2018 IEEE Conference on Decision and Control (CDC)
Miami Beach, FL, USA, Dec. 17-19, 2018

On the exponential convergence rate of proximal gradient flow algorithms

Sepideh Hassan-Moghaddam and Mihailo R. Jovanović

Abstract— **Many modern large-scale and distributed optimization problems can be cast into a form in which the objective function is a sum of a smooth term and a nonsmooth regularizer. Such problems can be solved via a proximal gradient method which generalizes standard gradient descent to a nonsmooth setup. In this paper, we leverage the tools from control theory to study global convergence of proximal gradient flow algorithms. We utilize the fact that the proximal gradient algorithm can be interpreted as a variable-metric gradient method on the forward-backward envelope. This continuously differentiable function can be obtained from the augmented Lagrangian associated with the original nonsmooth problem and it enjoys a number of favorable properties. We prove that global exponential convergence can be achieved even in the absence of strong convexity. Moreover, for in-network optimization problems, we provide a distributed implementation of the gradient flow dynamics based on the proximal augmented Lagrangian and prove global exponential stability for strongly convex problems.**

Index Terms— **Distributed optimization, forward-backward envelope, global exponential stability, gradient flow, large-scale systems, proximal algorithms, primal-dual method, proximal augmented Lagrangian.**

I. INTRODUCTION

We study a class of nonsmooth composite convex optimization problems in which the objective is a sum of a differentiable function and a possibly nondifferentiable regularizer. These problems emerge in compressive sensing, machine learning, and control. For example, structured feedback design can be cast as a nonsmooth composite optimization problem [1]–[3]. Standard descent methods cannot be used in the presence of nondifferentiable component. Proximal gradient algorithms [4], [5] offer viable alternatives to the generic descent methods for solving nonsmooth problems. Another effective strategy is to transform the associated augmented Lagrangian into the continuously differentiable proximal augmented Lagrangain [6] in which the former is restricted to the manifold that corresponds to the explicit minimization over the variable in the nonsmooth term.

Analysis of optimization algorithms from the system theoretic point of view has received significant recent attention [7]–[9]. In these references, the optimization algorithm is interpreted as a feedback interconnection in which the states converge to the optimal solution of the optimization problem.

In this paper, we utilize tools and ideas from control theory to study global convergence properties of proximal gradient flows in continuous time. We use the fact that the proximal gradient method can be interpreted as a variable-metric gradient method on the forward-backward envelope [10]–[12]. We illustrate that the forward-backward envelope can be obtained from the augmented Lagrangian associated with the original nonsmooth optimization problem. Specifically, we show that the forward-backward envelope can be achieved by restricting the augmented Lagrangian to the manifold in which the dual variable is given by the negative of the derivative of the smooth part of the objective function. By utilizing the theory of integral quadratic constraints (IQCs) [13], we prove exponential convergence when the smooth part of the objective function is strongly convex with a Lipschitz continuous gradient. We then propose a generalization of the Polyak-Lojasiewicz (PL) [14] condition that is well-suited to nonsmooth problems and study the convergence rate of the proximal gradient flow in the absence of strong convexity.

Distributed algorithms are critically important for solving large-scale optimization problems. The decentralized consensus problem in multi-agent networks [15]–[18] arises in many applications. Herein, we restrict our attention to distributed in-network optimization and show that the proximal gradient flow dynamics cannot be used for this class of problems. We next describe how primal-descent dual-ascent gradient flow dynamics based on the proximal augmented Lagrangian [6] can be used as an effective alternative. We provide a distributed implementation and prove global exponential stability in the presence of strong convexity.

The paper is structured as follows. In Section II, we formulate the nonsmooth composite optimization problem and provide the essential background. In Section III, we study the exponential convergence rate of the proximal gradient flows in continuous time in the presence of strong convexity. Moreover, we introduce the proximal PL condition and show exponential convergence in the value function. In Section IV, by restricting our attention to in-network optimization, we provide a distributed implementation based on the proximal augmented Lagrangian. Furthermore, by introducing an appropriate change of coordinates, we utilize the theory of IQCs to prove global exponential stability under the strong convexity assumption. Finally, in Section V, we offer concluding remarks.

II. PROBLEM FORMULATION AND BACKGROUND

We consider composite optimization problems,

$$\underset{x}{\text{minimize}} \;\; f(x) \; + \; g(Ex) \tag{1}$$

where $x \in \mathbb{R}^n$ is the optimization variable, $f\colon \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function with a Lipschitz

Financial support from the National Science Foundation under award ECCS-1809833 and from the Air Force Office of Scientific Research under award FA9550-16-1-0009 is gratefully acknowledged.

S. Hassan-Moghaddam and M. R. Jovanović are with the Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089. E-mails: hassanmo@usc.edu, mihailo@usc.edu.

continuous gradient ∇f, g: $\mathbb{R}^m \to \mathbb{R}$ is a possibly nondifferentiable convex regularization function, and $E \in \mathbb{R}^{m\times n}$ is a matrix that promotes structural properties in the desired set of coordinates. Such optimization problems arise in a number of different application domains. Depending on the structure of the functions f and g and the matrix E, different first- and second-order algorithms can be employed to solve them. Since most of these algorithms utilize proximal operator associated with the nonsmooth regularizer, we are interested in studying global convergence properties of first-order proximal gradient flow algorithms.

In what follows, we provide background material that we utilize in the rest of the paper.

A. Proximal operators and the associated envelopes

The proximal operator of a proper, lower semicontinuous, and convex function g is the solution of

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

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

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

Even for a nondifferentiable g, $M_{\mu g}$ is a continuously differentiable function and its gradient is given by [4],

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

The Moreau envelope of g can be used to introduce the forward-backward envelope [10]–[12] of

$$F(x) \;:=\; f(x) \;+\; g(x)$$

which is the value function of the optimization problem

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

where

$$\begin{aligned} J \;&:=\; f(x) \,+\, \langle \nabla f(x), v - x\rangle \,+\, \tfrac{1}{2\mu}\,\|v - x\|_2^2 \,+\, g(v) \\ &=\; g(v) \,+\, \tfrac{1}{2\mu}\,\|v \,-\, (x \,-\, \mu\nabla f(x))\|_2^2 \,+ \\ &\qquad f(x) \,-\, \tfrac{1}{2\mu}\,\|\nabla f(x)\|_2^2. \end{aligned} \tag{2b}$$

We note that J approximates F via a simple quadratic expansion of f around x and the optimal solution of (2) is

$$v^\star \;=\; \mathbf{prox}_{\mu g}(x \,-\, \mu\nabla f(x)).$$

This optimal solution is used to obtain the forward-backward envelope,

$$\begin{aligned} F_\mu(x) \;&:=\; J(x, v^\star) \;=\; J(x, \mathbf{prox}_{\mu g}(x \,-\, \mu\nabla f(x))) \\ &=\; f(x) \,+\, M_{\mu g}(x \,-\, \mu\,\nabla f(x)) \,-\, \frac{\mu}{2}\,\|\nabla f(x)\|_2^2 \end{aligned}$$

which is continuously differentiable with a gradient [10],

$$\nabla F_\mu(x) \;=\; (I \,-\, \mu\nabla^2 f(x))\, G_\mu(x). \tag{3}$$

Here, G_μ is the generalized gradient map,

$$G_\mu(x) \;:=\; \tfrac{1}{\mu}\,(x \,-\, \mathbf{prox}_{\mu g}(x \,-\, \mu\nabla f(x))) \tag{4}$$

which can be used to obtain an alternative expression for F_μ,

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

B. Proximal augmented Lagrangian

By introducing an auxiliary variable $z := Ex$, (1) can be rewritten as

$$\begin{array}{ll} \underset{x,\,z}{\text{minimize}} & f(x) \,+\, g(z) \\ \text{subject to} & Ex \,-\, z \,=\, 0. \end{array} \tag{6}$$

The augmented Lagrangian associated with constrained optimization problem (6) is given by,

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

and the completion of squares yields,

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

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

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

and the evaluation of $\mathcal{L}_\mu(x, z; y)$ along the manifold resulting from the explicit minimization over z yields the proximal augmented Lagrangian [6],

$$\begin{aligned} \mathcal{L}_\mu(x; y) \;&:=\; \mathcal{L}_\mu(x, z^\star(x,y); y) \\ &=\; f(x) \,+\, M_{\mu g}(Ex \,+\, \mu y) \,-\, \tfrac{\mu}{2}\,\|y\|_2^2 \end{aligned} \tag{7}$$

This function is continuously differentiable with respect to both x and y and it can be used as a foundation for the development of different first- and second-order primal-dual methods for nonsmooth composite optimization [6], [19]. It is noteworthy that, for $E = I$, forward-backward envelope $F_\mu(x)$ is obtained by restricting the proximal augmented Lagrangian $\mathcal{L}_\mu(x; y)$ along the manifold $y^\star(x) = -\nabla f(x)$,

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

C. Strong convexity and Lipschitz continuity

The function f is strongly convex with parameter m_f if for any x and y,

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

and the gradient of a continuously-differentiable function f is Lipschitz continuous with parameter L_f if for any x and y,

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

Furthermore, the subgradient ∂g of a nondifferentiable function g is the set of points $z \in \partial g(x)$ that for any x and y satisfy,

$$g(y) \;\geq\; g(x) \,+\, z^T(y \,-\, x). \tag{8}$$

D. Proximal Polyak-Lojasiewicz inequality

The Polyak-Lojasiewicz (PL) condition is an inequality that can be used to establish linear convergence of standard gradient descent method in the absence of strong convexity (or even convexity) [20]. For an unconstrained optimization problem with a non-empty solution set,

$$\underset{x}{\text{minimize}} \;\; f(x)$$

where f is a twice differentiable function with a Lipschitz continuous gradient, the PL condition is given by

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

where $\gamma > 0$ and $f^\star$ is the optimal value of the function f.

We next provide the generalization of the PL condition for nonsmooth composite optimization problems (1) with $E = I$. For this class of problems, the proximal PL inequality holds for $\mu < 1/L_f$ if there exist $\gamma > 0$ such that

$$\|G_\mu(x)\|^2 \;\geq\; \gamma\,(F_\mu(x) \,-\, F_\mu^\star). \tag{9}$$

Here, L_f is the Lipschitz constant of ∇f, F_μ is the forward-backward envelope, and G_μ is the generalized gradient map. It can be shown that the above condition is equivalent to the condition provided in [20]; the proof is omitted due to page limitations and it will be reported elsewhere.

III. Exponential stability of proximal algorithms

In this section, we briefly discuss the Arrow-Hurwicz-Uzawa gradient flow dynamics that can be used to solve (1) by computing the saddle points of the proximal augmented Lagrangian [6]. 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 (7) can be utilized to compute its saddle points via the Arrow-Hurwicz-Uzawa gradient flow dynamic,

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} \;=\; \begin{bmatrix} -\left(\nabla f(x) \,+\, E^T \nabla M_{\mu g}(Ex \,+\, \mu y)\right) \\ \mu\,(\nabla M_{\mu g}(Ex \,+\, \mu y) \,-\, y) \end{bmatrix}. \tag{10}$$

It was recently shown that these primal-descent dual-ascent dynamics are globally exponentially stable for convex problems in which the matrix EE^T is invertible and the smooth part of the objective function f is strongly convex [6].

For convex problems with $E = I$,

$$\underset{x}{\text{minimize}} \;\; f(x) \,+\, g(x) \tag{11}$$

the proximal gradient method,

$$x^{k+1} \,=\, \mathbf{prox}_{\mu g}(x^k \,-\, \alpha_k \nabla f(x^k)) \tag{12}$$

with the stepsize $\alpha_k \,\leq\, 1/L_f$ can be used to solve (11), where L_f is the Lipschitz constant of ∇f. In [10], it was demonstrated that (12) can be interpreted as a variable-metric gradient method on forward-backward envelope,

$$\begin{aligned} x^{k+1} \;&=\; x^k \,-\, \alpha_k \left(I \,-\, \alpha_k \nabla^2 f(x)\right)^{-1} \nabla F_{\alpha_k}(x^k) \\ &=\; x^k \,-\, \alpha_k\, G_{\alpha_k}(x^k). \end{aligned}$$

This interpretation can be utilized to solve (11) via the continuous-time proximal gradient flow dynamics

$$\begin{aligned} \dot{x} \;&=\; -G_\mu(x) \\ &=\; -\left(\nabla f(x) \,+\, \nabla M_{\mu g}(x \,-\, \mu \nabla f(x))\right) \\ &=\; -\tfrac{1}{\mu}\left(x \,-\, \mathbf{prox}_{\mu g}(x \,-\, \mu \nabla f(x))\right). \end{aligned} \tag{13}$$

Remark 1: Proximal gradient algorithm (12) can be obtained via explicit forward Euler discretization of (13) with the stepsize $\mu = \alpha_k$. This should be compared and contrasted to a standard interpretation [4] in which (12) results from implicit backward Euler discretization of the subgradient flow dynamics associated with (11). We also note that (13) 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 (10) with $E = I$.

We next study global stability of proximal gradient flow dynamics (13), first for strongly convex problems and then for the problems in which only the PL condition holds.

A. Strongly convex problems

Herein, we utilize the theory of integral quadratic constraints to establish global asymptotic stability of proximal gradient flow dynamics (13) under the following assumption.

Assumption 1: Let the differentiable part f of the objective function in (11) be strongly convex with parameter m_f, let ∇f be Lipschitz continuous with parameter L_f, and let the regularization function g be proper, lower semicontinuous, and convex.

Proximal gradient flow dynamics (13) can be expressed as a feedback interconnection of an LTI system

$$\begin{aligned} \dot{x} \;&=\; A\,x \,+\, B\,u \\ \xi \;&=\; C\,x \end{aligned} \tag{14a}$$

with a nonlinear term,

$$u \;=\; \mathbf{prox}_{\mu g}(x \,-\, \mu\,\nabla f(x)). \tag{14b}$$

Here,

$$A \,=\, -\tfrac{1}{\mu}\,I, \;\;\; B \,=\, \tfrac{1}{\mu}\,I, \;\;\; C \,=\, I \tag{14c}$$

and the corresponding transfer function is

$$H(s) \;=\; C\,(s\,I \,-\, A)^{-1} B \;=\; \frac{1}{\mu s \,+\, 1}\,I. \tag{14d}$$

Lemma 1 exploits properties of f and g to characterize nonlinear map (14b) via a quadratic inequality on x and u.

Lemma 1: Let Assumption 1 hold. Then, for any $x \in \mathbb{R}^n$ and $\hat{x} \in \mathbb{R}^n$ there exists a symmetric matrix $K_{x,\hat{x}}$ satisfying $m_f I \preceq K_{x,\hat{x}} \preceq L_f I$ such that the quadratic inequality

$$\begin{bmatrix} x \,-\, \hat{x} \\ u \,-\, \hat{u} \end{bmatrix}^T \Pi \begin{bmatrix} x \,-\, \hat{x} \\ u \,-\, \hat{u} \end{bmatrix} \;\geq\; 0$$

holds for $u := \mathbf{prox}_{\mu g}(x - \mu \nabla f(x))$ and $\hat{u} \in \mathbb{R}^n$, where

$$\Pi \; := \; \left[\begin{array}{cc} 0 & I - \mu K_{x,\hat{x}} \\ I - \mu K_{x,\hat{x}} & -2I \end{array} \right]. \tag{15}$$

Proof: Let $\zeta := x - \mu \nabla f(x)$. Since the proximal operator of g is firmly nonexpansive [4], we have

$$(\zeta - \hat{\zeta})^T (u(\zeta) - u(\hat{\zeta})) \;\geq\; \| u(\zeta) - u(\hat{\zeta}) \|_2^2. \tag{16}$$

Using the definition of ζ, the left-hand side in this inequality can be written as

$$(x - \hat{x})^T (u(\zeta) - u(\hat{\zeta})) - \\ \mu (\nabla f(x) - \nabla f(\hat{x}))^T (u(\zeta) - u(\hat{\zeta})).$$

For an m_f-strongly convex function f with an L_f-Lipschitz continuous gradient, there is a symmetric matrix $K_{x,\hat{x}}$ such that for any x and $\hat{x}$ [19, Lemma 3] ,

$$\nabla f(x) - \nabla f(\hat{x}) \;=\; K_{x,\hat{x}}(x - \hat{x}), \\ m_f I \;\preceq\; K_{x,\hat{x}} \;\preceq\; L_f I.$$

Thus, inequality (16) can be written as

$$(x - \hat{x})^T (I - \mu K_{x,\hat{x}})(u(\zeta) - u(\hat{\zeta})) \geq \| u(\zeta) - u(\hat{\zeta}) \|_2^2$$

which completes the proof. ■

Remark 2: Lemma 1 combines firm nonexpansiveness of $\mathbf{prox}_{\mu g}$, strong convexity of f, and Lipschitz continuity of ∇f to establish a quadratic inequality that $u = \mathbf{prox}_{\mu g}(x - \mu \nabla f(x))$ has to satisfy at any x and $\hat{x}$. Even though the matrix $K_{x,\hat{x}}$ depends on the operating point, its spectral properties, $m_f I \preceq K_{x,\hat{x}} \preceq L_f I$, hold for all x and $\hat{x}$.

We next use the KYP lemma [21]

$$\left[\begin{array}{c} H_\rho(j\omega) \\ I \end{array} \right]^* \Pi \left[\begin{array}{c} H_\rho(j\omega) \\ I \end{array} \right] \;\prec\; 0, \;\;\; \forall \, \omega \, \in \, \mathbb{R} \tag{17}$$

where Π is given by (15) and

$$H_\rho(j\omega) \;=\; C \, (j\omega I - (A + \rho I))^{-1} B \;=\; \frac{1}{j\mu\,\omega + 1 - \mu\rho} \, I$$

to establish global exponential stability of (14).

Theorem 2: Let Assumption 1 hold. Then proximal gradient flow dynamics (14) are globally exponentially stable, i.e., there is $\tau > 0$ and $0 < \rho < \min{(m_f, 1/\mu)}$ such that,

$$\| x(t) - x^\star \| \;\leq\; \tau \, \mathrm{e}^{-\rho t} \, \| x(0) - x^\star \|, \;\;\; \forall \, t \;\geq\; 0.$$

Proof: From the definition of Π in Lemma 1, we see that (17) holds for all $\omega \in \mathbb{R}$ if

$$2 \, I - H_\rho^*(j\omega) \, (I - \mu K_{x,\hat{x}}) - (I - \mu K_{x,\hat{x}}) H_\rho(j\omega) \;\succ\; 0.$$

This inequality can be equivalently written as

$$\mu^2 \omega^2 I + (1 - \mu\rho) \, ((1 - \mu\rho) \, I - (I - \mu K_{x,\hat{x}})) \;= \\ \mu^2 \omega^2 I + \mu(1 - \mu\rho) \, (K_{x,\hat{x}} - \rho I) \;\succ\; 0.$$

Since the stability of H_ρ requires $1 - \rho\mu > 0$, it holds if

$$\rho I \;\prec\; K_{x,\hat{x}}.$$

Finally, the spectral properties of the matrix $K_{x,\hat{x}}$

$$m_f I \;\preceq\; K_{x,\hat{x}} \;\preceq\; L_f I$$

imply exponential stability with the rate $\rho < \min{(m_f, 1/\mu)}$. ■

B. Proximal Polyak-Lojasiewicz condition

Next, we consider the problems in which the function f is not necessarily strongly convex but the function $F := f + g$ satisfies proximal PL condition (9).

Assumption 2: Let the regularization function g in (6) be proper, lower semicontinuous, and convex, let f be twice differentiable, let ∇f be Lipschitz continuous with parameter L_f, and let the generalized gradient map satisfy the proximal PL condition,

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

where $0 < \mu < 1/L_f$, $\gamma > 0$, and $F_\mu^\star$ is the optimal value of the forward-backward envelope F_μ.

Remark 3: We recall that the proximal gradient algorithm can be interpreted as a variable-metric gradient method on forward-backward envelope and that (13) can be equivalently written as

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

Under Assumption 2, the matrix $I - \mu \nabla^2 f(x)$ is invertible and the functions $F = f + g$ and F_μ have the same minimizers and the same optimal values [10],

$$\underset{x}{\operatorname{argmin}} \; F(x) \;=\; \underset{x}{\operatorname{argmin}} \; F_\mu(x), \;\;\; F^\star \;=\; F_\mu^\star.$$

This motivates study of the convergence properties of (13) in terms of the forward-backward envelope.

Theorem 3: Let Assumption 2 hold. Then the forward-backward envelope associated with proximal gradient flow dynamics (13) converge exponentially to the optimal function value $F_\mu^\star = F^\star$ with the rate $\rho = \gamma(1 - \mu L_f)$,

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

Proof: We introduce a Lyapunov function candidate,

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

where F_μ is the forward-backward envelope. The derivative of V along the solutions of (13) is given by

$$\dot{V}(x) = \; \langle \nabla F_\mu(x), \dot{x} \rangle \;=\; - \, G_\mu^T(x) \, (I - \mu \, \nabla^2 f(x)) \, G_\mu(x).$$

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

$$\begin{aligned} \dot{V}(x) \;&\leq\; -(1 - \mu L_f) \, \| G_\mu(x) \|_2^2 \\ &\leq\; -\gamma \, (1 - \mu L_f) \, (F_\mu(x) - F_\mu^\star) \end{aligned} \tag{18}$$

is non-positive for $\mu \in [0, 1/L_f)$. Moreover, from the definition of V and (18) we have,

$$\dot{V} \;\leq\; -\gamma \, (1 - \mu L_f) \, V$$

which yields

$$F_\mu(x(t)) - F_\mu^\star \le e^{-\gamma(1-\mu L_f)t}(F_\mu(x(0)) - F_\mu^\star).$$

■

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 $\|x(t) - x^\star\|$. Thus, in the absence of strong convexity, although the objective function converges exponentially fast, the solution to (13) does not enjoy this convergence rate. To the best of our knowledge, in the absence of strong convexity, the convergence rate of $x(t)$ to the set of optimal values $x^\star$ is unknown.

Proximal gradient flow dynamics (13) cannot be used for distributed optimization. We next describe how primal-descent dual-ascent gradient flow dynamics based on proximal augmented Lagrangian (10) can alleviate this challenge.

IV. Distributed optimization

Let us consider the unconstrained optimization problem,

$$\underset{x}{\text{minimize}} \quad \sum_{i=1}^{n} f_i(x)$$

where $x \in \mathbb{R}^n$ is the optimization variable and $f := \sum_i f_i$ is a strongly convex objective function. It is desired to solve this problem over an undirected connected network with the incidence matrix E^T and the graph Laplacian $L := E^T E$. To accomplish this objective, we reformulate it as,

$$\underset{x}{\text{minimize}} \quad \sum_{i=1}^{n} f_i(x_i) + g(Ex) \tag{19a}$$

where $x := [\, x_1 \;\cdots\; x_n \,]^T$ and $g(Ex)$ is an indicator function associated with the equality constraint $Ex = 0$,

$$g(Ex) = \begin{cases} 0, & Ex = 0, \\ \infty, & \text{otherwise.} \end{cases} \tag{19b}$$

This constraint is introduced to ensure asymptotic agreement between the node values $x_i(t) \in \mathbb{R}$.

As demonstrated in [6], the primal-descent dual-ascent gradient flow dynamics based on the proximal augmented Lagrangian (10) can be used to solve this problem. The resulting gradient flow dynamics are given by,

$$\begin{aligned} \dot{x} &= -\nabla f(x) - \tfrac{1}{\mu} L x - \tilde{y} \\ \dot{\tilde{y}} &= L x, \end{aligned} \tag{20}$$

where $L = E^T E$ is the Laplacian matrix of the underlying communication graph between neighboring nodes and the vector $\tilde{y} := E^T y$ belongs to the orthogonal complement of the vector of all ones. This setup is well-suited for distributed implementation in which each node only shares its state x_i with its neighbors and maintains the corresponding dual variable $\tilde{y}_i$. A Lyapunov-based argument was used in [22] to prove the exponential convergence of (20). Herein, we provide an alternative proof that utilizes the theory of IQCs to establish global exponential stability of (20) under the condition that $\mathbb{1}^T \tilde{y}(0) = 0$.

Assumption 3: Let the differentiable part $f := \sum_i f_i(x)$ of the objective function in (19) be strongly convex with parameter m_f, let ∇f be Lipschitz continuous with parameter L_f, let the regularization function g be proper, lower semicontinuous, and convex, let E^T be incidence matrix of a connected undirected network, and let $\mathbb{1}^T \tilde{y}(0) = 0$ in (20).

From Assumption 3 it follows that the graph Laplacian $L := E^T E$ is a positive semidefinite matrix with one zero eigenvalue. Thus, it can be decomposed as

$$L = V \Lambda V^T = \begin{bmatrix} U & \frac{1}{n}\mathbb{1} \end{bmatrix} \begin{bmatrix} \Lambda_0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U^T \\ \frac{1}{n}\mathbb{1}^T \end{bmatrix}$$

where the columns of V are the eigenvectors of L, Λ_0 is a diagonal matrix of the nonzero eigenvalues of L, and the matrix U satisfies,

$$U^T U = I, \quad U^T \mathbb{1} = 0, \quad U U^T = I - \tfrac{1}{n}\mathbb{1}\mathbb{1}^T. \tag{21}$$

By introducing a change of variables

$$\begin{bmatrix} \psi \\ \bar{x} \end{bmatrix} = \begin{bmatrix} U^T x \\ \frac{1}{n}\mathbb{1}^T x \end{bmatrix}, \quad \begin{bmatrix} \phi \\ \bar{y} \end{bmatrix} = \begin{bmatrix} U^T \tilde{y} \\ \frac{1}{n}\mathbb{1}^T \tilde{y} \end{bmatrix}$$

with $\psi \in \mathbb{R}^{n-1}$ and $\phi \in \mathbb{R}^{n-1}$, (20) can be written as

$$\begin{bmatrix} \dot{\psi} \\ \dot{\bar{x}} \\ \dot{\phi} \\ \dot{\bar{y}} \end{bmatrix} = \begin{bmatrix} -(\frac{1}{\mu}\Lambda_0 + m_f I)\psi - \phi - U^T u \\ -\frac{1}{n}\mathbb{1}^T u - m_f \bar{x} - \bar{y} \\ \Lambda_0 \psi \\ 0 \end{bmatrix} \tag{22}$$

where $u = \nabla f(x) - m_f x$, and from the properties (21) of the matrix U we have,

$$x = \begin{bmatrix} U & \mathbb{1} \end{bmatrix} \begin{bmatrix} \psi \\ \bar{x} \end{bmatrix}. \tag{23}$$

By choosing $\bar{y}(0) = 0$, we have $\bar{y} \equiv 0$. Thus, the $\bar{y}$-dynamics can be eliminated from (22), which yields

$$\begin{aligned} \dot{w} &= A w + B u \\ \xi &= C w \\ u &= \nabla f(\xi) - m_f \xi. \end{aligned} \tag{24a}$$

Here, $w := [\, \psi^T \; \bar{x} \; \phi^T \,]^T$, $\xi := x$,

$$A = \begin{bmatrix} -(\frac{1}{\mu}\Lambda_0 + m_f I) & 0 & -I \\ 0 & -m_f & 0 \\ \Lambda_0 & 0 & 0 \end{bmatrix}$$
$$B = \begin{bmatrix} -U^T \\ -\frac{1}{n}\mathbb{1}^T \\ 0 \end{bmatrix}, \quad C = \begin{bmatrix} U & \mathbb{1} & 0 \end{bmatrix} \tag{24b}$$

and the corresponding transfer function is

$$H(s) = -\begin{bmatrix} U & \mathbb{1} \end{bmatrix} \begin{bmatrix} H_1(s) & 0 \\ 0 & \dfrac{1}{s + m_f} \end{bmatrix} \begin{bmatrix} U^T \\ \frac{1}{n}\mathbb{1}^T \end{bmatrix} \tag{24c}$$

where

$$H_1(s) \;=\; \mathrm{diag}\left(\frac{s}{s^2\,+\,(\lambda_i/\mu\,+\,m_f)\,s\,+\,\lambda_i}\right).$$

Furthermore, under Assumption 3 with $u = \nabla f(\xi) - m_f\xi$, for any ξ and $\hat{\xi} \in \mathbb{R}^n$, we have [9],

$$\begin{bmatrix} \xi - \hat{\xi} \\ u - \hat{u} \end{bmatrix}^T \begin{bmatrix} 0 & (L_f - m_f)I \\ (L_f - m_f)I & -2I \end{bmatrix} \begin{bmatrix} \xi - \hat{\xi} \\ u - \hat{u} \end{bmatrix} \succeq 0.$$

We now employ the KYP lemma to establish global exponential stability of (24a).

Theorem 4: Let Assumption 3 hold. Then proximal gradient flow dynamics (24a) are globally exponentially stable, i.e., there is $\tau, 0 < \rho < m_f$ such that,

$$\|w(t)\,-\,w^\star\|\;\leq\;\tau\,\mathrm{e}^{-\rho t}\,\|w(0)\,-\,w^\star\|$$

.

Proof: The KYP lemma implies global exponential stability if

$$\begin{bmatrix} H_\rho(j\omega) \\ I \end{bmatrix}^* \Pi \begin{bmatrix} H_\rho(j\omega) \\ I \end{bmatrix} \;\prec\; 0,\;\;\forall\,\omega\,\in\,\mathbb{R} \qquad (25)$$

where

$$\begin{aligned} \Pi \;&=\; \begin{bmatrix} 0 & (L_f - m_f)I \\ (L_f - m_f)I & -2I \end{bmatrix} \\ H_\rho(j\omega) \;&=\; H(j\omega\,-\,\rho). \end{aligned}$$

It is easy to show that (25) holds for all $\omega \in \mathbb{R}$ if

$$2\,I\,-\,(L_f\,-\,m_f)\,(H_\rho^*(j\omega)\,+\,H_\rho(j\omega))\;\succ\;0.$$

This condition yields a decoupled family of inequalities,

$$\omega^2\,+\,(m_f\,-\,\rho)^2\,+\,(L_f\,-\,m_f)(m_f\,-\,\rho)\;>\;0 \qquad (26)$$

$$(\omega^2\,-\,b_i(\rho))^2\,+\,c_i(\rho)\,\omega^2\,+\,d_i(\rho)\;>\;0 \qquad (27)$$

which have to hold for all $\omega \in \mathbb{R}$ and for $i = 1, \ldots, n-1$. Condition (26) clearly holds if $\rho \in (0, m_f)$. On the other hand, checking (27) amounts to checking a decoupled family of quadratic inequalities in ω^2 where $b_i(\rho)$, $c_i(\rho)$, and $d_i(\rho)$ are parameters that depend on μ, L_f, m_f, λ_i, and ρ. At $\rho = 0$, these are given by

$$b_i(0) = \lambda_i/\mu,\; c_i(0) = (\lambda_i/\mu + m_f)\,(\lambda_i/\mu + L_f),\; d_i(0) = 0.$$

Positivity of $b_i(0)$ and $c_i(0)$ for each i and continuity of $b_i(\rho)$, $c_i(\rho)$, and $d_i(\rho)$ with respect to ρ imply the existence of $\rho > 0$ that guarantees (27) for each $\omega \in \mathbb{R}$ and each $i = 1, \ldots, n-1$, which completes the proof. ■

V. Concluding remarks

We studied a class of nonsmooth optimization problems in which it is desired to minimize a sum of differentiable and nondifferentiable functions. We employed the tools from control theory to prove exponential convergence of proximal gradient flows in continuous time in the presence of strong convexity. We also proposed a generalized version of the PL condition and established the global convergence of the first-order proximal algorithms in the absence of strong convexity. Moreover, by exploiting the structure of in-network optimization problems, we provided a gradient flow dynamics based on proximal augmented Lagrangian which is well-suited for distributed implementation and showed global exponential stability for strongly convex functions.

References

[1] F. Lin, M. Fardad, and M. R. Jovanović, "Design of optimal sparse feedback gains via the alternating direction method of multipliers," *IEEE Trans. Automat. Control*, vol. 58, no. 9, pp. 2426–2431, 2013.

[2] M. R. Jovanović and N. K. Dhingra, "Controller architectures: tradeoffs between performance and structure," *Eur. J. Control*, vol. 30, pp. 76–91, 2016.

[3] S. Hassan-Moghaddam and M. R. Jovanović, "Topology design for stochastically-forced consensus networks," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 3, pp. 1075–1086, September 2018.

[4] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 123–231, 2013.

[5] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[6] N. K. Dhingra, S. Z. Khong, and M. R. Jovanović, "The proximal augmented Lagrangian method for nonsmooth composite optimization," *IEEE Trans. Automat. Control*, 2018, doi:10.1109/TAC.2018.2867589.

[7] M. Fazlyab, A. Ribeiro, M. Morari, and V. M. Preciado, "Analysis of optimization algorithms via integral quadratic constraints: Nonstrongly convex problems," 2017, arXiv:1705.03615.

[8] B. Hu and P. Seiler, "Exponential decay rate conditions for uncertain linear systems using integral quadratic constraints," *IEEE Trans. Autom. Control*, vol. 61, no. 11, pp. 3631–3637, 2016.

[9] 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.

[10] P. Patrinos, L. Stella, and A. Bemporad, "Forward-backward truncated Newton methods for convex composite optimization," 2014, arXiv:1402.6655.

[11] L. Stella, A. Themelis, and P. Patrinos, "Forward–backward quasi-Newton methods for nonsmooth optimization problems," *Comput. Optim. Appl.*, vol. 67, no. 3, pp. 443–487, 2017.

[12] A. Themelis, L. Stella, and P. Patrinos, "Forward-backward envelope for the sum of two nonconvex functions: Further properties and nonmonotone line-search algorithms," *SIAM J. Optim.*, vol. 28, no. 3, pp. 2274–2303, 2018.

[13] A. Megretski and A. Rantzer, "System analysis via integral quadratic constraints," *IEEE Trans. Autom. Control*, vol. 42, no. 6, pp. 819–830, 1997.

[14] B. T. Polyak, "Gradient methods for minimizing functionals," *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, vol. 3, no. 4, pp. 643–653, 1963.

[15] W. Shi, Q. Ling, G. Wu, and W. Yin, "Extra: An exact first-order algorithm for decentralized consensus optimization," *SIAM J. Optim.*, vol. 25, no. 2, pp. 944–966, 2015.

[16] J. Wang and N. Elia, "A control perspective for centralized and distributed convex optimization," in *Proceedings of the 50th IEEE Conference on Decision and Control*, 2011, pp. 3800–3805.

[17] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multiagent optimization," *IEEE Trans. on Automat. Control*, vol. 54, no. 1, pp. 48–61, 2009.

[18] B. Gharesifard and J. Cortés, "Distributed continuous-time convex optimization on weight-balanced digraphs," *IEEE Trans. Autom. Control*, vol. 59, no. 3, pp. 781–786, 2014.

[19] N. K. Dhingra, S. Z. Khong, and M. R. Jovanović, "A second order primal-dual method for nonsmooth convex composite optimization," *IEEE Trans. Automat. Control*, 2017, submitted; also arXiv:1709.01610.

[20] H. Karimi, J. Nutini, and M. Schmidt, "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*, 2016, pp. 795–811.

[21] A. Rantzer, "On the Kalman-Yakubovich-Popov lemma," *Syst. Control Lett.*, vol. 28, no. 1, pp. 7–10, 1996.

[22] S. Kia, J. Cortés, and S. Martínez, "Distributed convex optimization via continuous-time coordination algorithms with discrete-time communication," *Automatica*, vol. 55, pp. 254–264, 2015.