Automatica 123 (2021) 109311

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

Proximal gradient flow and Douglas–Rachford splitting dynamics: Global exponential stability via integral quadratic constraints ☆

Sepideh Hassan-Moghaddam, Mihailo R. Jovanović *

Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA 90089, United States of America

ARTICLE INFO

Article history:
Received 22 August 2019
Received in revised form 24 June 2020
Accepted 25 August 2020
Available online xxxx

Keywords:
Control for optimization
Forward–backward envelope
Douglas–Rachford splitting
Global exponential stability
Integral quadratic constraints (IQCs)
Nonlinear dynamics
Nonsmooth optimization
Proximal algorithms
Primal–dual methods
Proximal augmented lagrangian

ABSTRACT

Many large-scale and distributed optimization problems can be brought into 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 a proximal gradient method and its variants, thereby generalizing gradient descent to a nonsmooth setup. In this paper, we view proximal algorithms as dynamical systems and leverage techniques from control theory to study their global properties. In particular, for problems with strongly convex objective functions, we utilize the theory of integral quadratic constraints to prove the global exponential stability of the equilibrium points of the differential equations that govern the evolution of proximal gradient and Douglas–Rachford splitting flows. In our analysis, we use the fact that these algorithms can be interpreted as variable-metric gradient methods on the suitable envelopes and exploit structural properties of the nonlinear terms that arise from the gradient of the smooth part of the objective function and the proximal operator associated with the nonsmooth regularizer. We also demonstrate that these envelopes can be obtained from the augmented Lagrangian associated with the original nonsmooth problem and establish conditions for global exponential convergence even in the absence of strong convexity.

1. Introduction

Structured optimal control and estimation problems typically lead to optimization of objective functions that consist of a sum of a smooth term and a nonsmooth regularizer. Such problems are of increasing importance in applications and it is thus necessary to develop efficient algorithms for distributed and embedded nonsmooth composite optimization (Latafat, Freris, & Patrinos, 2019; Latafat, Stella, & Patrinos, 2016; Nedić & Ozdaglar, 2009; Wang & Elia, 2011). The lack of differentiability in the objective function precludes the use of standard descent methods from smooth optimization. Proximal gradient method (Beck & Teboulle, 2009; Parikh & Boyd, 2013) generalizes gradient descent to nonsmooth context and provides a powerful tool for solving problems in which the nonsmooth term is separable over the optimization variable.

Examining optimization algorithms as continuous-time dynamical systems has been an active topic since the seminal work of Arrow, Hurwicz, and Uzawa (Arrow, Hurwicz, & Uzawa, 1958). This viewpoint can provide important insight into performance of optimization algorithms and streamline their convergence analysis. During the last decade, it has been advanced and extended to a broad class of problems including convergence analysis of primal–dual (Cherukuri, Mallada, & Cortés, 2016; Cherukuri, Mallada, Low, & Cortes, 2018; Dhingra, Khong, & Jovanović, 2019; Feijer & Paganini, 2010; Qu & Li, 2018; Wang & Elia, 2011) and accelerated (França, Robinson, & Vidal, 2018; Muehlebach & Jordan, 2019; Poveda & Li, 2019; Shi, Du, Jordan, & Su, 2018; Su, Boyd, & Candes, 2016; Wibisono, Wilson, & Jordan, 2016) 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 (Brown & Bartholomew-Biggs, 1989; Schropp & Singer, 2000); for recent efforts, see Wibisono et al. (2016) and Zhang, Mokhtari, Sra, and Jadbabaie (2018).

Optimization algorithms can be viewed as a feedback interconnection of linear dynamical systems with nonlinearities that possess certain structural properties. This system-theoretic interpretation was exploited in Lessard, Recht, and Packard (2016) and further advanced in recent papers (Dhingra et al., 2019; Ding, Hu, Dhingra, & Jovanović, 2018; Fazlyab, Ribeiro, Morari, &

☆ Supported in part by the National Science Foundation, USA under awards ECCS-1708906 and ECCS-1809833. The material in this paper was presented at the 57th IEEE Conference on Decision and Control, December 17–19, 2018, Miami Beach, Florida, USA. This paper was recommended for publication in revised form by Associate Editor Andrey V. Savkin under the direction of Editor Ian R. Petersen.

* Corresponding author.
E-mail addresses: hassanmo@usc.edu (S. Hassan-Moghaddam), mihailo@usc.edu (M.R. Jovanović).

https://doi.org/10.1016/j.automatica.2020.109311
0005-1098/

Preciado, 2018; Hassan-Moghaddam & Jovanović, 2018a, 2018b; Hu & Lessard, 2017; Hu, Seiler, & Rantzer, 2017; Seidman, Fazlyab, Preciado, & Pappas, 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 (Michalowsky, Scherer, & Ebenbauer, 2019; Mohammadi, Razaviyayn, & Jovanović, 2018, 2019, 2020) 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 properties of proximal gradient flow and Douglas–Rachford (DR) splitting dynamics. These algorithms 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. When the smooth term is strongly convex with a Lipschitz continuous gradient, we prove the global exponential stability of both the proximal gradient flow and the DR splitting dynamics by utilizing the theory of IQCs (Megretski & Rantzer, 1997). We also generalize the Polyak–Lojasiewicz (PL) (Polyak, 1963) condition to nonsmooth problems and show global exponential convergence of the forward–backward (FB) envelope (Patrinos, Stella, & Bemporad, 2014; Stella, Themelis, & Patrinos, 2017; Themelis, Stella, & Patrinos, 2018) even in the absence of strong convexity.

Although there are related approaches for studying optimization algorithms from a control-theoretic perspective, to the best of our knowledge, we are the first to introduce the continuous forms of proximal gradient and DR splitting algorithms. We use simple proofs to establish their global stability properties and provide explicit bounds on convergence rates. Furthermore, standard forms of these algorithms are obtained via explicit forward Euler discretization of continuous-time dynamics.

The paper is structured as follows. In Section 2, we formulate the nonsmooth composite optimization problem and provide background material. In Section 3, we establish the global exponential stability of the proximal gradient flow dynamics for a problem with strongly convex objective function. Moreover, by exploiting the problem structure, we demonstrate the global exponential convergence of the forward–backward envelope even in the absence of strong convexity. In Section 4, we introduce a continuous-time gradient flow dynamics based on the celebrated Douglas–Rachford splitting algorithm and utilize the theory of IQCs to prove global exponential stability for strongly convex problems. We offer concluding remarks in Section 5.

2. Problem formulation and background

We consider a composite optimization problem,

$$\underset{x}{\text{minimize}} \quad f(x) \, + \, g(Tx) \tag{1}$$

where $x \in \mathbb{R}^n$ is the optimization variable, $T \in \mathbb{R}^{m \times n}$ is a given matrix, $f\colon \mathbb{R}^n \to \mathbb{R}$ is a convex function with a Lipschitz continuous gradient, and $g\colon \mathbb{R}^m \to \mathbb{R}$ is a nondifferentiable convex function. Such optimization problems arise in a number of applications and depending on the structure of the functions f and g, different first- and second-order algorithms can be employed to solve them. We are interested in studying global convergence properties of methods based on proximal gradient flow algorithms. In what follows, we provide background material that we utilize in the rest of the paper.

2.1. Proximal operator and the associated envelopes

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

$$\mathbf{prox}_{\mu g}(v) \; := \; \underset{z}{\operatorname{argmin}} \left(g(z) \, + \, \frac{1}{2\mu} \, \| z \, - \, v \|_2^2 \right) \tag{2}$$

where μ is a positive parameter and v is a given vector. It is determined by the resolvent operator associated with $\mu \partial g$, $\mathbf{prox}_{\mu g} \; := \; (I \, + \, \mu \partial g)^{-1}$, and is a single-valued firmly nonexpansive mapping (Parikh & Boyd, 2013), i.e., for any u and v,

$$\|\mathbf{prox}_{\mu g}(u) \, - \, \mathbf{prox}_{\mu g}(v)\|_2^2 \; \leq \; \left\langle u \, - \, v, \mathbf{prox}_{\mu g}(u) \, - \, \mathbf{prox}_{\mu g}(v) \right\rangle.$$

The value function of the optimization problem (2) determines the associated Moreau envelope,

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

which is a continuously differentiable function even when g is not (Parikh & Boyd, 2013), with $\mu \nabla M_{\mu g}(v) = v - \mathbf{prox}_{\mu g}(v)$.

By introducing an auxiliary optimization variable z, problem (1) can be rewritten as follows,

$$\begin{array}{ll} \underset{x,\, z}{\text{minimize}} & f(x) \, + \, g(z) \\ \text{subject to} & Tx \, - \, z \, = \, 0 \end{array} \tag{3}$$

and the associated augmented Lagrangian is given by,

$$\mathcal{L}_\mu(x, z; y) \; := \; f(x) \, + \, g(z) \, + \, \langle y, Tx - z \rangle \, + \, \tfrac{1}{2\mu} \, \|Tx - z\|_2^2.$$

The completion of squares yields,

$$\mathcal{L}_\mu \; = f(x) \, + \, g(z) \, + \, \tfrac{1}{2\mu} \, \|z \, - \, (Tx \, + \, \mu y)\|_2^2 \, - \, \tfrac{\mu}{2} \, \|y\|_2^2$$

where y is the Lagrange multiplier. The minimizer of $\mathcal{L}_\mu$ with respect to z is

$$z^\star(x, y) \; = \; \mathbf{prox}_{\mu g}(Tx \, + \, \mu y)$$

and the evaluation of $\mathcal{L}_\mu$ along the manifold resulting from this explicit minimization yields the proximal augmented Lagrangian (Dhingra et al., 2019), $\mathcal{L}_\mu(x; y) := \mathcal{L}_\mu(x, z^\star(x, y); y)$,

$$\mathcal{L}_\mu(x; y) \; = \; f(x) \, + \, M_{\mu g}(Tx \, + \, \mu y) \, - \, \tfrac{\mu}{2} \, \|y\|_2^2. \tag{4}$$

This function is continuously differentiable with respect to both x and y and it can be used as a foundation for the development of first- and second-order primal–dual methods for nonsmooth composite optimization (Dhingra, Khong, & Jovanović, 2017; Dhingra et al., 2019). For $T \, = \, I$, the forward–backward envelope (Patrinos et al., 2014; Stella et al., 2017; Themelis et al., 2018) is obtained by restricting the proximal augmented Lagrangian $\mathcal{L}_\mu(x; y)$ along the manifold $y^\star(x) \, = \, -\nabla f(x)$ resulting from the KKT optimality conditions,

$$\begin{array}{rcl} F_\mu(x) & := & \mathcal{L}_\mu(x; y^\star(x)) \; = \; \mathcal{L}_\mu(x; y = -\nabla f(x)) \\ & = & f(x) \, + \, M_{\mu g}(x \, - \, \mu \nabla f(x)) \, - \, \tfrac{\mu}{2} \, \|\nabla f(x)\|_2^2. \end{array}$$

2.2. Strong convexity and Lipschitz continuity

The function f is m_f-strongly convex if

$$f(\hat{x}) \; \geq \; f(x) \, + \, \left\langle \nabla f(x), \hat{x} \, - \, x \right\rangle \, + \, \frac{m_f}{2} \, \|\hat{x} \, - \, x\|_2^2$$

and its gradient is L_f-Lipschitz continuous if

$$f(\hat{x}) \; \leq \; f(x) \, + \, \left\langle \nabla f(x), \hat{x} \, - \, x \right\rangle \, + \, \frac{L_f}{2} \, \|\hat{x} \, - \, x\|_2^2$$

for any x and $\hat{x}$. When both properties hold we have

$$m_f \, \|x \, - \, \hat{x}\|_2 \; \leq \; \|\nabla f(x) - \nabla f(\hat{x})\|_2 \; \leq \; L_f \, \|x \, - \, \hat{x}\|_2. \tag{5}$$

and the following inequality is satisfied (Nesterov, 2013),

$$\left\langle \nabla f(x) - \nabla f(\hat{x}), x - \hat{x} \right\rangle \;\geq\; \frac{m_f L_f}{m_f + L_f} \|x - \hat{x}\|_2^2 + \frac{1}{m_f + L_f} \|\nabla f(x) - \nabla f(\hat{x})\|_2^2. \tag{6}$$

Furthermore, the subgradient ∂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,

$$g(\hat{x}) \geq g(x) + z^T(\hat{x} - x). \tag{7}$$

2.3. Proximal Polyak–Lojasiewicz inequality

The Polyak–Lojasiewicz (PL) condition can be used to prove linear convergence of a gradient descent even in the absence of convexity (Karimi, Nutini, & Schmidt, 2016). For an unconstrained optimization problem with a non-empty solution set and a twice differentiable objective function f with a Lipschitz continuous gradient, the PL condition is given by

$$\|\nabla f(x)\|_2^2 \;\geq\; \gamma\,(f(x) - f^\star)$$

where $\gamma > 0$ and $f^\star$ is the optimal value of f. For nonsmooth optimization problem (3) with $T = I$, the proximal PL inequality holds for $\mu \in (0, 1/L_f)$ if there exist $\gamma > 0$ such that

$$\|G_\mu(x)\|_2^2 \;\geq\; \gamma\,(F_\mu(x) - F_\mu^\star). \tag{8}$$

Here, L_f is the Lipschitz constant of ∇f, F_μ is the FB envelope, and G_μ is the generalized gradient map,

$$G_\mu(x) \;:=\; \frac{1}{\mu}(x - \mathbf{prox}_{\mu g}(x - \mu\nabla f(x))). \tag{9}$$

When f is twice continuously differentiable, the FB envelope F_μ is continuously differentiable with (Patrinos et al., 2014),

$$\nabla F_\mu(x) \;=\; (I - \mu\nabla^2 f(x))\,G_\mu(x). \tag{10}$$

3. Exponential stability of proximal algorithms

In this section, we briefly discuss the Arrow–Hurwicz–Uzawa gradient flow dynamics that can be used to solve (3) by computing the saddle points of the proximal augmented Lagrangian (Dhingra et al., 2019). We then show that the proximal gradient method in continuous time can be obtained from the proximal augmented Lagrangian method by restricting the dual variable along the manifold $y = -\nabla f(x)$. Finally, we discuss global stability properties of proximal algorithms both in the presence and in the absence of strong convexity.

Continuous differentiability of the proximal augmented Lagrangian (4) can be utilized to compute its saddle points via the Arrow–Hurwicz–Uzawa dynamic,

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} -\mu\,(\nabla f(x) + T^T \nabla M_{\mu g}(Tx + \mu y)) \\ \mu\,(\nabla M_{\mu g}(Tx + \mu y) - y) \end{bmatrix}. \tag{11}$$

As shown in Dhingra et al. (2019), the optimal primal–dual pair $(x^\star, y^\star)$ is the globally exponentially stable equilibrium point of (11) and $x^\star$ is the solution of (1) for convex problems in which the matrix TT^T is invertible and the smooth part of the objective function f is strongly convex.

For convex problems with $T = I$ in (1),

$$\underset{x}{\text{minimize}} \;\; f(x) + g(x) \tag{12}$$

the optimality condition is given by

$$0 \in \nabla f(x^\star) + \partial g(x^\star) \tag{13}$$

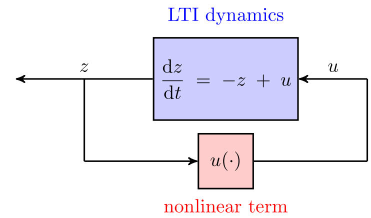

Fig. 1. Both the proximal gradient flow dynamics (15) and the DR splitting dynamics (26) can be represented via feedback interconnections of stable LTI systems with nonlinear terms that possess certain structural properties.

Multiplying by μ and adding/subtracting $x^\star$ yields,

$$0 \in [I + \mu\partial g](x^\star) + \mu\nabla f(x^\star) - x^\star.$$

Since $\mathbf{prox}_{\mu g}$ is determined by the resolvent operator associated with $\mu\partial g$ and is single-valued (Parikh & Boyd, 2013), we have

$$x^\star - \mathbf{prox}_{\mu g}(x^\star - \mu\nabla f(x^\star)) = 0. \tag{14}$$

We next demonstrate that (12) can be solved using the proximal gradient flow dynamics, $\dot{x} = -\mu\,G_\mu(x)$,

$$\begin{aligned} \dot{x} &= -\left(x - \mathbf{prox}_{\mu g}(x - \mu\nabla f(x))\right) \\ &= -\mu\left(\nabla f(x) + \nabla M_{\mu g}(x - \mu\nabla f(x))\right). \end{aligned} \tag{15}$$

Remark 1. Proximal gradient flow dynamics (15) are different from the subgradient flow dynamics associated with nonsmooth problem (12). Standard proximal gradient algorithm (Beck & Teboulle, 2009) can be obtained via explicit forward Euler discretization of (15) with the stepsize one, $x^{k+1} = \mathbf{prox}_{\mu g}(x^k - \mu\nabla f(x^k))$. This should be compared and contrasted with (Parikh & Boyd, 2013, Section 4.1.1) in which implicit backward Euler discretization of the subgradient flow dynamics associated with (12) was used. We also note that (15) can be obtained by substituting $-\nabla f(x)$ for the dual variable y in the x-update step of primal-descent dual-ascent gradient flow dynamics (11) with $T = I$.

In Sections 3.1 and 3.2 , we examine properties of system (15), first for strongly convex problems and then for the problems in which only the PL condition holds.

3.1. Strongly convex problems

We utilize the theory of integral quadratic constraints to prove global asymptotic stability of the proximal gradient flow dynamics (15) under the following assumption.

Assumption 1. Let f in (12) be m_f-strongly convex, let ∇f be L_f-Lipschitz continuous, and let the regularization function g be proper, closed, and convex.

As illustrated in Fig. 1, system (15) can be expressed as a feedback interconnection of an LTI system

$$\begin{aligned} \dot{z} &= Az + Bu, \quad \xi = Cz \\ A &= -I, \quad B = C = I \end{aligned} \tag{16a}$$

where $z := x$, with the nonlinear term,

$$u(\xi) := \mathbf{prox}_{\mu g}(\xi - \mu\nabla f(\xi)). \tag{16b}$$

Lemma 1 combines firm nonexpansiveness of $\mathbf{prox}_{\mu g}$, strong convexity of f, and Lipschitz continuity of ∇f to characterize nonlinear map (16b) by establishing a quadratic inequality for $u(\xi)$.

Lemma 1. *Let Assumption 1 hold. Then, for any $\xi \in \mathbb{R}^n$, $\hat{\xi} \in \mathbb{R}^n$, $u := \mathbf{prox}_{\mu g}(\xi - \mu \nabla f(\xi))$, and $\hat{u} := \mathbf{prox}_{\mu g}(\hat{\xi} - \mu \nabla f(\hat{\xi}))$, the pointwise quadratic inequality*

$$\begin{bmatrix} \xi - \hat{\xi} \\ u - \hat{u} \end{bmatrix}^T \underbrace{\begin{bmatrix} \sigma^2 I & 0 \\ 0 & -I \end{bmatrix}}_{\Pi} \begin{bmatrix} \xi - \hat{\xi} \\ u - \hat{u} \end{bmatrix} \geq 0 \tag{17a}$$

holds, where

$$\sigma = \max\left\{ |1 - \mu m_f|, |1 - \mu L_f| \right\}. \tag{17b}$$

Moreover, the nonlinear function $u(\xi) := \mathbf{prox}_{\mu g}(\xi - \mu \nabla f(\xi))$ is a contraction for $\mu \in (0, 2/L_f)$.

Proof. Since $\mathbf{prox}_{\mu g}$ is firmly nonexpansive (Parikh & Boyd, 2013), it is also Lipschitz continuous with parameter 1, i.e.,

$$\|u - \hat{u}\|_2^2 \leq \|(\xi - \mu \nabla f(\xi)) - (\hat{\xi} - \mu \nabla f(\hat{\xi}))\|_2^2. \tag{18}$$

Expanding the right-hand-side of (18) yields,

$$\begin{aligned} \|u - \hat{u}\|_2^2 \;\; \leq \;\; & \|\xi - \hat{\xi}\|_2^2 + \mu^2 \|\nabla f(\xi) - \nabla f(\hat{\xi})\|_2^2 - \\ & 2\mu \left\langle \xi - \hat{\xi}, \nabla f(\xi) - \nabla f(\hat{\xi}) \right\rangle \end{aligned}$$

and utilizing inequality (6) for an m_f-strongly convex function f with an L_f-Lipschitz continuous gradient, the last inequality can be further simplified to obtain,

$$\begin{aligned} \|u - \hat{u}\|_2^2 \;\; \leq \;\; & (1 - \frac{2\,\mu\, m_f L_f}{L_f + m_f}) \|\xi - \hat{\xi}\|_2^2 + \\ & (\mu^2 - \frac{2\,\mu}{L_f + m_f}) \|\nabla f(\xi) - \nabla f(\hat{\xi})\|_2^2. \end{aligned} \tag{19}$$

Depending on the sign of $\mu - 2/(L_f + m_f)$ either lower or upper bound in (5) can be used to upper bound the second term on the right-hand-side of (19), thereby yielding

$$\|u - \hat{u}\|_2^2 \leq \max\left\{ (1 - \mu L_f)^2, (1 - \mu m_f)^2 \right\} \|\xi - \hat{\xi}\|_2^2. \tag{20}$$

Thus, for σ given by (17b) the nonlinear function $u(\xi)$ is a contraction if and only if $-1 < 1 - \mu L_f < 1$ and $-1 < 1 - \mu m_f < 1$. Since $m_f \leq L_f$, these conditions hold for $\mu \in (0, 2/L_f)$ which completes the proof.

We next employ (Hu & Seiler, 2016, Theorem 3) to prove the global exponential stability of the equilibrium point $z^\star$ of (16) with the rate $\rho > 0$ by verifying the existence of a positive definite matrix P such that,

$$\begin{bmatrix} A_\rho^T P + P A_\rho & PB \\ B^T P & 0 \end{bmatrix} + \begin{bmatrix} C^T & 0 \\ 0 & I \end{bmatrix} \Pi \begin{bmatrix} C & 0 \\ 0 & I \end{bmatrix} \preceq 0, \tag{21}$$

where $A_\rho := A + \rho I$ and Π is given by (17a).

Theorem 2. *Let Assumption 1 hold and let $\mu \in (0, 2/L_f)$. Then, the equilibrium point $z^\star$ of the proximal gradient flow dynamics (16) is globally ρ-exponentially stable, i.e., there is $c > 0$ and $\rho \in (0, 1 - \sigma]$ such that,*

$$\|z(t) - z^\star\|_2 \leq c\, \mathrm{e}^{-\rho t} \|z(0) - z^\star\|_2, \quad \forall\, t \geq 0$$

where σ is given by (17b). Moreover, $x^\star = z^\star$ is the optimal solution of (12).

Proof. Substituting Π given by (17a) into (21) implies that the condition (21) holds if there exists a positive scalar p such that

$$\begin{bmatrix} 2(1-\rho)p - \sigma^2 & -p \\ -p & 1 \end{bmatrix} \succeq 0 \tag{22}$$

where the block-diagonal structure of A, B, C, and Π allows us to choose $P = pI$ without loss of generality. Condition (22) is satisfied if there is $p > 0$ such that

$$p^2 - 2(1 - \rho)p + \sigma^2 \leq 0 \tag{23}$$

where $\rho < 1$ guarantees positivity of the first element on the main diagonal of the matrix in (22). For $\mu \in (0, 2/L_f)$, Lemma 1 implies $\sigma < 1$ and $\rho \leq 1 - \sigma$ is required for the existence of $p > 0$ such that (23) holds. Thus, $z^\star$ is globally exponentially stable with the rate $\rho \leq 1 - \sigma$. The result follows because the equilibrium point $z^\star = x^\star$ of (16) satisfies the optimality condition (14) for optimization problem (12).

Remark 2. For $\mu = 2/(L_f + m_f)$, the second term on the right-hand-side in (19) disappears and σ is given by $\sigma = (L_f - m_f)/(L_f + m_f) = (\kappa - 1)/(\kappa + 1)$ where $\kappa := L_f/m_f$ is the condition number of the function f and ρ is upper bounded by $2/(\kappa + 1)$. In fact, this is the best achievable convergence rate for system (15).

3.2. Proximal Polyak–Lojasiewicz condition

Next, we consider the problems in which the function f is not strongly convex but the function $F := f + g$ satisfies the proximal PL condition (8).

Assumption 2. Let the regularization function g in (12) be proper, closed, and convex, let f be twice continuously differentiable with $\nabla^2 f(x) \preceq L_f I$, and let the generalized gradient map satisfy the proximal PL condition,

$$\|G_\mu(x)\|_2^2 \geq \gamma\,(F_\mu(x) - F_\mu^\star)$$

where $\mu \in (0, 1/L_f)$, $\gamma > 0$, and $F_\mu^\star$ is the optimal value of the FB envelope F_μ.

Remark 3. The proximal gradient algorithm can be interpreted as a variable-metric gradient method on FB envelope and (15) can be equivalently written as

$$\dot{x} = -\mu\,(I - \mu \nabla^2 f(x))^{-1}\, \nabla F_\mu(x).$$

Under Assumption 2, $I - \mu \nabla^2 f(x)$ is invertible and the functions F and F_μ have the same minimizers and the same optimal values (Patrinos et al., 2014), i.e., $\operatorname{argmin}_x F(x) = \operatorname{argmin}_x F_\mu(x)$ and $F^\star = F_\mu^\star$. This motivates the analysis of the convergence properties of (15) in terms of the FB envelope.

Theorem 3. *Let Assumption 2 hold. Then the forward–backward envelope associated with the proximal gradient flow dynamics (15) converge exponentially to $F_\mu^\star = F^\star$ with the rate $\rho = \gamma\mu(1 - \mu L_f)$, i.e.,*

$$F_\mu(x(t)) - F_\mu^\star \leq \mathrm{e}^{-\rho t}\,(F_\mu(x(0)) - F_\mu^\star), \quad \forall\, t \geq 0.$$

Proof. For a Lyapunov function candidate,

$$V(x) = F_\mu(x) - F_\mu^\star$$

the derivative of V along the solutions of (15) is given by

$$\begin{aligned} \dot{V}(x) &= \left\langle \nabla F_\mu(x), \dot{x} \right\rangle \\ &= -\left\langle \nabla F_\mu(x), \mu(I - \mu \nabla^2 f(x))^{-1} \nabla F_\mu(x) \right\rangle \\ &= -\left\langle G_\mu(x), \mu(I - \mu \nabla^2 f(x))\, G_\mu(x) \right\rangle. \end{aligned}$$

Since the gradient of f is L_f-Lipschitz continuous, i.e., $\nabla^2 f(x) \preceq L_f I$ for all $x \in \mathbb{R}^n$, Assumption 2 implies $-(I - \mu \nabla^2 f(x)) \preceq -(1 - \mu L_f) I$, and, thus,

$$\begin{aligned} \dot{V}(x) &\leq -\mu(1 - \mu L_f)\, \|G_\mu(x)\|_2^2 \\ &\leq -\gamma\mu(1 - \mu L_f)(F_\mu(x) - F_\mu^\star) \end{aligned} \tag{24}$$

is non-positive for $\mu \in (0, 1/L_f)$. Moreover, combining the last inequality with the definition of V yields $\dot{V} \leq -\gamma\mu(1-\mu L_f)V$, which implies

$$F_\mu(x(t)) - F_\mu^\star \;\leq\; \mathrm{e}^{-\gamma\mu(1-\mu L_f)t}\,(F_\mu(x(0)) - F_\mu^\star).$$

Remark 4. When the proximal PL condition is satisfied, $F_\mu(x(t)) - F_\mu^\star$ converges exponentially but, in the absence of strong convexity, the exponential convergence rate cannot be established for $\|x(t) - x^\star\|_2$. Thus, although the objective function converges exponentially fast, the solution to (15) does not enjoy this convergence rate. To the best of our knowledge, the convergence rate of $x(t)$ to the set of optimal values $x^\star$ is not known in this case.

4. Global exponential stability of the Douglas-Rachford splitting dynamics

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

4.1. Non-smooth composite optimization problem

The optimality condition for (12) is given by (13), i.e., $0 \in \nabla f(x^\star) + \partial g(x^\star)$. Multiplication by μ and addition of x to the both sides yields $0 \in [I + \mu\nabla f](x^\star) + \mu\partial g(x^\star) - x^\star$. Since $\mathbf{prox}_{\mu f} := (I + \mu\nabla f)^{-1}$ is single-valued (Parikh & Boyd, 2013), introducing $z := x - \mu\partial g(x)$ leads to,

$$x^\star \;=\; \mathbf{prox}_{\mu f}(x^\star - \mu\partial g(x^\star)) \;=\; \mathbf{prox}_{\mu f}(z^\star). \tag{25a}$$

Now, adding x to the both sides of the defining equation for z gives $[I + \mu\partial g](x^\star) = 2\mathbf{prox}_{\mu f}(z^\star) - z^\star$, i.e.,

$$x^\star \;=\; \mathbf{prox}_{\mu g}(2\,\mathbf{prox}_{\mu f}(z^\star) - z^\star). \tag{25b}$$

Combining (25a) and (25b) results in the following optimality condition,

$$\mathbf{prox}_{\mu f}(z^\star) - \mathbf{prox}_{\mu g}(2\,\mathbf{prox}_{\mu f}(z^\star) - z^\star) \;=\; 0. \tag{25c}$$

Furthermore, the reflected proximal operators (Giselsson & Boyd, 2017), $R_{\mu f}(z) := [2\,\mathbf{prox}_{\mu f} - I](z)$ and $R_{\mu g} := [2\,\mathbf{prox}_{\mu g} - I](z)$, can be used to rewrite optimality condition (25c) as

$$z^\star - [R_{\mu g}R_{\mu f}](z^\star) \;=\; 0. \tag{25d}$$

We are now ready to introduce the continuous-time DR gradient flow dynamics to compute $z^\star$,

$$\dot{z} \;=\; -z \;+\; [R_{\mu g}R_{\mu f}](z). \tag{26}$$

Note that the explicit forward Euler discretization of (26) yields the standard DR splitting algorithm (Eckstein & Bertsekas, 1992). We view (26) as a feedback interconnection of an LTI system (16a) with the nonlinear term,

$$u(\xi) \;:=\; [R_{\mu g}R_{\mu f}](\xi). \tag{27}$$

We first characterize properties of nonlinearity u in (27) and then, similar to the previous section, establish global exponential stability of nonlinear system (26).

Lemma 4. *Let Assumption 1 hold and let $\mu \in (0, 2/L_f)$. Then, the operator $R_{\mu f}$ is σ-contractive,*

$$\|R_{\mu f}(x) - R_{\mu f}(y)\|_2 \;\leq\; \sigma\,\|x - y\|_2$$

where σ is given by

$$\sigma \;=\; \max\left\{|1 - \mu m_f|, |1 - \mu L_f|\right\} \;<\; 1. \tag{28}$$

Proof. Given $z_x := \mathbf{prox}_{\mu f}(x)$ and $z_y := \mathbf{prox}_{\mu f}(y)$, 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,

$$\begin{aligned}
&\|R_{\mu f}(x) - R_{\mu f}(y)\|^2 \;=\; \|2(z_x - z_y) - (x-y)\|^2 \;=\\
&\|(z_x - z_y) - \mu\,(\nabla f(z_x) - \nabla f(z_y))\|^2 \;=\; \|z_x - z_y\|^2 +\\
&\|\mu(\nabla f(z_x) - \nabla f(z_y))\|^2 - 2\,\mu\left\langle \nabla f(z_x) - \nabla f(z_y), z_x - z_y\right\rangle\\
&\leq \max\left\{(1 - \mu L_f)^2, (1 - \mu m_f)^2\right\}\,\|z_x - z_y\|^2\\
&\leq \sigma^2\,\|x - y\|^2.
\end{aligned}$$

where the firm non-expansiveness of $\mathbf{prox}_{\mu f}$ is used in the last step. Moreover, according to Lemma 1, for $\mu \in (0, 2/L_f)$ we have $\sigma < 1$, which completes the proof.

Lemma 5. *Let Assumption 1 hold and let $\mu \in (0, 2/L_f)$. Then, the operator $R_{\mu g}$ is firmly non-expansive.*

Proof. $\|R_{\mu g}(x) - R_{\mu g}(y)\|_2^2 = 4\|\mathbf{prox}_{\mu f}(x) - \mathbf{prox}_{\mu f}(y)\|_2^2 + \|x - y\|_2^2 - 4\left\langle x - y, \mathbf{prox}_{\mu f}(x) - \mathbf{prox}_{\mu f}(y)\right\rangle \leq \|x - y\|_2^2$.

Remark 5. Since $R_{\mu g}$ is firmly non-expansive and $R_{\mu f}$ is σ-contractive, the composite operator $R_{\mu g}R_{\mu f}$ is also σ-contractive. Moreover, since the operator $R_{\mu f}$ and nonlinearity u in (16b) have the same contraction parameters, the quadratic inequality that describes (16b) can be also used to characterize the composite operator $R_{\mu g}R_{\mu f}$.

Theorem 6. *Let Assumption 1 hold and let $\mu \in (0, 2/L_f)$. Then, the equilibrium point $z^\star$ of the DR splitting dynamics (26) is globally ρ-exponentially stable, i.e., there is $c > 0$ and $\rho \in (0, 1-\sigma)$ such that,*

$$\|z(t) - z^\star\| \;\leq\; c\,\mathrm{e}^{-\rho t}\,\|z(0) - z^\star\|,\quad \forall\, t \,\geq\, 0$$

where σ is given by (28). Moreover, $x^\star = \mathbf{prox}_{\mu f}(z^\star)$ is the optimal solution of (12).

Proof. Although the nonlinear terms in systems (11) and (26) are different, they share quadratic characterization (17a) and the LTI dynamics (16a). Thus, the result follows from the proof of Theorem 2 and the fact that $x^\star = \mathbf{prox}_{\mu f}(z^\star)$ satisfies optimality condition (25c).

4.2. Douglas-Rachford splitting on the dual problem

Even though the DR splitting algorithm cannot be directly used to solve a problem with a more general linear equality constraint,

$$\begin{array}{ll}
\underset{x,\,z}{\text{minimize}} & f(x) + g(z)\\
\text{subject to} & Tx + Sz = r
\end{array} \tag{29}$$

it can be utilized to solve the dual problem,

$$\underset{\zeta}{\text{minimize}} \quad f_1(\zeta) + g_1(\zeta). \tag{30}$$

Here, $T \in \mathbb{R}^{m\times n}$, $S \in \mathbb{R}^{m\times n}$, and $r \in \mathbb{R}^m$ are the problem parameters, $f_1(\zeta) := f^\star(-T^T\zeta) + r^T\zeta$, $g_1(\zeta) := g^\star(-S^T\zeta)$, and $h^\star(\zeta) := \sup_x(\zeta^T x - h(x))$ is the conjugate of the function h. It is a standard fact (Eckstein & Bertsekas, 1992; Gabay, 1983) that solving the dual problem (30) via the DR splitting algorithm is equivalent to using ADMM for the original problem (29). If Assumption 1 holds and if T is a full row rank matrix, the global exponentially stability of the DR gradient flow dynamics associated with (30), $\dot{\zeta} = -\zeta + [R_{\mu g_1}R_{\mu f_1}](\zeta)$, is readily established.

5. Concluding remarks

We study a class of nonsmooth optimization problems in which it is desired to minimize the sum of a continuously differentiable function with a Lipschitz continuous gradient and a nondifferentiable function. For strongly convex problems, we employ the theory of integral quadratic constraints to prove global exponential stability of proximal gradient flow and Douglas–Rachford splitting dynamics. We also utilize a generalized Polyak–Lojasiewicz condition for nonsmooth problems to demonstrate the global exponential convergence of the forward–backward envelope for the proximal gradient flow algorithm even in the absence of strong convexity.

Appendix. Proximal PL condition

The generalization of the PL condition to nonsmooth problems was introduced in Karimi et al. (2016) and is given by

$$\mathcal{D}_g(x, L_f) \;\geq\; 2\kappa\,(F(x) \,-\, F^\star) \tag{A.1}$$

where κ is a positive constant, L_f is the Lipschitz constant of ∇f, and $\mathcal{D}_g(x, \alpha)$ is determined by

$$-2\alpha \min_y \left(\langle \nabla f(x), y - x\rangle + \frac{\alpha}{2}\,\|y - x\|_2^2 + g(y) - g(x)\right). \tag{A.2}$$

Herein, we show that if proximal PL condition (A.1) holds, there is a lower bound given by (8) on the norm of the generalized gradient map $G_\mu(x)$. For $\mu \in (0, 1/L_f)$, $\mathcal{D}_g(x, 1/\mu) \geq \mathcal{D}_g(x, L_f)$, and, thus, inequality (A.1) also holds for $\mathcal{D}_g(x, 1/\mu)$. Moreover, from the definition (A.2) of $\mathcal{D}_g(x, \alpha)$, it follows that

$$\mathcal{D}_g(x, 1/\mu) \;=\; \frac{2}{\mu}\,(F(x) \,-\, F_\mu(x))$$

where $F := f + g$ and F_μ is the FB envelope. Substituting this expression for $\mathcal{D}_g(x, 1/\mu)$ to (A.1) yields,

$$\frac{1}{\mu}\,(F(x) \,-\, F_\mu(x)) \;\geq\; \kappa\,(F(x) \,-\, F^\star). \tag{A.3}$$

The smooth part of the objective function f can be written as (Patrinos et al., 2014),

$$\begin{aligned} f(x) \;\;=\;\; & F_\mu(x) \,-\, g(\mathbf{prox}_{\mu g}(x \,-\, \mu\,\nabla f(x))) \,+ \\ & \mu\,\left\langle \nabla f(x), G_\mu(x)\right\rangle \,-\, \frac{\mu}{2}\,\|G_\mu(x)\|_2^2 \end{aligned}$$

and substituting this expression for f to (A.3) yields

$$\begin{aligned} &\tfrac{(\mu\kappa\,-\,1)}{2}\,\|G_\mu(x)\|_2^2 \,\geq\, \kappa\,(F_\mu(x) - F^\star) \,+\, \tfrac{(\mu\kappa-1)}{\mu}\,g(x) \,- \\ &(\mu\kappa \,-\, 1)(\tfrac{1}{\mu}\,g(\mathbf{prox}_{\mu g}(x - \mu\,\nabla f(x))) + \left\langle \nabla f(x), G_\mu(x)\right\rangle). \end{aligned} \tag{A.4}$$

Since $G_\mu(x) \,-\, \nabla f(x) \;\in\; \partial g(x)$, the subgradient inequality (7) implies

$$\begin{aligned} 0 \leq \mu\,\|G_\mu(x)\|_2^2 \;\;\leq\;\; & g(x) \,-\, g(\mathbf{prox}_{\mu g}(x \,-\, \mu\,\nabla f(x))) \,+ \\ & \mu\,\left\langle \nabla f(x), G_\mu(x)\right\rangle. \end{aligned} \tag{A.5}$$

Combining (A.4) and (A.5) and taking the sign of $\mu\kappa \,-\, 1$ into account yields,

$$\frac{\alpha}{2}\,\|G_\mu(x)\|_2^2 \;\geq\; \kappa\,(F_\mu(x) \,-\, F^\star), \quad \alpha \;:=\; |\mu\kappa \,-\, 1|.$$

Furthermore, since (Patrinos et al., 2014), $\operatorname{argmin}_x F(x) = \operatorname{argmin}_x F_\mu(x)$ and $F^\star = F_\mu^\star$, $F_\mu^\star$ can be substituted for $F^\star$ and we have $\|G_\mu(x)\|_2^2 \geq \gamma\,(F_\mu(x) \,-\, F_\mu^\star)$ with $\gamma := 2\kappa/|\mu\kappa - 1|$.

References

Arrow, K. J., Hurwicz, L., & Uzawa, H. (1958). Studies in linear and non-linear programming.

Beck, A., & Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, *2*(1), 183–202.

Brown, A., & Bartholomew-Biggs, M. (1989). Some effective methods for unconstrained optimization based on the solution of systems of ordinary differential equations. *Journal of Optimization Theory and Applications*, *62*(2), 211–224.

Cherukuri, A., Mallada, E., & Cortés, J. (2016). Asymptotic convergence of constrained primal–dual dynamics. *Systems & Control Letters*, *87*, 10–15.

Cherukuri, A., Mallada, E., Low, S., & Cortes, J. (2018). The role of convexity on saddle-point dynamics: Lyapunov function and robustness. *IEEE Transactions on Automatic Control*, *63*(8), 2449–2464.

Dhingra, N. K., Khong, S. Z., & Jovanović, M. R. (2017). A second order primal–dual method for nonsmooth convex composite optimization. arXiv:1709.01610.

Dhingra, N. K., Khong, S. Z., & Jovanović, M. R. (2019). The proximal augmented Lagrangian method for nonsmooth composite optimization. *IEEE Transactions on Automatic Control*, *64*(7), 2861–2868.

Ding, D., Hu, B., Dhingra, N. K., & Jovanović, M. R. (2018). An exponentially convergent primal–dual algorithm for nonsmooth composite minimization. In *Proceedings of the 57th IEEE Conference on Decision and Control* (pp. 4927–4932). Miami, FL.

Douglas, J., & Rachford, H. (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.

Eckstein, J., & Bertsekas, D. P. (1992). On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, *55*(1–3), 293–318.

Fazlyab, M., Ribeiro, A., Morari, M., & Preciado, V. M. (2018). Analysis of optimization algorithms via integral quadratic constraints: Nonstrongly convex problems. *SIAM Journal on Optimization*, *28*(3), 2654–2689.

Feijer, D., & Paganini, F. (2010). Stability of primal–dual gradient dynamics and applications to network optimization. *Automatica*, *46*(12), 1974–1981.

França, G., Robinson, D., & Vidal, R. (2018). ADMM and accelerated ADMM as continuous dynamical systems. arXiv:1805.06579.

Gabay, D. (1983). Applications of the method of multipliers to variational inequalities. In *Studies in mathematics and its applications, Vol. 15* (pp. 299–331).

Giselsson, P., & Boyd, S. (2017). Linear convergence and metric selection for Douglas-Rachford splitting and ADMM. *IEEE Transactions on Automatic Control*, *62*(2), 532–544.

Hassan-Moghaddam, S., & Jovanović, M. R. (2018). Distributed proximal augmented Lagrangian method for nonsmooth composite optimization. In *Proceedings of the 2018 American Control Conference* (pp. 2047–2052). Milwaukee, WI.

Hassan-Moghaddam, S., & Jovanović, M. R. (2018). On the exponential convergence rate of proximal gradient flow algorithms. In *Proceedings of the 57th IEEE Conference on Decision and Control* (pp. 4246–4251). Miami, FL.

Hu, B., & Lessard, L. (2017). Dissipativity theory for Nesterov's accelerated method. In *Proceedings of the 34th International Conference on Machine Learning* (pp. 1549–1557).

Hu, B., & Seiler, P. (2016). Exponential decay rate conditions for uncertain linear systems using integral quadratic constraints. *IEEE Transactions on Automatic Control*, *61*(11), 3631–3637.

Hu, B., Seiler, P., & Rantzer, A. (2017). A unified analysis of stochastic optimization methods using jump system theory and quadratic constraints. In *Proceedings of the 2017 Conference on Learning Theory* (pp. 1157–1189).

Karimi, H., Nutini, J., & Schmidt, M. (2016). Linear convergence of gradient and proximal-gradient methods under the Polyak-Lojasiewicz condition. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 795–811).

Latafat, P., Freris, N., & Patrinos, P. (2019). A new randomized block-coordinate primal–dual proximal algorithm for distributed optimization. *IEEE Transactions on Automatic Control*, http://dx.doi.org/10.1109/TAC.2019.2906924.

Latafat, P., Stella, L., & Patrinos, P. (2016). New primal–dual proximal algorithm for distributed optimization. In *Proceedings of the 55th IEEE Conference on Decision and Control* (pp. 1959–1964).

Lessard, L., Recht, B., & Packard, A. (2016). Analysis and design of optimization algorithms via integral quadratic constraints. *SIAM Journal on Optimization*, *26*(1), 57–95.

Megretski, A., & Rantzer, A. (1997). System analysis via integral quadratic constraints. *IEEE Transactions on Automatic Control*, *42*(6), 819–830.

Michalowsky, S., Scherer, C., & Ebenbauer, C. (2019). Robust and structure exploiting optimization algorithms: An integral quadratic constraint approach. arXiv:1905.00279.

Mohammadi, H., Razaviyayn, M., & Jovanović, M. R. (2018). Variance amplification of accelerated first-order algorithms for strongly convex quadratic optimization problems. In *Proceedings of the 57th IEEE Conference on Decision and Control* (pp. 5753–5758). Miami, FL.

Mohammadi, H., Razaviyayn, M., & Jovanović, M. R. (2019). Performance of noisy Nesterov's accelerated method for strongly convex optimization problems. In *Proceedings of the 2019 American Control Conference* (pp. 3426–3431). Philadelphia, PA.

Mohammadi, H., Razaviyayn, M., & Jovanović, M. R. (2020). Robustness of accelerated first-order algorithms for strongly convex optimization problems. *IEEE Transactions on Automatic Control*, http://dx.doi.org/10.1109/TAC.2020.3008297.

Muehlebach, M., & Jordan, M. I. (2019). A dynamical systems perspective on Nesterov acceleration. arXiv:1905.07436.

Nedić, A., & Ozdaglar, A. (2009). Distributed subgradient methods for multiagent optimization. *IEEE Transactions on Automatic Control*, *54*(1), 48–61.

Nesterov, Y. (2013). Introductory lectures on convex optimization: A basic course. (p. 87).

Parikh, N., & Boyd, S. (2013). Proximal algorithms. *Foundations and Trends® in Optimization*, *1*(3), 123–231.

Patrinos, P., Stella, L., & Bemporad, A. (2014). Forward-backward truncated Newton methods for convex composite optimization. arXiv:1402.6655.

Polyak, B. T. (1963). Gradient methods for minimizing functionals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, *3*(4), 643–653.

Poveda, J. I., & Li, N. (2019). Inducing uniform asymptotic stability in time-varying accelerated optimization dynamics via hybrid regularization. arXiv:1905.12110.

Qu, G., & Li, N. (2018). On the exponential stability of primal–dual gradient dynamics. *IEEE Control Systems Letters*, *3*(1), 43–48.

Schropp, J., & Singer, I. (2000). A dynamical systems approach to constrained minimization. *Numerical Functional Analysis and Optimization*, *21*(3–4), 537–551.

Seidman, J., Fazlyab, M., Preciado, V., & Pappas, G. (2019). A control-theoretic approach to analysis and parameter selection of Douglas-Rachford splitting. arXiv:1903.11525.

Shi, B., Du, S., Jordan, M. I., & Su, W. (2018). Understanding the acceleration phenomenon via high-resolution differential equations. arXiv:1810.08907.

Stella, L., Themelis, A., & Patrinos, P. (2017). Forward–backward quasi-Newton methods for nonsmooth optimization problems. *Computational Optimization and Applications*, *67*(3), 443–487.

Su, W., Boyd, S., & Candes, E. (2016). A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights. *Journal of Machine Learning Research*, *17*, 1–43.

Themelis, A., Stella, L., & Patrinos, P. (2018). Forward-backward envelope for the sum of two nonconvex functions: Further properties and nonmonotone line-search algorithms. *SIAM Journal on Optimization*, *28*(3), 2274–2303.

Wang, J., & Elia, N. (2011). A control perspective for centralized and distributed convex optimization. In *Proceedings of the 50th IEEE Conference on Decision and Control* (pp. 3800–3805).

Wibisono, A., Wilson, A. C., & Jordan, M. I. (2016). A variational perspective on accelerated methods in optimization. *Proceedings of the National Academy of Sciences*, *113*(47), E7351–E7358.

Zhang, J., Mokhtari, A., Sra, S., & Jadbabaie, A. (2018). Direct Runge-Kutta discretization achieves acceleration. In *Advances in Neural Information Processing Systems* (pp. 3900–3909).

Sepideh Hassan-Moghaddam received her B.Sc. in Electrical Engineering from Sharif University of Technology, Tehran, Iran, in 2013, M.S. in Electrical and Computer Engineering from the University of Minnesota, Minneapolis, MN, in 2016, and Ph.D. in Electrical and Computer Engineering from the University of Southern California, Los Angeles, CA, in 2019. She is currently Data and Applied Scientist with Microsoft in Seattle, WA. Her primary research interests are in optimization, inference, and control of large-scale networks.

Mihailo R. Jovanović received the Dipl. Ing. and M.S. degrees from the University of Belgrade, Serbia, in 1995 and 1998, respectively, and the Ph.D. degree from the University of California, Santa Barbara, in 2004. He was a visiting researcher with the Department of Mechanics, the Royal Institute of Technology, Stockholm, Sweden, from September to December 2004, and a faculty member in the Department of Electrical and Computer Engineering at the University of Minnesota, Minneapolis, from 2004 until 2017. He has held visiting positions with Stanford University and the Institute for Mathematics and its Applications.

Prof. Jovanovic's current research focuses on large-scale and distributed optimization, design of controller architectures, sparsity-promoting optimal control, fundamental limitations in the design of large dynamic networks, and dynamics/control of fluid flows. He currently serves as an Associate Editor of the IEEE Transactions on Control of Network Systems and had served as a Guest Editor (of the Special Issue on Analysis, Control and Optimization of Energy System Networks in the IEEE Transactions on Control of Network Systems), the Chair of the APS External Affairs Committee, a Program Vice-Chair of the 55th IEEE Conference on Decision and Control, an Associate Editor of the SIAM Journal on Control and Optimization, and an Associate Editor of the IEEE Control Systems Society Conference Editorial Board.

Prof. Jovanovic is a fellow of the American Physical Society (APS) and the Institute of Electrical and Electronics Engineers (IEEE). He received a CAREER Award from the National Science Foundation in 2007, the George S. Axelby Outstanding Paper Award from the IEEE Control Systems Society in 2013, and the Distinguished Alumnus Award from the Department of Mechanical Engineering at UC Santa Barbara in 2014. Papers of his students were finalists for the Best Student Paper Award at the American Control Conference in 2007 and 2014.