A Control-Theoretic Approach to Analysis and Parameter Selection of Douglas-Rachford Splitting

Jacob H. Seidman¹, Mahyar Fazlyab², Victor M. Preciado², George J. Pappas²

Abstract—Douglas-Rachford splitting, and its equivalent dual formulation ADMM, are widely used iterative methods in composite optimization problems arising in control and machine learning applications. The performance of these algorithms depends on the choice of step size parameters, for which the optimal values are known in some specific cases, and otherwise are set heuristically. We provide a new unified method of convergence analysis and parameter selection by interpreting the algorithm as a linear dynamical system with nonlinear feedback. This approach allows us to derive a dimensionally independent matrix inequality whose feasibility is sufficient for the algorithm to converge at a specified rate. By analyzing this inequality, we are able to give performance guarantees and parameter settings of the algorithm under a variety of assumptions regarding the convexity and smoothness of the objective function. In particular, our framework enables us to obtain a new and simple proof of the $O(1/k)$ convergence rate of the algorithm when the objective function is not strongly convex.

I. INTRODUCTION

In this paper, we consider problems of the form

$$\min_{x \in \mathbb{R}^d} \{F(x) = f(x) + g(x)\},$$

where $f, g : \mathbb{R}^d \to \mathbb{R}$ are convex and $g$ is non-differentiable. Douglas-Rachford splitting solves problem (1) with the following iterations:

$$y_k = \text{prox}_{\alpha f}(x_k),$$

$$z_k = \text{prox}_{\alpha g}(2y_k - x_k),$$

$$x_{k+1} = x_k + \lambda(z_k - y_k),$$

where $\text{prox}$ is the proximal operator (see Definition 3) and $\alpha$ and $\lambda$ are user specified parameters known as the proximal step size and relaxation parameter, respectively. The goal of this work is to provide convergence rates for Douglas-Rachford splitting over various assumptions on the function $f$, and optimize these rates with respect to the algorithm parameters $\alpha$ and $\lambda$ using semidefinite programming.

The algorithm was first proposed in [1], and since then has found application for general separable optimization problems [2]. Its dual formulation, ADMM, has been particularly useful for its decentralized implementation in distributed optimization problems [3]. Since the iterates of ADMM can be written as applying Douglas-Rachford splitting to the dual problem [4], convergence results for one algorithm are valid for the other as well.

The convergence of Douglas-Rachford splitting has previously been analyzed using monotone operator theory and variational inequalities, see [5], [6]. These techniques have led to proofs of a $O(1/k)$ convergence rate for the non-strongly convex case [7], and linear convergence when $f$ is smooth and strongly convex [8], [9].

Recently, there has been a growing interest in automating the analysis and design of optimization algorithms via semidefinite programming (SDP). [10]–[16]. In particular, through the method of integral quadratic constraints proposed in [11], the authors of [17] derive a SDP for choosing the parameters of ADMM in the case of smooth and strongly convex $f$. Using a similar framework, the authors of [18] provide evidence that as $\lambda$ approaches 2 from below, the linear convergence rate is close to being optimal. The work in [19] gives an optimal choice for the relaxation parameter in the case of quadratic objective functions. Furthermore, [20] gives a set of assumptions in which a bound on the linear convergence rate is minimized by setting $\lambda = 2$.

Our Contribution: By viewing Douglas-Rachford splitting as a linear system with non-linear feedback, we derive a dimensionally independent matrix inequality which gives convergence guarantees via Lyapunov functions. Whereas such an approach was previously applied in [17] to the case of smooth and strongly convex $f$, our framework is novel in that it encompasses varying assumptions on the smoothness and convexity of $f$. By changing a single term in the Lyapunov function for each scenario, we are able to relate the satisfaction of a matrix inequality to the convergence of the algorithm. In particular, we give a new and simple proof of $O(1/k)$ convergence in the non-strongly convex case in terms of both the distance to the fixed point condition and the suboptimality of the objective function. These symbolic results can then be used to select step sizes that optimize the derived rates.

In the strongly convex case, the corresponding matrix inequality has a nonlinear dependence on the relaxation parameter $\lambda$. We are able to modify the matrix inequality to linearize the dependence on $\lambda$, allowing us to numerically optimize its value for the convergence rate directly. While previous work derived SDP’s which can verify the performance of the algorithm for a given parameter setting, to the best of our knowledge this is the first time such a method immediately gives an optimal relaxation parameter when solved numerically, as opposed to having to search over a range of values for $\lambda$. 

1: Department of Applied Mathematics and Computational Science, University of Pennsylvania. Email: seidj@sas.upenn.edu. 2: Department of Electrical and Systems Engineering, University of Pennsylvania. Email: {mahyarfa, preciado, pappasg}@seas.upenn.edu. Work was supported by ARL CRA DCIST W911NF-17-2-0181 and the NSF under grants CAREER-ECCS-1651433.
II. Preliminaries

We denote the set of real numbers by \( \mathbb{R} \), the set of real \( n \)-dimensional vectors by \( \mathbb{R}^n \), the set of \( m \times n \)-dimensional matrices by \( \mathbb{R}^{m \times n} \), and the \( n \)-dimensional identity matrix by \( I_n \). For a function \( f : \mathbb{R}^d \to \mathbb{R} \), we denote by \( \text{dom} f = \{ x \in \mathbb{R}^n : f(x) < \infty \} \) the effective domain of \( f \). The subdifferential of a convex function \( f \) at a point \( x \) is \( \partial f(x) := \{ g \mid f(y) - f(x) \geq g^T (y - x), \forall y \in \text{dom} f \} \). By abuse of notation we will also refer to a subgradient, that is an element of the set \( \{ g \mid f(y) - f(x) \geq g^T (y - x), \forall y \in \text{dom} f \} \) by \( \partial f(x) \) as well. The indicator function of a set \( C \) is given by \( 1_C(x) = 0 \) if \( x \in C \) and \( 1_C(x) = \infty \) if \( x \notin C \).

For two matrices \( A \in \mathbb{R}^{m \times n} \) and \( B \in \mathbb{R}^{p \times q} \) of arbitrary dimensions, we denote their Kronecker product by \( A \otimes B \).

**Definition 1 (Smoothness)** A differentiable function \( f : \mathbb{R}^d \to \mathbb{R} \) is \( L_f \)-smooth on \( \mathcal{S} \subseteq \text{dom} f \) if
\[
\| \nabla f(x) - \nabla f(y) \|_2 \leq L_f \| x - y \|_2,
\](3a)
for some \( L_f > 0 \) and all \( x, y \in \mathcal{S} \). (3a) implies
\[
f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{L_f}{2} \| y - x \|_2^2,
\](3b)
for all \( x, y \in \mathcal{S} \).

**Definition 2 (Strong convexity)** A differentiable function \( f : \mathbb{R}^d \to \mathbb{R} \) is \( m_f \)-strongly convex on \( \mathcal{S} \subseteq \text{dom} f \)
\[
m_f \| x - y \|_2^2 \leq (x - y)^T (\nabla f(x) - \nabla f(y)),
\](4a)
for some \( m_f > 0 \) and all \( x, y \in \mathcal{S} \). for all \( x, y \in \mathcal{S} \).

We denote the class of functions satisfying (3a) and (4a) by \( \mathcal{F}(m_f, L_f) \).

**Definition 3 (Proximal Operator)** Given a convex function \( f : \mathbb{R}^n \to \mathbb{R} \), the proximal operator \( \text{prox}_{\alpha f} : \mathbb{R}^n \to \mathbb{R}^n \) is defined as
\[
\text{prox}_{\alpha f}(x) = \arg \min_y \left\{ f(y) + \frac{1}{2\alpha} \| x - y \|_2^2 \right\}.
\](5)
The point \( y = \text{prox}_{\alpha f}(x) \) also is given by the implicit solution to the sub-gradient equation
\[
y = x - \alpha \partial f(y).
\](6)

A. Quadratic Constraints

We say that a nonlinear function \( \phi : \mathbb{R}^n \to \mathbb{R}^n \) satisfies a quadratic constraint defined by \( Q \) if for all \( x, y \),
\[
\begin{bmatrix} x - y \\ \phi(x) - \phi(y) \end{bmatrix}^T Q \begin{bmatrix} x - y \\ \phi(x) - \phi(y) \end{bmatrix} \geq 0,
\](7)
This type of constraint is also referred to as an incremental quadratic constraint [21], or point-wise integral quadratic constraint [11] in the literature.

It was noted in [11], [22] that a differentiable function \( f \) belongs to the class \( \mathcal{F}(m_f, L_f) \) on \( \mathcal{S} \) if and only if the (sub)gradient \( \nabla f \) satisfies the quadratic constraint defined by \( Q_f \), where
\[
Q_f = \begin{bmatrix} -\frac{m_f L_f}{m_f + L_f} & 1/2 \\ 1/2 & -\frac{1}{m_f + L_f} \end{bmatrix} \otimes I_d.
\](8)
The corresponding quadratic constraints for functions which are not strongly convex \( (m_f = 0) \) or non-smooth \( (L_f = \infty) \) are found by taking the appropriate limit above and replacing the gradients with sub-gradients. We define \( Q_g \) for \( g \) analogously.

If we define
\[
Q_p = \begin{bmatrix} 0 & I_d \\ \alpha I_d & -I_d \end{bmatrix} Q_f \begin{bmatrix} 0 & \alpha I_d \\ I_d & -I_d \end{bmatrix},
\](9)
then the proximal operator of a function \( f \in \mathcal{F}(m, L) \), \( \text{prox}_{\alpha f} \), satisfies the quadratic constraint defined by \( Q_p \) [12].

III. Analysis of Douglas-Rachford Splitting via Matrix Inequalities

A. Douglas-Rachford Splitting as a Dynamical System

If we define the variable \( u_k = z_k - y_k \), we can view the updates in (2) as a linear system with state \( x_k \), input \( u_k \) and feedback nonlinearity \( \phi(x_k) \), where
\[
\phi(x_k) := \text{prox}_{\alpha g}(2\text{prox}_{\alpha f}(x_k) - x_k) - \text{prox}_{\alpha g}(x_k),
\]
as shown in Figure 1.

![Fig. 1. Block diagram of Douglas-Rachford splitting](image)

Our main technique is to describe the nonlinearity \( \phi \) with quadratic constraints representing the prox operator. This allows us to derive a matrix inequality as a sufficient condition for closed loop stability of the system via a Lyapunov function argument. We perform this derivation in the following three cases:

- **Case 1**: \( f \in \mathcal{F}(0, \infty) \) and \( g \in \mathcal{F}(0, \infty) \),

- **Case 2**: \( f \in \mathcal{F}(0, L_f) \) and \( g \in \mathcal{F}(0, \infty) \), with \( 0 < L_f < \infty \),

- **Case 3**: \( f \in \mathcal{F}(m_f, L_f) \) and \( g \in \mathcal{F}(0, \infty) \), with \( 0 < m_f \leq L_f < \infty \).

We will see that for each case only one term in the Lyapunov function needs to be modified to obtain the convergence result. We then use the matrix inequality condition for each case to obtain information about optimal choices of the algorithm parameters both symbolically and numerically.
B. Characterization of Fixed Points

We first note that from relation (6), we may rewrite the first two iterates of (2) in terms of subgradients of \( f \) and \( g \),

\[
y_k = x_k - \alpha \partial f(y_k), \quad z_k = 2y_k - x_k - \alpha \partial g(z_k).
\]

After simplifications, we see that the fixed points of the algorithm satisfy

\[
x_* = y_* + \alpha \partial f(y_*), \quad y_* = z_* + \partial f(y_*) + \partial g(z_*) = 0.
\]

Hence, a fixed point of the algorithm satisfies \( \partial f(y_*) + \partial g(z_*) = 0 \), which is the optimality condition for the problem in (1).

We will also make use of the following relation, obtained from adding the equations in (10) and the definition of \( \phi \),

\[
\phi(x_k) = z_k - y_k = -\alpha(\partial f(y_k) + \partial g(z_k)).
\]

From this, we can interpret the feedback nonlinearity \( \phi \) as the optimality residual of the optimization problem, which is driven to zero by the linear system in the feedback interconnection, as shown in Figure 1.

C. Convergence Certificates via Matrix Inequalities

1) Case 1: Non-strongly convex and non-smooth case: In this section we assume that \( f, g \in \mathcal{F}(0, \infty) \). We propose the following family of Lyapunov functions with \( \theta > 0 \)

\[
V_k = \|x_k - x_*\|^2 + \theta \sum_{i=0}^{k-1} \|\partial f(y_i) + \partial g(z_i)\|^2.
\]

for all \( k \geq 0 \). The presence of the running sum of subgradients is reminiscent of the Popov criterion [23]. It can also be interpreted as the running sum of fixed point residuals (see (11)). The next lemma shows how this Lyapunov function can ensure an \( O(1/k) \) convergence rate in the non-strongly convex case.

**Lemma 1** Consider the algorithm in (2). Suppose there exists a \( \theta > 0 \) such that the Lyapunov function \( V_k \) defined by (13) is non-increasing, i.e., \( V_{k+1} \leq V_k \) for all \( k \). Then

\[
\frac{1}{k} \sum_{i=0}^{k-1} \|\partial f(y_i) + \partial g(z_i)\|^2 \leq \frac{\|x_0 - x_*\|^2}{\theta k}.
\]

**Proof:** Since \( V_{k+1} \leq V_k \) for all \( k \), in particular we have that \( V_k \leq V_0 \), or

\[
\|x_k - x_*\|^2 + \theta \sum_{i=0}^{k-1} \|\partial f(y_i) + \partial g(z_i)\|^2 \leq \|x_0 - x_*\|^2.
\]

We may ignore the first term on the left, and divide through by \( \theta k \) to obtain the desired result.

In the following theorem, we derive a matrix inequality in terms of \( \alpha, \lambda, \) and \( \theta \) as a sufficient condition to guarantee \( V_{k+1} \leq V_k \) and hence, an \( O(1/k) \) convergence rate.

**Theorem 1** Consider the following matrix inequality

\[
W_0 + \sigma_1 Q_1 + \sigma_2 Q_2 \preceq 0,
\]

where

\[
W_0 = \begin{bmatrix} 0 & -\lambda \\ -\lambda & \lambda^2 + \frac{\theta}{\alpha^2} - \frac{(\lambda^2 + \theta)}{\alpha^2} \end{bmatrix} \otimes I_d,
\]

\[
Q_1 = \begin{bmatrix} 0 & \alpha I_d & -I_d \\ \alpha I_d & -I_d & 0 \\ -I_d & 0 & 0 \end{bmatrix},
\]

\[
Q_2 = \begin{bmatrix} 0 & 0 & \alpha I_d \\ 0 & 2I_d & -I_d \\ \alpha I_d & -I_d & -I_d \end{bmatrix},
\]

If \( \sigma_1, \sigma_2, \alpha, \lambda, \theta > 0 \) are chosen so that (16) is satisfied, then for all \( f \in \mathcal{F}(0, \infty) \) and \( g \in \mathcal{F}(0, \infty) \) the iterates in (2) have an optimality residual whose running average converges as

\[
\frac{1}{k} \sum_{i=0}^{k-1} \|\partial f(y_i) + \partial g(z_i)\|^2 \leq \frac{1}{k \theta} \|x_0 - x_*\|^2.
\]

**Proof:** We first see that \( V_{k+1} - V_k \) can be written as a quadratic form. Define the error signal

\[
e_k := \begin{bmatrix} (x_k - x_*)^T \\ (y_k - y_*)^T \\ (z_k - z_*)^T \end{bmatrix}^T.
\]

Using the updates in (10), the fact that \( z_* = y_* \) (see (11)), and the relation (12), it can be verified that

\[
V_{k+1} - V_k = e_k^T W_0 e_k,
\]

where \( W_0 \) is given by (17a). Next, note that

\[
e_k^T Q_1 e_k = \begin{bmatrix} x_k - x_* \\ y_k - y_* \end{bmatrix}^T Q_1 \begin{bmatrix} x_k - x_* \\ y_k - y_* \end{bmatrix} + \begin{bmatrix} x_k - x_* \end{bmatrix}^T q_2 \begin{bmatrix} x_k - x_* \end{bmatrix},
\]

where \( Q_2 \) is defined in (7). Since \( y_k = \text{prox}_{\alpha f}(x_k) \) and \( y_* = \text{prox}_{\alpha f}(x_*) \), this is exactly the quadratic constraint that the \( \text{prox}_{\alpha f} \) operator satisfies. Thus, we have for all \( k \),

\[
e_k^T Q_1 e_k \geq 0.
\]

We also note that

\[
0 \leq \begin{bmatrix} 0 & \alpha I_d \\ -I_d & 2I_d & -I_d \end{bmatrix} e_k = \begin{bmatrix} 0 & \alpha I_d \\ -I_d & 2I_d & -I_d \end{bmatrix} e_k = \begin{bmatrix} (2y_k - x_k) - (2y_* - x_*) \\ z_k - z_* \end{bmatrix},
\]

As \( z_k = \text{prox}_{\alpha g}(2y_k - x_k) \) and \( z_* = \text{prox}_{\alpha g}(2y_* - x_*) \), we similarly conclude that \( e_k^T Q_2 e_k \geq 0 \) is implied from the quadratic constraint that \( \text{prox}_{\alpha g} \) satisfies. Returning to (16), if we multiply from the left and right by \( e_k^T \) and \( e_k \) respectively, we obtain

\[
e_k^T W_0 e_k + \sigma_1 e_k^T Q_1 e_k + \sigma_2 e_k^T Q_2 e_k \leq 0.
\]

Since \( \sigma_1, \sigma_2 > 0 \) and we have shown that \( e_k^T Q_1 e_k \geq 0 \) and \( e_k^T Q_2 e_k \geq 0 \), it must be that \( e_k^T W_0 e_k \leq 0 \). Hence, \( V_{k+1} - V_k \leq 0 \), and the result now follows from Lemma 1.

2) Case 2: Non-strongly convex and smooth \( f \): If \( f \in \mathcal{F}(0, L_f) \) with \( 0 < L_f < \infty \), we may leverage the smoothness of \( f \) to refine the result of the previous section. In this case, we can give a convergence rate of the average sub-optimality of the objective function. Define the new Lyapunov function

\[
V_k = \|x_k - x_*\|^2 + \theta \sum_{i=0}^{k-1} \|F(z_i) - F(z_*)\|.
\]
When this Lyapunov function is non-increasing we can obtain the following convergence rate in terms of the function values, the proof of which is identical to that of Lemma 1.

**Lemma 2** Consider the algorithm in 2. Suppose there exists $\theta > 0$ such the Lyapunov function $V_k$ defined by (23) is non-increasing for all $k$, i.e., $V_{k+1} \leq V_k$ for all $k$. Then

$$
\frac{1}{k} \sum_{i=0}^{k-1} [F(z_i) - F(z_*)] \leq \frac{1}{\theta k} \|x_0 - x_*\|_2^2.
$$

This allows us to prove the following theorem for when $f$ is non-strongly convex and smooth.

**Theorem 2** Consider the following matrix inequality

$$
W_1 + \sigma_1 Q_1 + \sigma_2 Q_2 \preceq 0,
$$

where

$$
W_1 = \begin{bmatrix}
0 & -\alpha L_f - \lambda \\
-\alpha L_f & \alpha L_f - \lambda - \lambda^2
\end{bmatrix} \otimes I_d,
$$

and $Q_1$ and $Q_2$ are defined in (17). If $\sigma_1, \sigma_2, \alpha, \lambda, \theta > 0$ are chosen so that (24) is satisfied, then for all $f \in F(0, L_f)$ with $0 < L_f < \infty$ and $g \in F(0, \infty)$ the iterates in (2) satisfy

$$
\frac{1}{k} \sum_{i=0}^{k-1} [F(z_i) - F(z_*)] \leq \frac{1}{\theta k} \|x_0 - x_*\|_2^2.
$$

**Proof:** We begin by bounding the difference of the Lyapunov function defined in (23), $V_{k+1} - V_k$, by a quadratic form in the error signal $e_k$ (see (19)). From the convexity and smoothness of $f$, we can write

$$
f(z_k) - f(y_k) \leq \nabla f(y_k)^T (z_k - y_k) + \frac{L_f}{2} \|z_k - y_k\|_2^2
$$

and

$$
f(y_k) - f(z_*) \leq \nabla f(y_k)^T (y_k - z_*),
$$

where we have used the fact that $z_* = y_*$. From the convexity of $g$ we have

$$
g(z_k) - g(z_k) \leq \partial g(z_*)^T (z_k - z_*).
$$

Adding these three inequalities together and using the relation (12) allows us to conclude

$$
F(z_k) - F(z_*) \leq \frac{L_f}{2} \|y_k - y_*\|_2^2 + \left( \frac{L_f}{2} - \frac{1}{\alpha} \right) \|z_k - z_*\|_2^2
$$

$$
+ \left( \frac{1}{\alpha} - L \right) (y_k - y_*)^T (z_k - z_*).
$$

Using the recursion for $x_{k+1}$, we then find that

$$
V_{k+1} - V_k \leq e_k^T W_1 e_k.
$$

The proof now proceeds identically as in the proof of Theorem 1 up to the statement that (24) implies $e_k^T W_1 e_k \leq 0$. Then by (11), we have that $V_{k+1} - V_k \leq 0$, and the result follows from Lemma 2.

3) Case 3: Strongly convex and smooth $f$: We now assume that $f \in F(m_f, L_f)$ and $g \in F(0, \infty)$, with $0 < m_f \leq L_f < \infty$. For this scenario we propose the Lyapunov function

$$
V_k = \|x_k - x_*\|_2^2.
$$

The following lemma characterizes when we can extract a linear convergence rate from this Lyapunov function.

**Lemma 3** Consider the algorithm in (2). Suppose there exists $\rho \in (0, 1)$ such the Lyapunov function $V_k$ defined by (13) satisfies

$$
V_{k+1} \leq \rho^2 V_k, \quad \text{for all } k.
$$

Then we have the following linear convergence rate

$$
\|x_k - x_*\|_2^2 \leq \rho^{2k} \|x_0 - x_*\|_2^2.
$$

**Proof:** The proof follows immediately from (33), the definition of $V_k$, and induction. We again see that the difference $V_{k+1} - \rho^2 V_k$ can be written as a quadratic form acting on the error signal $e_k$ as defined in (19). Using the definition for $x_{k+1}$ in terms of the previous iterates, we can write

$$
V_{k+1} - \rho^2 V_k = e_k^T Q e_k,
$$

where $Q$ is given by

$$
Q = \begin{bmatrix}
1 - \rho^2 & -\lambda & \lambda \\
-\lambda & \lambda & -\lambda^2 \\
\lambda & -\lambda^2 & \lambda^2
\end{bmatrix} \otimes I_d.
$$

Now that the function $V_k$ is represented as a quadratic form in the error state $e_k$, we can use the exact same reasoning developed for the non-strongly convex case to arrive at the following theorem.

**Theorem 3** Consider the following matrix inequality

$$
Q + \sigma_1 Q_1 + \sigma_2 Q_2 \preceq 0,
$$

where $Q$ is given in (36) and $Q_1$ and $Q_2$ are given in (17). If $\sigma_1, \sigma_2, \alpha, \lambda, \rho \in (0, 1)$ are chosen so that (37) is satisfied, then for all $f \in F(m_f, L_f)$ and $g \in F(0, \infty)$ with $0 < m_f \leq L_f < \infty$, the iterates in (2) satisfy the following linear convergence rate

$$
\|x_k - x_*\|_2^2 \leq \rho^{2k} \|x_0 - x_*\|_2^2.
$$

**Proof:** We proceed identically as in the proof of Theorems 1 and 2. If (37) is satisfied, then $e_k^T Q e_k \leq 0$ for all $k$. This is equivalent to $V_{k+1} - \rho^2 V_k \leq 0$, by which (38) follows from Lemma 3.

**D. Optimizing the Bound and Relaxation Parameter**

In this section, for each of the three cases presented above we explore how the associated matrix inequality can give optimal settings for the algorithm parameters and, in the case of non-strongly convex objective functions, explicit symbolic convergence rates.
1) Case 1: Non-strongly convex and non-smooth case:
We now select algorithm parameters that satisfy the matrix inequality in (10). In doing so, we arrive at a new and simple proof of the $O(1/k)$ convergence of Douglas-Rachford splitting in the non-strongly convex and non-smooth case.

**Theorem 4** If $f \in F(0, \infty)$ and $g \in F(0, \infty)$, then for any choice of $\lambda \in (0, 2)$ and $\alpha > 0$, if we set $\sigma_1 = \sigma_2 = \sigma$, with
\[
\sigma := 2\lambda/\alpha, \quad \theta_0 := \alpha^2\lambda(2 - \lambda),
\]
then $\sigma_1, \sigma_2, \alpha, \lambda$, and $\theta_0$ satisfy the matrix inequality (16).

**Proof:** Making these substitutions results in $W_0 + \sigma_1 Q_1 + \sigma_2 Q_2$ being equal to the zero matrix, which is indeed negative semidefinite.

**Remark 1** We can maximize $\theta$ by setting $\lambda = 1$ which gives us the convergence rate
\[
\frac{1}{k} \sum_{i=0}^{k-1} \| \partial f(y_i) + \partial g(z_i) \|_2^2 \leq \frac{1}{\alpha^2 k} \| x_0 - x^* \|_2^2.
\]

**Remark 2** Using the relation (12) we can rewrite (40) as
\[
\frac{1}{k} \sum_{i=0}^{k-1} \| y_i - z_i \|_2^2 \leq \frac{1}{k} \| x_0 - x^* \|_2^2.
\]
Thus we see that Theorem 2 also gives a $O(1/k)$ rate toward the iterates being a fixed point of the algorithm.

2) Case 2: Non-strongly convex and smooth case: When $f \in F(0, L_f)$ with $0 < L_f < \infty$ and $g \in F(0, \infty)$, we have the following result on a feasible set of parameters for the matrix inequality (24).

**Theorem 5** For any $\alpha > 0$ and $0 < \lambda < 2$, if we set $\sigma_1 = \sigma_2 = \sigma$ and define
\[
\sigma := \frac{2\lambda}{\alpha} \left( \sqrt{\left(\frac{2 - \lambda}{\alpha L_f}\right)^2 + 1} - \left(\frac{2 - \lambda}{\alpha L_f}\right) \right)
\]
and
\[
\theta_1 := 2\alpha \lambda \left( 1 + \frac{2 - \lambda}{\alpha L_f} \right) - \sqrt{\left(\frac{2 - \lambda}{\alpha L_f}\right)^2 + 1}.
\]
Then $\sigma_1, \sigma_2, \alpha, \lambda$, and $\theta_1$ satisfy the matrix inequality (24).

**Proof:** This can be verified by substituting the expressions for $\sigma_1, \sigma_2$, and $\theta_1$ into the minors of $W_1 + \sigma_1 Q_2 + \sigma_2 Q_2$ and seeing that Sylvester's criterion is satisfied (24).

**Remark 3** For moderate values of $\alpha L_f$, we can take a second order Taylor expansion of the rightmost term in (42b) and maximize the resulting expression with respect to $\lambda$. This suggests that for the maximum $\theta_1$ in the convergence rate (26) we should set $\lambda$ to
\[
\lambda = \frac{2}{3} \left( 2 - \alpha L_f + \sqrt{1 - \alpha L_f + \alpha^2 L_f^2} \right).
\]

3) Case 3: Strongly convex and smooth case: When $f \in F(m_f, L_f)$, $g \in F(0, \infty)$, with $0 < m_f \leq L_f < \infty$, we can modify the matrix inequality (37) to get a linear dependence on the relaxation parameter $\lambda$. If we define $\Lambda := \begin{bmatrix} 0 & -\lambda \\ -\lambda & \Lambda^\top \end{bmatrix}$ and $M := (Q - \Lambda \Lambda^\top) + \sigma_1 Q_1 + \sigma_2 Q_2$, (44) then (16) is equivalent to
\[
M - \Lambda[-1] \Lambda^\top \preceq 0.
\]
As $\Lambda^\top \Lambda \succeq 0$, if (45) is satisfied then it must be the case that $M \preceq 0$. We now recognize that $M - \Lambda[-1] \Lambda^\top$ is the Schur Complement of the bottom right entry in the matrix
\[
\Sigma := \begin{bmatrix} M & \Lambda \\ \Lambda^\top & -1 \end{bmatrix}.
\]
By the properties of the Schur complement [25], we can conclude that (16) is satisfied if and only if $\Sigma \preceq 0$. The advantage of using $\Sigma \preceq 0$ instead of (37), is that now both the convergence rate $\rho^2$ and the relaxation parameter $\lambda$ appear linearly. Therefore, their optimal values can be given as the solution to the following SDP:
\[
\begin{align*}
\text{minimize} \quad & \rho^2, \\
\text{subject to} \quad & \Sigma \preceq 0.
\end{align*}
\]
The optimal $\rho^2$ from solving this program over a range of step sizes $\alpha$ and condition numbers $K_f = L_f/m_f$ is shown in Figure 2. We see that with increasing $K_f$, the optimal choice of $\alpha$ decreases. The corresponding optimal value of the relaxation parameter is $\lambda = 2$ in almost all cases.

**IV. A Basis Pursuit Problem**

In this section we investigate the optimal relaxation parameter for a basis pursuit problem. More details on the problem can be found in [3].

![Fig. 2. Optimal upper bound to linear convergence rate $\rho$ over $f \in F(m_f, L_f)$, $g \in F(0, \infty)$ as a function of step size $\alpha$ and condition number $K_f = L_f/m_f$.](image-url)
The basis pursuit problem can be written as
\[
\begin{align*}
\text{minimize} & \quad 1_{\{y \in \mathbb{R}^n \mid Ay = b\}}(x) + \|z\|_1 \\
\text{subject to} & \quad x - z = 0,
\end{align*}
\]
with data $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. We will run Douglas-Rachford splitting on the dual of this problem (ADMM). The objective function is neither smooth nor strongly convex, so the dual function will not be either. The previous sections suggest that $\lambda = 1$ should result in the fastest convergence.

We test the convergence with $n = 10000$ and $m = 300$ over a range of values of $\lambda$ with $\alpha = 1$. We see in Figure 3 that setting $\lambda = 1$ gives close to optimal performance, while setting $\lambda = 2$ makes the algorithm appear to not converge at all. Theorem 1 predicts that this would be the case, as our convergence bound becomes arbitrarily large as $\lambda \to 2$ for the weakly convex case.

**Fig. 3.** Convergence of Basis Pursuit problem for varying values of relaxation parameter $\lambda$.

**V. CONCLUSION**

We have presented a unified framework for deriving convergence bounds and optimal parameter settings for Douglas-Rachford splitting. With minor modifications, our framework encompasses different sets of assumptions on the smoothness and convexity parameters of $f$. We are able to give simple proofs of symbolic rates for the non-strongly convex case and, in the strongly convex case, find optimal choices for the relaxation parameter $\lambda$ by solving a small convex program for each given step size $\alpha$. This is in contrast to previous numerical work where, instead, a grid search is done over possible values of $\lambda$ to find an optimal corresponding rate. For future work, this framework will be extended to encompass accelerated variants of Douglas-Rachford splitting, as well as three or more operator splitting and multi-block ADMM.

**REFERENCES**

[1] J. Douglas and H. H. Rachford, “On the numerical solution of heat conduction problems in two and three space variables,” *Transactions of the American Mathematical Society*, vol. 82, no. 2, pp. 421–439, 1956.

[2] G. Stathopoulos, H. Shukla, A. Szucs, Y. Pu, C. N. Jones et al., “Operator splitting methods in control,” *Foundations and Trends® in Systems and Control*, vol. 3, no. 3, pp. 249–362, 2016.

[3] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein et al., “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[4] D. Gabay, “Chapter ix applications of the method of multipliers to variational inequalities,” in *Studies in mathematics and its applications*. Elsevier, 1983, vol. 15, pp. 299–331.

[5] P.-L. Lions and B. Mercier, “Splitting algorithms for the sum of two nonlinear operators,” *SIAM Journal on Numerical Analysis*, vol. 16, no. 6, pp. 964–979, 1979.

[6] J. Eckstein and D. P. Bertsekas, “On the douglasrachford splitting method and the proximal point algorithm for maximal monotone operators,” *Mathematical Programming*, vol. 55, no. 1-3, pp. 293–318, 1992.

[7] B. He and X. Yuan, “On the $O(\ln)$ convergence rate of the douglas–rachford alternating direction method,” *SIAM Journal on Numerical Analysis*, vol. 50, no. 2, pp. 700–709, 2012.

[8] W. Deng and W. Yin, “On the global and linear convergence of the generalized alternating direction method of multipliers,” *Journal of Scientific Computing*, vol. 66, no. 3, pp. 889–916, 2016.

[9] P. Giselsson and S. Boyd, “Linear convergence and metric selection for douglas-rachford splitting and admm,” *IEEE Transactions on Automatic Control*, vol. 62, no. 2, pp. 532–544, 2017.

[10] Y. Drori and M. Teboulle, “Performance of first-order methods for smooth convex minimization: a novel approach,” *Mathematical Programming*, vol. 145, no. 1-2, pp. 451–482, 2014.

[11] L. Lessard, B. Recht, and A. Packard, “Analysis and design of optimization algorithms via integral quadratic constraints,” *SIAM Journal on Optimization*, vol. 26, no. 1, pp. 57–95, 2016.

[12] M. Fazlyab, A. Ribeiro, M. Morari, and V. M. Preciado, “Analysis of optimization algorithms via integral quadratic constraints: Nonstrongly convex problems,” *SIAM Journal on Optimization*, vol. 28, no. 3, pp. 2654–2689, 2018.

[13] B. Hu and L. Lessard, “Dissipativity theory for nesterovs accelerated method,” in *International Conference on Machine Learning*, 2017, pp. 1549–1557.

[14] B. Van Scoy, R. A. Freeman, and K. M. Lynch, “The fastest known globally convergent first-order method for minimizing strongly convex functions,” *IEEE Control Systems Letters*, vol. 2, no. 1, pp. 49–54, 2018.

[15] E. K. Ryu, A. B. Taylor, C. Bergeling, and P. Giselsson, “Operator splitting performance estimation: Tight contraction factors and optimal parameter selection,” *arXiv preprint arXiv:1812.00146*, 2018.

[16] M. Fazlyab, M. Morari, and V. M. Preciado, “Design of first-order optimization algorithms via sum-of-squares programming,” in *2018 IEEE Conference on Decision and Control (CDC)*. IEEE, 2018, pp. 4445–4452.

[17] R. Nishihara, L. Lessard, B. Recht, A. Packard, and M. Jordan, “A general analysis of the convergence of adm,” in *International Conference on Machine Learning*, 2015, pp. 343–352.

[18] G. Franca and J. Bento, “An explicit rate bound for over-relaxed admm,” in *Information Theory (ISIT)*, 2016 *IEEE International Symposium on*. IEEE, 2016, pp. 2104–2108.

[19] E. Ghadimi, A. Teixeira, I. Shames, and M. Johansson, “Optimal parameter selection for the alternating direction method of multipliers (admm): quadratic problems,” *IEEE Transactions on Automatic Control*, vol. 60, no. 3, pp. 644–658, 2015.

[20] P. Giselsson and S. Boyd, “Diagonal scaling in douglas-rachford splitting and admm,” in *Decision and Control (CDC)*, 2014 IEEE 53rd Annual Conference on. IEEE, 2014, pp. 5033–5039.

[21] B. Açıkmeşe and M. Corless, “Observers for systems with nonlinearities satisfying incremental quadratic constraints,” *Automatica*, vol. 47, no. 7, pp. 1339–1348, 2011.

[22] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, 1st ed. Springer Publishing Company, Incorporated, 2014.

[23] W. M. Haddad and V. Chellaboina, *Nonlinear dynamical systems and control: a Lyapunov-based approach*. Princeton University Press, 2011.

[24] R. A. Horn, R. A. Horn, and C. R. Johnson, *Matrix analysis*. Cambridge university press, 1990.

[25] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.