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

Jacob H. Seidman, Mahyar Fazlyab, Victor M. Preciado, and George J. Pappas

***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.**

***Index Terms*—Optimization algorithms, Lyapunov methods.**

I. Introduction

IN THIS letter, we consider problems of the form

$$\text{minimize}_{x\in\mathbb{R}^d}\ \{F(x) = f(x) + g(x)\}, \tag{1}$$

where $f, g : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ are convex, closed, and proper (c.c.p.). Douglas-Rachford splitting (DRS) solves problem (1) with the following iterations:

$$y_k = \text{prox}_{\alpha f}(x_k), \tag{2a}$$

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

$$x_{k+1} = x_k + \lambda_k(z_k - y_k), \tag{2c}$$

where prox is the proximal operator (see Definition 1) and α and λ_k are known as the proximal step size and relaxation parameter, respectively. For a proper selection of these parameters, the limiting values of both y_k and z_k will be a solution to (1). The goal of this letter is to provide convergence rates for DRS over various assumptions on f and g, and optimize these rates with respect to the algorithm parameters α and λ_k using semidefinite programs (SDPs).

The algorithm was first proposed in [1], and has since found application in general separable optimization problems [2]. Its dual formulation, ADMM, has been particularly useful in distributed optimization problems [3]. Since the iterates of ADMM can be written as applying DRS to the dual problem [4], [5], convergence results for one algorithm are valid for the other as well when strong duality holds.

The convergence of DRS has previously been analyzed using monotone operator theory and variational inequalities, see [6], [7]. These techniques have led to proofs of a $O(1/k)$ convergence rate for the non-strongly convex case [8], and linear convergence when f is smooth and strongly convex [9], [10]. In the most general case, a condition for convergence is that $\lambda_k \in (0, 2)$ with $\sum_{k=0}^{\infty} \lambda_k(2 - \lambda_k) = \infty$ [11], though there exist cases where the algorithm converges with $\lambda_k > 2$ [12].

Recently, there has been interest in automating the analysis and design of optimization algorithms via SDPs, [13]–[19]. In particular, through the method of integral quadratic constraints proposed in [14], the authors of [12] derive an SDP for choosing the parameters of ADMM in the case of smooth and strongly convex f. Using a similar framework, the authors of [20] provide evidence that as the relaxation parameter approaches 2 from below, the linear convergence rate is close to being optimal and in [21] are able to analytically solve the SDP to give a convergence rate. The work in [22] gives an optimal choice for the relaxation parameter when f is quadratic. Furthermore, [23] gives a set of assumptions in which a bound on the linear convergence rate is minimized by setting $\lambda_k = 2$.

Our Contribution: By viewing DRS as a linear system with non-linear feedback, we derive a dimensionally

Manuscript received March 1, 2019; revised May 7, 2019; accepted May 27, 2019. Date of publication June 13, 2019; date of current version June 25, 2019. This work was supported in part by Army Research Laboratory under Grant CRA DCIST W911NF-17-2-0181, and in part by NSF under Grant CAREER-ECCS-1651433. Recommended by Senior Editor J.-F. Zhang. *(Corresponding author: Jacob H. Seidman.)*

J. H. Seidman is with the Department of Applied Mathematics and Computational Science, University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: seidj@sas.upenn.edu).

M. Fazlyab, V. M. Preciado, and G. J. Pappas are with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: mahyarfa@seas.upenn.edu; preciado@seas.upenn.edu; pappasg@seas.upenn.edu).

Digital Object Identifier 10.1109/LCSYS.2019.2922669

independent matrix inequality which gives convergence guarantees via Lyapunov functions. Whereas such an approach was previously applied in [12] 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. 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 is sufficient to guarantee a linear convergence rate. We are able to modify the matrix inequality to linearize the dependence on λ_k, 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 λ_k.

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 real $m\times n$-dimensional matrices by $\mathbb{R}^{m\times n}$, and the n-dimensional identity matrix and zero matrix by I_n and 0_n, respectively. For a function $f\colon \mathbb{R}^d \to \mathbb{R}$, we denote by $\mathrm{dom} f = \{x \in \mathbb{R}^d \colon f(x) < \infty\}$ the effective domain of f. The subdifferential of a function f at a point x is $\partial f(x) := \{g \mid f(y) - f(x) \geq g^\top(y-x), \forall y \in \mathrm{dom}(f)\}$. By abuse of notation we will also refer to a subgradient, that is an element of the subdifferential 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 $B \in \mathbb{R}^{p\times q}$ their Kronecker product is $A \otimes B$.

We say a differentiable function $f\colon \mathbb{R}^d \to \mathbb{R}\cup\{+\infty\}$ is L_f-smooth on $\mathcal{S} \subseteq \mathrm{dom} f$ if $\|\nabla f(x) - \nabla f(y)\|_2 \leq L_f\|x-y\|_2$ for some $L_f > 0$ and all $x, y \in \mathcal{S}$. This also implies for all $x, y \in \mathcal{S}$, $f(y) \leq f(x) + \nabla f(x)^\top(y-x) + (L_f/2)\|y-x\|_2^2$. A differentiable function $f\colon \mathbb{R}^d \to \mathbb{R}$ is m_f-strongly convex on $\mathcal{S} \subseteq \mathrm{dom} f$ if $m_f\|x-y\|_2^2 \leq (x-y)^\top(\nabla f(x) - \nabla f(y))$ for some $m_f > 0$ and all $x, y \in \mathcal{S}$. The class of functions which are L_f-smooth and m_f-strongly convex is denoted by $\mathcal{F}(m_f, L_f)$.

Definition 1 (Proximal Operator): Given a c.c.p. function $f : \mathbb{R}^d \to \mathbb{R}\cup\{+\infty\}$ and $\alpha > 0$, the proximal operator $\mathrm{prox}_{\alpha f} : \mathbb{R}^d \to \mathbb{R}^d$ is defined as

$$\mathrm{prox}_{\alpha f}(x) = \mathrm{argmin}_y \left\{ f(y) + \frac{1}{2\alpha}\|x-y\|_2^2 \right\}. \tag{3}$$

The point $y = \mathrm{prox}_{\alpha f}(x)$ also is given by the implicit solution to the subgradient equation

$$y = x - \alpha\partial f(y). \tag{4}$$

We say that a nonlinear function $\phi : \mathbb{R}^d \to \mathbb{R}^d$ satisfies the incremental quadratic constraint [24] (or point-wise integral quadratic constraint [14]) defined by $Q \in \mathbb{R}^{2n\times 2n}$ if for all x, y,

$$\begin{bmatrix} x-y \\ \phi(x)-\phi(y) \end{bmatrix}^\top Q \begin{bmatrix} x-y \\ \phi(x)-\phi(y) \end{bmatrix} \geq 0. \tag{5}$$

For $0 \leq m, L < \infty$ define

$$Q(m, L) = \begin{bmatrix} -\frac{mL}{m+L} & 1/2 \\ 1/2 & -\frac{1}{m+L} \end{bmatrix} \otimes I_d, \tag{6}$$

and define $Q(m, \infty)$ as $\lim_{L\to\infty} Q(m, L)$. It was noted in [14], [25] that a differentiable function f belongs to the class $\mathcal{F}(m_f, L_f)$ on $\mathcal{S}$ if and only if the gradient ∇f satisfies the incremental QC defined by $Q(m_f, L_f)$. When $L_f = \infty$, the subgradient ∂f satisfies the QC defined by $Q(m_f, \infty)$. If we define

$$Q_p(m, L, \alpha) = \begin{bmatrix} 0 & I_d \\ \alpha I_d & -I_d \end{bmatrix} Q(m, L) \begin{bmatrix} 0 & \alpha I_d \\ I_d & -I_d \end{bmatrix}, \tag{7}$$

then the proximal operator of a function $f \in \mathcal{F}(m, L)$, $\mathrm{prox}_{\alpha f}$, satisfies the incremental QC defined by $Q_p(m, L, \alpha)$ [15].

III. Analysis of Douglas-Rachford Splitting via Matrix Inequalities

A. Douglas-Rachford Splitting as a Dynamical System

We can write the updates in (2) as a linear system with state x_k and feedback nonlinearity $\phi(x_k)$,

$$x_{k+1} = x_k + \lambda_k\phi(x_k), \tag{8}$$

where

$$\phi(x_k) := \mathrm{prox}_{\alpha g}(2\mathrm{prox}_{\alpha f}(x_k) - x_k) - \mathrm{prox}_{\alpha g}(x_k). \tag{9}$$

Our main technique is to describe the nonlinearity ϕ with incremental QCs 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

From relation (4), the iterates (2) can be rewritten as

$$y_k = x_k - \alpha\partial f(y_k), \tag{10a}$$
$$z_k = 2y_k - x_k - \alpha\partial g(z_k), \tag{10b}$$
$$x_{k+1} = x_k + \lambda_k(z_k - y_k). \tag{10c}$$

The fixed points of (10) satisfy

$$x_\star = y_\star + \alpha\partial f(y_\star), \;\; y_\star = z_\star, \;\; \partial f(y_\star) + \partial g(z_\star) = 0. \tag{11}$$

Since $y_\star = z_\star$, the rightmost equality is exactly the optimality condition for (1).

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

$$\phi(x_k) = z_k - y_k = -\alpha(\partial f(y_k) + \partial g(z_k)). \tag{12}$$

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

C. Convergence Certificates via Matrix Inequalities

1) Case 1 (Non-Strongly Convex and Non-Smooth Case): We first assume that $f, g \in \mathcal{F}(0, \infty)$. We propose the following family of Lyapunov functions parameterized by a sequence $\{\theta_i\}_{i=0}^{\infty}$ with $\theta_i > 0$,

$$V_k = \|x_k - x_\star\|_2^2 + \sum_{i=0}^{k-1} \theta_i \|\partial f(y_i) + \partial g(z_i)\|_2^2, \tag{13}$$

for all $k > 0$ and $V_0 = \|x_0 - x_\star\|_2^2$. For notational convenience define the partial sums $\Theta_k = \sum_{i=0}^{k-1} \theta_i$. The presence of the running sum of subgradients is reminiscent of the Popov criterion [26]. It can also be interpreted as the running weighted sum of fixed point residuals (see (11)). The next lemma shows how this Lyapunov function can ensure a convergence rate in terms of the growth of Θ_k.

Lemma 1: Consider the algorithm in (2). Suppose there exists a sequence $\{\theta_i\}_{i=0}^{\infty}$ with $\theta_i > 0$ such that $V_{k+1} \leq V_k$ for all $k \geq 0$. Then

$$\min_{i=0,\ldots,k-1} \|\partial f(y_i) + \partial g(z_i)\|_2^2 \leq \frac{1}{\Theta_k} \|x_0 - x_\star\|_2^2. \tag{14}$$

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

$$\|x_k - x_\star\|_2^2 + \sum_{i=0}^{k-1} \theta_i \|\partial f(y_i) + \partial g(z_i)\|_2^2 \leq \|x_0 - x_\star\|_2^2. \tag{15}$$

Removing the first term on the left, and dividing through by Θ_k gives

$$\sum_{i=0}^{k-1} \frac{\theta_i}{\Theta_k} \|\partial f(y_i) + \partial g(z_i)\|_2^2 \leq \frac{\|x_0 - x_\star\|_2^2}{\Theta_k}. \tag{16}$$

The result follows from the fact that the left side is a weighted average, as $\theta_i > 0$ and $\sum_{i=0}^{k-1} \theta_i / \Theta_k = 1$. ■

In the following theorem, we derive a matrix inequality in terms of α, λ, and $\{\theta_i\}_{i=0}^{\infty}$ as a sufficient condition to guarantee $V_{k+1} \leq V_k$, which in turn implies (14).

Theorem 1: Let $m_f = 0$, $L_f = \infty$, and consider the following matrix inequality

$$W_k^{(0)} + \sigma_k^{(1)} Q^{(1)} + \sigma_k^{(2)} Q^{(2)} \preceq 0, \tag{17}$$

where

$$W_k^{(0)} = \begin{bmatrix} 0 & -\lambda_k & \lambda_k \\ -\lambda_k & \lambda_k^2 + \frac{\theta_k}{\alpha^2} & -\left(\lambda_k^2 + \frac{\theta_k}{\alpha^2}\right) \\ \lambda_k & -\left(\lambda_k^2 + \frac{\theta_k}{\alpha^2}\right) & \lambda_k^2 + \frac{\theta_k}{\alpha^2} \end{bmatrix} \otimes I_d, \tag{18a}$$

$$Q^{(1)} = \begin{bmatrix} 0 & I_d \\ \alpha I_d & -I_d \\ 0 & 0 \end{bmatrix} Q(m_f, L_f) \begin{bmatrix} 0 & \alpha I_d & 0 \\ I_d & -I_d & 0 \end{bmatrix}, \tag{18b}$$

$$Q^{(2)} = \begin{bmatrix} 0 & -I_d \\ 0 & 2I_d \\ \alpha I_d & -I_d \end{bmatrix} Q(0, \infty) \begin{bmatrix} 0 & 0 & \alpha I_d \\ -I_d & 2I_d & -I_d \end{bmatrix}. \tag{18c}$$

If $\alpha, \lambda_k, \theta_k > 0$ and $\sigma_k^{(1)}, \sigma_k^{(2)} \geq 0$ are chosen so that (17) is satisfied for all $k \geq 0$, then for all $f \in \mathcal{F}(0, \infty)$ and $g \in \mathcal{F}(0, \infty)$ the iterates in (2) satisfy

$$\min_{i=0,\ldots,k-1} \|\partial f(y_i) + \partial g(z_i)\|_2^2 \leq \frac{1}{\Theta_k} \|x_0 - x_\star\|_2^2. \tag{19}$$

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

$$e_k := \left[(x_k - x_\star)^\top \quad (y_k - y_\star)^\top \quad (z_k - z_\star)^\top\right]^\top. \tag{20}$$

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

$$V_{k+1} - V_k = e_k^\top W_k^{(0)} e_k, \tag{21}$$

where $W_k^{(0)}$ is given by (18a). Next, note that

$$e_k^\top Q^{(1)} e_k = \begin{bmatrix} x_k - x_\star \\ y_k - y_\star \end{bmatrix}^\top Q_p(m_f, L_f, \alpha) \begin{bmatrix} x_k - x_\star \\ y_k - y_\star \end{bmatrix}, \tag{22}$$

where $Q_p(m_f, L_f, \alpha)$ is defined in (7). Since $y_k = \mathrm{prox}_{\alpha f}(x_k)$ and $y_\star = \mathrm{prox}_{\alpha f}(x_\star)$, this is exactly the incremental QC that the $\mathrm{prox}_{\alpha f}$ operator satisfies. Thus, we have for all k, $e_k^\top Q^{(1)} e_k \geq 0$. We also note that

$$\begin{bmatrix} 0 & 0 & \alpha I_d \\ -I_d & 2I_d & -I_d \end{bmatrix} e_k = \begin{bmatrix} 0 & \alpha I_d \\ I_d & -I_d \end{bmatrix} \begin{bmatrix} (2y_k - x_k) - (2y_\star - x_\star) \\ z_k - z_\star \end{bmatrix}.$$

As $z_k = \mathrm{prox}_{\alpha g}(2y_k - x_k)$ and $z_\star = \mathrm{prox}_{\alpha g}(2y_\star - x_\star)$, we similarly conclude that $e_k^\top Q^{(2)} e_k \geq 0$ is implied from the incremental QC that $\mathrm{prox}_{\alpha g}$ satisfies. Returning to (17), if we multiply from the left and right by $e_k^\top$ and e_k respectively, we obtain

$$e_k^\top W_k^{(0)} e_k + \sigma_k^{(1)} e_k^\top Q^{(1)} e_k + \sigma_k^{(2)} e_k^\top Q^{(2)} e_k \leq 0. \tag{23}$$

Since $\sigma_k^{(1)}, \sigma_k^{(2)} \geq 0$ and we have shown that $e_k^\top Q^{(1)} e_k \geq 0$ and $e_k^\top Q^{(2)} e_k \geq 0$, it must be that $e_k^\top W_k^{(0)} e_k \leq 0$. 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 particular, we can use the inequality for L_f-smooth functions (see Preliminaries) to relate the behavior of the subgradients to the objective values, whereas in the previous section this inequality was not available. Let

$$V_k = \|x_k - x_\star\|_2^2 + \sum_{i=0}^{k-1} \theta_i [F(z_i) - F(z_\star)], \tag{24}$$

with V_0 defined as in the previous case.

This Lyapunov function leads to the following Lemma, the proof of which is identical to that of Lemma 1.

Lemma 2: Consider the algorithm in (2). Suppose there exists a sequence $\{\theta_i\}_{i=0}^{\infty}$ with $\theta_i > 0$ such that $V_{k+1} \leq V_k$ for all $k \geq 0$. Then

$$\min_{i=0,\ldots,k-1} [F(z_i) - F(z_\star)] \leq \frac{1}{\Theta_k} \|x_0 - x_\star\|_2^2.$$

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

Theorem 2: Let $0 = m_f < L_f < \infty$ and consider the following matrix inequality

$$W_k^{(1)} + \sigma_k^{(1)} Q^{(1)} + \sigma_k^{(2)} Q^{(2)} \preceq 0, \tag{25}$$

where

$$W_k^{(1)} = \begin{bmatrix} 0 & -\lambda_k & \lambda_k \\ -\lambda_k & \frac{\theta_k L_f}{2} + \lambda_k^2 & \frac{\theta_k}{2}\left(\frac{1}{\alpha} - L_f\right) - \lambda_k^2 \\ \lambda_k & \frac{\theta_k}{2}\left(\frac{1}{\alpha} - L_f\right) - \lambda_k^2 & \theta_k\left(\frac{L_f}{2} - \frac{1}{\alpha}\right) + \lambda_k^2 \end{bmatrix} \otimes I_d, \tag{26}$$

and $Q^{(1)}$ and $Q^{(2)}$ are defined in (18). If $\alpha, \lambda_k, \theta_k > 0$ and $\sigma_k^{(1)}, \sigma_k^{(2)} \geq 0$ are chosen so that (25) is satisfied for all $k \geq 0$, then for all $f \in \mathcal{F}(0, L_f)$ with $0 < L_f < \infty$ and $g \in \mathcal{F}(0, \infty)$ the iterates in (2) satisfy

$$\min_{i=0,\ldots,k-1} [F(z_i) - F(z_\star)] \leq \frac{1}{\Theta_k} \|x_0 - x_\star\|_2^2. \tag{27}$$

Proof: We begin by bounding the difference of the Lyapunov function defined in (24), $V_{k+1} - V_k$, by a quadratic form in the error signal e_k (see (20)). From the convexity and smoothness of f, we can write

$$f(z_k) - f(y_k) \leq \nabla f(y_k)^\top (z_k - y_k) + \frac{L_f}{2}\|z_k - y_k\|_2^2, \tag{28}$$

$$f(y_k) - f(z_\star) \leq \nabla f(y_k)^\top (y_k - y_\star), \tag{29}$$

where we have used that $z_\star = y_\star$. From the convexity of g

$$g(z_k) - g(z_\star) \leq \partial g(z_k)^\top (z_k - z_\star). \tag{30}$$

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

$$F(z_k) - F(z_\star) \leq \frac{L_f}{2}\|y_k - y_\star\|_2^2 + \left(\frac{L_f}{2} - \frac{1}{\alpha}\right)\|z_k - z_\star\|_2^2 + \left(\frac{1}{\alpha} - L_f\right)(y_k - y_\star)^\top (z_k - z_\star). \tag{31}$$

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

$$V_{k+1} - V_k \leq e_k^\top W_k^{(1)} e_k. \tag{32}$$

The proof now proceeds identically as in the proof of Theorem 1 up to the statement that (25) implies $e_k^\top W_k^{(1)} e_k \leq 0$. Then by (32), 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 $f \in \mathcal{F}(m_f, L_f)$ and $g \in \mathcal{F}(0, \infty)$, with $0 < m_f \leq L_f < \infty$. For this scenario let

$$V_k = \|x_k - x_*\|_2^2. \tag{33}$$

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 (33) satisfies $V_{k+1} \leq \rho^2 V_k$ for all $k \geq 0$. Then

$$\|x_k - x_*\|_2^2 \leq \rho^{2k}\|x_0 - x_*\|_2^2. \tag{34}$$

Proof: The proof follows immediately $V_{k+1} \leq \rho^2 V_k$, the definition (33) 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 (20). Using the definition for x_{k+1} in terms of the previous iterates, we can write

$$V_{k+1} - \rho^2 V_k = e_k^\top Q_k e_k, \tag{35}$$

where Q_k is given by

$$Q_k = \begin{bmatrix} 1 - \rho^2 & -\lambda_k & \lambda_k \\ -\lambda_k & \lambda_k^2 & -\lambda_k^2 \\ \lambda_k & -\lambda_k^2 & \lambda_k^2 \end{bmatrix} \otimes I_d. \tag{36}$$

From (35), we can use the same reasoning developed in the non-strongly convex case to arrive at the following theorem.

Theorem 3: Let $0 < m_f \leq L_f < \infty$ and consider the following matrix inequality

$$Q_k + \sigma_k^{(1)} Q^{(1)} + \sigma_k^{(2)} Q^{(2)} \preceq 0, \tag{37}$$

where Q_k is given in (36) and $Q^{(1)}$ and $Q^{(2)}$ are given in (18). If $\alpha, \lambda_k > 0$, $\sigma_k^{(1)}, \sigma_k^{(2)} \geq 0$, and $\rho \in (0, 1)$ are chosen so that (37) is satisfied for all $k \geq 0$, then for all $f \in \mathcal{F}(m_f, L_f)$ and $g \in \mathcal{F}(0, \infty)$ with $0 < m_f \leq L_f < \infty$, the iterates in (2) satisfy the following linear convergence rate

$$\|x_k - x_*\|_2^2 \leq \rho^{2k}\|x_0 - x_*\|_2^2. \tag{38}$$

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

IV. Optimizing the Bound and Relaxation Parameter

For each of the cases presented in the previous section we use the associated matrix inequality to optimize our bounds on the convergence rate. In the case of non-strongly convex f we provide analytic convergence rates, while for strongly convex f we optimize our bounds numerically.

A. Case 1: Non-Strongly Convex and Non-Smooth Case

We now select algorithm parameters that satisfy the matrix inequality in (17). In doing so, we arrive at a new and simple proof of the $O(1/k)$ convergence of DRS in the non-strongly convex and non-smooth case.

Theorem 4: If $f \in \mathcal{F}(0, \infty)$ and $g \in \mathcal{F}(0, \infty)$, then for any choice of $\lambda_k = \lambda \in (0, 2)$ and $\alpha > 0$, if we set $\sigma_k^{(1)} = \sigma_k^{(2)} = \sigma_k$, with

$$\sigma_k := 2\lambda_k/\alpha, \quad \theta_k := \alpha^2 \lambda_k (2 - \lambda_k), \tag{39}$$

then $\sigma_k^{(1)}, \sigma_k^{(2)}, \alpha, \lambda_k$, and θ_k satisfy the matrix inequality (17).

Proof: Making these substitutions gives $W_k^{(0)} + \sigma_k^{(1)} Q^{(1)} + \sigma_k^{(2)} Q^{(2)} = 0_n \preceq 0$. ∎

Remark 1: The convergence rate bound provided by the parameter choices in Theorem 4 guarantees convergence only if $\lim_{k\to\infty} \Theta_k = \infty$, which in this case means $\sum_{i=0}^{\infty} \lambda_k (2 - \lambda_k) = \infty$. This is consistent with the conditions on the relaxation parameter found in [11].

Remark 2: After setting θ_k as in (39), we can maximize Θ_k by setting $\lambda_k = 1$ which results in $\Theta_k = \alpha^2 k$ and the following result

$$\min_{i=0,\ldots,k-1} \|\partial f(y_i) + \partial g(z_i)\|_2^2 \leq \frac{1}{\alpha^2 k}\|x_0 - x_\star\|_2^2. \tag{40}$$

Remark 3: Using the relation (12) we can rewrite (40) as

$$\min_{i=0,\ldots,k-1} \|z_i - y_i\|_2^2 \leq \frac{1}{k}\|x_0 - x_\star\|_2^2. \tag{41}$$

Thus we see that Theorem 4 also gives a $O(1/k)$ rate toward the iterates being a fixed point of the algorithm.

B. Case 2: Non-Strongly Convex and Smooth Case

When $f \in \mathcal{F}(0, L_f), g \in \mathcal{F}(0, \infty)$ with $0 < L_f < \infty$, we have the following result on feasibility of the matrix inequality (25).

Theorem 5: For any $\alpha > 0$ and $\lambda_k = \lambda$ with $0 < \lambda < 2$, if we set $\sigma_k^{(1)} = \sigma_k^{(2)} = \sigma_k$,

$$\sigma_k := \frac{2\lambda_k}{\alpha}\left(\sqrt{\left(\frac{2-\lambda_k}{\alpha L_f}\right)^2 + 1} - \left(\frac{2-\lambda_k}{\alpha L_f}\right)\right), \tag{42a}$$

$$\theta_k := 2\lambda_k \alpha\left(1 + \left(\frac{2-\lambda_k}{\alpha L_f}\right) - \sqrt{\left(\frac{2-\lambda_k}{\alpha L_f}\right)^2 + 1}\right). \tag{42b}$$

Then $\sigma_k^{(1)}, \sigma_k^{(2)}, \alpha, \lambda_k$, and θ_k satisfy the matrix inequality (25).

Proof: This can be verified by substituting the expressions for $\sigma_k^{(1)}$, $\sigma_k^{(2)}$, and θ_k into the minors of $W_k^{(1)} + \sigma_k^{(1)} Q^{(2)} + \sigma_k^{(2)} Q^{(2)}$ and seeing that Sylvester's criterion is satisfied [27]. ■

Remark 4: Note that for a constant selection of $\theta_k = \theta$, $\Theta_k = \theta k$. If (25) holds then

$$\min_{i=0,\ldots,k-1} [F(z_i) - F(z_\star)] \leq \frac{1}{\theta k}\|x_0 - x_\star\|_2^2. \tag{43}$$

Remark 5: For moderate values of αL_f, we can take a second order Taylor expansion of the rightmost term in (42b) and maximize the resulting expression with respect to λ_k. This suggests that we should set λ_k to

$$\lambda_k = \frac{2}{3}\left(2 - \alpha L_f + \sqrt{1 - \alpha L_f + \alpha^2 L_f^2}\right). \tag{44}$$

C. Case 3: Strongly Convex and Smooth Case

When $f \in \mathcal{F}(m_f, L_f)$, $g \in \mathcal{F}(0, \infty)$, with $0 < m_f \leq L_f < \infty$, we can modify the matrix inequality (37) to get a linear dependence on the relaxation parameter. If we define $\Lambda_k := \begin{bmatrix} 0 & -\lambda_k & \lambda_k \end{bmatrix}^\top \otimes I_d$ and

$$M_k := (Q_k - \Lambda_k \Lambda_k^\top) + \sigma_k^{(1)} Q^{(1)} + \sigma_k^{(2)} Q^{(2)}, \tag{45}$$

then (37) is equivalent to

$$M_k - \Lambda_k[-1]\Lambda_k^\top \preceq 0. \tag{46}$$

As $\Lambda_k \Lambda_k^\top \succeq 0$, if (46) is satisfied then it must be the case that $M_k \preceq 0$. We now recognize that $M_k - \Lambda_k[-1]\Lambda_k^\top$ is the Schur Complement of the bottom right entry in the matrix

$$\Sigma_k := \begin{bmatrix} M_k & \Lambda_k \\ \Lambda_k^\top & -1 \end{bmatrix}. \tag{47}$$

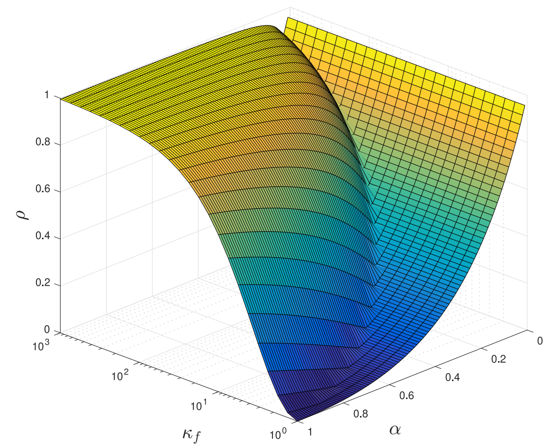

Fig. 1. Optimal upper bound to linear convergence rate ρ over $f \in \mathcal{F}(m_f, L_f)$, $g \in \mathcal{F}(0, \infty)$ as a function of step size α and condition number $\kappa_f = L_f/m_f$.

By the properties of the Schur complement [28], we can conclude that (37) is satisfied if and only if $\Sigma_k \preceq 0$. The advantage of using $\Sigma_k \preceq 0$ instead of (37), is that now both the convergence rate ρ^2 and the relaxation parameter λ_k appear linearly. If we set $\sigma_k^{(1)} = \sigma^{(1)}$, $\sigma_k^{(2)} = \sigma^{(2)}$, and $\lambda_k = \lambda$ for all k, the corresponding bound in (38) can be optimized by solving the following SDP, where λ is now a decision variable,

$$\text{minimize} \quad \rho^2, \quad \text{subject to} \quad \Sigma_k \preceq 0, \tag{48}$$

where the decision variables are $\rho^2, \lambda > 0$ and $\sigma^{(1)}, \sigma^{(2)} \geq 0$. The optimal ρ^2 from solving this program over a range of step sizes α and condition numbers $\kappa_f = L_f/m_f$ is shown in Figure 1. We see that with increasing κ_f, the optimal choice of α decreases. Across the range of values of κ_f and α, the SDP (48) returns $\lambda = 2$ as the optimal relaxation parameter.

Remark 6: We may repeat the same derivation for the inequalities (17) and (25) to linearize their dependence on λ_k as well.

V. Numerical Experiments

We investigate how our theoretical results compare with the experimental performance of DRS in the three scenarios described above. For the non-smooth and non-strongly convex case, we consider a basis pursuit problem (see [3]),

$$\begin{aligned} &\underset{x,z \in \mathbb{R}^n}{\text{minimize}} \quad \mathbb{1}_{\{y \in \mathbb{R}^n | Ay = b\}}(x) + \|z\|_1 \\ &\text{subject to} \quad x - z = 0, \end{aligned}$$

with data $A \in \mathbb{R}^{300 \times 10000}$ and $b \in \mathbb{R}^{300}$. We run DRS on the dual of this problem (ADMM) which is non-smooth and non-strongly convex. We test the convergence over a range of values of $\lambda_k = \lambda$, including $\lambda^\star = 1$ (see Remark 2), and fixed $\alpha = 1$.

For the smooth cases we consider a LASSO problem,

$$\text{minimize}_x \quad \frac{1}{2}\|Ax - b\|_2^2 + \gamma\|x\|_1.$$

with $A \in \mathbb{R}^{300 \times 200}$, $b \in \mathbb{R}^{300}$ and $\gamma = 0.1$. For the non-strongly convex case we set A to be rank deficient and plot the

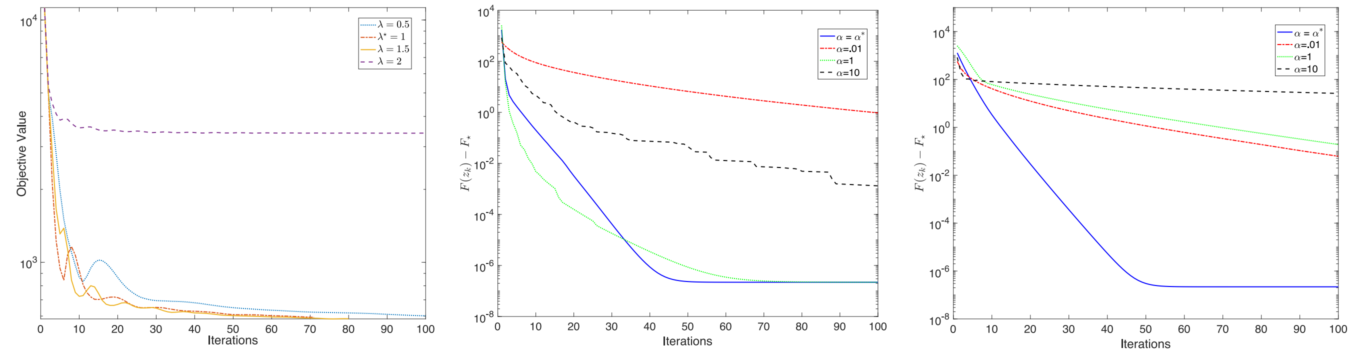

Fig. 2. From left to right: convergence of DRS on basis pursuit problem with $\alpha = 1$ and varying λ, convergence of DRS on LASSO problem with row-rank deficient A with $\lambda = 1$ and varying α, convergence of DRS on LASSO problem with full row-rank A with fixed $\lambda = 2$ and varying α.

convergence of DRS over a range of α with $\lambda_k = 1$ fixed. The value $\alpha^\star$ is found by performing a grid search over possible values and choosing that which maximizes the rate given by Theorem 2. For the strongly convex case A is set to be full rank and plot the convergence of DRS over a range of α with $\lambda_k = 2$ fixed. Again, $\alpha^\star$ is the value of α which gives the best rate as provided by Theorem 3. The results are presented in Figure 2.

VI. Conclusion

We presented a unified framework for deriving convergence bounds for DRS and parameter settings that optimize these bounds. Our framework encompasses different assumptions on the smoothness and convexity of f. We are able to give simple proofs of convergence and find optimal choices for the relaxation parameter by solving a small convex program for a fixed α. It is important to note that the parameter selections optimize our bounds in the sense of the best worst-case convergence rate over the entire class of objective functions with $f \in \mathcal{F}(m_f, L_f)$. While there are scenarios where additional structure in the problem might make alternate parameter settings more effective, the settings we see here bound the worst-case performance agnostic of any additional problem structure. For future work, this framework will be extended to encompass accelerated variants of DRS, 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," *Trans. Amer. Math. Soc.*, vol. 82, no. 2, pp. 421–439, 1956.

[2] G. Stathopoulos *et al.*, "Operator splitting methods in control," *Found. Trends® Syst. Control*, vol. 3, no. 3, pp. 249–362, 2016.

[3] S. Boyd *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends® Syst. Control*, 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*, vol. 15. Amsterdam, The Netherlands: Elsevier, 1983, pp. 299–331.

[5] J. Eckstein, "Splitting methods for monotone operators with applications to parallel optimization," Ph.D. dissertation, Dept. Civil Eng., Massachusetts Inst. Technol., Cambridge, MA, USA, 1989.

[6] P.-L. Lions and B. Mercier, "Splitting algorithms for the sum of two nonlinear operators," *SIAM J. Numer. Anal.*, vol. 16, no. 6, pp. 964–979, 1979.

[7] J. Eckstein and D. P. Bertsekas, "On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, nos. 1–3, pp. 293–318, 1992.

[8] B. He and X. Yuan, "On the O(1/n) convergence rate of the Douglas–Rachford alternating direction method," *SIAM J. Numer. Anal.*, vol. 50, no. 2, pp. 700–709, 2012.

[9] W. Deng and W. Yin, "On the global and linear convergence of the generalized alternating direction method of multipliers," *J. Sci. Comput.*, vol. 66, no. 3, pp. 889–916, 2016.

[10] P. Giselsson and S. Boyd, "Linear convergence and metric selection for Douglas–Rachford splitting and ADMM," *IEEE Trans. Autom. Control*, vol. 62, no. 2, pp. 532–544, Feb. 2017.

[11] P. L. Combettes, "Solving monotone inclusions via compositions of nonexpansive averaged operators," *Optimization*, vol. 53, nos. 5–6, pp. 475–504, 2004.

[12] R. Nishihara, L. Lessard, B. Recht, A. Packard, and M. Jordan, "A general analysis of the convergence of ADMM," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 343–352.

[13] Y. Drori and M. Teboulle, "Performance of first-order methods for smooth convex minimization: A novel approach," *Math. Program.*, vol. 145, nos. 1–2, pp. 451–482, 2014.

[14] L. Lessard, B. Recht, and A. Packard, "Analysis and design of optimization algorithms via integral quadratic constraints," *SIAM J. Optim.*, vol. 26, no. 1, pp. 57–95, 2016.

[15] M. Fazlyab, A. Ribeiro, M. Morari, and V. M. Preciado, "Analysis of optimization algorithms via integral quadratic constraints: Nonstrongly convex problems," *SIAM J. Optim.*, vol. 28, no. 3, pp. 2654–2689, 2018.

[16] B. Hu and L. Lessard, "Dissipativity theory for Nesterov's accelerated method," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1549–1557.

[17] 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 Syst. Lett.*, vol. 2, no. 1, pp. 49–54, Jan. 2018.

[18] 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.

[19] M. Fazlyab, M. Morari, and V. M. Preciado, "Design of first-order optimization algorithms via sum-of-squares programming," in *Proc. IEEE Conf. Decis. Control (CDC)*, 2018, pp. 4445–4452.

[20] G. França and J. Bento, "An explicit rate bound for over-relaxed ADMM," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2016, pp. 2104–2108.

[21] G. França and J. Bento, "Tuning over-relaxed ADMM," in *Proc. Optim. Optim. Workshop NIPS*, 2016.

[22] E. Ghadimi, A. Teixeira, I. Shames, and M. Johansson, "Optimal parameter selection for the alternating direction method of multipliers (ADMM): Quadratic problems," *IEEE Trans. Autom. Control*, vol. 60, no. 3, pp. 644–658, Mar. 2015.

[23] P. Giselsson and S. Boyd, "Diagonal scaling in Douglas–Rachford splitting and ADMM," in *Proc. IEEE 53rd Annu. Conf. Decis. Control (CDC)*, 2014, pp. 5033–5039.

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

[25] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, vol. 87. New York, NY, USA: Springer, 2013.

[26] W. M. Haddad and V. Chellaboina, *Nonlinear Dynamical Systems and Control: A Lyapunov-Based Approach*. Princeton, NJ, USA: Princeton Univ. Press, 2011.

[27] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1990.

[28] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.