

# Global exponential stability of primal-dual gradient flow dynamics based on the proximal augmented Lagrangian: A Lyapunov-based approach

Dongsheng Ding and Mihailo R. Jovanović

**Abstract**—For a class of nonsmooth composite optimization problems with linear equality constraints, we utilize a Lyapunov-based approach to establish the global exponential stability of the primal-dual gradient flow dynamics based on the proximal augmented Lagrangian. The result holds when the differentiable part of the objective function is strongly convex with a Lipschitz continuous gradient; the non-differentiable part is proper, lower semi-continuous, and convex; and the matrix in the linear constraint is full row rank. Our quadratic Lyapunov function generalizes recent result from strongly convex problems with either affine equality or inequality constraints to a broader class of composite optimization problems with nonsmooth regularizers and it provides a worst-case lower bound of the exponential decay rate. Finally, we use computational experiments to demonstrate that our convergence rate estimate is less conservative than the existing alternatives.

## I. INTRODUCTION

Primal-dual gradient flow dynamics belong to a class of Lagrangian-based methods for constrained optimization problems. Among other applications, such dynamics have found use in network utility maximization [1], resource allocation [2], distributed optimization [3], and feedback-based online optimization [4] problems. Stability conditions for various forms of the gradient flow dynamics have been proposed since their introduction in the 1950's [5].

Lyapunov-based approach has been an effective tool for studying the stability of primal-dual algorithms starting with the seminal paper of Arrow, Hurwicz, and Uzawa [5]. They utilized a quadratic Lyapunov function to establish the global asymptotic stability of the primal-dual dynamics for strictly convex-concave Lagrangians. This early result was extended to the problems in which the Lagrangian is either strictly convex or strictly concave [6]. A simplified Lyapunov function was also proposed for linearly-convex or linearly-concave Lagrangians. For a projected variant of the primal-dual dynamics that could account for inequality constraints, a Krasovskii-based Lyapunov function was combined with LaSalle's invariance principle to show the global asymptotic stability in [1]. The invariance principle was also specialized to discontinuous Carathéodory systems and a quadratic Lyapunov function was used to show the global asymptotic stability of projected primal-dual gradient flow dynamics under globally-strict or locally-strong convexity-concavity assumptions [7], [8]. Additional information about the utility

of a Lyapunov-based analysis in optimization can be found in a recent reference [9].

In [10], the theory of proximal operators was combined with the augmented Lagrangian approach to solve optimization problems in which the objective function can be decomposed into the sum of the strongly convex term with a Lipschitz continuous gradient and a convex non-differentiable term. By evaluating the augmented Lagrangian along a certain manifold, a continuously differentiable function of both primal and dual variables was obtained. This function was named the proximal augmented Lagrangian and the theory of integral quadratic constraints (IQCs) in the frequency domain was employed to prove the global exponential stability of the resulting primal-dual dynamics [10]. This method yields an evolution model with a continuous right-hand-side even for nonsmooth problems and it avoids the explicit construction of a Lyapunov function. In [11], a quadratic Lyapunov function was used to prove similar properties for a narrower class of problems that involve strongly convex and smooth objective functions with either affine equality or inequality constraints. More recently, this Lyapunov-based result was extended to account for variations in the constraints [12]–[14]. Finally, in [15] for problems with strongly convex objective functions, the theory of IQCs was used 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.

Herein, we utilize a Lyapunov-based approach to establish the global exponential stability of the primal-dual gradient flow dynamics resulting from the proximal augmented Lagrangian [10]. Instead of utilizing the frequency-domain condition [10], we employ the theory of IQCs in the time domain to obtain a quadratic Lyapunov function that establishes the global exponential stability. One benefit is that the Lyapunov function allows us to derive a worst-case lower bound on the exponential decay rate which the frequency-domain characterization [10] does not provide. Furthermore, our Lyapunov function employs a more general parameterization than the one used in [11] or [16] and it yields less conservative convergence rate estimates. We also extend the results of [11] from strongly convex problems with affine equality or inequality constraints to a broader class of optimization problems with nonsmooth regularizers. In contrast to [1], [7], [8], [14], [17], our gradient flow dynamics are projection-free and there are no nonsmooth terms in the Lyapunov function.

The remainder of the paper is organized as follows. In

Financial support from the National Science Foundation under awards ECCS-1708906 and ECCS-1809833 is gratefully acknowledged.

D. Ding and M. R. Jovanović are with the Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA 90089. E-mails: dongshed@usc.edu, mihailo@usc.edu.

Section II, we provide background material, formulate the nonsmooth composite optimization problem, and describe the proximal augmented Lagrangian as well as the resulting primal-dual gradient flow dynamics. In Section III, we construct a quadratic Lyapunov function for verifying the global exponential stability of the primal-dual dynamics. In Section IV, we use computational experiments to illustrate the utility of our results. In Section V, we close the paper with concluding remarks.

## II. PROBLEM FORMULATION AND BACKGROUND

We consider convex composite optimization problems in which the objective function consists of a continuously differentiable term  $f$  and a non-differentiable term  $g$

$$\begin{aligned} & \underset{x, z}{\text{minimize}} && f(x) + g(z) \\ & \text{subject to} && Tx - z = 0 \end{aligned} \quad (1)$$

where  $T \in \mathbb{R}^{m \times n}$  is a matrix that relates the optimization variables  $x \in \mathbb{R}^n$  and  $z \in \mathbb{R}^m$ .

*Assumption 1:* Problem (1) is feasible and its minimum is finite.

*Assumption 2:* The continuously differentiable function  $f$  is  $m_f$ -strongly convex with an  $L_f$ -Lipschitz continuous gradient and the non-differentiable function  $g$  is proper, lower semi-continuous, and convex.

*Assumption 3:* The matrix  $T \in \mathbb{R}^{m \times n}$  has a full row rank.

### A. Proximal augmented Lagrangian

The proximal operator of the function  $g$  is given by [18]

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

and the associated value function is Moreau envelope,

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

where  $\mu$  is a positive parameter. The Moreau envelope is continuously differentiable, even when  $g$  is not, and its gradient is determined by,

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

The augmented Lagrangian of the constrained optimization problem (1) is given by

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

where  $x \in \mathbb{R}^n$  and  $z \in \mathbb{R}^m$  are the primal variables,  $y \in \mathbb{R}^m$  is a dual variable, and  $\mu$  is a positive parameter. Completion of squares brings  $\mathcal{L}(x, z; y)$  into the following form

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

The minimizer of the augmented Lagrangian with respect to  $z$  is

$$z_{\mu}^*(x; y) = \mathbf{prox}_{\mu g}(Tx + \mu y)$$

where  $\mathbf{prox}_{\mu g}$  denotes the proximal operator of the function  $g$ . Restriction of  $\mathcal{L}$  along the manifold determined by  $z_{\mu}^*(x; y)$  yields the proximal augmented Lagrangian [10],

$$\begin{aligned} \mathcal{L}_{\mu}(x; y) &:= \mathcal{L}(x, z_{\mu}^*(x; y); y) \\ &= f(x) + M_{\mu g}(Tx + \mu y) - \frac{\mu}{2} \|y\|^2 \end{aligned} \quad (2)$$

where  $M_{\mu g}$  is the Moreau envelope of the function  $g$ . Continuous differentiability of the proximal augmented Lagrangian  $\mathcal{L}_{\mu}(x; y)$  with respect to both  $x$  and  $y$  follows from continuous differentiability of  $M_{\mu g}$  and Lipschitz continuity of the gradient of  $f$ .

### B. Examples

We next provide examples of convex optimization problems that can be brought into the form (1). For instance, the problem with linear equality constraints,

$$\begin{aligned} & \underset{x}{\text{minimize}} && f(x) \\ & \text{subject to} && Tx = b \end{aligned} \quad (3)$$

where  $b \in \mathbb{R}^m$  is a given vector can be cast as (1) by choosing  $g(z)$  to be an indicator function,  $g(z_i) := \{0, z_i = b_i; \infty, \text{otherwise}\}$ . In this case, the proximal operator is given by  $\mathbf{prox}_{\mu g}(v_i) = b_i$ , the associated Moreau envelope is  $M_{\mu g}(v_i) = \frac{1}{2\mu} (v_i - b_i)^2$ , and the gradient of the Moreau envelope is  $\nabla M_{\mu g}(v_i) = (v_i - b_i)/\mu$ .

The problem with linear inequality constraints,

$$\begin{aligned} & \underset{x}{\text{minimize}} && f(x) \\ & \text{subject to} && Tx \leq b \end{aligned} \quad (4)$$

where  $b \in \mathbb{R}^m$  is a given vector can be cast as (1) by choosing  $g(z)$  to be an indicator function,  $g(z_i) := \{0, z_i \leq b_i; \infty, \text{otherwise}\}$ . The proximal operator is  $\mathbf{prox}_{\mu g}(v_i) = \min\{v_i, b_i\}$ , the associated Moreau envelope is  $M_{\mu g}(v_i) = \{\frac{1}{2\mu} (v_i - b_i)^2, v_i > b_i; 0, \text{otherwise}\}$ , and the gradient of the Moreau envelope is  $\nabla M_{\mu g}(v_i) = \max(0, (v_i - b_i)/\mu)$ .

Unconstrained optimization problems with nonsmooth regularizers can be also represented by (1). For example, the logistic regression with elastic net regularization [19]

$$\underset{x}{\text{minimize}} \quad \ell(x) + \frac{1}{2} \|x\|^2 + \|x\|_1 \quad (5)$$

where the logistic loss  $\ell(x)$  is given by  $\sum_{i=1}^d (\log(1 + e^{a_i^T x}) - y_i a_i^T x)$  where  $a_i$  is the feature vector and  $y_i \in \{0, 1\}$  is the corresponding label. Choosing  $f(x) := \ell(x) + \frac{1}{2} \|x\|^2$ ,  $g(z) := \|z\|_1$ , and  $T := I$  brings (5) into (1). The proximal operator is the soft-thresholding  $\mathbf{prox}_{\mu g}(v_i) = \operatorname{sign}(v_i) \max\{|v_i| - \mu, 0\}$ , the associated Moreau envelope is the Huber function  $M_{\mu g}(v_i) = \{\frac{1}{2\mu} v_i^2, |v_i| \leq \mu; |v_i| - \frac{\mu}{2}, |v_i| \geq \mu\}$ , and the gradient of the Moreau envelope is the saturation function  $\nabla M_{\mu g}(v_i) = \operatorname{sign}(v_i) \min(|v_i|/\mu, 1)$ .

### C. Primal-dual gradient flow dynamics

The primal-dual gradient flow dynamics can be used to compute the saddle points of (2),

$$\dot{w} = F(w) \quad (6a)$$

where  $w := [x^T \ y^T]^T$  and

$$\begin{aligned} F(w) &:= \begin{bmatrix} -\nabla_x \mathcal{L}_\mu(x; y) \\ \nabla_y \mathcal{L}_\mu(x; y) \end{bmatrix} \\ &= \begin{bmatrix} -(\nabla f(x) + T^T \nabla M_{\mu g}(Tx + \mu y)) \\ \mu (\nabla M_{\mu g}(Tx + \mu y) - y) \end{bmatrix}. \end{aligned} \quad (6b)$$

Let  $\bar{w} := [\bar{x}^T \ \bar{y}^T]^T$  denote the equilibrium points of (6), i.e., the solutions to  $F(\bar{w}) = 0$ . Let  $(x^*, z^*)$  and  $y^*$  be the optimal primal and dual points. The Lagrangian of the optimization problem (1) is given by  $f(x) + g(z) + y^T(Tx - z)$  and the associated KKT optimality condition are,

$$\begin{aligned} 0 &= \nabla f(x^*) + T^T y^* \\ 0 &\in \partial g(z^*) - y^* \\ 0 &= Tx^* - z^* \end{aligned} \quad (7)$$

where  $\partial g$  is the subgradient of  $g$ . The following lemma establishes the relation between  $\bar{w}$  and the optimality conditions (7); see [10] for details.

*Lemma 1:* Let Assumptions 1 and 2 hold. The equilibrium point  $\bar{w} := [\bar{x}^T \ \bar{y}^T]^T$  of the primal-dual gradient flow dynamics (6) satisfies optimality conditions (7) with  $\bar{z} := \text{prox}_{\mu g}(T\bar{x} + \mu\bar{y})$ . Moreover,  $(\bar{x}, \bar{z})$  is the optimal solution of nonsmooth composite optimization problem (1).

Under Assumptions 1-3, the global exponential stability of the primal-dual gradient flow dynamics (6) was established in [10] by employing the theory of IQCs in the frequency domain. An upper bound on the convergence rate was also obtained but the explicit form for the quadratic Lyapunov function was not provided. Recent reference [11] used a Lyapunov-based approach to show the global exponential stability for a class of problems with a strongly convex and smooth objective function  $f$  subject to either affine equality or inequality constraints. In our preliminary work [16], a similar quadratic Lyapunov function was used to prove global exponential stability of the primal-dual gradient flow dynamics (6). In what follows, we employ the theory of IQCs in the time domain to obtain a quadratic Lyapunov function that establishes the global exponential stability of (6) and yields less conservative convergence rate estimates.

### III. GLOBAL EXPONENTIAL STABILITY VIA QUADRATIC LYAPUNOV FUNCTION

In this section, we identify a quadratic Lyapunov function that can be used to establish the global exponential stability of the primal-dual gradient flow dynamics (6) for strongly convex problems (1) with full row rank matrix  $T$  and provide an estimate of the convergence rate.

#### A. A system-theoretic viewpoint of primal-dual dynamics

Inspired by [20], [21], we view (6) as a feedback interconnection of an LTI system with static nonlinearities; see Fig. 1. These are determined by the gradient of the smooth part of the objective function  $\nabla f$  and the proximal operator  $\text{prox}_{\mu g}$ . Structural properties of nonlinear terms that we exploit in our analysis are specified in Assumption 2.

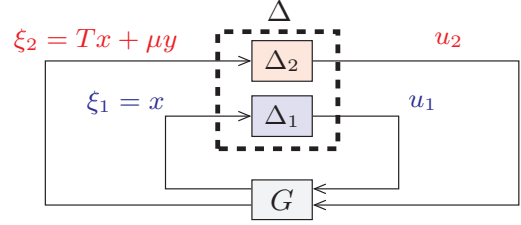


Fig. 1: Block diagram of primal-dual gradient flow dynamics (6):  $G$  is an exponentially stable LTI system in (8a) and  $\Delta$  is a static nonlinear map that satisfies quadratic constraint (9).

Let  $u = [u_1^T \ u_2^T]^T$  and  $\xi = [\xi_1^T \ \xi_2^T]^T$ , with

$$\begin{aligned} \xi_1 &:= x \\ \xi_2 &:= Tx + \mu y \\ u_1 &:= \Delta_1(\xi_1) = \nabla f(x) - m_f x \\ u_2 &:= \Delta_2(\xi_2) = \text{prox}_{\mu g}(Tx + \mu y). \end{aligned}$$

For strongly convex  $f$ , the primal-dual dynamics (6) can be cast as an LTI system  $G$  in feedback with a nonlinear block  $\Delta$ , where

$$\begin{aligned} \dot{w} &= Aw + Bu \\ \xi &= Cw, \quad u = \Delta(\xi) \end{aligned} \quad (8a)$$

with

$$\begin{aligned} A &= \begin{bmatrix} -(m_f I + \frac{1}{\mu} T^T T) & -T^T \\ T & 0 \end{bmatrix} \\ B &= \begin{bmatrix} -I & \frac{1}{\mu} T^T \\ 0 & -I \end{bmatrix}, \quad C = \begin{bmatrix} I & 0 \\ T & \mu I \end{bmatrix}. \end{aligned} \quad (8b)$$

The input is given by  $u = \Delta(\xi)$  where  $\Delta(\xi)$  is a  $2 \times 2$  block-diagonal matrix with the diagonal blocks  $\Delta_1(\xi_1)$  and  $\Delta_2(\xi_2)$ . These nonlinearities satisfy the pointwise quadratic inequalities [10]

$$\begin{bmatrix} \xi_i - \bar{\xi}_i \\ u_i - \bar{u}_i \end{bmatrix}^T \begin{bmatrix} 0 & \hat{L}_i I \\ \hat{L}_i I & -2I \end{bmatrix} \begin{bmatrix} \xi_i - \bar{\xi}_i \\ u_i - \bar{u}_i \end{bmatrix} \geq 0$$

where  $\bar{w} := [\bar{x}^T \ \bar{y}^T]^T$  is the equilibrium point of system (8),  $\bar{\xi}_1 = \bar{x}$ ,  $\bar{\xi}_2 = T\bar{x} + \mu\bar{y}$ ,  $\hat{L}_1 := L_f - m_f$ , and  $\hat{L}_2 = 1$ . This is because  $\Delta_1$  is the gradient of the convex function  $f(\xi_1) - (m_f/2)\|\xi_1\|^2$  and, thus, it is Lipschitz continuous with parameter  $L_f - m_f$  [20, Proposition 5]; and  $\Delta_2$  is given by the proximal operator of the function  $g$  and, thus, it is firmly non-expansive (i.e., Lipschitz continuous with parameter one) [18]. These quadratic constraints can be combined into,

$$\begin{bmatrix} \xi - \bar{\xi} \\ u - \bar{u} \end{bmatrix}^T \underbrace{\begin{bmatrix} 0 & \Pi_0 \\ \Pi_0 & -2\Lambda \end{bmatrix}}_{\Pi} \begin{bmatrix} \xi - \bar{\xi} \\ u - \bar{u} \end{bmatrix} \geq 0 \quad (9)$$

where

$$\Pi_0 = \begin{bmatrix} \lambda_1 \hat{L}_1 I & 0 \\ 0 & \lambda_2 I \end{bmatrix}, \quad \Lambda = \begin{bmatrix} \lambda_1 I & 0 \\ 0 & \lambda_2 I \end{bmatrix}$$

and  $\lambda_1, \lambda_2$  are non-negative scalars.

### B. Lyapunov-based analysis for global exponential stability

For the primal-dual gradient flow dynamics (6) with equilibrium point  $\bar{w}$ , we propose a quadratic Lyapunov function candidate

$$V(\tilde{w}) = \tilde{w}^T P \tilde{w} \quad (10a)$$

with  $\tilde{w} := w - \bar{w}$  and

$$P = \alpha \begin{bmatrix} I & (1/\mu) T^T \\ (1/\mu) T & (1 + m_f/\mu) I + (1/\mu^2) T T^T \end{bmatrix} \quad (10b)$$

where  $\alpha$  is a positive parameter,  $m_f$  is the strong convexity module of the function  $f$ ,  $\mu$  is the augmented Lagrangian parameter, and  $T$  is the full rank matrix associated with the linear equality constraint in (1). The matrix  $P$  is positive definite and, for  $A$  in (8b), we have

$$A^T P + P A = -2\alpha \begin{bmatrix} m_f I & 0 \\ 0 & (1/\mu) T T^T \end{bmatrix} \prec 0. \quad (11)$$

Thus,  $A$  is a Hurwitz matrix and the LTI system in Fig. 1 is exponentially stable. Furthermore, the derivative of  $V$  along the solutions of (8a) is determined by

$$\dot{V} = \begin{bmatrix} \tilde{w} \\ \tilde{u} \end{bmatrix}^T \begin{bmatrix} A^T P + P A & P B \\ B^T P & 0 \end{bmatrix} \begin{bmatrix} \tilde{w} \\ \tilde{u} \end{bmatrix} \quad (12a)$$

where  $\tilde{u} := u - \bar{u}$ , and the substitution of the output equation  $\xi = C w$  in (8a) to (9) yields the quadratic inequality,

$$\begin{bmatrix} \tilde{w} \\ \tilde{u} \end{bmatrix}^T \begin{bmatrix} 0 & C^T \Pi_0 \\ \Pi_0 C & -2\Lambda \end{bmatrix} \begin{bmatrix} \tilde{w} \\ \tilde{u} \end{bmatrix} \geq 0. \quad (12b)$$

The sufficient condition for the global exponential stability of (8) is obtained by adding (12b) to (12a) and it amounts to the existence of a positive constant  $\rho$  such that

$$\begin{bmatrix} -(A^T P + P A + 2\rho P) & -(P B + C^T \Pi_0) \\ -(P B + C^T \Pi_0)^T & 2\Lambda \end{bmatrix} \succ 0. \quad (13)$$

If this condition holds, we have  $\dot{V} \leq -2\rho V$ . Thus,  $V(\tilde{w}(t)) \leq V(\tilde{w}(0)) e^{-2\rho t}$  and since  $P \succ 0$ ,

$$\|\tilde{w}(t)\| \leq \sqrt{\kappa_p} \|\tilde{w}(0)\| e^{-\rho t}, \text{ for all } t \geq 0 \quad (14)$$

where  $\kappa_p$  is the condition number of the matrix  $P$ . Since  $\Lambda \succ 0$ , the remaining task is to verify the existence of the positive parameters  $\alpha, \mu, \lambda_1, \lambda_2$ , and  $\rho$  such that

$$\begin{aligned} & -(A^T P + P A + 2\rho P) \\ & - \frac{1}{2} (P B + C^T \Pi_0) \Lambda^{-1} (P B + C^T \Pi_0)^T \succ 0 \end{aligned} \quad (15)$$

which follows from the application of the Schur complement to (13).

We are now ready to prove the global exponential stability of the primal-dual gradient flow dynamics (6) and provide estimates of the convergence rate  $\rho$  for  $L_f > m_f$ . Similar result can be established for  $L_f = m_f$ .

**Theorem 2:** Let Assumptions 1-3 hold, let  $L_f > m_f$ , and let  $\sigma_{\max}(T)$  be the largest singular value of the matrix  $T$ . Then, the global exponential stability of the primal-dual gradient flow dynamics (6) can be established with Lyapunov

function (10) if the augmented Lagrangian parameter satisfies

$$\mu > \max \left( \frac{L_f - m_f}{4}, \frac{\sigma_{\max}^2(T)}{8m_f} \left( 1 + \sqrt{1 + \frac{16m_f^2}{\sigma_{\max}^2(T)}} \right) \right). \quad (16)$$

*Proof:* If (15) holds for  $\rho = 0$ , the continuity of the left-hand side of (15) with respect to  $\rho$  implies the existence of  $\rho > 0$  such that (15) holds. For  $\rho = 0$ , (15) becomes

$$-(A^T P + P A) - \frac{1}{2} (P B + C^T \Pi_0) \Lambda^{-1} (P B + C^T \Pi_0)^T \succ 0 \quad (17)$$

where  $A^T P + P A$  is given by (11), and  $P B + C^T \Pi_0$  reads

$$\begin{bmatrix} (\lambda_1 \hat{L}_1 - \alpha) I & \lambda_2 T^T \\ -(\alpha/\mu) T & (\mu\lambda_2 - \alpha(1 + \alpha m_f/\mu)) I \end{bmatrix}$$

where  $\hat{L}_1 := L_f - m_f > 0$ . Thus, the matrix  $M := \frac{1}{2} (P B + C^T \Pi_0) \Lambda^{-1} (P B + C^T \Pi_0)^T$  is given by

$$M = \begin{bmatrix} M_1 & M_0^T \\ M_0 & M_2 \end{bmatrix} \quad (18)$$

where

$$\begin{aligned} M_1 &= \frac{1}{2} \left( \frac{(\alpha - \lambda_1 \hat{L}_1)^2}{\lambda_1} I + \lambda_2 T^T T \right) \\ M_2 &= \frac{1}{2} \left( \frac{\alpha^2}{\lambda_1 \mu^2} T T^T + \frac{1}{\lambda_2} (\mu\lambda_2 - \alpha(1 + \frac{m_f}{\mu}))^2 I \right) \\ M_0 &= \frac{1}{2} \left( \frac{\alpha(\alpha - \lambda_1 \hat{L}_1)}{\lambda_1 \mu} + \mu\lambda_2 - \alpha(1 + \frac{m_f}{\mu}) \right) T \end{aligned}$$

and  $\alpha, \lambda_1, \lambda_2$ , and  $\mu$  are positive parameters that have to be selected such that (17) holds. Setting  $\lambda_1 := \alpha/\hat{L}_1$  and  $\lambda_2 := (\alpha/\mu)(1 + m_f/\mu)$  yields  $M_0 = 0$  and (17) simplifies to

$$\begin{bmatrix} 2\alpha m_f I - \frac{\lambda_2}{2} T T^T & 0 \\ 0 & \frac{\alpha}{\mu} (2 - \frac{\alpha}{2\lambda_1 \mu}) T T^T \end{bmatrix} \succ 0$$

or, equivalently,

$$4\alpha m_f > \lambda_2 \sigma_{\max}^2(T) \text{ and } 4\mu\lambda_1 > \alpha.$$

Combining these two conditions with the above definitions of  $\lambda_1$  and  $\lambda_2$  yields (16). ■

We next utilize the choices of parameters  $\lambda_1$  and  $\lambda_2$  in Theorem 2 to estimate the convergence rate  $\rho$ .

**Proposition 3:** Let Assumptions 1-3 hold, let  $L_f > m_f$ , and let  $\sigma_{\min}(T)$  and  $\sigma_{\max}(T)$  be the smallest and the largest singular values of the matrix  $T$ . Then, the primal-dual gradient flow dynamics (6) are globally exponentially stable with the rate

$$\rho \geq \rho_0(\mu) := \frac{\sigma_{\min}^2(T)}{2(\mu + m_f + \sigma_{\max}^2(T)/\mu)} \quad (19a)$$

if  $\mu > \max(L_f - m_f, \hat{\mu})$ , where

$$\hat{\mu} := \inf \{ \mu \in [\sigma_{\max}(T), \infty), \beta(\mu) < 2m_f \} \quad (19b)$$

$$\beta(\mu) := \frac{(m_f + \mu)\sigma_{\max}^2(T)}{2\mu^2} + \frac{2\rho_0(\mu)(\mu + 4\rho_0(\mu))}{\mu}. \quad (19c)$$

*Proof:* See Appendix A. ■

In Proposition 3, we show that for given problem parameters  $\{m_f, L_f, \sigma_{\max}(T), \sigma_{\min}(T)\}$ , if the augmented La-

grangian parameter  $\mu$  is properly selected, then primal-dual gradient flow dynamics (6) converge to the optimal solution with an exponential rate (see (14)). This holds for any initial point and the worst-case rate is given by  $\rho_0(\mu)$ . This complements the result of [10, Remark 4] that provides an upper bound on the rate. In Section IV, for the quadratic program we demonstrates that our estimated convergence rate is less conservative that those provided in [11], [16]. This improved rate may provide better convergence estimates for the discretized implementation of the continuous-time gradient flow dynamics (6) (see [22, Theorem 1]).

#### IV. COMPUTATIONAL EXPERIMENTS

We next provide an example to demonstrate the merits of our approach. Let us consider optimization problem (1) with,

$$\begin{aligned} f(x) &= \frac{1}{2} x^T Q x + q^T x \\ g(z) &= \begin{cases} 0, & z \leq b \\ \infty, & \text{otherwise} \end{cases} \end{aligned} \quad (20)$$

where  $x$  and  $q$  are the  $n$ -dimensional vectors,  $Q \in \mathbb{R}^{n \times n}$  is a positive definite matrix,  $T \in \mathbb{R}^{m \times n}$  if a full row rank matrix, and  $b \in \mathbb{R}^m$  is a given vector. The gradient of the Moreau envelope is determined by  $\nabla M_{\mu g}(v_i) = \max(0, (v_i - b_i)/\mu)$  and  $(L_f, m_f)$  are the largest and the smallest eigenvalues of the matrix  $Q$ , respectively.

We use Matlab ODE solver ode45 to simulate the primal-dual gradient flow dynamics (6) and set  $n = m = 10$ ,  $q = 10 \times \text{randn}(n, 1)$ , and  $Q = HH^T + K$ , where  $H = \text{randn}(n, n)$  and  $K = \text{diag}(\exp(\text{randn}(n, 1)))$ . We choose  $b$  to be a vector of all ones, set  $T = I$ , and report results for  $(L_f, m_f) = (1.24, 1.03)$  in Fig. 2 and Fig. 3.

Fig. 2 demonstrates the exponential convergence of primal dynamics in (6) with  $(L_f, m_f) = (1.24, 1.03)$  for different values of  $\mu$ , where  $x^*$  is obtained via quadprog. We note that the convergence rate decreases when  $\mu$  becomes larger than 2. For a given value of  $\mu$  that satisfies Proposition 3, we use formula (19a) to estimate the lower bound on the convergence rate  $\rho$ . We compare our estimate with [11, Theorem 2] and [16, Theorem 6]. As shown in Fig. 3, Proposition 3 provides a less conservative estimate of the convergence rate than the existing methods. Similar observations can be made for a larger condition number.

The condition number of the matrix  $Q$  is denoted by  $\kappa_q := L_f/m_f$ . To see effect of the condition number, we fix  $m_f$  and increase  $L_f$  to generate five instances of problem (20) with different values of  $\kappa_q$ . We set  $\mu = 50$  and compare three estimates of the convergence rate  $\rho$  in Figure 4. For this value of  $\mu$ , our estimates outperform the alternatives for different condition numbers.

#### V. CONCLUDING REMARKS

In this paper, we use a Lyapunov-based approach to establish global exponential stability of the primal-dual gradient flow dynamics resulting from the proximal augmented Lagrangian framework for nonsmooth composite optimization. We provide a worst-case estimate of the exponential

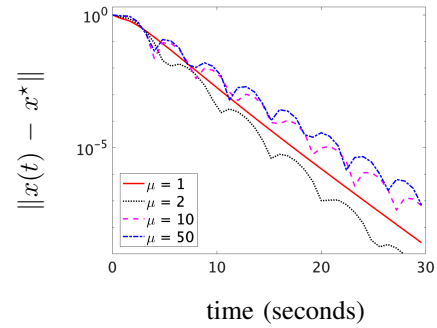


Fig. 2: Convergence of the primal-dual gradient flow dynamics (6) for problem (20).

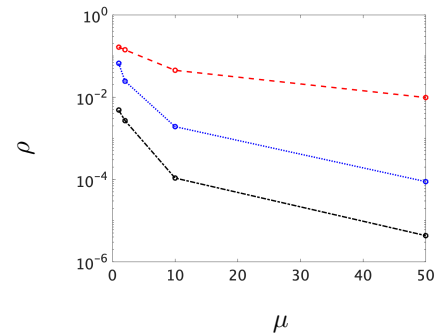


Fig. 3: Convergence rate estimates, as a function of  $\mu$ , resulting from (19a) (---), [16, Theorem 6] (⋯), and [11, Theorem 2] (-·-) for problem (1) with (20).

decay rate when the differentiable part of the objective function is strongly convex and its gradient is Lipschitz continuous. For a quadratic programming problem, computational experiments are used to show that our estimate of the convergence rate is less conservative compared to the existing literature. Our ongoing work focuses on identifying a quadratic Lyapunov function that can certify the global exponential stability of a second-order primal-dual method for nonsmooth composite optimization [23].

#### APPENDIX

##### A. Proof of Proposition 3

We show that (15) holds for  $\rho = \rho_0(\mu)$ . Substitution of the expressions for  $A^T P + PA$  and  $\frac{1}{2}(PB + C^T \Pi_0) \Lambda^{-1} (B^T P + \Pi_0 C)$  given by (11) and (18) into (15) yields

$$R = \begin{bmatrix} R_1 & R_0^T \\ R_0 & R_2 \end{bmatrix} \succ 0 \quad (21)$$

where

$$\begin{aligned} R_1 &= 2\alpha(m_f - \rho)I - \frac{(\alpha - \lambda_1 \hat{L}_1)^2}{2\lambda_1} I - \frac{\lambda_2}{2} T^T T \\ R_2 &= \left(\frac{2\alpha}{\mu} - \frac{\alpha^2}{2\lambda_1 \mu^2}\right) T^T T - 2\rho\alpha\left(1 + \frac{m_f}{\mu}\right)I + \frac{1}{\mu^2} T^T T \\ &\quad - \frac{1}{2\lambda_2} (\mu\lambda_2 - \alpha(1 + \frac{m_f}{\mu}))^2 I \\ R_0 &= -\frac{\alpha}{\mu} \left(\frac{\alpha - \lambda_1 \hat{L}_1}{2\lambda_1} + 2\rho\right) T - \frac{1}{2} (\mu\lambda_2 - \alpha(1 + \frac{m_f}{\mu})) T. \end{aligned}$$

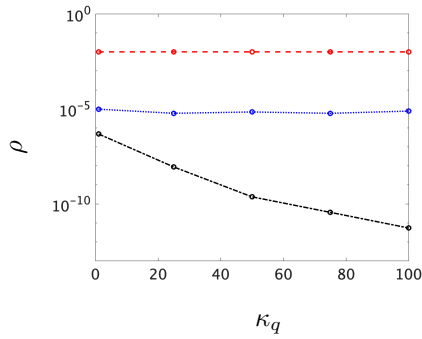


Fig. 4: Convergence rate estimates, as a function of  $\kappa_q$ , resulting from (19a) (---), [16, Theorem 6] (...), and [11, Theorem 2] (—) for problem (1) with (20).

Here,  $\hat{L}_1 := L_f - m_f > 0$ , and  $\alpha$ ,  $\lambda_1$ ,  $\lambda_2$ , and  $\mu$  are positive parameters that have to be selected such that (15) holds for  $\rho = \rho_0(\mu)$ .

We set  $\lambda_1 := \alpha/\hat{L}_1$ ,  $\lambda_2 := \alpha(1 + m_f/\mu)/\mu$ , and add/subtract  $3\alpha TT^T/(2\mu)$  to  $R_2$  to obtain

$$R_2 = \alpha\left(\frac{2}{\mu} - \frac{\hat{L}_1}{2\mu^2}\right)TT^T - \frac{3\alpha}{2\mu}TT^T + \frac{\alpha}{\mu}TT^T - 2\rho\alpha\left(1 + \frac{m_f}{\mu}\right)I + \frac{1}{\mu^2}TT^T + \frac{\alpha}{2\mu}TT^T.$$

If  $\mu \geq \hat{L}_1$ , then

$$\alpha\left(\frac{2}{\mu} - \frac{\hat{L}_1}{2\mu^2}\right)TT^T - \frac{3\alpha}{2\mu}TT^T \succeq 0. \quad (22)$$

Furthermore, for  $\rho = \rho_0$ , we have

$$\begin{aligned} \frac{\alpha}{\mu}TT^T - 2\rho\alpha\left(1 + \frac{m_f}{\mu}\right)I + \frac{1}{\mu^2}TT^T &\succeq \\ \frac{\alpha}{\mu}\sigma_{\min}^2(T)I - 2\rho\alpha\left(1 + \frac{m_f}{\mu} + \frac{\sigma_{\max}^2(T)}{\mu^2}\right)I &= 0. \end{aligned} \quad (23)$$

Combining (22) and (23) with the definition of  $R_2$  yields  $R_2 \succeq \alpha TT^T/(2\mu)$  and the positive definiteness of  $R_2$  follows from the fact that  $T$  is a full row rank matrix.

The application of the Schur complement requires  $R_1 - R_0^T R_2^{-1} R_0 \succ 0$ . Using  $R_2 \succeq \alpha TT^T/(2\mu)$ , we can rewrite this condition as  $R_1 - (2\mu/\alpha)R_0^T (TT^T)^{-1} R_0 \succ 0$ . For  $\beta(\mu)$  and  $\hat{\mu}$  given by (19c) and (19b), respectively, if  $\mu > \hat{\mu}$ , we have

$$\begin{aligned} R_1 - (2\mu/\alpha)R_0^T (TT^T)^{-1} R_0 &\succeq \\ \alpha\left(2m_f - (2\rho_0 + \frac{\sigma_{\max}^2(T)}{2\mu})\left(1 + \frac{m_f}{\mu}\right) + \frac{8\rho_0^2}{\mu}\right)I & \\ = \alpha\left(2m_f - \beta(\mu)\right)I &\succ 0. \end{aligned}$$

We now prove the existence of such  $\hat{\mu}$ . Since  $\rho_0(\mu)$  is monotonically decreasing for  $\mu \geq \sigma_{\max}(T)$ ,  $\beta(\mu)$  monotonically decreases to zero on the interval  $\mu \in [\sigma_{\max}(T), \infty)$ . There are two cases:

(i) if  $\beta(\sigma_{\max}(T)) > 2m_f$ , then  $\beta(\bar{\mu}) = 2m_f$  for some  $\bar{\mu} \in [\sigma_{\max}(T), \infty)$  and  $\hat{\mu} = \bar{\mu}$ ;

(ii) if  $\beta(\sigma_{\max}(T)) \leq 2m_f$ , then  $\beta(\mu) < \beta(\sigma_{\max}(T)) \leq 2m_f$  for all  $\mu \in (\sigma_{\max}(T), \infty)$ . Thus,  $\hat{\mu} = \sigma_{\max}(T)$ , and the set in (19b) is nonempty and such  $\hat{\mu}$  always exists.

## REFERENCES

- [1] D. Feijer and F. Paganini, "Stability of primal-dual gradient dynamics and applications to network optimization," *Automatica*, vol. 46, no. 12, pp. 1974–1981, 2010.
- [2] D. Ding and M. R. Jovanović, "A primal-dual Laplacian gradient flow dynamics for distributed resource allocation problems," in *Proceedings of the 2018 American Control Conference*, 2018, pp. 5316–5320.
- [3] 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.
- [4] M. Colombino, E. Dall'Anese, and A. Bernstein, "Online optimization as a feedback controller: stability and tracking," *IEEE Trans. Control Netw. Syst.*, vol. 7, no. 1, pp. 422–432, 2020.
- [5] K. J. Arrow, L. Hurwicz, and H. Uzawa, *Studies in linear and non-linear programming*. Stanford University Press, 1958.
- [6] A. Cherukuri, B. Ghahsifard, and J. Cortés, "Saddle-point dynamics: conditions for asymptotic stability of saddle points," *SIAM J. Control Optim.*, vol. 55, no. 1, pp. 486–511, 2017.
- [7] A. Cherukuri, E. Mallada, and J. Cortés, "Asymptotic convergence of constrained primal-dual dynamics," *Syst. Control Lett.*, vol. 87, pp. 10–15, 2016.
- [8] A. Cherukuri, E. Mallada, S. Low, and J. Cortés, "The role of convexity on saddle-point dynamics: Lyapunov function and robustness," *IEEE Trans. Automat. Control*, vol. 63, no. 8, pp. 2449–2464, 2018.
- [9] B. Polyak and P. Shcherbakov, "Lyapunov functions: An optimization theory perspective," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 7456–7461, 2017.
- [10] N. K. Dhingra, S. Z. Khong, and M. R. Jovanović, "The proximal augmented Lagrangian method for nonsmooth composite optimization," *IEEE Trans. Automat. Control*, vol. 64, no. 7, pp. 2861–2868, 2019.
- [11] G. Qu and N. Li, "On the exponential stability of primal-dual gradient dynamics," *IEEE Control Syst. Lett.*, vol. 3, no. 1, pp. 43–48, 2019.
- [12] Y. Tang, G. Qu, and N. Li, "Semi-global exponential stability of primal-dual gradient dynamics for constrained convex optimization," *Syst. Control Lett.*, vol. 144, p. 104754, 2020.
- [13] X. Chen and N. Li, "Exponential stability of primal-dual gradient dynamics with non-strong convexity," in *Proceedings of the 2020 American Control Conference*, 2020, pp. 1612–1618.
- [14] Z. Wang, W. Wei, C. Zhao, Z. Zheng, Y. Zhang, and F. Liu, "Exponential stability of partial primal-dual gradient dynamics with nonsmooth objective functions," *arXiv preprint arXiv:2003.07714*, 2020.
- [15] S. Hassan-Moghaddam and M. R. Jovanović, "Proximal gradient flow and Douglas-Rachford splitting dynamics: global exponential stability via integral quadratic constraints," *Automatica*, 2020, to appear; also arXiv:1908.09043.
- [16] D. Ding and M. R. Jovanović, "Global exponential stability of primal-dual gradient flow dynamics based on the proximal augmented Lagrangian," in *Proceedings of the 2019 American Control Conference*, 2019, pp. 3414–3419.
- [17] J. Cortés and S. K. Niederländer, "Distributed coordination for nonsmooth convex optimization via saddle-point dynamics," *J. Nonlinear Sci.*, pp. 1–26, 2018.
- [18] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.
- [19] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. R. Stat. Soc. B*, vol. 67, no. 2, pp. 301–320, 2005.
- [20] 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.
- [21] B. Hu and P. Seiler, "Exponential decay rate conditions for uncertain linear systems using integral quadratic constraints," *IEEE Trans. Automat. Control*, vol. 61, no. 11, pp. 3631–3637, 2016.
- [22] D. Ding, B. Hu, N. K. Dhingra, and M. R. Jovanović, "An exponentially convergent primal-dual algorithm for nonsmooth composite minimization," in *Proceedings of the 57th IEEE Conference on Decision and Control*, Miami, FL, 2018, pp. 4927–4932.
- [23] 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, conditionally accepted; also arXiv:1709.01610.