Global exponential stability of the Douglas-Rachford splitting dynamics*

Sepideh Hassan-Moghaddam and Mihailo R. Jovanović

Ming Hsieh Department of Electrical and Computer Engineering,
University of Southern California, Los Angeles, CA 90089;
(e-mails: hassanmo@usc.edu, mihailo@usc.edu)

Abstract: Many modern optimization problems admit a composite form in which the objective function is given by the sum of a smooth term and a nonsmooth regularizer. Such problems can be solved via proximal methods and their variants, including the Douglas-Rachford (DR) splitting algorithm. In this paper, we view the DR splitting flow as a dynamical system and leverage techniques from control theory to study its global stability properties. In particular, for problems with strongly convex objective functions, we utilize the theory of integral quadratic constraints to prove global exponential stability of the ordinary differential equation that governs the evolution of the DR splitting flow. In our analysis, we use the fact that this algorithm can be interpreted as a variable-metric gradient method on the DR envelope and exploit structural properties of nonlinear terms that arise from composition of reflected proximal operators.

Keywords: Control for optimization, Douglas-Rachford splitting, global exponential stability, integral quadratic constraints, nonlinear dynamics, nonsmooth optimization, proximal algorithms.

1. INTRODUCTION

Examining optimization algorithms as continuous-time dynamical systems has been an active topic since the seminal work of Arrow, Hurwicz, and Uzawa (Arrow et al., 1958). This viewpoint can provide important insight into performance of optimization algorithms and, during the last decade, it has been advanced and extended to a broad class of problems including convergence analysis of primal-dual (Feijer and Paganini, 2010; Wang and Elia, 2011; Cherukuri et al., 2016, 2018; Dhirag et al., 2019; Li and Li, 2018) and accelerated (Su et al., 2016; Wibisono et al., 2016; Franca et al., 2018; Shi et al., 2018; Muehlebach and Jordan, 2019; Poveda and Li, 2019) first-order methods. Furthermore, establishing the connection between theory of ordinary differential equations (ODEs) and numerical optimization algorithms has been a topic of many studies, including Schropp and Singer (2000); for recent efforts, see Wibisono et al. (2016); Zhang et al. (2018).

Most algorithms can be viewed as a feedback interconnection of linear dynamical systems with nonlinearities that posses certain structural properties. This system-theoretic interpretation was exploited in Lessard et al. (2016) and further advanced in a number of recent papers (Dhingra et al., 2019; Hu et al., 2017; Hu and Lessard, 2017; Fazlyab et al., 2018; Hatanaka et al., 2018; Hassan-Moghaddam and Jovanović, 2018a,b; Seidman et al., 2019). The key idea is to exploit structural features of linear and nonlinear terms and utilize theory and techniques from stability analysis of nonlinear dynamical systems to study properties of optimization algorithms. This approach provides new methods for studying not only convergence rate but also robustness of optimization routines (Mohammadi et al., 2018, 2019a,b; Michalowsky et al., 2019) and can lead to new classes of algorithms that strike a desired tradeoff between the speed and robustness.

In this paper, we utilize techniques from control theory to establish global stability properties of an ordinary differential equation that describes the Douglas-Rachford (DR) splitting flows. This algorithm provide an effective tool for solving nonsmooth convex optimization problems in which the objective function is given by a sum of a differentiable term and a nondifferentiable regularizer. For strongly convex problems, we exploit the fact that the DR splitting algorithm (Douglas and Rachford, 1956) can be interpreted as a variable-metric gradient method on DR envelope (Patrinos et al., 2014) and prove global exponential stability by utilizing the theory of integral quadratic constraints (IQCs).

The paper is structured as follows. In Section 2, we formulate the nonsmooth composite optimization problem and provide background material. In Section 3, we introduce a continuous-time gradient flow dynamics based on the celebrated DR splitting algorithm and utilize the theory of IQCs to prove global exponential stability for strongly convex problems. We conclude the paper in Section 4.

2. PROBLEM FORMULATION AND BACKGROUND

We consider a composite optimization problem,

$$\minimize_{x} f(x) + g(x)$$

(1)

where $x \in \mathbb{R}^n$ is the optimization variable, $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function with a Lipschitz continuous gradient, and $g : \mathbb{R}^m \to \mathbb{R}$ is a nondifferentiable convex function. Such optimization problems arise in a
number of different applications and different first- and second-order algorithms can be employed to solve them. We are interested in studying global stability properties of the method based on Douglas-Rachford splitting dynamics. In the rest of this section, we provide background material that we utilize in the paper.

2.1 Proximal operators and the associated envelopes

The proximal operator of a proper, closed, and convex function $g$ is defined as

$$
\text{prox}_{\mu g}(v) := \underset{z}{\text{argmin}} \left( g(z) + \frac{1}{2\mu} \|z - v\|_2^2 \right).
$$

The value function of this optimization problem determines the associated Moreau envelope,

$$
M_{\mu g}(v) := g(\text{prox}_{\mu g}(v)) + \frac{1}{2\mu} \|\text{prox}_{\mu g}(v) - v\|_2^2,
$$

which is a continuously differentiable function even when $g$ is not (Parikh and Boyd, 2013),

$$
\nabla M_{\mu g}(v) = \frac{1}{\mu} (v - \text{prox}_{\mu g}(v)).
$$

Combining the last two expressions yields,

$$
M_{\mu g}(v) = g(\text{prox}_{\mu g}(v)) + \frac{\mu}{2} \|\nabla M_{\mu g}(v)\|_2^2.
$$

The Moreau envelope of $g$ can be used to introduce the forward-backward (FB) envelope (Patrinos et al., 2014; Stella et al., 2017; Themelis et al., 2018) of the composite function

$$
F(x) := f(x) + g(x).
$$

The FB envelope is determined by the value function of the problem

$$
\underset{v}{\text{minimize}} \ J(x, v) \quad (2a)
$$

where $J$ approximates $F$ via a quadratic expansion of the function $f$ around $x$,

$$
J(x, v) := f(x) + \langle \nabla f(x), v - x \rangle + \frac{1}{2\mu} \|v - x\|_2^2 + g(v) + \frac{1}{2\mu} \|v - (x - \mu \nabla f(x))\|_2^2 + f(x) - \frac{\mu}{2} \|\nabla f(x)\|_2^2. \quad (2b)
$$

The optimal solution of (2) is determined by $v_{\mu}^*(x) = \text{prox}_{\mu g}(x - \mu \nabla f(x))$ and it can be used to obtain the FB envelope of the function $F$,

$$
F_{\mu}(x) := J(x, v_{\mu}^*(x)) = J(x, \text{prox}_{\mu g}(x - \mu \nabla f(x)))
= f(x) - \mu \langle \nabla f(x), G_{\mu}(x) \rangle + \frac{\mu}{2} \|G_{\mu}(x)\|_2^2 + g(\text{prox}_{\mu g}(x - \mu \nabla f(x))), \quad (3)
$$

where $G_{\mu}$ is the generalized gradient map,

$$
G_{\mu}(x) := \frac{1}{\mu} (x - \text{prox}_{\mu g}(x - \mu \nabla f(x))). \quad (4)
$$

Alternatively, the FB envelope $F_{\mu}$ can be also expressed as

$$
F_{\mu}(x) = f(x) + M_{\mu g}(x - \mu \nabla f(x)) - \frac{\mu}{2} \|\nabla f(x)\|_2^2. \quad (5)
$$

Moreover, when $f$ is twice continuously differentiable, $F_{\mu}$ is continuously differentiable and its gradient is determined by (Patrinos et al., 2014),

$$
\nabla F_{\mu}(x) = (I - \mu \nabla^2 f(x)) G_{\mu}(x). \quad (6)
$$

The Douglas-Rachford (DR) envelope is another useful object that is obtained by evaluating the FB envelope at $\text{prox}_{\mu f}(x)$ (Themelis and Patrinos, 2017),

$$
F_{\mu}^{DR}(x) := F_{\mu}(\text{prox}_{\mu f}(x)). \quad (7)
$$

Alternatively, the DR envelope can be expressed as

$$
F_{\mu}^{DR}(x) = M_{\mu g}(x - 2\mu \nabla M_{\mu f}(x)) + M_{\mu f}(x) - \mu \|\nabla M_{\mu f}(x)\|_2^2. \quad (8)
$$

From the definition of the proximal operator of the continuously differentiable function $f$, we have

$$
\mu \nabla f(\text{prox}_{\mu f}(x)) + \text{prox}_{\mu f}(x) - x = 0 \quad (9)
$$

and, thus,

$$
\nabla M_{\mu f}(x) = \nabla f(\text{prox}_{\mu f}(x)). \quad (10)
$$

Equality (7) follows from substituting the expression for $\nabla M_{\mu f}(x)$ into (8), using equation (9), and leveraging the properties of the Moreau envelope,

$$
F_{\mu}^{DR}(x) = M_{\mu g}(x - 2\mu \nabla f(\text{prox}_{\mu f}(x))) + f(\text{prox}_{\mu f}(x)) + \frac{1}{2\mu} \|\nabla f(\text{prox}_{\mu f}(x))\|_2^2
= M_{\mu g}(\text{prox}_{\mu f}(x)) - \mu \nabla f(\text{prox}_{\mu f}(x)) + f(\text{prox}_{\mu f}(x)) - \frac{\mu}{2} \|\nabla f(\text{prox}_{\mu f}(x))\|_2^2
= F_{\mu}(\text{prox}_{\mu f}(x)).
$$

If $f$ is twice continuously differentiable with $\nabla^2 f(x) \preceq L_f I$ for all $x$, the DR envelope is continuously differentiable and its gradient is given by (Themelis and Patrinos, 2017)

$$
\nabla F_{\mu}^{DR}(x) = \frac{1}{\mu} (2\mu \nabla\text{prox}_{\mu f}(x) - I) G_{\mu}^{DR}(x) \quad (11)
$$

where

$$
\nabla\text{prox}_{\mu f}(x) = (I + \mu \nabla^2 f(\text{prox}_{\mu f}(x)))^{-1}
$$

and

$$
G_{\mu}^{DR}(x) := \text{prox}_{\mu f}(x) - \text{prox}_{\mu g}(2\text{prox}_{\mu f}(x) - x). \quad (12)
$$

2.2 Strong convexity and Lipschitz continuity

The function $f$ is strongly convex with parameter $m_f$ if for any $x$ and $\hat{x}$,

$$
f(\hat{x}) \geq f(x) + \langle \nabla f(x), \hat{x} - x \rangle + \frac{m_f}{2} \|\hat{x} - x\|_2^2 \quad \text{and equivalently,}
$$

$$
\|\nabla f(\hat{x}) - \nabla f(x)\|_2 \geq m_f \|\hat{x} - x\|_2. \quad (13)
$$

The gradient of a continuously-differentiable function $f$ is Lipschitz continuous with parameter $L_f$ if for any $x$ and $\hat{x}$,

$$
f(\hat{x}) \leq f(x) + \langle \nabla f(x), \hat{x} - x \rangle + \frac{L_f}{2} \|\hat{x} - x\|_2^2 \quad \text{and equivalently,}
$$

$$
\|\nabla f(x) - \nabla f(\hat{x})\|_2 \leq L_f \|x - \hat{x}\|_2. \quad (14)
$$

Moreover, if an $m_f$-strongly convex function $f$ has an $L_f$-Lipschitz continuous gradient, the following inequality holds for any $x$ and $\hat{x}$,
\[
\langle \nabla f(x) - \nabla f(\hat{x}), x - \hat{x} \rangle \geq \frac{m_f L_f}{m_f + L_f} \|x - \hat{x}\|^2 + \frac{1}{m_f + L_f} \|\nabla f(x) - \nabla f(\hat{x})\|^2. 
\]

Furthermore, the subgradient \( \partial g \) of a nondifferentiable function \( g \) is defined as the set of points \( z \in \partial g(x) \) that for any \( x \) and \( \hat{x} \), satisfy,
\[
g(\hat{x}) \geq g(x) + z^T(\hat{x} - x). 
\]

### 3. Global exponential stability

We next introduce a continuous-time gradient flow dynamics based on the celebrated Douglas-Rachford splitting algorithm (Douglas and Rachford, 1956) and establish global exponential stability for strongly convex \( f \).

#### 3.1 Douglas-Rachford splitting dynamics

The optimality condition for non-smooth composite optimization problem (1) is given by
\[
0 \in \nabla f(x) + \partial g(x).
\]

Multiplication by \( \mu \) and addition/subtraction of \( x \) yields,
\[
0 \in [I + \mu \nabla f](x) + \mu \partial g(x) - x.
\]

Since the proximal operator associated with \( \mu f \) is determined by the resolvent operator of \( \mu \nabla f \), we have
\[
x = (I + \mu \nabla f)^{-1}(x - \mu \partial g(x)) = \text{prox}_{\mu f}(x - \mu \partial g(x)).
\]

Introducing a new variable \( z := x - \mu \partial g(x) \) allows us to bring the optimality condition into the following form
\[
x = \text{prox}_{\mu f}(z)
\]
or, equivalently,
\[
\mu \partial g(x) = \text{prox}_{\mu f}(z) - z.
\]

Now, adding \( x \) to both sides of this equation yields
\[
(I + \mu \partial g)(x) = x + \text{prox}_{\mu f}(z) - z = 2 \text{prox}_{\mu f}(z) - z
\]

which leads to,
\[
x^* = \text{prox}_{\mu g}(2 \text{prox}_{\mu f}(z^*) - z^*) = \text{prox}_{\mu f}(z^*). 
\]

Furthermore, the reflected proximal operators (Giselsson and Boyd, 2017) of the functions \( f \) and \( g \),
\[
R_{\mu f}(z) := [2 \text{prox}_{\mu f} - I](z),
\]
\[
R_{\mu g}(z) := [2 \text{prox}_{\mu g} - I](z)
\]
can be used to write optimality condition (17a) as
\[
[R_{\mu g}R_{\mu f}](z^*) = z^*. 
\]

This follows from (17a) and
\[
[R_{\mu g}R_{\mu f}](z) = z + [2 \text{prox}_{\mu g}(2 \text{prox}_{\mu f}(z) - z) - \text{prox}_{\mu f}(z)].
\]

Building on the optimality conditions, the DR splitting algorithm consists of the following iterative steps,
\[
x^{k+1} = \text{prox}_{\mu f}(z^k)
\]
\[
y^{k+1} = \text{prox}_{\mu g}(2z^k - z^k)
\]
\[
z^{k+1} = z^k + 2\alpha(y^k - x^k).
\]

Under standard convexity assumptions (Eckstein and Bertsekas, 1992), the DR splitting algorithm converges for \( \alpha \in (0, 1) \). Combining all the steps in (18) yields the first-order recurrence,
\[
z^{k+1} = z^k + 2\alpha \left( \text{prox}_{\mu g}(2z^k - z^k) - x^k \right) = [(1 - \alpha)I + \alpha R_{\mu g}R_{\mu f}](z^k)
\]

where \( z^k \) converges to the fixed point of the operator \( R_{\mu f}R_{\mu g} \) and \( x^k \) converges to the optimal solution of (1).

Optimality conditions (17) can be used to obtain the continuous-time gradient flow dynamics to compute \( z^k \),
\[
\dot{z} = -\mu (2\text{prox}_{\mu f}(z) - I)^{-1} \nabla F_{\mu}^{DR}(z)
\]

where the inverse is well-defined for \( \mu \in (0, 1/L_f) \).

Thus, the DR splitting algorithm can be interpreted as a variable-metered gradient method on the DR envelope
\[
F_{\mu}^{DR}(z) \text{ (Patrinos et al., 2014)}.
\]

#### 3.2 Global exponential stability via theory of IQCs

The continuous-time dynamics (20) can be also seen as a feedback interconnection of an LTI system
\[
\dot{z} = Az + Bu \\
\xi = Cz 
\]

with a nonlinear term,
\[
u(z) := [R_{\mu g}R_{\mu f}](z).
\]

Herein, the matrices in (21a) are given by
\[
A = -I, B = I, C = I 
\]

and the corresponding transfer function is
\[
H(s) = C(sI - A)^{-1}B = \frac{1}{s + 1} I.
\]

We first characterize properties of nonlinearity \( u \) in (21b) and then utilize the theory of integral quadratic constraints to establish the conditions for global exponential stability of (21) under the following assumption.

**Assumption 1.** Let the differentiable part \( f \) of the objective function in (1) be strongly convex with parameter \( m_f \), let \( \nabla f \) be Lipschitz continuous with parameter \( L_f \), and let the regularization function \( g \) be proper, lower semicontinuous, and convex.

**Lemma 1.** Let Assumption 1 hold and let \( \mu \in (0, 2/L_f) \). Then, the operator \( R_{\mu f} \) is \( \sigma \)-contractive,
\[
\|R_{\mu f} (x) - R_{\mu f}(y)\|_2 \leq \sigma \|x - y\|_2
\]

where \( \sigma \) is given by
\[
\sigma = \max \{|1 - \mu m_f|, |1 - \mu L_f|\} < 1. 
\]

**Proof.** Given \( z_x := \text{prox}_{\mu f}(x) \) and \( z_y := \text{prox}_{\mu f}(y) \), \( x \) and \( y \) can be computed as follows
\[
x = z_x + \mu \nabla f(z_x), \ y = z_y + \mu \nabla f(z_y).
\]

Thus,
where \( \rho \) holds, where \( \sigma < 1 \) if and only if,
\[
-1 < 1 - \mu L_f < 1 \quad \text{and} \quad -1 < 1 - \mu m_f < 1.
\]
Since \( m_f \leq L_f \), these conditions hold for \( \mu \in (0, 2/L_f) \), which completes the proof.

**Lemma 2.** Let Assumption 1 hold and let \( \mu \in (0, 2/L_f) \). Then, the operator \( R_{\mu g} \) is firmly non-expansive.

**Proof.**
\[
\|R_{\mu g}(x) - R_{\mu g}(y)\|^2 = 4\|\text{prox}_{\mu g}(x) - \text{prox}_{\mu g}(y)\|^2 + \|x - y\|^2 - 4(x - y, \text{prox}_{\mu g}(x) - \text{prox}_{\mu g}(y)) \leq \|x - y\|^2.
\]

**Remark 2.** Since \( R_{\mu g} \) is firmly non-expansive and \( R_{\mu f} \) is \( \sigma \)-contractive, the composite operator \( R_{\mu g} R_{\mu f} \) is also \( \sigma \)-contractive. This follows the preceding lemma.

**Lemma 3.** Let Assumption 1 hold. Then, for any \( z \in \mathbb{R}^n \), \( \tilde{z} \in \mathbb{R}^n \), \( u := [R_{\mu g} R_{\mu f}](z) \), and \( \tilde{u} := [R_{\mu g} R_{\mu f}](\tilde{z}) \), the pointwise quadratic inequality
\[
\begin{bmatrix}
z - \tilde{z} \\
u - \tilde{u}
\end{bmatrix}^T \begin{bmatrix}
\sigma^2 I & 0 \\
0 & -I
\end{bmatrix} \begin{bmatrix}
z - \tilde{z} \\
u - \tilde{u}
\end{bmatrix} \geq 0
\]
holds, where
\[
\sigma = \max \{|1 - \mu m_f|, |1 - \mu L_f|\}.
\]

We next employ the KYP lemma in the frequency domain (Rantzer, 1996)
\[
\left[ H_{\rho}(j\omega) \right]^* \Pi \left[ H_{\rho}(j\omega) \right] < 0, \quad \forall \omega \in \mathbb{R}
\]
where \( \Pi \) is given by (23), \( \omega \) is the temporal frequency, and
\[
H_{\rho}(j\omega) := C(j\omega I - (A + \rho I))^{-1} B = \frac{1}{j\omega + 1 - \rho} I
\]
to establish the global exponential stability of (21).

**Theorem 4.** Let Assumption 1 hold. Then, the DR splitting dynamics (21) are globally exponentially stable, i.e., there is \( c > 0 \) and \( \rho (0, 1 - \sigma) \) such that,
\[
\|z(t) - z^*\| \leq c e^{-\rho t}\|z(0) - z^*\|, \quad \forall t \geq 0.
\]

**Proof.** The KYP lemma implies the global exponential stability of (21) if there exists \( \rho \in (0, 1) \) such that
\[
\sigma^2 H_{\rho}(j\omega) H_{\rho}(j\omega) - I \prec 0, \quad \forall \omega \in \mathbb{R},
\]
where \( H_{\rho}(j\omega) \) is given by (26). Inequality (27) is satisfied if
\[
\sigma^2 - (1 - \rho)^2 - \omega^2 < 0, \quad \forall \omega \in \mathbb{R}
\]
which proves \( \rho < 1 - \sigma \).

### 3.3 Douglas-Rachford splitting on the dual problem

The DR splitting algorithm cannot be used to directly solve a problem with a more general linear constraint,


