THE FUN IS FINITE: DOUGLAS-RACHFORD AND SUDOKU PUZZLE – FINITE TERMINATION AND LOCAL LINEAR CONVERGENCE

ROBERT TOVEY[1], JINGWEI LIANG[2,*]

[1]*INRIA Paris, France*
[2]*Institute of Natural Sciences and School of Mathematical Sciences,*
Shanghai Jiao Tong University, Shanghai, China

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, a local convergence analysis for non-convex feasibility problems is presented and both finite termination and local linear convergence are obtained. For a generalization of the Sudoku puzzle, it is proved 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, it is proved 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.

Keywords. Douglas-Rachford; Feasibility problem; Finite termination; Local Linear Convergence; Sudoku puzzle.

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 were developed based on projections. Among them, the alternating projection is the fundamental one. The method of the alternating projection was first introduced by von Neumann for the case of two linear subspaces [1], then it was extended to closed convex sets by Bregman [2]. Relaxation is a standard approach to speed up the alternating projection. For related work, we refer to [3, 4] and the references therein.

Proximal splitting methods, such as Forward-Backward [5] splitting and Backward-Backward splitting [6], and Peaceman-Rachford/Douglas-Rachford splitting [7, 8], can also be applied to solve feasibility problems either directly or up to reformulation. Moreover, the equivalence between projection-based methods and proximal splitting methods can be established, such as the alternating projection is equivalent to Backward-Backward splitting while the relaxed alternating relaxed projection covers Peaceman-Rachford/Douglas-Rachford splitting as special cases [6].

Our focus in this paper is the Douglas-Rachford splitting method, which has shown to be effective for solving feasibility problems, particularly in the non-convex setting [9]. However,

the convergence property is rather less understood than its convex counterpart. One reason for this is that the Douglas-Rachford splitting method is not symmetric and non-descent, when compared to the (proximal) gradient descent whose non-convex case is much easier studied [10]. 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, the authors [11] considered the Douglas-Rachford splitting for solving the feasibility problem of a line intersecting with a circle, and the conditions for convergence. In [12], the authors proposed a damped Douglas-Rachford splitting method for general non-convex optimization problems under the condition that one function has a Lipschitz continuous gradient.

The study of this paper is motivated by applying the Douglas-Rachford to solve Sudoku puzzle, for which three different convergence behaviors are observed

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

Finite termination and local linear convergence were reported in the literatures [9, 13], however, the conditions in respective work either were designed for convex setting or cannot be satisfied by Sudoku puzzle. Therefore, a new analysis is needed for the Douglas-Rachford splitting which is the aim of this paper:

(1) **Finite termination** Under a non-degeneracy condition (see (4.1)), we show in Section 4 that one sequence generated by the 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 the Douglas-Rachford splitting method. Particularly, for Sudoku puzzle, we prove that the linear rate of convergence of Douglas-Rachford splitting method is precisely $\frac{\sqrt{5}}{5}$. Moreover, such a rate is independent of the 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 on the finite termination property of the standard Douglas-Rachford splitting method. In [13], the authors established the finite convergence of Douglas-Rachford in the presence of Slater's condition for solving the 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 [14] under the notion of partial smoothness [15]. In [16], finite termination was 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 [13] was later extended to the non-convex case in [9], where one of the two sets can be finite.

For the local linear convergence, results can be found in, for instance, [17], where the linear convergence of the Douglas-Rachford splitting method was established under a regularity condition. Similar results can be found in [18, 19]. Under a constraint qualification condition, the local linear convergence property of the damped Douglas-Rachford splitting method was discussed in [12].

---

https://en.wikipedia.org/wiki/Sudoku

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. The global convergence is also discussed in this section. 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. In Section 6, the last section, we give the concluding remarks.

2. PRELIMINARIES

Throughout the paper, $\mathbb{N}$ is the set of nonnegative integers, and $\mathbb{R}^n$ is a finite n-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

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

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 & 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

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

The mapping $\mathscr{P}_C \colon \mathbb{R}^n \rightrightarrows C$ 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 $x \in C$, then C is said to be *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

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

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.*

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

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

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

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

$$x_{k+1} = x_k + \lambda \left(\mathscr{P}_{C_2}^{\lambda_2} \mathscr{P}_{C_1}^{\lambda_1}(x_k) - x_k \right),$$

where $\lambda, \lambda_1,$ and λ_2 are relaxation parameters. The iteration 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 [20] 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 [21].

Definition 2.5 (Convergent matrices). A matrix $M \in \mathbb{R}^{n \times n}$ is said to be *convergent* to $M^\infty \in \mathbb{R}^{n \times n}$ if and only if $\lim_{k \to +\infty} \|M^k - M^\infty\| = 0$. 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 $\|M^k - M^\infty\| = O(\eta^k)$. If M does not converge at any rate $\eta' \in [0, \eta[$, then η is called the *optimum convergence rate*.

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

Theorem 2.1 (Limits of powers). *For $M \in \mathbb{R}^{n \times n}$, the power of M converges to M^∞ 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 M^∞, and we have the following lemma.

Lemma 2.1 (Convergence rate). *If $M \in \mathbb{R}^{n \times n}$ is convergent to some $M^\infty \in \mathbb{R}^{n \times n}$, then*
  (i) *for any $k \in \mathbb{N}$,*

$$M^k - M^\infty = (M - M^\infty)^k \quad and \quad \|M^k - M^\infty\| \leq \|M - M^\infty\|^k.$$

  The equality holds only when M is normal.
 (ii) *We have that $\rho(M - M^\infty) < 1$, and M is linearly convergent for any $\eta \in]\rho(M - M^\infty), 1[$.*
(iii) *$\rho(M - M^\infty)$ 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 $|\eta| = \rho(M - M^\infty)$ are semi-simple.*

Proof. See Theorems 2.12, 2.13, 2.15 and 2.16 of [22]. $\qquad\qquad\qquad\qquad\qquad\square$

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

Definition 2.7 (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 $u_0 = v_0 \overset{\text{def}}{=} 0$ and inductively

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

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

Definition 2.8 (Friedrichs angle). The *Friedrichs angle* $\theta_F \in [0, \frac{\pi}{2}]$ between T_1 and T_2 is

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

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

Lemma 2.2 ([22, Proposition 3.3]). *We have* $\theta_F(T_1, T_2) = \theta_{d+1} > 0$, *where* $d \overset{\text{def}}{=} \dim(T_1 \cap T_2)$.

3. THE PROBLEM AND ALGORITHMS

The formal statement of the feasibility problem is written below

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

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**) $C \cap S \neq \emptyset$, *i.e.* the intersection is non-empty.

Note that problem (3.1) is not necessarily convex as we suppose that C is only non-empty and closed. The 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 the Douglas-Rachford (DR) splitting method [8] dates back to 1950s for solving numerical PDEs. In recently years, the method has also been shown to be effective for non-convex feasibility problems; see, e.g., [23, 24] and the references therein. 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

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

until *convergence*;

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

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

Then, $z_{k+1} = \mathscr{F}_{\text{DR}}(z_k)$. The other two variables u_k, x_k are called the *shadow sequences* [25].

Determining the convergence properties of the Douglas-Rachford splitting for the non-convex setting is a challenging problem since the non-descent property of the method makes it much harder to obtain convergence results than the descent-type methods, which include (proximal) gradient descent [10].

Moreover, since the method has three different sequences u_k, x_k, and z_k, various different convergence behaviors may occur. We refer to [9] for more detailed discussions. In Example 3.1, we demonstrate a case of a circle intersecting with a line where:

- the shadow sequences $\{u_k\}_{k\in\mathbb{N}}$, $\{x_k\}_{k\in\mathbb{N}}$ converge to $u^\star$ and $x^\star$, respectively. But $u^\star \neq x^\star$;
- the fixed-point sequence $\{z_k\}_{k\in\mathbb{N}}$ diverges.

As our main interest in this paper is to study the local 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

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

To avoid assumption (**A.4**), people either turn to specific cases [26] or imposing stronger assumptions, such as smoothness [27]. Modifications to the original Douglas-Rachford splitting method were also considered in the literature. Below we describe a damped version of the Douglas-Rachford splitting proposed in [12].

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

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\text{dist}^2(x,S) \quad \text{s.t.} \quad x \in C. \tag{3.3}$$

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

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

we then obtain

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

As a result, we obtain the algorithm proposed in [12].

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

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

repeat

$$\begin{aligned}
x_{k+1} &= \mathscr{P}_S^{\frac{\gamma}{1+\gamma}}(z_k), \\
u_{k+1} &\in \mathscr{P}_C(2x_{k+1} - z_k), \\
z_{k+1} &= z_k + u_{k+1} - x_{k+1},
\end{aligned} \tag{3.4}$$

until *convergence*;

We refer to the original work [12] for a more detailed discussion of Algorithm 2. When $\gamma = +\infty$, Algorithm 2 recovers standard Douglas-Rachford splitting method (3.2). The fixed-point operator of dDR reads

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

We have the following convergence result of the dDR from [12].

Lemma 3.1 (Global convergence of dDR [12, Theorem 5]**).** *For non-convex feasibility problem (3.1), suppose that Assumptions (A.1)-(A.3) hold and moreover C is compact. Choose $\gamma \in {]}0, \sqrt{3/2} - 1{[}$ for Douglas-Rachford splitting method (3.4). Then $\{u_k, x_k, z_k\}_{k\in\mathbb{N}}$ 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 problem (3.3).*

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

Example 3.1 (A circle intersects with a line). Let $C = \{x \in \mathbb{R}^2 : \|x\| = 1\}$ be the unit circle, and let $S = \{x \in \mathbb{R}^2 : \langle x, \left(\begin{smallmatrix}1\\2\end{smallmatrix}\right)\rangle = \sqrt{2}\}$ be a line that intersects with C at two different points. For both methods, same initial point $z_0 = (-10, -8)$ is chosen. For damped DR, we set $\gamma = \frac{1}{5}$. 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 [11] for a detailed discussion of the convergence properties of the standard DR for solving this feasibility problem.

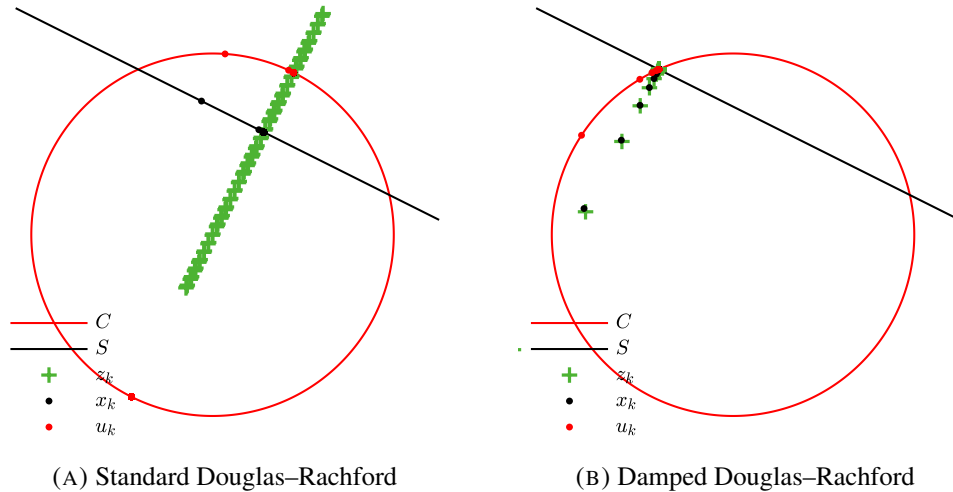(A) Standard Douglas–Rachford   (B) Damped Douglas–Rachford

FIGURE 1. Convergence behavior of the standard and damped Douglas–Rachford for solving the problem of a line intersecting with a circle.

Though dDR solves the problem, as remarked in the original paper [12], it may also converges to some stationary point of (3.3), which is not a solution. In fact, as we shall see in the numerical

experiments, for the Sudoku puzzle and the *s*-queens puzzle, dDR with $\gamma = \frac{1}{5} \in]0, \sqrt{3/2} - 1[$ fails all tests while sDR achieves very good performance; see Table 1.

3.2. **Problems with more than two sets.** Up to now, we have been dealing with 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, and let C_i be a non-empty closed set for each $i \in \{1, \ldots, m\}$. Consider the following feasibility problem

$$\text{find} \quad x \in \mathbb{R}^n \quad \text{s.t.} \quad x \in \bigcap_{i=1}^m C_i. \tag{3.5}$$

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

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

Let $\mathcal{C} \stackrel{\text{def}}{=} C_1 \times \cdots \times C_m$. Then $\mathcal{C} \subset \mathcal{H}$. Denote the subspace $\mathcal{S} \stackrel{\text{def}}{=} \{x = (x_i)_i \in \mathcal{H} : x_1 = \cdots = x_m\}$. The feasibility problem (3.5) can be reformulated into the following form

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

The projection operator of $\mathcal{C}$ is component-wise for each set $C_i, i = 1, \ldots, m$,

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

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

Adapting the standard Douglas-Rachford to the case of (3.5), we obtain

$$\begin{aligned}
&x_{k+1} = \frac{1}{m} \sum_{i=1}^m z_{i,k}, \\
&\text{For } i = 1, \ldots, m: \\
&\left| \begin{aligned}
&u_{i,k+1} \in \mathscr{P}_{C_i}(2x_{k+1} - z_{i,k}), \\
&z_{i,k+1} = z_{i,k} + u_{i,k+1} - x_{k+1}.
\end{aligned} \right.
\end{aligned} \tag{3.6}$$

Note that for the standard DR, there is no need to store 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:

$$\begin{aligned}
&\text{For } i = 1, \ldots, m: \\
&\left| \begin{aligned}
&x_{i,k+1} = \frac{1}{1+\gamma} \left(z_{i,k} + \gamma \frac{1}{m} \sum_{j=1}^m z_{j,k} \right), \\
&u_{i,k+1} \in \mathscr{P}_{C_i}(2x_{i,k+1} - z_{i,k}), \\
&z_{i,k+1} = z_{i,k} + u_{i,k+1} - x_{i,k+1}.
\end{aligned} \right.
\end{aligned} \tag{3.7}$$

4. LOCAL CONVERGENCE OF THE DOUGLAS-RACHFORD SPLITTING

In this section, we present our main result, the local convergence of Douglas–Rachford splitting. 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

$$x^\star - z^\star \in \text{int}\left(\mathcal{N}_C(x^\star)\right) \tag{4.1}$$

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

Remark 4.1. The non-degeneracy condition (4.1) requires $\mathcal{N}_C(x^\star)$ 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.
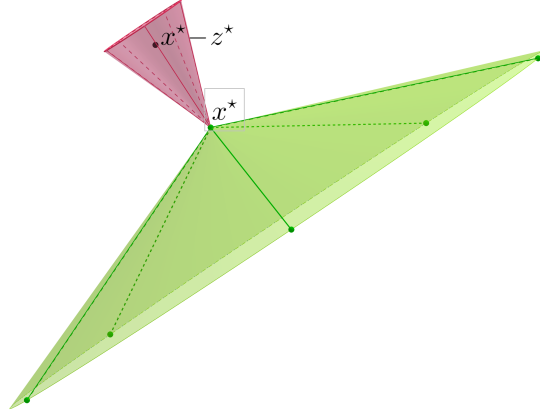
FIGURE 2. Normal cone (red) at a point x^* in the polytope C (green).

4.1. **Local convergence of Douglas-Rachford splitting.** We start with the standard Douglas-Rachford splitting 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 the standard Douglas-Rachford splitting method, for what follows, we impose the global convergence as an assumption, *i.e.* (A.4) holds.

Theorem 4.1 (Finite termination of DR). *For feasibility problem* (3.1) *and Douglas-Rachford iteration* (3.2)*, suppose that 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* $z^\star \in \text{Fix}(\mathscr{F}_{\text{DR}})$ *being a fixed point and* $x^\star = \mathscr{P}_S(z^\star)$. *If, moreover, non-degeneracy condition* (4.1) *holds, then* $\{u_k, x_k, z_k\}_{k\in\mathbb{N}}$ *converges to* $(x^\star, x^\star, z^\star)$ *finitely.*

Remark 4.2. It is worth noting that Theorem 4.1 also holds true for the convex setting. In [13], the authors studied the DR for solving convex affine-polyhedral feasibility problems, and imposed the following condition for finite convergence

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

which does not hold for the non-convex case as the interior of C in (3.1) may be empty (see also Section 4.2 the puzzles for which (4.2) fails). In [9], when the non-convex set is finite, the finite termination was 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 [13], as we characterize the situation where finite convergence happens but (4.2) fails.

Proof. The imposed global convergence of (3.2) means

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

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 $\mathscr{B}$ such that

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

By the definition of convergence, one concludes that there exists $K \in \mathbb{N}$ such that $2x_{k+1} - z_{k+1} \in \mathscr{B}$ for all $k \geq K$. Consequently, by the update of u_{k+1} in (3.4), one has

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

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 $k \geq K$. It follows that

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

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 $k > K$, the finite termination of z_k. The finite convergence of x_k follows naturally that of z_k, and we conclude the proof. $\square$

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

$$\begin{aligned}
x_{k+1} &\in \mathscr{P}_C(z_k), \\
u_{k+1} &= \mathscr{P}_S(2x_{k+1} - z_k), \\
z_{k+1} &= z_k + u_{k+1} - x_{k+1}.
\end{aligned} \tag{4.3}$$

The corollary below shows that the finite termination holds for (4.3).

Corollary 4.1. *For feasibility problem* (3.1) *and Douglas-Rachford iteration* (4.3), *suppose that 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,*

$$-(x^\star - z^\star) \in \mathrm{int}\big(\mathscr{N}_C(x^\star)\big), \tag{4.4}$$

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 4.1, we can easily derive the finite termination of x_k under new non-degeneracy condition (4.4). 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 4.1. Hence we conclude the proof immediately. $\qquad\square$

4.1.2. *The damped Douglas–Rachford splitting.* We now turn to the local convergence analysis of damped Douglas-Rachford splitting (3.4), for which we have the following result.

Theorem 4.2 (Local convergence of dDR). *For feasibility problem* (3.1) *and damped Douglas-Rachford iteration* (3.4), *suppose that Assumptions* (**A.1**)-(**A.3**) *hold, and* (3.4) *is ran under the conditions of Theorem 3.1. Then* $(u_k, x_k, z_k) \to (x^\star, x^\star, z^\star)$ *with* $z^\star \in \mathrm{Fix}(\mathscr{F}_{\mathrm{dDR}})$ *being a fixed point and* $x^\star$ *a stationary point of* (3.3). *If, moreover, condition* (4.1) *holds, then*
   (i) *u_k converges in finite number of iterations. $u_k = x^\star$.*
  (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 4.1. For the update of x_k in (3.4), since S is an affine subspace, $\mathscr{P}_S$ is linear

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

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

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

Note that the spectral radius of the matrix appears above is

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

which is combined with the fact the matrix is symmetric and normal, we conclude from Lemma 2.1 that $\frac{\gamma}{1+\gamma}$ is the local linear convergence rate of $\|z_k - z^\star\|$. $\qquad\square$

Remark 4.3. In [12], the authors also discussed the local linear convergence of the damped DR under the following constraint qualification condition

$$\mathscr{N}_S\big(\mathscr{P}_S(x^\star)\big) \bigcap -\mathscr{N}_C(x^\star) = 0. \tag{4.5}$$

As shown in [12, Proposition 2], such a condition allows to show $x^\star \in C \cap S$ and $z^\star = x^\star$; See Example 3.1 which satisfies the above condition. The update of x_{k+1} in (3.4) yields

$$\frac{1+\gamma}{\gamma}(z^\star - x^\star) = z^\star - \mathscr{P}_S(z^\star) \in \mathscr{N}_S\big(\mathscr{P}_S(x^\star)\big),$$

$$z^\star - x^\star \in -\mathscr{N}_C(x^\star).$$

This implies that only the fixed-points $z^\star$ such that $z^\star = x^\star$ satisfy qualification condition (4.5). In comparison, our non-degeneracy condition is more general than (4.5) 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. Our result holds for all fixed-points of $\mathrm{Fix}(\mathscr{F}_{\mathrm{dDR}})$.

Remark 4.4. If S, instead of being an affine subspace, has locally smooth curvature around $x^\star$, then, according to the result of [14], 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. The examples of these two puzzles are provided in Figure 3 below.

(A) Sudoku puzzle   (B) Eight queens puzzle
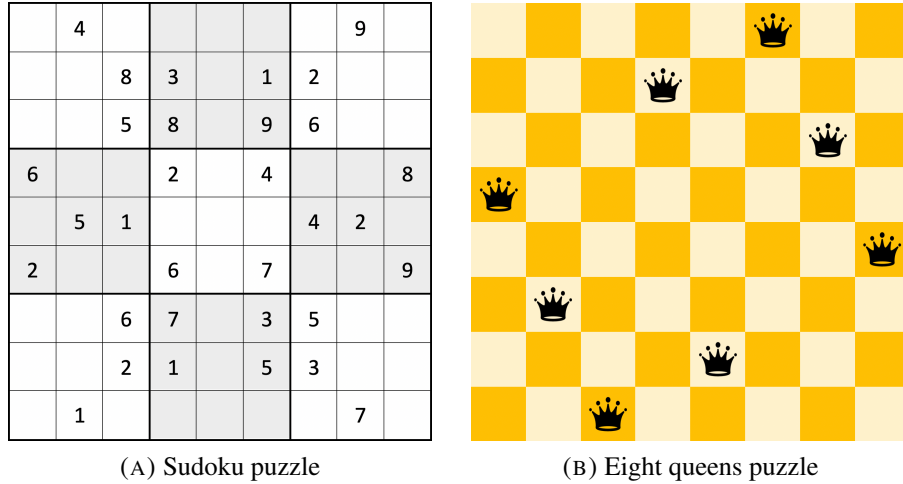
FIGURE 3. Examples of Sudoku and eight queens. The goal of Sudoku is to complete the grid such that each row, column, and 3×3 square contains all the digits from 1 to 9. The goal of eight queens is to place eight chess queens on an 8×8 board such that no two queens share the same row/column/diagonal.

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 a feasibility problem. Here we consider the formulation proposed in [28], which formulates Sudoku as a binary feasibility problem. We also refer to [24] for studies on the Sudoku puzzle and the 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 $e = \{1,0,\ldots,0\}$. This leads to four Sudoku feasibility constraints:

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

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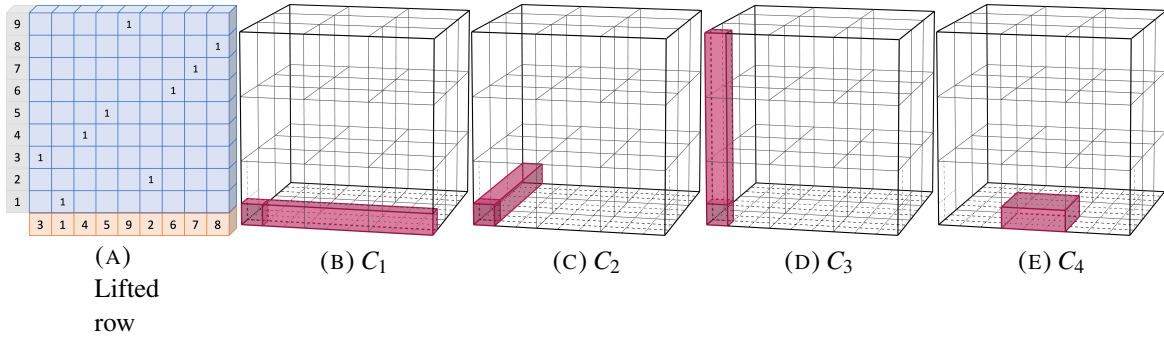
FIGURE 4. Lifted Sudoku problem. (a) the lifted representation of a row of numbers. (b)-(e) show what is a *lifted row/column/pillar/sub-grid* respectively.

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

$$\text{find} \quad x \in \mathbb{R}^{s \times s \times s} \text{ s.t. } x \in C_1 \cap C_2 \cap C_3 \cap C_4 \cap C_5. \tag{4.6}$$

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

Proposition 4.1 (Local convergence of DR). *For Sudoku puzzle* (4.6) *and Douglas-Rachford splitting* (3.6)*, suppose that Assumptions* (A.1)-(A.4) *hold. Then* $\{\boldsymbol{u}_k, x_k, \boldsymbol{z}_k\}_{k \in \mathbb{N}}$ *converges to* $(\mathcal{K}(x^\star), x^\star, \boldsymbol{z}^\star)$ *with* $\boldsymbol{z}^\star \in \text{Fix}(\mathscr{F}_{\text{DR}})$ *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*

$$x^\star - z_i^\star \in \text{int}\left(\mathscr{N}_{C_i}(x^\star)\right), \tag{4.7}$$

then, for all k large enough, there holds
- $u_{i,k} = x^\star$ *for* $i = 1, \ldots, 4$,
- $\|z_k - z^\star\| = O(\eta^k)$ *with* $\eta = \frac{\sqrt{5}}{5}$.

Proof. Denote $x_{k+1} = \mathcal{K}(x_{k+1})$. From the updates of x_k, we have that

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

with $\mathscr{P}_{\mathcal{S}} = \frac{1}{5} \mathbf{1}_{5 \times 5} \otimes \mathrm{Id}_{s^3 \times s^3}$, where $\mathbf{1}_{5 \times 5}$ stands for matrix of all 1 and $\otimes$ for Kronecker product. The separability of $\mathscr{P}_{\mathcal{C}}$ and the definition of projection operator lead to, for each $i = 1, \ldots, 5$,

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

Under non-degeneracy condition (4.7), apply the argument of Theorem 4.1 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

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

As a result, for k large enough, there holds

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

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

$$z_{k+1} - z^\star = (z_k - z^\star) + (u_{k+1} - u^\star) - (x_{k+1} - x^\star)$$
$$= (\mathbf{Id} + 2\mathscr{P}_{\mathcal{C}}\mathscr{P}_{\mathcal{S}} - \mathscr{P}_{\mathcal{C}} - \mathscr{P}_{\mathcal{S}})(z_k - z^\star),$$

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

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

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 immediately. $\qquad \square$

Next, we present the result for damped Douglas-Rachford splitting (3.4).

Proposition 4.2 (Local convergence of dDR). *For Sudoku puzzle* (4.6) *and damped Douglas-Rachford splitting* (3.7), *suppose that Assumptions* (A.1)-(A.3) *hold, and* (3.7) *is ran under the conditions of Theorem 3.1. Then* (u_k, x_k, z_k) *converges to* $(x^\star, x^\star, z^\star)$ *with* $z^\star$ *being a fixed point and* $x^\star$ *a stationary point of* $\min_x \{\mathrm{dist}^2(x, \mathcal{S}) \text{ s.t. } x \in \mathcal{C}\}$. *If, moreover, non-degeneracy condition* (4.7) *holds for* $C_{1,\ldots,4}$, *then, for all k large enough, it holds*
- $u_{i,k} = x^\star$ *for* $i = 1, \ldots, 4$,

- $\|z_k - z^\star\| = O(\eta^k)$ with $\eta = \frac{2\gamma + 5 + \sqrt{25 - 16\gamma^2}}{10(1 + \gamma)}$.

Proof. From the updates of x_{k+1}, we have that

$$x_{k+1} - x^\star = \frac{1}{1+\gamma}\big(z_k + \gamma \mathscr{P}_{\mathcal{S}}(z_k)\big) - \frac{1}{1+\gamma}\big(z^\star + \gamma \mathscr{P}_{bcS}(z^\star)\big)$$

$$= \frac{1}{1+\gamma}(z_k - z^\star) + \frac{\gamma}{1+\gamma}\mathscr{P}_{\mathcal{S}}(z_k - z^\star)$$

with $\mathscr{P}_{\mathcal{S}} = \frac{1}{5}\mathbf{1}_{5\times 5} \otimes \mathrm{Id}_{s^3 \times s^3}$. For u_k, the finite termination of $u_{i,k}, i = 1,\dots,4$ follows from the proof of Proposition 4.1. 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 $\mathscr{P}_{\mathcal{C}} \overset{\mathrm{def}}{=} \begin{bmatrix} \mathbf{0}_{4s^3 \times 4s^3} & \\ & \mathscr{P}_{C_5} \end{bmatrix}$. It follows that

$$u_{k+1} - u^\star = 2\mathscr{P}_{\mathcal{C}}(x_{k+1} - x^\star) - \mathscr{P}_{\mathcal{C}}(z_k - z^\star)$$

$$= \frac{2}{1+\gamma}\mathscr{P}_{\mathcal{C}}(z_k - z^\star) + \frac{2\gamma}{1+\gamma}\mathscr{P}_{\mathcal{C}}\mathscr{P}_{\mathcal{S}}(z_k - z^\star) - \mathscr{P}_{\mathcal{C}}(z_k - z^\star)$$

$$= \frac{1-\gamma}{1+\gamma}\mathscr{P}_{\mathcal{C}}(z_k - z^\star) + \frac{2\gamma}{1+\gamma}\mathscr{P}_{\mathcal{C}}\mathscr{P}_{\mathcal{S}}(z_k - z^\star).$$

Back to $z_{k+1} - z^\star$, we see that

$$z_{k+1} - z^\star = (z_k - z^\star) + (u_{k+1} - u^\star) - (x_{k+1} - x^\star)$$

$$= (z_k - z^\star) + \frac{1-\gamma}{1+\gamma}\mathscr{P}_{\mathcal{C}}(z_k - z^\star) + \frac{2\gamma}{1+\gamma}\mathscr{P}_{\mathcal{C}}\mathscr{P}_{\mathcal{S}}(z_k - z^\star)$$

$$\quad - \frac{1}{1+\gamma}(z_k - z^\star) - \frac{\gamma}{1+\gamma}\mathscr{P}_{\mathcal{S}}(z_k - z^\star)$$

$$= \frac{1}{1+\gamma}\big(\gamma\mathbf{Id} + 2\gamma\mathscr{P}_{\mathcal{C}}\mathscr{P}_{\mathcal{S}} + (1-\gamma)\mathscr{P}_{\mathcal{C}} - \gamma\mathscr{P}_{\mathcal{S}}\big)(z_k - z^\star)$$

$$= \frac{1}{1+\gamma}\big(\gamma(\mathbf{Id} + 2\mathscr{P}_{\mathcal{C}}\mathscr{P}_{\mathcal{S}} - \mathscr{P}_{\mathcal{C}} - \mathscr{P}_{\mathcal{S}}) + \mathscr{P}_{\mathcal{C}}\big)(z_k - z^\star).$$

Denote $M_\gamma = \frac{1}{1+\gamma}\big(\gamma(\mathbf{Id} + 2\mathscr{P}_{\mathcal{C}}\mathscr{P}_{\mathcal{S}} - \mathscr{P}_{\mathcal{C}} - \mathscr{P}_{\mathcal{S}}) + \mathscr{P}_{\mathcal{C}}\big)$. Let p, q be the rank of $\mathscr{P}_{\mathcal{C}}$ and $\mathscr{P}_{\mathcal{S}}$, respectively, and assume $p \leq q$ (for the case $p \geq q$, similar result can be obtained). Based on [22], there exists an orthogonal matrix U such that

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

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

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

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

$$
0 = \prod_{i=1}^{p} \left(\left(\frac{\gamma\alpha_i^2 + 1}{1+\gamma} - \lambda \right) \left(\frac{\gamma\alpha_i^2}{1+\gamma} - \lambda \right) + \frac{\gamma^2 \alpha_i^2 \beta_i^2}{(1+\gamma)^2} \right).
$$

Solving the quadratic equation for each i, we have

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

As in the proof of Proposition 4.1, we have that $\alpha_i = \frac{\sqrt{5}}{5}$ for all $i = 1, \dots, p$. Therefore, M_γ has only 4 distinct eigenvalues, which are

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

We also have

$$
\frac{\gamma}{1+\gamma} \le \frac{2\gamma + 5 + \sqrt{25 - 16\gamma^2}}{10(1+\gamma)} : \gamma \in]0,1],
$$

$$
\frac{\gamma}{1+\gamma} \ge \left| \frac{2\gamma + 5 + \sqrt{25 - 16\gamma^2}}{10(1+\gamma)} \right| : \gamma \in]1, +\infty[
$$

Next, we focus on $\gamma \in]0,1]$, and show that $\eta = \frac{2\gamma + 5 + \sqrt{25 - 16\gamma^2}}{10(1+\gamma)}$ is the convergence rate. To this end, we need to show that η 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:

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

As a result, we have

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

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

which means that η is semi-simple by Definition 2.6, and then we conclude the linear rate of convergence. $\qquad\square$

Remark 4.5. The proof of the two propositions above is dimension independent, which means the results hold true for all puzzle sizes of perfect squares s with $s \ge 4$; 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 8×8 chessboard so that no two queens threaten each other. The size of the 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.

We follow the setting of [28]. 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 a feasibility problem of four sets

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

Since all the sets above are binary, so is the set $\mathcal{C} \overset{\text{def}}{=} C_1 \times \cdots \times C_4$, as a result finite convergence can be obtained under the conditions of Theorems 4.1 and 4.2, for the standard Douglas–Rachford and the damped one, respectively.

5. NUMERICAL RESULTS

We now provide numerical results on Sudoku and *s*-queens puzzles to support our theoretical findings. Before analyzing the convergence rates, we first compare the performance of standard Douglas-Rachford splitting method (3.2) and damped one (3.4) on how successful are they when applied to solve these two puzzles. *i.e.* how often each method finds a feasible point.

The comparison is shown in Table 1. For (3.4), two choices of γ are considered: $\gamma = \frac{1}{5} \in \left]0, \sqrt{3/2} - 1\right[$ suggested by Lemma 3.1 and $\gamma = 99$ with online tracking rule suggested in [12, Remark 4]. 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. Also, a minimal 100 number of iteration is set. For a given puzzle, each method is repeated 10^3 times with different initialization for each running.

For Sudoku, the size of both puzzles are 9×9: "Puzzle 1" is provided with 37 digits, hence is easy; "Puzzle 2" has 22 given digits and is more difficult than "Puzzle 1"

- the standard DR solves both puzzles with 100% success rate, while the dDR with $\gamma = \frac{1}{5}$ fails all tests. The dDR with $\gamma = 99$ succeeds on "Puzzle 1", and the rate drops to about 88% for "Puzzle 2";
- in terms of number of iterations, the sDR needs much less number of iterations compared to those of the dDR with $\gamma = 99$.

For the *s*-queens puzzle, two different sizes are considered: $s = 8$ for "Puzzle 1" and $s = 16$ for "Puzzle 2".

---

https://en.wikipedia.org/wiki/Eight_queens_puzzle

For 9×9 Sudoku, to ensure the uniqueness of solutions, the smallest number of given digits of the puzzle is 17. See https://www.technologyreview.com/2012/01/06/188520/mathematicians-solve-minimum-sudoku-problem/

- similar to Sudoku, the dDR with $\gamma = \frac{1}{5}$ fails all tests. This time, between the sDR and the dDR with $\gamma = 99$, neither achieves 100% success rate with the dDR being better than sDR.
- in terms of number of iteration, same as the Sudoku case, the sDR is better.

The above observation, in particular, the failure of dDR with $\gamma = \frac{1}{5}$, is in contrast to Example 3.1. One possible reason leads to the failure of the dDR with $\gamma = \frac{1}{5}$ is that set C is finite and the dDR can not escape bad local stationary point with small value of γ.

TABLE 1. Comparison of success rate of standard DR and damped DR for solving Sudoku and *s*-queens puzzles over 1,000 random initializations.

		Puzzle 1			Puzzle 2		
		sDR	dDR $\gamma = \frac{1}{5}$	dDR $\gamma = 99$	sDR	dDR $\gamma = \frac{1}{5}$	dDR $\gamma = 99$
Sudoku	success rate	100%	0	100%	100%	0	89.7%
	avg. # of itr.	114	184	2710	408	184	5409
s-queens	success rate	94.8%	0	98.0%	90.2%	0	92.2%
	avg. # of itr.	653	100	2812	1286	100	3618

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 the 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 ℓ_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 $\|z_k - z^\star\|$. Taking $\|z_k - z^\star\|$ for example, we have that its convergence has two different regimes: *sub-linear* 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–achford 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 with $\gamma = 99$. The results on the Sudoku puzzle of size 9×9 and eight queens puzzle of size 8×8 are

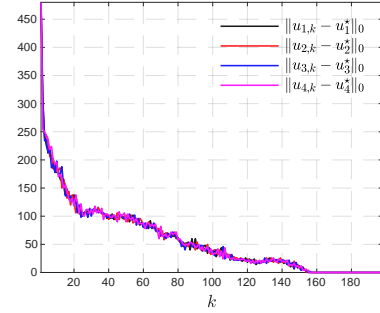(A) Size 4×4
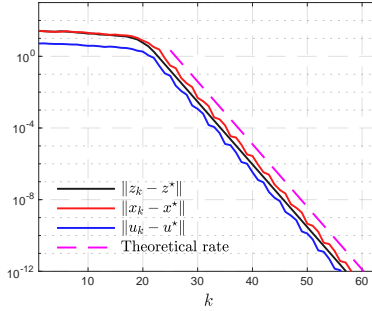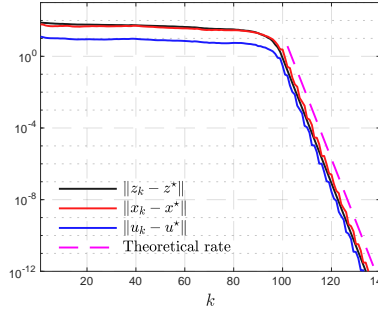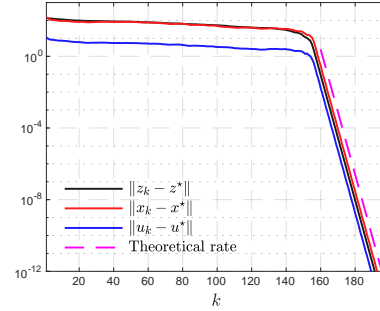

(B) Size 9×9

(C) Size 16×16

(D) Size 4: convergence of $u_{i,k}$
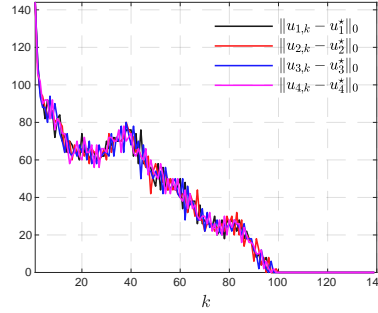
(E) Size 9: convergence of $u_{i,k}$

(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.

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. We omit the plots of the dDR with $\gamma = 99$ as they are very similar to those of Figure 7, except the different rates of local linear convergence.
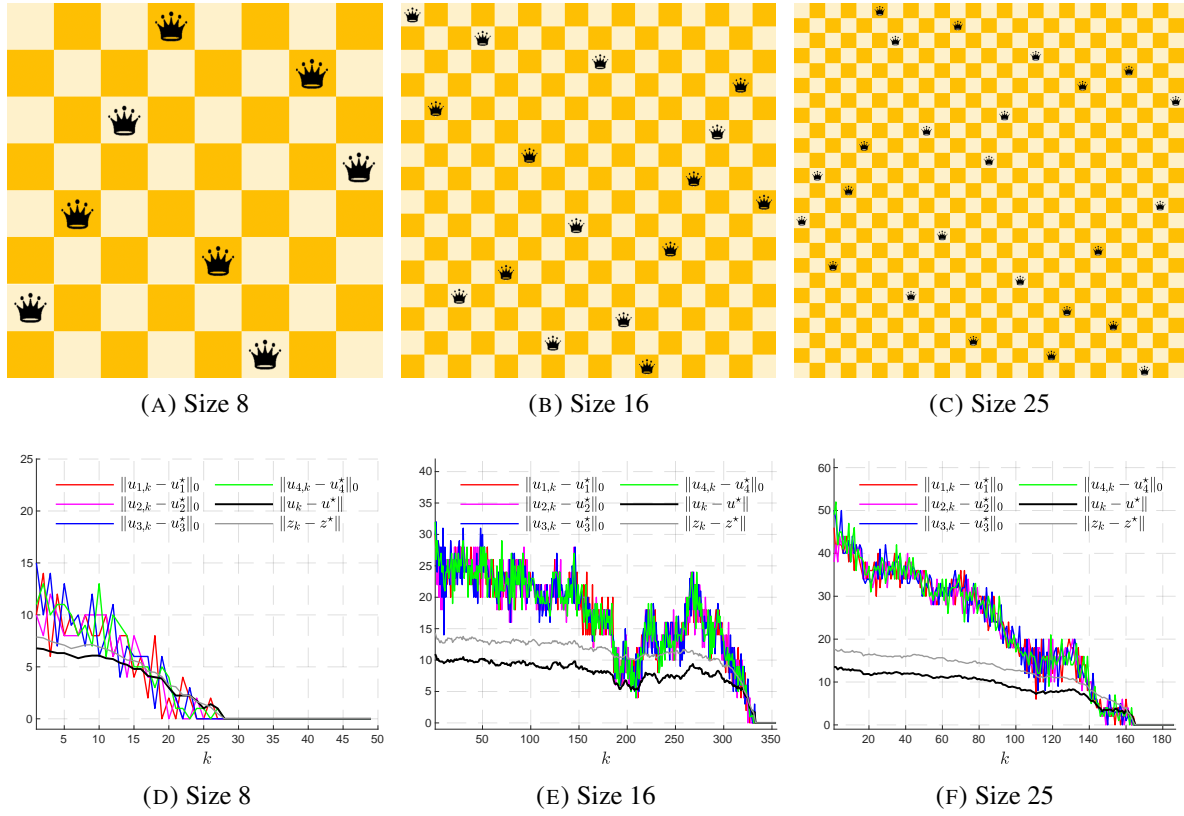
(A) Size 8   (B) Size 16   (C) Size 25

(D) Size 8   (E) Size 16   (F) Size 25

FIGURE 6. Different sizes of queens puzzles and convergence observations.
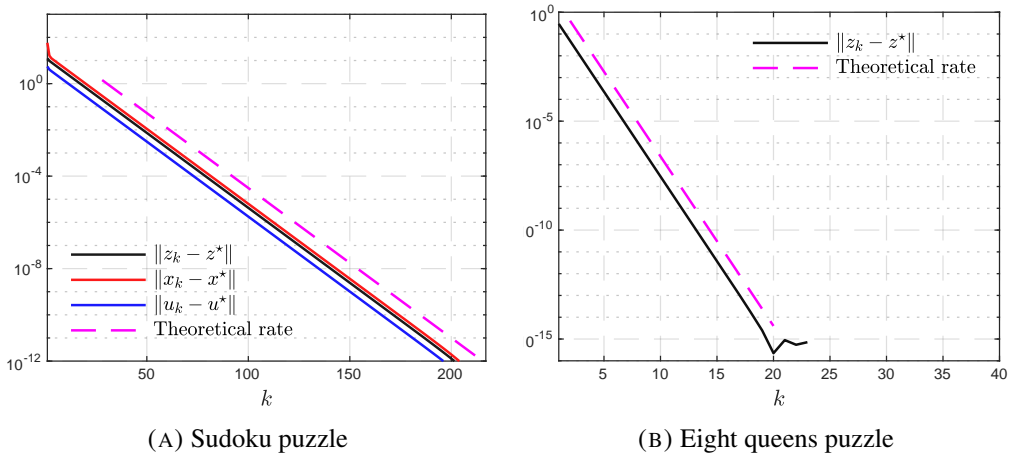

(A) Sudoku puzzle   (B) Eight queens puzzle

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

6. CONCLUSIONS

In this paper, we studied local convergence properties of the 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 were 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.

Acknowledgments

References

[1] J. von Neumann, Functional operators, vol. 2 (Annals of Mathematics Studies, no. 22), Princeton, NJ, 1950.

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

[3] A. Cegielski, A. Suchocka, Relaxed alternating projection methods, SIAM J. Optim. 19 (2008), 1093–1106.

[4] D.R. Luke, Relaxed averaged alternating reflections for diffraction imaging, Inverse Probl. 21 (2004), 37.

[5] P.L. Lions, B. Mercier, Splitting algorithms for the sum of two nonlinear operators, SIAM J. Nume. Anal. 16 (1979), 964–979.

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

[7] D.W. Peaceman, H.H. Rachford, Jr., The numerical solution of parabolic and elliptic differential equations, J. Soc. Indust. Appl. Math. 3 (1955), 28–41.

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

[9] H.H. Bauschke, M.N. Dao, On the finite convergence of the douglas–rachford algorithm for solving (not necessarily convex) feasibility problems in euclidean spaces, SIAM J. Optim. 27 (2017), 507–537.

[10] H. Attouch, J. Bolte, P. Redont, A. Soubeyran, Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality, Math. Oper. Res. 35 (2010), 438–457.

[11] F.J.A. Artacho, J.M. Borwein, Global convergence of a non-convex douglas–rachford iteration, J. Global Optim. 57 (2013), 753–769.

[12] G. Li, T.K. Pong, Douglas–rachford splitting for nonconvex optimization with application to nonconvex feasibility problems, Math. Program. 159 (2016), 371–401.

[13] H.H. Bauschke, M.N. Dao, D. Noll, H.M. Phan, On slater's condition and finite convergence of the douglas–rachford algorithm for solving convex feasibility problems in euclidean spaces, J. Global Optim. 65 (2016), 329–349.

[14] J. Liang, J. Fadili, G. Peyré, Local convergence properties of Douglas–Rachford and alternating direction method of multipliers, J. Optim. Theory Appl. 172 (2017), 874–913.

[15] A.S. Lewis, Active sets, nonsmoothness, and sensitivity, SIAM J. Optim. 13 (2003), 702–725.

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

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

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

[19] R. Hesse, D.R. Luke, P. Neumann, Projection methods for sparse affine feasibility: Results and counterex-
     amples, Technical Report, 2013.

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

[21] C.D. Meyer, Matrix Analysis and Applied Linear Algebra, vol. 2. SIAM, 2000.

[22] H.H. Bauschke, J.Y. Bello Cruz, T.T.A. Nghia, H.M. Pha, X. Wang, Optimal rates of linear convergence of
     relaxed alternating projections and generalized douglas-rachford methods for two subspaces, Numer. Algo.
     73 (2016), 33–76.

[23] R. Hesse, D. R. Luke, P. Neumann, Alternating projections and Douglas-Rachford for sparse affine feasibil-
     ity, IEEE Trans. Signal Process. 62 (2014), 4868–4881.

[24] F.J.A. Artacho, J.M. Borwein, M.K. Tam, Recent results on douglas–rachford methods for combinatorial
     optimization problems, J. Optim. Theory Appl. 163 (2014), 1–30.

[25] H.H. Bauschke, D. Noll, On the local convergence of the douglas–rachford algorithm, Archiv der Math. 102
     (2014), 589–600.

[26] F.J.A. Artacho, J.M. Borwein, M.K. Tam, Global behavior of the douglas–rachford method for a nonconvex
     feasibility problem, J. Global Optim. 65 (2016), 309–327.

[27] A. Themelis, P. Patrinos, Douglas–rachford splitting and admm for nonconvex optimization: Tight conver-
     gence results, SIAM J. Optim. 30 (2020), 149–181.

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