On the convergence rate of the Douglas-Rachford splitting algorithm

Hadi Abbaszadehpeivasti

Department of Econometrics and Operations Research, Tilburg University, Tilburg, The Netherlands, h.peivasti@tilburguniversity.edu

Moslem Zamani

Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, Eindhoven, The Netherlands, m.zamani@tue.nl

Abstract. This work is concerned with the convergence rate analysis of the Douglas–Rachford splitting (DRS) method for finding a zero of the sum of two maximally monotone operators. We obtain an exact rate of convergence for the DRS algorithm and demonstrate its sharpness in the setting of convex feasibility problems. Furthermore, we investigate the linear convergence of the DRS algorithm, providing both necessary and sufficient conditions that characterize this behavior. We further examine the performance of the DRS method when applied to convex composite optimization problems. The paper concludes with several conjectures on the convergence behavior of the DRS algorithm for this class of problems.

Key words: Douglas-Rachford splitting method, Convergence rate, Performance estimation

1. Introduction

We consider the following monotone inclusion problem

$$\underset{x \in \mathbb{R}^n}{\text{find}} \quad 0 \in Ax + Bx, \tag{1}$$

where $A, B : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ are maximally monotone operators. The monotone inclusion problem provides a unifying framework that captures numerous fundamental problems in convex optimization, variational inequalities, and equilibrium models [5, 7, 24].

Among the iterative schemes devised for solving monotone inclusions, the Douglas–Rachford splitting (DRS) algorithm [7] has emerged as one of the most effective and versatile methods. Its popularity stems from its ability to decompose complex problems into simpler subproblems. Furthermore, several important algorithms, including the alternating direction method of multipliers (ADMM), can be analyzed through the lens of the DRS algorithm; see [24, Chapter 3] for further

details. Indeed, the close connection of the DRS method to other fundamental operator-splitting techniques, notably ADMM, further highlights its central role in modern optimization [24].

The DRS method is originally introduced for solving linear systems that appeared in the numerical treatment of partial differential equations (heat equation) [8]. Lions and Mercier [18] established that the method is also convergent for a general maximal monotone inclusion problem in the form of (1). Eckstein and Bertsekas [10] showed that the DRS method can be interpreted as a proximal point method. We refer the interested reader to [13] for a historical review of the DRS. The method is given in Algorithm 1; see Section 2 for the notations used in Algorithm 1.

Algorithm 1 The Douglas-Rachford splitting method

Parameters: number of iterations N, positive stepsize $\gamma > 0$ and relaxation factor $\lambda \in (0, 2)$.

Inputs: maximally monotone operators A and B, initial iterate $w^1 \in \mathbb{R}^n$.

For $k = 1, 2, \ldots, N$ perform the following steps:

 i) $x^k = J_{\gamma B}(w^k)$.

 ii) $y^k = J_{\gamma A}(2x^k - w^k)$.

 iii) $w^{k+1} = w^k + \lambda(y^k - x^k)$.

Algorithm 1 can be written as a fixed point iteration $w^{k+1} = Tw^k$, where

$$T = (1 - \tfrac{\lambda}{2})I + \tfrac{\lambda}{2}R_{\gamma A}R_{\gamma B}. \tag{2}$$

and R denotes the reflected resolvent transformation. We call T the Douglas-Rachford (DR) operator throughout the text. It is worth noting that $x^\star$ is a solution of problem (1) if and only if there exists $w^\star$ such that $x^\star = J_{\gamma B}w^\star$ and $w^\star$ is a fixed point of the DR operator. We denote the fixed points of operator T by $W^\star$.

We adopt the following assumptions to establish convergence rate results.

ASSUMPTION 1. *Operators A and B are maximally monotone operators and the DR operator* (2) *for the given parameters admits a fixed point w^*.*

Note that under these assumptions, $w^k \to w^\star$ for some $w^\star \in W^\star$; see [5, Theorem 26.11]. In this manuscript, we employ the performance estimation framework to establish our main results. Performance estimation provides a strong and unifying approach for the analysis and design of first-order methods. The framework is introduced in the seminal work of Drori and Teboulle [9],

and has since attracted significant attention in the optimization community. For a comprehensive review of the performance estimation, we refer the interested reader to [1, 25, 26].

This paper is structured as follows. Section 2 introduces the definitions and fundamental concepts that will be used throughout the subsequent sections. In Section 3, we analyze the sublinear convergence of the DRS method. Section 4 is dedicated to the study of its linear convergence. Finally, in Section 5, we investigate the behavior of the DRS method in the context of convex composite optimization problems, and we conclude the paper with some conjectures on the convergence rate of the DRS algorithm for this class of problems.

2. Terminology and notation

We briefly recall the key definitions employed in this paper. For further details, the reader is referred to standard references [5, 24].

We denote the Euclidean inner product and norm by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively. We use I to denote the identity operator. For a (set-valued) operator $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, we define its graph by $\operatorname{gra} A = \{(x, u) : u \in Ax\}$ and we denote its inverse by A^{-1}. An operator $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is called monotone if

$$\langle u - v, x - y \rangle \geq 0 \quad \forall (x, u), (y, v) \in \operatorname{gra} A,$$

and it is called maximally monotone if its graph is not properly contained in that of any other monotone operator. The operator A is β-cocoercive if for some $\beta > 0$, we have

$$\langle u - v, x - y \rangle \geq \beta \|u - v\|^2 \quad \forall (x, u), (y, v) \in \operatorname{gra} A.$$

It is seen that cocoercivity implies Lipschitz continuity but its converse does not necessarily hold. Consequently, a cocoercive operator is single-valued. The operator A is called nonexpansive if it is Lipschitz continuous with modulus one. An operator A is called μ-strongly monotone if its inverse is μ-cocoercive. The resolvent of operator A is denoted and defined as $J_A = (I + A)^{-1}$. In addition, we use $R_A = 2J_A - I$ to denote the reflected resolvent of A.

Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be an extended convex function. The function f is called closed if its epigraph is closed, i.e. $\{(x, r) : f(x) \leq r\}$ is a closed subset of $\mathbb{R}^{n+1}$. The function f is said to be proper if there exists $x \in \mathbb{R}^n$ with $f(x) < \infty$. The subgradients of f at x are denoted and defined as

$$\partial f(x) = \{u : f(y) \geq f(x) + \langle u, y - x \rangle, \forall y \in \mathbb{R}^n\}.$$

When f is a closed proper convex function, ∂f is a maximally monotone operator. The convex function f is called L-smooth if ∂f is $\frac{1}{L}$-cocoercive. The function f is said to be μ-strongly convex

if ∂f is μ-strongly monotone. The proximal operator of a closed proper convex function f is defined and denoted as $\text{prox}_f = (I + \partial f)^{-1}$.

Given closed convex set C, the normal cone operator N_C of C is defined by

$$N_C(x) = \{u \in \mathbb{R}^n : \langle u, y - x \rangle \le 0, \ \forall y \in C\}, \quad x \in C,$$

with $N_C(x) = \emptyset$ otherwise. The projection onto C is written as Π_C. The corresponding distance function to C is denoted by D_C.

3. Sublinear convergence of the DRS method

In this section, we study the convergence of the DRS algorithm for a general monotone inclusion problem. Regarding the DRS algorithm, we have

$$\left\| Tw^N - w^N \right\|^2 \le \frac{\lambda (N-1)^{(N-1)}}{(2-\lambda)N^N} \left\| w^1 - w^\star \right\|^2, \tag{3}$$

when $\gamma > 0$ and $\lambda \in [1, 1 + \sqrt{\frac{k-1}{k}})$ in Algorithm 1; see [17, Theorem 4.9]. Indeed, this result is established for the Krasnoselski-Mann iteration for a general nonexpansive operator, and the convergence rate of the DRS algorithm emerges as a specific instance. Since the proof in [17] is rather technical and lengthy (spanning seven pages), we provide here a proof for the case $\lambda = 1$ to keep the presentation self-contained. We present the theorem for a general nonexpansive operator and the convergence rate of the DRS follows from it as $R_{\gamma A} R_{\gamma B}$, the Peaceman–Rachford operator, is nonexpansive [24]. Note that the DR operator for $\lambda = 1$ is $T = \frac{1}{2}I + \frac{1}{2}R_{\gamma A} R_{\gamma B}$.

THEOREM 1. *Let $S : \mathbb{R}^n \to \mathbb{R}^n$ be a nonexpansive operator and let S have a fixed point $w^\star$, i.e. $Sw^\star = w^\star$. Suppose that $w^1 \in \mathbb{R}^n$ the sequence $\{w^k\}$ is generated by $w^{k+1} = (\frac{1}{2}I + \frac{1}{2}S)w^k$. Then*

$$\left\| \frac{1}{2}Sw^N - \frac{1}{2}w^N \right\|^2 \le \frac{(N-1)^{(N-1)}}{N^N} \left\| w^1 - w^\star \right\|^2. \tag{4}$$

Proof. To establish the convergence rate, we demonstrate its validity by summing a series of valid inequalities. Without loss of generality, we assume that $w^\star = 0$ (consider the nonexpansive operator given by $S(w + w^\star) - w^\star$). Due to nonexpansivity of S, we have

$$\left\| w^{k+1} - w^k \right\|^2 - \left\| 2w^{k+2} - w^{k+1} - (2w^{k+1} - w^k) \right\|^2 \ge 0, \quad \left\| w^k - w^\star \right\|^2 - \left\| 2w^{k+1} - w^k - w^\star \right\|^2 \ge 0,$$

where we use $Sw^k = 2w^{k+1} - w^k$ and $Sw^\star = w^\star$. Summing these inequalities with the indicated multipliers yields

$$\sum_{k=1}^{N-1} \frac{k(N-1)^{N-k-1}}{2N^{N-k}} \left(\left\| w^{k+1} - w^k \right\|^2 - \left\| 2w^{k+2} - w^{k+1} - (2w^{k+1} - w^k) \right\|^2 \right) + \frac{1}{2N} \left(\left\| w^N \right\|^2 - \left\| 2w^{N+1} - w^N \right\|^2 \right)$$

$$+ \sum_{k=1}^{N-2} \frac{(N-k-1)(N-1)^{N-k-1}}{2N^{N+1-k}} \left(\left\| w^k \right\|^2 - \left\| 2w^{k+1} - w^k \right\|^2 \right) = \frac{(N-1)^{N-1}}{N^N} \left\| w^1 - w^\star \right\|^2 - \left\| w^{N+1} - w^N \right\|^2$$

$$- \left\| w^{N+1} - \frac{2(N-1)}{N} w^N + \frac{N-1}{N} w^{N-1} \right\|^2 - \sum_{k=1}^{N-2} \frac{k(N-1)^{N-2-k}}{N^{N-k-1}} \left\| w^{k+2} - \frac{2(N-1)}{N} w^{k+1} + \frac{N-1}{N} w^k \right\|^2 \geq 0.$$

The above inequality implies the convergence rate (4) as $w^{N+1} = \frac{1}{2} Sw^N + \frac{1}{2} w^N$ and the proof is complete. □

In the following, we demonstrate that convergence rate (3) is tight by constructing a convex feasibility problem that achieves this rate.

THEOREM 2. *Let $\gamma > 0$, $\lambda = 1$, and $N \geq 2$ in Algorithm 1. Then there exist maximally monotone operators A and B such that after N iterations of the algorithm, we have*

$$\left\| T(w^N) - w^N \right\|^2 = \frac{(N-1)^{N-1}}{N^N} \left\| w^1 - w^\star \right\|^2,$$

where $w^\star$ denotes a fixed point of the DR operator.

Proof. Consider the following subspaces in $\mathbb{R}^2$,

$$P = \{(x, 0) : x \in \mathbb{R}\}, \quad Q = \{(x, \tfrac{x}{\sqrt{N-1}}) : x \in \mathbb{R}\}.$$

Consider the following problem,

$$\underset{x \in \mathbb{R}^2}{\text{find}} \quad 0 \in N_P(x) + N_Q(x).$$

Assume A and B denote the normal cone mappings of Q and P, respectively. For the given problem, the DR operator is

$$T = \sqrt{\frac{N-1}{N}} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

where $\theta = \arcsin\left(\frac{1}{\sqrt{N}}\right)$. It is easily seen that zero is the solution to the problem and the fixed point of operator T. Consider $w^1 = \begin{pmatrix} \cos\phi & \sin\phi \end{pmatrix}^T$, for some arbitrary ϕ. It is seen

$$\left\| w^N \right\| = \left(\frac{N-1}{N} \right)^{\frac{N-1}{2}}, \qquad \left\| w^{N+1} \right\| = \left(\frac{N-1}{N} \right)^{\frac{N}{2}},$$

and the angle between them is θ. By the law of cosines, we get

$$\left\| w^{N+1} - w^N \right\|^2 = \frac{(N-1)^{N-1}}{N^N}, \tag{5}$$

and the proof is complete. □

It is worth noting that θ used in the above theorem is equivalent to the Friedrichs angle between P and Q. A notable implication of Theorem 2 is that the convergence rate of the DRS algorithm for convex feasibility problem cannot exceed that of a general monotone inclusion problem. Moreover, in the proof of Theorem 2, if we select ϕ such that $w^N \in P$, we have

$$D_P(y^N) = \sqrt{\frac{(N-1)^{N-1}}{N^N}}.$$

This bound resembles the rate established in [30, Theorem 4.4] for the alternating projection method.

One may wonder if convergence rate (3) can be improved in the presence of cocoercivity. The answer is negative. Assume that $N \geq 2$ is given. Consider the following problem,

$$\operatorname*{find}_{x \in \mathbb{R}^2} \quad 0 \in Ax + Bx,$$

where $A, B : \mathbb{R}^2 \to \mathbb{R}^2$, B is the zero operator and $Ax = \left(\frac{-x_2}{\sqrt{N-1}} \quad \frac{x_1}{\sqrt{N-1}} \right)^T$. It is readily seen that both A and B are monotone and B is cocoercive. In addition, the DR operator is the same as that given in Theorem 2 for $\lambda = 1$. Therefore, it follows from the theorem we have (5) for any unit vector $w^1 \in \mathbb{R}^2$.

As the aforementioned example illustrates, cocoercivity of B is not a sufficient condition to improve convergence rate (3). However, it appears that if A is a subdifferential operator, the rate can be improved. We conclude this section with a conjecture regarding this topic, which is informed by numerical experiments conducted within the performance estimation framework.

CONJECTURE 1. *Let Assumption 1 hold. Assume that B is a β-cocoercive operator and $A = \partial g$ for some closed proper convex function g. If the sequence $\{w^k\}$ is generated by Algorithm 1 with stepsize $\gamma \in (0, \beta)$ and relaxation parameter $\lambda \in (0, 2)$, then*

$$\left\| T(w^N) - w^N \right\|^2 \leq \frac{\lambda^2}{((N-1)\lambda+1)^2} \left\| w^1 - w^\star \right\|^2.$$

4. Linear convergence of the DRS method

In this section, we study the linear convergence of the DRS algorithm. The linear convergence rate of the DRS algorithm under strong monotonicity has been extensively studied in recent years, and exact rates have been established. Interested readers may refer to [11, 12, 23] for further details.

As strong convexity is not commonly satisfied in many applications, alternative assumptions have been considered in the study of the linear convergence rate of first-order methods; see e.g. [4, 16, 28]. We study the linear convergence of the DRS algorithm under the error bound condition introduced in [22].

DEFINITION 1. Let $W \subseteq R^n$ and let T denote the DR operator for some $\gamma > 0$ and $\lambda \in (0, 2)$. The operator T satisfies the error bound condition on W if there exists $\mu \geq 0$ such that

$$\mathrm{D}_{W^\star}(w) \leq \mu \left\| (I - T)w \right\|, \quad \forall w \in W. \tag{6}$$

It is worth noting that this definition resembles the error bound condition introduced in [19]. The following theorem investigates the convergence rate of the DRS algorithm in terms of the distance of w^k to the fixed point set of the DR operator.

THEOREM 3. *Let Assumption 1 hold and let $\gamma > 0$ and $\lambda \in (0, 2)$. Suppose that $\{(x^k, y^k, w^k)\}$ is generated by Algorithm 1 with initial point w^1. If T satisfies error bound condition (6) on $W = \{w : \mathrm{D}_{W^\star}(w) \leq \mathrm{D}_{W^\star}(w^1)\}$, then*

$$\mathrm{D}_{W^\star}(w^{k+1}) \leq \sqrt{1 - \tfrac{1}{\mu^2}(\tfrac{2}{\lambda} - 1)} \, \mathrm{D}_{W^\star}(w^k).$$

Proof. Let $w^\star$ denote the nearest point to w^k in $W^\star$, that is $\mathrm{D}_{W^\star}(w^k) = \|w^k - w^\star\|$. By [22, Theorem 1], we have $w^k \in W$. Assume that $x^\star = J_{\gamma B}(w^\star)$. Therefore, we have $\frac{1}{\gamma}(w^\star - x^\star) \in Bx^\star$ and $\frac{1}{\gamma}(x^\star - w^\star) \in Ax^\star$. In addition, $\frac{1}{\gamma}(w^k - x^k) \in Bx^k$ and $\frac{1}{\gamma}(2x^k - y^k - w^k) \in Ay^k$. By the monotonicity, we have

$$\langle w^k - x^k - (w^\star - x^\star), x^\star - x^k \rangle \leq 0,$$

$$\langle 2x^k - y^k - w^k - (x^\star - w^\star), x^\star - y^k \rangle \leq 0.$$

By summing these inequalities, we get

$$\langle w^k + y^k - x^k - w^\star, y^k - x^k \rangle =$$

$$\tfrac{1}{\lambda} \langle w^k - w^\star + \tfrac{1}{\lambda}(w^{k+1} - w^k), w^{k+1} - w^k \rangle \leq 0, \tag{7}$$

where the last equality follows from $w^{k+1} = w^k + \lambda(y^k - x^k)$. By the error bound condition, we have

$$\left\| w^k - w^\star \right\|^2 - \mu^2 \left\| w^{k+1} - w^k \right\|^2 \leq 0. \tag{8}$$

By multiplying (7) and (8) by 2λ and $\frac{2-\lambda}{\lambda\mu^2}$, respectively, we get

$$\left\| w^{k+1} - w^\star \right\|^2 \leq \left(1 - \frac{1}{\mu^2}(\frac{2}{\lambda} - 1) \right) \left\| w^k - w^\star \right\|^2 .$$

The above inequality implies the desired inequality since $\| w^{k+1} - w^\star \| \geq \mathrm{D}_{W^\star}(w^{k+1})$ and the proof is complete. $\quad\square$

It is seen from the theorem that μ cannot take a value less than $\frac{2}{\lambda} - 1$. Peña et al. established the linear convergence of the DRS algorithm when $\gamma = \lambda = 1$; see [22, Theorem 1]. In comparison with the above-mentioned result, Theorem 3 studies the convergence rate for all possible stepsizes and relaxation parameters. Moreover, it coincides with their result for $\gamma = \lambda = 1$.

In the following proposition, we establish that error bound condition (6) is a necessary condition for the linear convergence in terms of $\mathrm{D}_{W^\star}(w^k)$.

PROPOSITION 1. *Let Assumption 1 hold and let $\gamma > 0$ and $\lambda \in (0, 2)$. If Algorithm 1 is linearly convergent with rate $r \in (0, 1)$ in terms of $\mathrm{D}_{W^\star}(w^k)$, then T satisfies error bound condition (6) on $W = \{w : \mathrm{D}_{W^\star}(w) \leq \mathrm{D}_{W^\star}(w^1)\}$ with $\mu = \frac{1}{1-r}$.*

Proof. Let $w \in W \setminus W^\star$ and $\|Tw - w^\star\| = \mathrm{D}_{W^\star}(Tw)$. As the algorithm is linearly convergent, we have

$$\|Tw - w^\star\|^2 \leq r^2 \, \mathrm{D}_{W^\star}^2(w) \leq r^2 \|w - w^\star\|^2.$$

Therefore,

$$(1 - r^2)\|w - w^\star\|^2 \leq \|w - w^\star\|^2 - \|Tw - w^\star\|^2$$

$$= \left(\|w - w^\star\| + \|Tw - w^\star\| \right) \left(\|w - w^\star\| - \|Tw - w^\star\| \right)$$

$$\leq (1 + r) \left\| w - w^\star \right\| \left\| Tw - w \right\|,$$

where the last inequality follows from the assumptions and the triangle inequality. By dividing both sides of the inequality by $\|w - w^\star\|$, we get the desired inequality and complete the proof. $\quad\square$

Peña et al.[22, Theorem 2] proved that error bound condition (6) holds under the following assumptions:

i) $B = \partial f$ and $A = \partial g$ for some proper convex functions f and g;

ii) f or g is strongly convex;

iii) f or g is L-smooth.

This result demonstrates that the error bound condition is less restrictive than the strong convexity assumption, which is commonly employed to establish linear convergence rate. In the following proposition, we show that the error bound condition may hold under assumptions more relaxed than those stated above. To this end, we employ restricted strong monotonicity notion.

DEFINITION 2. Let A and B be maximal monotone operators. We say that the problem (1) satisfies restricted strong monotonicity if there exists $\mu_f > 0$ with

$$\langle u + v, x - x^\star \rangle \geq \mu_f \|x - x^\star\|^2, \quad \forall u \in Ax, \forall v \in Bx, \tag{9}$$

where $x^\star = \Pi_{X^\star}(x)$.

PROPOSITION 2. *Let Assumption 1 hold and let B be β-cocoercive. If problem (1) satisfies restricted strong monotonicity with modulus μ_f, then T satisfies error bound condition (6) on $W = \{w : D_{W^\star}(w) \leq D_{W^\star}(w^1)\}$ with*

$$\mu = \frac{\gamma + \gamma \min(\mu_f \beta, 1) + \beta}{\lambda \gamma \min(\mu_f \beta, 1)},$$

for any $\lambda \in (0, 2)$ and $0 < \gamma \leq \beta$.

Proof. Without loss of generality, we assume $\beta = 1$ and $x^\star = 0$, since problem (1) is equivalent to $0 \in \frac{1}{\beta} A(x - x^\star) + \frac{1}{\beta} B(x - x^\star)$. Assume that $x^\star = J_{\gamma B}(w^\star)$. So $Bx^\star = \frac{1}{\gamma}(w^\star - x^\star)$ and $\frac{-1}{\gamma}(w^\star - x^\star) \in Ax^\star$. Suppose that $w^k \in W$ and

$$x^k = J_{\gamma B} w^k, \qquad y^k = J_{\gamma A}(2x^k - w^k), \qquad w^{k+1} = Tw^k = w^k + \lambda(y^k - x^k).$$

We define $\bar{\mu}_f = \min(\mu_f, 1)$, $u_y = By^k$ and

$$u_x = \frac{1}{\gamma}(w^k - x^k) = Bx^k, \qquad v_y = \frac{1}{\gamma}(2x^k - w^k - y^k) \in Ay^k. \tag{10}$$

Suppose that $S = 1 + \gamma(1 + \bar{\mu}_f)$ and

$$\alpha_1 = \frac{2}{\bar{\mu}_f^2}\left(\gamma\left(1 - \bar{\mu}_f^2\right) + \frac{1}{\gamma} + 2\right), \qquad \alpha_2 = 2\gamma S, \qquad \alpha_3 = \frac{2(\gamma + 1)S}{\bar{\mu}_f}.$$

Due to β-cocoercivity, we have

$$\langle u_x - u_y, x^k - y^k \rangle - \|u_x - u_y\|^2 \geq 0, \quad \langle u_y - Bx^\star, y^k - x^\star \rangle - \|u_y - Bx^\star\|^2 \geq 0.$$

By restricted strong monotonicity, we get

$$\langle u_y + v_y, y^k - x^\star \rangle - \bar{\mu}_f \|y^k - x^\star\|^2 \geq 0.$$

Upon multiplying these inequalities by the specified multipliers and invoking (10), we derive

$$\alpha_3 \left(\langle u_y + v_y, y^k - x^\star \rangle - \mu_f \|y^k - x^\star\|^2 \right) + \alpha_1 \left(\langle u_x - u_y, x^k - y^k \rangle - \|u_x - u_y\|^2 \right)$$

$$+ \alpha_2 \left(\langle u_y - Bx^\star, y^k - x^\star \rangle - \|u_y - Bx^\star\|^2 \right) = \mu^2 \|Tw^k - w^k\|^2 - \|w^k - w^\star\|^2$$

$$- D \left\| w - w^\star - \frac{4\gamma C}{D} u_x - \frac{\gamma(\gamma + \bar{\mu}_f \gamma + 1) + \frac{C}{\bar{\mu}_f}}{D} u_y - \frac{2(\bar{\mu}_f \gamma + \frac{1}{2})C}{\bar{\mu}_f D} v_y \right\|^2$$

$$- E \left\| u_x - \frac{P_1}{D'} u_y + \frac{2\bar{\mu}_f \gamma^3 (2\gamma + 3)}{D'} v_y \right\|^2 - \frac{\gamma(\bar{\mu}_f - 1)^2 P_2}{\bar{\mu}_f^2 D'} \left\| u_y - \frac{1}{\bar{\mu}_f - 1} v_y \right\|^2,$$

where

$$C = (\gamma + 1)\big((\bar{\mu}_f + 1)\gamma + 1\big), \qquad D = 2C - 1,$$

$$E = \frac{(\gamma + \bar{\mu}_f \gamma + 1)^2}{\bar{\mu}_f^2} + 8\gamma^2 C - \frac{\left(-2\gamma \bar{\mu}_f^2 + \frac{2\gamma^2 + 4\gamma + 2}{\gamma} \right)(\gamma - 1)}{\bar{\mu}_f^2} - \frac{16\gamma^2 C^2}{D},$$

$$D' = 2 - 2\bar{\mu}_f^2 \gamma^4 - 6\bar{\mu}_f^2 \gamma^3 - 4\bar{\mu}_f^2 \gamma^2 + 4\bar{\mu}_f \gamma^4 + 8\bar{\mu}_f \gamma^3 + \bar{\mu}_f \gamma^2 + 2\bar{\mu}_f \gamma - 2\gamma^4 - 2\gamma^3 + 7\gamma^2 + 9\gamma,$$

$$P_1 = 2\bar{\mu}_f^2 \gamma^4 - 4\bar{\mu}_f^2 \gamma^2 + 2\bar{\mu}_f \gamma^3 + \bar{\mu}_f \gamma^2 + 2\bar{\mu}_f \gamma - 2\gamma^4 - 2\gamma^3 + 7\gamma^2 + 9\gamma + 2,$$

$$P_2 = -3\bar{\mu}_f^3 \gamma^5 - 10\bar{\mu}_f^3 \gamma^4 - 8\bar{\mu}_f^3 \gamma^3 - 3\bar{\mu}_f^2 \gamma^5 - 7\bar{\mu}_f^2 \gamma^4 - 8\bar{\mu}_f^2 \gamma^3 - 4\bar{\mu}_f^2 \gamma^2 - \bar{\mu}_f \gamma^5 + 4\bar{\mu}_f \gamma^4$$

$$+ 19\bar{\mu}_f \gamma^3 + 22\bar{\mu}_f \gamma^2 + 8\bar{\mu}_f \gamma - \gamma^5 - 3\gamma^4 + \gamma^3 + 11\gamma^2 + 12\gamma + 4.$$

By doing some algebra, one can show that $D, E, P_2 \geq 0$. Therefore,

$$0 \leq \mu^2 \|Tw^k - w^k\|^2 - \|w^k - w^\star\|^2,$$

where completes the proof. □

5. Convex composite optimization problem

In this section, we study the DRS algorithm for the following convex composite optimization problem,

$$\min \ f(x) + g(x), \tag{11}$$

where $f, g : \mathbb{R}^n \to \mathbb{R}$ are closed proper convex functions. This problem may be regarded as a special case of problem (1), since it is equivalent to the monotone inclusion problem

$$\operatorname*{find}_{x \in \mathbb{R}^n} \quad 0 \in \partial f(x) + \partial g(x).$$

Therefore, all the results established above remain valid for this setting. However, because problem (11) is itself a particular instance of a monotone inclusion problem, one can derive sharper results in this setting.

The DRS algorithm for the convex composite optimization problem is given in Algorithm 2.

Algorithm 2 The DRS algorithm for the composite optimization problem

---

    Parameters: number of iterations N, positive stepsize $\gamma > 0$ and relaxation factor $\lambda \in (0, 2)$.

    Inputs: closed proper convex functions f and g, initial iterate $w^1 \in \mathbb{R}^n$.

    For $k = 1, 2, \ldots, N$ perform the following steps:

      i) $x^k = \operatorname{prox}_{\gamma f}(w^k)$.

      ii) $y^k = \operatorname{prox}_{\gamma g}(2x^k - w^k)$.

      iii) $w^{k+1} = w^k + \lambda(y^k - x^k)$.

Throughout the section, we use the following assumptions.

ASSUMPTION 2. *We consider the following assumptions.*

  i) f and g are closed proper convex functions.

  ii) The DR operator T admits a fixed point $w^\star$.

 iii) f is L-smooth.

Note that Assumption 2 implies that $x^\star = \operatorname{prox}_{\gamma f}(w^\star)$ is a solution to problem (11). In the following conjecture, we give the exact convergence rate of the DRS algorithm. The rate is based on numerical experiments conducted within the performance estimation framework. [1]

CONJECTURE 2. *Let Assumption 2 hold. If the sequence $\{(x^k, y^k, w^k)\}$ is generated by the Algorithm 2 with stepsize $\gamma \in (0, \frac{2\sqrt{2}-1}{L})$, relaxation parameter $\lambda \in (0, \frac{1+\sqrt{5}}{2})$ and the initial point w^1, then*

$$f(y^N) + g(y^N) - f(x^\star) - g(x^\star) \leq \frac{1}{4\gamma((N-1)\lambda + 1)} \|w^1 - w^*\|^2.$$

It is worth noting that the convergence rate given in this conjecture for $\lambda = 1$ is similar to that established in [29] for the ADMM.

---

[1] We hope that the conjectures presented in this paper will contribute to advancing research on the topic. Owing to external constraints, the authors have faced challenges in dedicating sufficient time to further investigation of this topic.

5.1. Silver Stepsize Schedule

In this subsection, we study the convergence rate of the DRS algorithm when using the silver stepsize schedule as a relaxation parameter. Let $\rho = 1 + \sqrt{2}$ denote the silver ratio. The silver stepsize schedule $\pi_k \in \mathbb{R}^{2^k-1}$ for $k \geq 1$ is defined as

$$\pi_{k+1} = [\pi_k, \ 1 + \rho^{k-1}, \ \pi_k],$$

where $\pi_1 = \sqrt{2}$. In recent seminal work, Altschuler and Parrillo [2, 3] established that the convergence rate of the gradient method for smooth convex problems can be improved by employing the silver stepsize schedule. They consider the convex optimization problem

$$\min_{x \in \mathbb{R}^n} F(x), \tag{12}$$

where $F : \mathbb{R}^n \to \mathbb{R}$ is L-smooth. They demonstrate that, by using stepsizes $h = \frac{1}{L}\pi_k$, the gradient descent

$$x^{i+1} = x^i - h_i \nabla F(x^i) \tag{GD}$$

satisfies the bound

$$F(x^N) - F(x^\star) \leq \frac{L\|x^0 - x^\star\|^2}{\sqrt{1 + 4\rho^{2k} - 3}} \approx \frac{L\|x^0 - x^\star\|^2}{2N^{\log_2(\rho)}}, \tag{13}$$

where $N = 2^k - 1$, $x^\star \in \arg\min F(x)$ and x^0 is the initial point; see [2, Theorem 1]. In what follows, we give a tighter rate for this stepszies[2]. Before we get to the proof we need to present a lemma. Moreover, we assume that F is 1-smooth without loss of generality and $F^\star = F(x^\star)$. In the lemma, we employ the following interpolation inequality [20, Theorem 2.1.5],

$$F(y) \geq F(x) + \langle \nabla F(x), y - x \rangle + \tfrac{1}{2}\|\nabla F(y) - \nabla F(x)\|^2.$$

LEMMA 1. *Let F be 1-smooth and $N = 2^k - 1$. Assume that $x^1, \ldots, x^N$ is generated by the GD with the silver stepsize schedule, $h = \pi_k$, and the initial point x^0. Then*

$$\left(2\rho^k - 1\right)\left(F^\star - F(x^N)\right) + \tfrac{1}{2}\left\|x^0 - x^\star\right\|^2 - \sum_{i=0}^{N-1} h_i\left(F^\star - F(x^i) - \langle \nabla F(x^i), x^\star - x^i \rangle - \tfrac{1}{2}\left\|\nabla F(x^i)\right\|^2\right)$$

$$- \rho^k\left(F^\star - F(x^N) - \langle \nabla F(x^N), x^\star - x^N \rangle - \tfrac{1}{2}\left\|\nabla F(x^N)\right\|^2\right) - \tfrac{1}{2}\left\|x^N - \rho^k \nabla F(x^N) - x^\star\right\|^2 \geq 0. \tag{14}$$

---

[2] The proof was prepared in the fall of 2023, when the second author was a researcher at UCLouvain. The argument was developed jointly by the second author and François Glineur.

Proof. We prove by induction on k. Let $k = 1$. By doing some algebra, one can show that

$$\rho \left(F(x^0) - F(x^1) - \langle \nabla F(x^1), x^0 - x^1 \rangle - \tfrac{1}{2} \left\| \nabla F(x^1) - \nabla F(x^0) \right\|^2 \right) +$$

$$\left(F(x^1) - F(x^0) - \langle \nabla F(x^0), x^1 - x^0 \rangle - \tfrac{1}{2} \left\| \nabla F(x^1) - \nabla F(x^0) \right\|^2 \right) =$$

$$(2\rho - 1) \left(F^\star - F(x^1) \right) + \tfrac{1}{2} \left\| x^0 - x^\star \right\| - h_1 \left(F^\star - F(x^0) - \tfrac{1}{h_1} \langle x^0 - x^1, x^\star - x^0 \rangle - \tfrac{1}{2h_0^2} \left\| x^0 - x^1 \right\|^2 \right)$$

$$- \rho \left(F^\star - F(x^1) - \langle \nabla F(x^1), x^\star - x^1 \rangle - \tfrac{1}{2} \left\| \nabla F(x^1) \right\|^2 \right) - \tfrac{1}{2} \left\| x^1 - \rho \nabla F(x^1) - x^\star \right\|^2 \geq 0,$$

where the equality follows from $\nabla F(x^0) = \frac{1}{h_0}(x^0 - x^1)$ and the inequality follows from the interpolation inequality. Assume that (14) holds for k. As $h_i = h_{N+1+i}$ for $i \in \{0, \ldots, N-1\}$, by setting x^{N+1} as the initial point in the GD, we get

$$\left(2\rho^k - 1 \right) \left(F^\star - F(x^{2N+1}) \right) + \tfrac{1}{2} \left\| x^{N+1} - x^\star \right\|^2 - \sum_{i=N}^{2N} h_i \left(F^\star - F(x^i) - \langle \nabla F(x^i), x^\star - x^i \rangle - \tfrac{1}{2} \left\| \nabla F(x^i) \right\|^2 \right)$$

$$- \rho^k \left(F^\star - F(x^{2N+1}) - \langle \nabla F(x^{2N+1}), x^\star - x^{2N+1} \rangle - \tfrac{1}{2} \left\| \nabla F(x^{2N+1}) \right\|^2 \right) - \tfrac{1}{2} \left\| x^{2N+1} - \rho^k \nabla F(x^{2N+1}) - x^\star \right\|^2$$

$$+ h_N \left(F^\star - F(x^N) - \langle \nabla F(x^N), x^\star - x^N \rangle - \tfrac{1}{2} \left\| \nabla F(x^N) \right\|^2 \right) \geq 0. \tag{15}$$

In inequality (15), we add and subtract the interpolation inequality between x^N and $x^\star$ to facilitate subsequent computations. By the interpolation inequalities, we have

$$\rho \sum_{i=N}^{2N} h_i \left(F(x^N) - F(x^i) - \langle \nabla F(x^i), x^N - x^i \rangle - \tfrac{1}{2} \left\| \nabla F(x^i) - \nabla F(x^N) \right\|^2 \right)$$

$$+ \rho \sum_{i=N}^{2N} h_i \left(F(x^{2N+1}) - F(x^i) - \langle \nabla F(x^i), x^{2N+1} - x^i \rangle - \tfrac{1}{2} \left\| \nabla F(x^i) - \nabla F(x^{2N+1}) \right\|^2 \right) \tag{16}$$

$$+ \rho \left(F(x^N) - F(x^{2N+1}) - \langle \nabla F(x^{2N+1}), x^N - x^{2N+1} \rangle - \tfrac{1}{2} \left\| \nabla F(x^{2N+1}) - \nabla F(x^N) \right\|^2 \right)$$

$$- \rho \left(F(x^{2N+1}) - F(x^N) - \langle \nabla F(x^N), x^{2N+1} - x^N \rangle - \tfrac{1}{2} \left\| \nabla F(x^N) - \nabla F(x^{2N+1}) \right\|^2 \right)$$

$$= \rho \sum_{i=N}^{2N} h_i \left(F(x^N) - F(x^i) - \langle \nabla F(x^i), x^N - x^i \rangle - \tfrac{1}{2} \left\| \nabla F(x^i) - \nabla F(x^N) \right\|^2 \right)$$

$$+ \rho \sum_{i=N+1}^{2N} h_i \left(F(x^{2N+1}) - F(x^i) - \langle \nabla F(x^i), x^{2N+1} - x^i \rangle - \tfrac{1}{2} \left\| \nabla F(x^i) - \nabla F(x^{2N+1}) \right\|^2 \right)$$

$$+ \rho \left(F(x^N) - F(x^{2N+1}) - \langle \nabla F(x^{2N+1}), x^N - x^{2N+1} \rangle - \tfrac{1}{2} \left\| \nabla F(x^{2N+1}) - \nabla F(x^N) \right\|^2 \right)$$

$$+ \rho^k \left(F(x^{2N+1}) - F(x^N) - \langle \nabla F(x^N), x^{2N+1} - x^N \rangle - \tfrac{1}{2} \left\| \nabla F(x^N) - \nabla F(x^{2N+1}) \right\|^2 \right) \geq 0,$$

Note that we use $h_N = \rho^{k-1} + 1$. By multiplying (15) by $2\rho + 1$ and then summing it with inequalities (14) and (16), we obtain

$$(2\rho + 1)\left(2\rho^k - 1\right)\left(F^\star - F(x^{2N+1})\right) + \tfrac{1}{2}\left\|x^0 - x^\star\right\|^2 - \sum_{i=0}^{2N} h_i\left(F^\star - F(x^i) - \langle\nabla F(x^i), x^\star - x^i\rangle - \tfrac{1}{2}\left\|\nabla F(x^i)\right\|^2\right)$$

$$-\rho^{k+1}\left(F^\star - F(x^{2N+1}) - \langle\nabla F(x^{2N+1}), x^\star - x^{2N+1}\rangle - \tfrac{1}{2}\left\|\nabla F(x^{2N+1})\right\|^2\right) - \tfrac{1}{2}\left\|x^N - \rho^k\nabla F(x^N) - x^\star\right\|^2$$

$$+\left(2\rho^k - 1\right)\left(F^\star - F(x^N)\right) + (h_{N+1}(2\rho + 1) - \rho^k)\left(F^\star - F(x^N) - \langle\nabla F(x^N), x^\star - x^N\rangle - \tfrac{1}{2}\left\|\nabla F(x^N)\right\|^2\right)$$

$$+(\rho + \tfrac{1}{2})\left\|x^N - h_N\nabla f(x^N) - x^\star\right\|^2 - (\rho + \tfrac{1}{2})\left\|x^{2N+1} - \rho^k\nabla F(x^{2N+1}) - x^\star\right\|^2$$

$$-(\rho + 1)\rho^k\left(F^\star - F(x^{2N+1}) - \langle\nabla F(x^{2N+1}), x^\star - x^{2N+1}\rangle - \tfrac{1}{2}\left\|\nabla F(x^{2N+1})\right\|^2\right)$$

$$+\rho\sum_{i=N}^{2N} h_i\left(F(x^N) - F^\star - \langle\nabla F(x^i), x^N - \nabla F(x^N) - x^\star\rangle - \tfrac{1}{2}\left\|\nabla F(x^N)\right\|^2\right)$$

$$+\rho\sum_{i=N}^{2N} h_i\left(F(x^{2N+1}) - F^\star - \langle\nabla F(x^i), x^{2N+1} - \nabla F(x^{2N+1}) - x^\star\rangle - \tfrac{1}{2}\left\|\nabla F(x^{2N+1})\right\|^2\right)$$

$$+\rho\left(F(x^N) - F(x^{2N+1}) - \langle\nabla F(x^{2N+1}), x^N - x^{2N+1}\rangle - \tfrac{1}{2}\left\|\nabla F(x^{2N+1}) - \nabla F(x^N)\right\|^2\right)$$

$$-\rho\left(F(x^{2N+1}) - F(x^N) - \langle\nabla F(x^N), x^{2N+1} - x^N\rangle - \tfrac{1}{2}\left\|\nabla F(x^N) - \nabla F(x^{2N+1})\right\|^2\right) \geq 0.$$

By using the following identities,

$$\sum_{i=N}^{2N} h_i = \rho^k + \rho^{k-1}, \qquad \sum_{i=N}^{2N} h_i\nabla F(x^i) = x^N - x^{2N+1}, \qquad \rho^2 - 2\rho - 1 = 0,$$

we get

$$\left(2\rho^{k+1} - 1\right)\left(F^\star - F(x^{2N+1})\right) - \sum_{i=0}^{2N} h_i\left(F^\star - F(x^i) - \langle\nabla F(x^i), x^\star - x^i\rangle - \tfrac{1}{2}\left\|\nabla F(x^i)\right\|^2\right)$$

$$-\rho^{k+1}\left(F^\star - F(x^{2N+1}) - \langle\nabla F(x^{2N+1}), x^\star - x^{2N+1}\rangle - \tfrac{1}{2}\left\|\nabla F(x^{2N+1})\right\|^2\right) + \tfrac{1}{2}\left\|x^0 - x^\star\right\|^2 -$$

$$\tfrac{1}{2}\left\|x^{2N+1} - \rho^{k+1}\nabla F(x^{2N+1}) - x^\star\right\|^2 \geq 0,$$

which completes the proof.   □

THEOREM 4. *Let F be 1-smooth and $N = 2^k - 1$. If $x^1, \ldots, x^N$ is generated by the GD with the silver stepsize schedule and the initial point x^0, then*

$$F(x^N) - F^\star \leq \tfrac{1}{4\rho^k - 2}\left\|x^0 - x^\star\right\|^2. \tag{17}$$

Proof.   The theorem is derived from Lemma 1.   □

One can show the rate given in the theorem is tight by selecting a proper Huber function. After the publication of [3], several interesting developments have appeared. Grimmer et al.[15] improved the convergence rate by slightly modifying the silver stepsize schedule and also investigated the schedule for other performance measures. Bok and Altschuler [6] studied the schedule in the context of the proximal gradient method. Wang et al.[27] analyzed it for the proximal point algorithm. We refer the interested reader to [6] for a review of related work. The following conjecture concerns the convergence rate of the DRS algorithm when the silver stepsize schedule is employed to tune the relaxation parameter.

CONJECTURE 3. *Let Assumption 2 hold and let* $\lambda = (\pi_k, 1)$. *If the sequence* $\{(x^k, y^k, w^k)\}$ *is generated by Algorithm 2 with stepsize* $\gamma \in (0, \frac{2\sqrt{2}-1}{L})$, *relaxation parameter* λ_i *at iterate i and the initial point* w^1, *then*

$$f(y^N) + g(y^N) - f(x^\star) - g(x^\star) \leq \frac{1}{4\gamma\rho^k} \|w^1 - w^*\|^2,$$

where $N = 2^k$.

Note that since λ_N does not influence y^N in the DRS algorithm, it may be assigned arbitrarily as any positive scalar.

5.2. Accelerated DRS algorithm

This subsection is devoted to an accelerated DRS algorithm. Several accelerated methods for problem (11) have been proposed in the literature; see, for example, [14, 21]. However, their analyses establish an accelerated convergence rate of $O(\frac{1}{N^2})$ only in the quadratic setting. To address problem (11) in greater generality, we introduce Algorithm 3.

Algorithm 3 Accelerated DRS algorithm

Parameters: number of iterations N, positive stepsize $\gamma > 0$ and relaxation factor $\lambda \in (0, 2)$.

Inputs: closed proper convex functions f and g, initial iterate $w^1 \in \mathbb{R}^n$.

Set $u^1 = w^1$ and for $k = 1, 2, \ldots, N$ perform the following steps:

  i) $x^k = \text{prox}_{\gamma f}(w^k)$.

  ii) $y^k = \text{prox}_{\gamma g}(2x^k - w^k)$.

  iii) $u^{k+1} = w^k + \lambda(y^k - x^k)$.

  iv) $w^{k+1} = u^{k+1} + \frac{k}{k+3}(u^{k+1} - u^k)$.

15

Algorithm 3 closely resembles the method proposed in [21]. The key distinction lies in the choice of the momentum coefficient: their algorithm, at iterate $k \geq 3$, employs $\frac{k-2}{k+1}$ while Algorithm 3 uses $\frac{k}{k+3}$. It is also worth noting that [21] sets the coefficient to zero for the first two iterates. Based on our numerical experiments, we propose the following conjecture regarding the convergence rate of Algorithm 3.

CONJECTURE 4. *Let Assumption 2 hold. If the sequence $\{(x^k, y^k, w^k, u^k)\}$ is generated by Algorithm 3 with stepsize $\gamma \in (0, \frac{1}{L}]$, relaxation parameter $\lambda \in (0, 1]$ and the initial point w^1, then*

$$f(y^N) + g(y^N) - f(x^\star) - g(x^\star) \leq \frac{2}{\gamma((N^2+7N-8)\lambda+8)} \|w^1 - w^*\|^2.$$

References

[1] Abbaszadehpeivasti H (2024) *Performance analysis of optimization methods for machine learning.* Ph.D. thesis, Tilburg University.

[2] Altschuler JM, Parrilo PA (2024) Acceleration by stepsize hedging: Silver stepsize schedule for smooth convex optimization. *Mathematical Programming* 1–14.

[3] Altschuler JM, Parrilo PA (2025) Acceleration by stepsize hedging: Multi-step descent and the silver stepsize schedule. *Journal of the ACM* 72(2):1–38.

[4] Aspelmeier T, Charitha C, Luke DR (2016) Local linear convergence of the ADMM/Douglas–Rachford algorithms without strong convexity and application to statistical imaging. *SIAM Journal on Imaging Sciences* 9(2):842–868.

[5] Bauschke HH, Combettes PL (2017) *Convex Analysis and Monotone Operator Theory in Hilbert Spaces* (Springer), 2nd edition.

[6] Bok J, Altschuler JM (2025) Optimized methods for composite optimization: a reduction perspective. *arXiv preprint arXiv:2506.23756* .

[7] Combettes PL (2018) Monotone operator theory in convex optimization. *Mathematical Programming* 170(1):177–206.

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

[9] Drori Y, Teboulle M (2014) Performance of first-order methods for smooth convex minimization: a novel approach. *Mathematical Programming* 145(1):451–482.

[10] Eckstein J, Bertsekas DP (1992) On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming* 55(1):293–318.

[11] Giselsson P (2017) Tight global linear convergence rate bounds for Douglas–Rachford splitting. *Journal of Fixed Point Theory and Applications* 19(4):2241–2270.

[12] Giselsson P, Boyd S (2016) Linear convergence and metric selection for Douglas-Rachford splitting and ADMM. *IEEE Transactions on Automatic Control* 62(2):532–544.

[13] Glowinski R, Osher SJ, Yin W (2017) *Splitting methods in communication, imaging, science, and engineering* (Springer).

[14] Goldstein T, O'Donoghue B, Setzer S, Baraniuk R (2014) Fast alternating direction optimization methods. *SIAM Journal on Imaging Sciences* 7(3):1588–1623.

[15] Grimmer B, Shu K, Wang AL (2025) Accelerated objective gap and gradient norm convergence for gradient descent via long steps. *INFORMS Journal on Optimization* 7(2):156–169.

[16] Karimi H, Nutini J, Schmidt M (2016) Linear convergence of gradient and proximal-gradient methods under the Polyak-Lojasiewicz condition. *Joint European conference on machine learning and knowledge discovery in databases*, 795–811 (Springer).

[17] Lieder F (2018) *Projection based methods for conic linear programming—optimal first order complexities and norm constrained quasi newton methods*. Ph.D. thesis, Dissertation, Düsseldorf, Heinrich-Heine-Universität, 2018.

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

[19] Luo ZQ, Tseng P (1992) On the linear convergence of descent methods for convex essentially smooth minimization. *SIAM Journal on Control and Optimization* 30(2):408–425.

[20] Nesterov Y, et al. (2018) *Lectures on convex optimization*, volume 137 (Springer).

[21] Patrinos P, Stella L, Bemporad A (2014) Douglas-Rachford splitting: Complexity estimates and accelerated variants. *53rd IEEE Conference on Decision and Control*, 4234–4239 (IEEE).

[22] Peña J, Vera JC, Zuluaga LF (2021) Linear convergence of the Douglas-Rachford algorithm via a generic error bound condition. *arXiv preprint arXiv:2111.06071* .

[23] Ryu EK, Taylor AB, Bergeling C, Giselsson P (2020) Operator splitting performance estimation: Tight contraction factors and optimal parameter selection. *SIAM Journal on Optimization* 30(3):2251–2271.

[24] Ryu EK, Yin W (2022) *Large-scale convex optimization: algorithms & analyses via monotone operators* (Cambridge University Press).

[25] Taylor A (2024) *Towards principled and systematic approaches to the analysis and design of optimization algorithms*. Ph.D. thesis, PSL Research University.

[26] Taylor AB, Hendrickx JM, Glineur F (2017) Smooth strongly convex interpolation and exact worst-case performance of first-order methods. *Mathematical Programming* 161(1):307–345.

[27] Wang B, Ma S, Yang J, Zhou D (2024) Relaxed proximal point algorithm: Tight complexity bounds and acceleration without momentum. *arXiv preprint arXiv:2410.08890* .

[28] Zamani M, Abbaszadehpeivasti H, de Klerk E (2024) Convergence rate analysis of the gradient descent–ascent method for convex–concave saddle-point problems. *Optimization Methods and Software* 39(5):967–989.

[29] Zamani M, Abbaszadehpeivasti H, de Klerk E (2024) The exact worst-case convergence rate of the alternating direction method of multipliers. *Mathematical Programming* 208(1):243–276.

[30] Zamani M, Glineur F (2024) Exact convergence rate of the subgradient method by using Polyak step size. *arXiv preprint arXiv:2407.15195* .