: Different sizes of queens puzzles and convergence observations.




(a) Sudoku puzzle (b) Eight queens puzzle

Figure 7: Local linear convergence of the damped Douglas-Rachford splitting method for Sudoku puzzle and eight queens puzzle. Note again that convergence does not imply finding a solution to the feasibility problem.


vergence of relaxed alternating projections and generalized douglas-rachford methods for two subspaces. *Numerical Algorithms*, 73(1):33–76, 2016.

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

[7] H. H. Bauschke and M. N. Dao. On the finite convergence of the douglas–rachford algorithm for solving (not

necessarily convex) feasibility problems in euclidean spaces. *SIAM Journal on Optimization*, 27(1):507–537, 2017.

[8] H. H. Bauschke and D. Noll. On the local convergence of the douglas–rachford algorithm. *Archiv der Mathematik*, 102(6):589–600, 2014.

[9] Heinz H Bauschke, JY Bello Cruz, Tran TA Nghia, Hung M Pha, and Xianfu Wang. Optimal rates of linear convergence of relaxed alternating projections and generalized douglas-rachford methods for two subspaces. *Numerical Algorithms*, 73(1):33–76, 2016.

[10] Heinz H Bauschke, Minh N Dao, Dominikus Noll, and Hung M Phan. On slater's condition and finite convergence of the douglas–rachford algorithm for solving convex feasibility problems in euclidean spaces. *Journal of Global Optimization*, 65(2):329–349, 2016.

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

[12] A. Cegielski and A. Suchocka. Relaxed alternating projection methods. *SIAM Journal on Optimization*, 19(3):1093–1106, 2008.

[13] P. L. Combettes and J. C. Pesquet. Proximal splitting methods in signal processing. In *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer, 2011.

[14] J. Douglas and H. H. Rachford. On the numerical solution of heat conduction problems in two and three space variables. *Transactions of the American mathematical Society*, 82(2):421–439, 1956.

[15] R. Hesse and D. R. Luke. Nonconvex notions of regularity and convergence of fundamental algorithms for feasibility problems. *SIAM Journal on Optimization*, 23(4):2397–2419, 2013.

[16] R. Hesse, D. R. Luke, and P. Neumann. Projection methods for sparse affine feasibility: Results and counterexamples. Technical report, 2013.

[17] R. Hesse, D. R. Luke, and P. Neumann. Alternating projections and douglas-rachford for sparse affine feasibility. *IEEE Transactions on Signal Processing*, 62(18):4868–4881, 2014.

[18] A. S. Lewis. Active sets, nonsmoothness, and sensitivity. *SIAM Journal on Optimization*, 13(3):702–725, 2003.

[19] G. Li and T. Kei. Pong. Douglas–rachford splitting for nonconvex optimization with application to nonconvex feasibility problems. *Mathematical programming*, 159(1-2):371–401, 2016.

[20] J. Liang, J. Fadili, and G. Peyré. Local convergence properties of douglas–rachford and alternating direction method of multipliers. *Journal of Optimization Theory and Applications*, 172(3):874–913, 2017.

[21] P. L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979.

[22] D. R. Luke. Relaxed averaged alternating reflections for diffraction imaging. *Inverse problems*, 21(1):37, 2004.

[23] S.-Y. Matsushita and L. Xu. On the finite termination of the douglas-rachford method for the convex feasibility problem. *Optimization*, 65(11):2037–2047, 2016.

[24] C. D. Meyer. *Matrix analysis and applied linear algebra*, volume 2. SIAM, 2000.

[25] D. W. Peaceman and H. H. Rachford, Jr. The numerical solution of parabolic and elliptic differential equations. *Journal of the Society for Industrial and Applied Mathematics*, 3(1):28–41, 1955.

[26] H. M. Phan. Linear convergence of the douglas–rachford method for two closed sets. *Optimization*, 65(2):369–385, 2016.

[27] J. Schaad. *Modeling the 8-queens problem and sudoku using an algorithm based on projections onto nonconvex sets*. PhD thesis, University of British Columbia, 2010.

[28] A. Themelis and P. Patrinos. Douglas–rachford splitting and admm for nonconvex optimization: Tight convergence results. *SIAM Journal on Optimization*, 30(1):149–181, 2020.

[29] J von Neumann. Functional operators, vol. 2 (annals of mathematics studies, no. 22), princeton, nj, 1950. *Reprinted from mimeographed lecture notes first distributed in*, 1933.