2019 American Control Conference (ACC)
Philadelphia, PA, USA, July 10-12, 2019

Global exponential stability of primal-dual gradient flow dynamics based on the proximal augmented Lagrangian

Dongsheng Ding and Mihailo R. Jovanović

Abstract— **The primal-dual gradient flow dynamics based on the proximal augmented Lagrangian were introduced in [1] to solve nonsmooth composite optimization problems with a linear equality constraint. We use a Lyapunov-based approach to demonstrate global exponential stability of the underlying dynamics when the differentiable part of the objective function is strongly convex and its gradient is Lipschitz continuous. This also allows us to determine a bound on the stepsize that guarantees linear convergence of the discretized algorithm.**

Index Terms— **Convex optimization, global exponential stability, Lyapunov functions, non-smooth optimization, primal-dual gradient flow dynamics, proximal augmented Lagrangian.**

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 [2], resource allocation [3], and distributed optimization [4] problems. Stability conditions for various forms of the gradient flow dynamics have been proposed since their introduction in the 1950's [5].

For strictly convex-concave Lagrangians, the authors of [5] analyzed the global asymptotic stability of the primal-dual dynamics. Recently, this result was extended to cases in which the Lagrangian is either strictly convex or strictly concave [6]. A relaxed condition was also proposed for linearly-convex or linearly-concave Lagrangians. The invariance principle was also employed to prove global asymptotic stability of a projected variant of the primal-dual dynamics that could handle inequality constraints [2]. In [7], [8], the invariance principle was specialized to discontinuous Caratheodory systems, and was used to show global asymptotic stability of projected primal-dual gradient flow dynamics under globally-strict or locally-strong convexity-concavity assumptions.

Recently, focus has shifted to establishing conditions for the exponential stability of primal-dual gradient flow dynamics. The authors of [9] proposed a nonsmooth Lyapunov function to demonstrate the global exponential stability of discontinuous primal-dual gradient flow dynamics under mild convexity and regularity conditions. In [1], the theory of Integral Quadratic Constraints (IQCs) was employed to prove global exponential stability of the primal-dual dynamics for composite optimization problems that involve strongly convex smooth components with Lipschitz continuous gradients. This method avoids the explicit construction of Lyapunov functions and restricts the choice of the augmented Lagrangian parameter in algorithmic design. In [10], a similar result was presented using a quadratic Lyapunov function for a narrower class of problems that involve strongly convex and smooth objective functions with either affine equality or inequality constraints. For equality constrained optimization problems, the primal-dual dynamics were shown to be strictly contractive using a Riemannian metric in [11].

Herein, we use a Lyapunov-based approach to establish the global exponential stability of the primal-dual gradient flow dynamics resulting from the proximal augmented Lagrangian. This method was proposed in [1] to solve nonsmooth composite optimization problems with a linear equality constraint. When the objective function is strongly convex and its gradient is Lipschitz continuous, we build on [10] to provide explicit construction of Lyapunov functions for the primal-dual dynamics resulting from the proximal augmented Lagrangian. We also provide an estimate for the exponential decay rate. In contrast to [2], [7]–[9], our gradient flow dynamics are projection-free and we do not involve nonsmooth terms in constructing Lyapunov functions. Moreover, we guarantee linear convergence of the discretized dynamics by providing an upper bound for stepsize selection.

This work builds on recent references [1], [10] and utilizes a Lyapunov-based approach to demonstrate global exponential stability of the primal-dual gradient flow dynamics based on the proximal augmented Lagrangian. In [1], the global exponential stability of such dynamics was established using frequency-domain equivalent of the condition developed in [12, Theorem 3] but no Lyapunov function was provided. Our paper complements these earlier results and provides a quadratic Lyapunov function that certifies global exponential stability. This extends recent result [10] from strongly convex optimization problems with either affine equality or inequality constraints to a broader class of composite optimization problems with nonsmooth regularizers.

The remainder of the paper is organized as follows. In Section II, we describe the proximal augmented Lagrangian and the resulting primal-dual gradient flow dynamics. In Section III, we propose a Lyapunov function to establish the global exponential stability of the primal-dual dynamics. We also consider the discretized dynamics and provide an upper bound for stepsize selection. In Section IV, we use computational experiments to illustrate our findings. We close the paper in Section V with concluding remarks.

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.

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.

978-1-5386-7928-9/$31.00 ©2019 AACC

II. Problem formulation and background

We consider the nonsmooth convex optimization problem

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

where the objective function consists of a continuously differentiable term f and a non-differentiable term g, and $T \in \mathbb{R}^{m\times n}$ is a matrix that relates the optimization variables $x \in \mathbb{R}^n$ and $z \in \mathbb{R}^m$. We assume that the problem (1) is feasible and that its minimum is finite.

Assumption 1: Let f be an m_f-strongly convex continuously differentiable function with an L_f-Lipschitz continuous gradient ∇f and let g be a proper, lower semi-continuous, and convex non-differentiable function.

Assumption 2: Let $T \in \mathbb{R}^{m\times n}$ be a full row rank matrix with $l_0 I \preceq TT^T \preceq u_0 I$.

The augmented Lagrangian of (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 primal variables, $y \in \mathbb{R}^m$ is a dual variable, and μ is a positive parameter. The minimizer of the augmented Lagrangian with respect to z is

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

where $\mathbf{prox}_{\mu g}$ denotes the proximal operator of the function g. Restriction of $\mathcal{L}$ along the manifold determined by $z^\star_\mu(x;y)$ yields the proximal augmented Lagrangian [1],

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

where $M_{\mu g}$ is the Moreau envelope of the function g,

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

The Moreau envelope is continuously differentiable, even when g is not, and its gradient is determined by [13],

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

This implies continuous differentiability, with respect to both x and y, of the proximal augmented Lagrangian [1].

For example, when $g(z) = \|z\|_1$, the proximal operator is the soft-thresholding $\mathbf{prox}_{\mu g}(v_i) = \text{sign}(v_i)\max\{|v_i| - \mu, 0\}$, the associated Moreau envelope is the Huber function $M_{\mu g}(v_i) = \{\frac{1}{2\mu}v_i^2, |v_i| \,\le\, \mu; |v_i| - \frac{\mu}{2}, |v_i| \,\ge\, \mu\}$, and the gradient of the Moreau envelope is the saturation function, $\nabla M_{\mu g}(v_i) = \text{sign}(v_i)\min\left(|v_i|/\mu,\, 1\right)$.

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

$$\dot{w} \,=\, F(w) \tag{3}$$

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

$$\begin{aligned} F(w) \,&:=\, \left[\begin{array}{c} -\nabla_x \mathcal{L}_\mu(x;y) \\ \nabla_y \mathcal{L}_\mu(x;y) \end{array}\right] \\ &=\, \left[\begin{array}{c} -(\nabla f(x) + T^T \nabla M_{\mu g}(Tx + \mu y)) \\ \mu\,(\nabla M_{\mu g}(Tx + \mu y) \,-\, y) \end{array}\right]. \end{aligned}$$

Let $\bar{w} := [\, \bar{x}^T \; \bar{y}^T \,]^T$ denote the equilibrium point of primal-dual dynamics (3), i.e., the solution to $F(\bar{w}) \,=\, 0$. The following lemma characterizes the relation between $\bar{w}$ and the KKT optimality conditions for (1) and it is borrowed from [1].

Lemma 1: Let Assumption 1 hold. The equilibrium point $\bar{w}$ of primal-dual gradient flow dynamics (3) satisfies the KKT condition of problem (1). Moreover, $(\bar{x}, \bar{z}) \,=\, (\bar{x}, \mathbf{prox}_{\mu g}(T\bar{x} + \mu\bar{y}))$ is the optimal solution of (1).

Under Assumptions 1 and 2, the global exponential stability of the primal-dual gradient flow dynamics (3) was established in [1] using the theory of IQCs in the frequency domain. The convergence rate estimates were also provided. The recent reference [10] 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 this paper, we provide a quadratic Lyapunov function that can be used to prove global exponential stability of the primal-dual gradient flow dynamics (3).

III. Global exponential stability

Herein, we employ the Lyapunov-based approach to establish the global exponential stability of primal-dual gradient flow dynamics (3). We propose a quadratic Lyapunov function and identify conditions under which its derivative is upper bounded by a negative definite quadratic form.

We first state two technical lemmas that are useful for proving the main result.

Lemma 2: Let Assumption 1 hold. Then, for any $x, \bar{x} \in \mathbb{R}^n$, there exists a symmetric matrix $D_{x,\bar{x}}$ satisfying $0 \preceq D_{x,\bar{x}} \preceq I$ such that $\mathbf{prox}_{\mu g}(x) - \mathbf{prox}_{\mu g}(\bar{x}) = D_{x,\bar{x}}(x - \bar{x})$.

Proof: See [14, Lemma 5]. ■

Lemma 3: Let Assumption 1 hold. Then, for any $x, \bar{x} \in \mathbb{R}^n$ there exists a symmetric matrix $B_{x,\bar{x}}$ satisfying $m_f I \preceq B_{x,\bar{x}} \preceq L_f I$ such that $\nabla f(x) - \nabla f(\bar{x}) = B_{x,\bar{x}}\,(x - \bar{x})$.

Proof: See [14, Lemma 6]. ■

Using Lemmas 2 and 3, we rewrite the primal-dual gradient flow dynamics (3),

$$\dot{\tilde{w}} \,=\, F(w) \,-\, F(\bar{w}) \,=\, G_{\tilde{w}}\,\tilde{w} \tag{4a}$$

where $\tilde{w} := w - \bar{w}$ and the dependence of the matrix

$$G_{\tilde{w}} \,=\, \left[\begin{array}{cc} -B - \frac{1}{\mu}T^T(I-D)T & -T^T(I-D) \\ (I-D)T & -\mu D \end{array}\right] \tag{4b}$$

on $\tilde{w}$ originates from the dependence of B and D on x and $\bar{x}$. For notational convenience, we have suppressed the dependence of matrices B and D on the operating point. We will do the same in the rest of the paper for the matrix G.

We propose a quadratic Lyapunov function candidate for system (4),

$$V(\tilde{w}) \;=\; \tilde{w}^T P\, \tilde{w} \tag{5a}$$

with

$$P \;:=\; \begin{bmatrix} \alpha I & T^T \\ T & \alpha I \,+\, \beta T T^T \end{bmatrix}. \tag{5b}$$

When Assumption 2 holds, a simple sufficient condition for $P \succ 0$ is given by

$$\alpha \;>\; \sqrt{u_0}, \quad \beta \;\geq\; 0. \tag{6}$$

We next show how to choose parameters α and β so that the Lyapunov function candidate V satisfies

$$\dot{V}(\tilde{w}) \;=\; \tilde{w}^T(G^T P \,+\, PG)\,\tilde{w} \;\leq\; -\rho\, V(\tilde{w}) \tag{7}$$

which establishes a sufficient condition for global exponential stability of nonlinear dynamics (4) at a convergence rate $\rho > 0$.

Let

$$Q \;:=\; -(G^T P \,+\, PG \,+\, \rho P) \;=\; \begin{bmatrix} Q_1 & Q_0^T \\ Q_0 & Q_2 \end{bmatrix} \tag{8}$$

where the conformable partitioning of Q follows the partitioning of matrices G and P. Here,

$$Q_0 \;:=\; TQ_4 + Q_5$$
$$Q_1 \;:=\; 2\alpha B + 2(\alpha/\mu - 1)T^T(I-D)T - \alpha\rho I$$
$$Q_2 \;:=\; Q_3 - \alpha\rho I - \beta\rho TT^T$$

where the matrices Q_3, Q_4, and Q_5 are given by

$$Q_3 \;:=\; TT^T(I-D) + (I-D)TT^T + \mu(\alpha I + \beta TT^T)D + \mu D(\alpha I + \beta TT^T)$$
$$Q_4 \;:=\; B + (1/\mu - \beta)T^T(I-D)T - \rho I$$
$$Q_5 \;:=\; \mu DT.$$

If parameters α, β, and ρ are chosen such that $Q \succeq 0$, then inequality (7) will clearly be satisfied.

Lemma 4: Let Assumption 2 hold. If $\beta \in [0, 1/\mu]$ and $\alpha \geq u_0(1-\mu\beta)/\mu$, then

$$Q_3 \;\succeq\; \frac{\mu\beta \,+\, 3}{2}\, TT^T. \tag{9}$$

Proof: See Appendix A. ■

Lemma 5: Let Assumptions 1 and 2 hold. If $\beta \in [0, 1/\mu]$ and

$$\alpha \;>\; \max\big(u_0(1-\mu\beta)/\mu,\; h_0(\beta,\mu)\big) \tag{10}$$

then $Q \succeq 0$ with

$$\rho \;=\; \frac{l_0\,(1 \,+\, \beta\mu)}{2\,(\alpha \,+\, \beta u_0)}. \tag{11}$$

The positive function $h_0(\beta,\mu)$ is defined in Appendix B.

Proof: See Appendix C. ■

We now combine Lemma 5 with condition (6) to prove the main result.

Theorem 6: Let Assumptions 1 and 2 hold. If $\beta \in [0, 1/\mu]$ and

$$\alpha \;>\; \max\big(\sqrt{u_0},\; u_0(1-\mu\beta)/\mu,\; h_0(\beta,\mu)\big) \tag{12}$$

then the primal-dual gradient flow dynamics (3) are globally exponentially stable with a rate no smaller than ρ given by (11). The definition of the positive function $h_0(\beta,\mu)$ is given in Appendix B.

Proof: Under Assumptions 1 and 2, condition (12) is obtained by combining condition (6) with conditions on α and β from Lemmas 4 and 5. Now $Q \succeq 0$ and (7) imply

$$\dot{V}(\tilde{w}) \;\leq\; -\rho\, V(\tilde{w}).$$

Since $P \succ 0$, $V(\tilde{w}) = \tilde{w}^T P \tilde{w}$ is indeed a Lyapunov function of primal-dual gradient flow dynamics (3) that certifies the global exponential stability with the rate no smaller than ρ given by (11). ■

Remark 1: Note that the conditions in Theorem 6 are sufficient. The convergence rate provided in Eq. (11) identifies the smallest value for a given parameter set $\{\mu, \alpha, \beta\}$. While this rate can be maximized over various parameter values, establishing an explicit form for its maximum value is challenging. Depending on the choice of μ, we consider two extreme cases:

- When $\beta = 1/\mu$, the convergence rate ρ simplifies to
$$\rho \;=\; \frac{\mu l_0}{(\mu\alpha \,+\, u_0)}$$
where $\alpha > \max\big(\sqrt{u_0},\; h_0(1/\mu,\mu)\big)$.
- When $\beta = 0$, the convergence rate ρ becomes
$$\rho \;=\; \frac{l_0}{2\,\alpha}$$
where $\alpha > \max\big(\sqrt{u_0},\; u_0/\mu,\; h_0(0,\mu)\big)$.

Remark 2: For strongly convex optimization problem with linear equality constraints

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & Tx \,-\, b \,=\, 0 \end{array} \tag{13}$$

less conservative conditions for global exponential stability can be derived. Problem (13) can be brought to (1) by defining

$$g(z) \;=\; \begin{cases} 0 & z \,=\, b \\ \infty & \text{otherwise.} \end{cases}$$

The proximal operator is given by $\mathbf{prox}_{\mu g}(v) = b$ and, thus, $D = 0$ in Lemma 2. Under this restriction, the Lyapunov function candidate (5) for system (4) yields $Q_3 = 2TT^T$. Assumption 2 implies $Q_3 \succ 0$, and Lemma 5 can be simplified: if $\beta \geq 0$ and $\alpha > h_0'(\beta,\mu)$, then $Q \succeq 0$ with

$$\rho \;=\; \frac{l_0}{\alpha + \beta u_0} \tag{14}$$

where the positive function $h_0'(\beta,\mu)$ is defined in Appendix B. Using similar argument to Theorem 6, we conclude that the primal-dual gradient flow dynamics (3) are globally exponentially stable with a rate no smaller than ρ given by (14) if $\alpha > \max\left(\sqrt{u_0},\, h_0'(\beta, \mu)\right)$.

Remark 3: The forward Euler discretization of the continuous gradient flow dynamics (3) with stepsize $\delta > 0$ is given by

$$w^{k+1} \;=\; w^k \,+\, \delta\, F(w^k) \tag{15}$$

where k is the iteration index. Since F is Lipschitz continuous, we can use [15, Theorem 1] to determine a range for the stepsize $\delta \in (0, \delta_0)$ that warrants linear convergence. Based on [15, Theorem 1], since δ_0 is obtained from solving

$$\delta \nu^2 \kappa_p \,-\, \rho \mathrm{e}^{-\rho\delta} \;=\; 0$$

then the discretized primal-dual gradient flow dynamics (15) satisfies

$$\|w^k \,-\, \bar{w}\| \;\leq\; \sqrt{\kappa_p}\, r^k\, \|w^0 \,-\, \bar{w}\| \tag{16}$$

where $\bar{w}$ is the equilibrium point of (15),

$$r \;=\; \frac{\delta^2 \nu^2 \kappa_p}{2} \,+\, \mathrm{e}^{-\rho\delta} \tag{17}$$

and ρ is the decay rate estimated by (11). The Lipschitz constant of F is given by [15],

$$\nu \;=\; L_f \,+\, 2\lambda_{\max} \,+\, 2\mu \,+\, \frac{\lambda_{\max}^2}{\mu}$$

where $\lambda_{\max}$ is the largest eigenvalue of TT^T. The condition number of P can be calculated from

$$\kappa_p \;=\; (\alpha \,+\, \bar{\gamma})/(\alpha \,+\, \underline{\gamma}) \tag{18}$$

where

$$\bar{\gamma} \;=\; \max_i \frac{\beta\lambda_i + \sqrt{\beta^2\lambda_i^2 + 4\lambda_i}}{2}$$

$$\underline{\gamma} \;=\; \min_i \frac{\beta\lambda_i - \sqrt{\beta^2\lambda_i^2 + 4\lambda_i}}{2}$$

and λ_i is the ith eigenvalues of TT^T.

IV. Computational Experiments

In this section, we provide examples to demonstrate the convergence of the primal-dual gradient flow dynamics (3) and the discretized algorithm (15).

A. Quadratic optimization problem

Consider a quadratic optimization problem [16],

$$\begin{array}{ll} \underset{x,\, z}{\text{minimize}} & f(x) \,+\, g(z) \\ \text{subject to} & Tx \,-\, z \,=\, 0. \end{array} \tag{19}$$

Here, $f(x) = (1/2)x^T A x + a^T x$, x and a are n-dimensional vectors, $A \in \mathbb{R}^{n\times n}$ is a positive definite matrix, $T \in \mathbb{R}^{m\times n}$, and g is the indicator function given by $g(z) = \{0,\, z \leq b;\ \infty, \text{otherwise}\}$ with $b \in \mathbb{R}^m$. We choose L_f and m_f to be the largest and the smallest eigenvalue of A and note that the gradient of the Moreau envelope is given by $\nabla M_{\mu g}(v_i) = \max(0,\, (v_i - b_i)/\mu)$.

We use Matlab ODE solver ode45 to simulate the continuous-time primal-dual gradient flow dynamics (3). This example is also used to show the linear convergence of the discretized algorithm (15) in [15]. We generate problem instances as follows: we set $n = m = 10$, $a = 10 \times \mathsf{randn}(n, 1)$, and $A = HH^T + K$, where $H = \mathsf{randn}(n, n)$ and $K = \mathsf{diag}(\mathsf{exp}(\mathsf{randn}(n, 1)))$. We choose b to be a vector with all ones, and set $T = I$. We report one problem instance with $L_f = 54.17$ and $m_f = 1.95$ and demonstrate the linear convergence for different μ in Fig. 1.

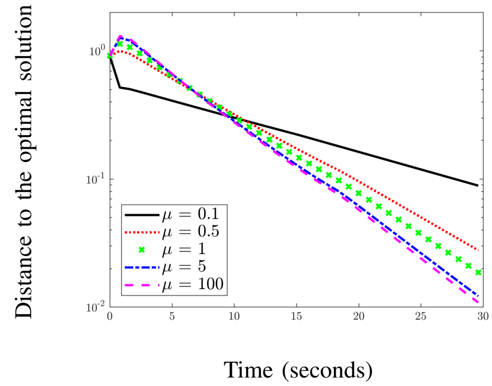

Fig. 1: Problem instance with $L_f = 54.17$ and $m_f = 1.95$.

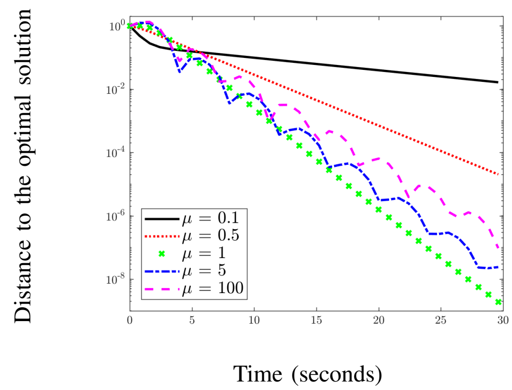

Fig. 2: Problem instance with $L_f = 1.24$ and $m_f = 1.03$.

For comparison, we test some well-conditioned problem instances. We generate problem data in a similar manner except that we rescale the singular values of A to obtain a well-conditioned problem. For $L_f = 1.24$ and $m_f = 1.03$, we demonstrate the linear convergence for different values of μ in Fig. 2.

As shown in Fig. 1, the convergence rate improves when we increase μ. However, in Fig. 2 when μ is increased beyond 1, the convergence becomes slow. For a given $\mu > 0$, we can use results of Theorem 6 to estimate the convergence rate ρ. For instance, when $\mu = 1$, for $L_f = 1.24$ and $m_f = 1.03$, we can use (11) to compute rate ρ as a function of α and β. The estimated rate ρ is given as its peak 0.13. Thus, practical rate should be larger than 0.13 and it is about 0.58 as shown in Fig. 2.

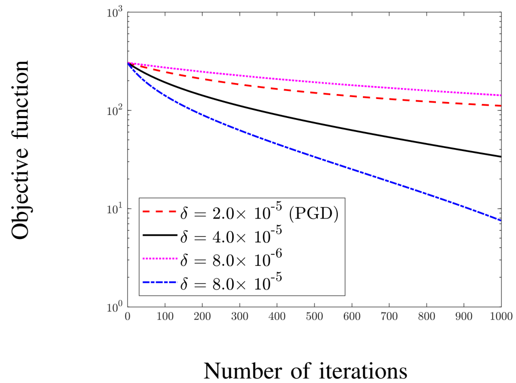

Fig. 3: Objective function for a problem instance.

B. Elastic net logistic regression

Consider the logistic regression with elastic net regularization [17],

$$\underset{x}{\text{minimize}} \quad \ell(x) \,+\, (1/2)\lambda_2\|x\|^2 \,+\, \lambda_1\|x\|_1 \tag{20}$$

where $x \in \mathbb{R}^n, \lambda_1, \lambda_2 > 0$ and the logistic loss $\ell(x)$ is $\sum_{i=1}^{d} -y_i a_i^T x + \log(1 + \mathrm{e}^{a_i^T x})$ where the elements of the vectors a_i are i.i.d. with $\mathcal{N}(0,1)$, and labels $y_i \in \{0,1\}$ are generated by the logistic model: $P(Y_i = 1) = 1/(1 + \mathrm{e}^{-a_i^T x_i^0})$, where x_i^0 are realizations of i.i.d. with $\mathcal{N}(0, 1/100)$. If we denote $T = I$, $f(x) = \ell(x) + (1/2)\lambda_2\|x\|^2$, and $g(z) = \lambda_1\|z\|_1$, this problem takes the form of (1). We can estimate parameters $m_f = \lambda_2$ and $L_f = \lambda_2 + \sum_{i=1}^{d}\|a_i\|^2$, and we choose $\mu = L_f - m_f$.

We use the discretized algorithm (15) to solve this problem with different step sizes. We report one instance here. We can estimate the step size upper bound δ_0 using Remark 3. In this estimation, we assume $\beta = 0$ and calculate the κ_p and ρ from (18) and (11) respectively. The step size upper bound is 2.1×10^{-8}. Another approach was taken in [15, Theorem 3] which leads to $\delta_0 = 4.0 \times 10^{-5}$. The criterion for designing the stepsize discussed in Remark 3 is conservative and may lead to slow convergence. Figure 3 illustrates the convergence of the algorithm for three different stepsizes and compares them to proximal gradient algorithm [13]. The black line corresponds to the stepsize designed based on [15, Theorem 3] which outperforms our design criteria for this particular example. Our ongoing efforts are directed towards providing less conservative Lyapunov-based characterization for global exponential stability of primal-dual gradient flow dynamics.

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. We provide an estimate of the exponential rate of decay and, for the discretized implementation of the continuous-time dynamics, we derive an upper bound for stepsize that guarantees linear convergence. Computational experiments are used to verify our theoretical findings.

Appendix

A. Proof of Lemma 4

The proof idea comes from [10, Lemma 6]. From Lemma 2, we notice that $0 \preceq D \preceq I$ and it has the eigenvalue value decomposition: $D = U\Sigma U^T$, where Σ is a diagonal matrix with diagonals $\sigma_i \in [0,1]$ and U is an unitary matrix. If we denote $\Gamma := U^T T T^T U$, we have

$$\bar{Q}_3 \;:=\; U^T Q_3 U \;=\; 2\Gamma + 2\mu\alpha\Sigma + (\mu\beta - 1)(\Gamma\Sigma + \Sigma\Gamma).$$

We treat $\bar{Q}_3(\Sigma)$ as a function of Σ and it is a convex combination of elements in $\{\bar{Q}_3(R), R$ is a diagonal matrix and diagonals are $r_i = 0$ or $1\}$. Therefore, it is sufficient to show that

$$\bar{Q}_3(R) \;\succeq\; \frac{\mu\beta + 3}{2}\Gamma \tag{21}$$

holds for any R with $r_i = 0$ or 1.

When $R = I$ or 0, it is easy to check the above formula using the conditions of Lemma 4. Since there exists a permutation that sorts the non-zero entries of R, without loss of generality, we can assume that R is given by $r_1 = \cdots = r_k = 1$ and $r_{k+1} = \cdots = r_n = 0$, with $0 < k < n$.

Let

$$\Gamma \;=\; \begin{bmatrix} \Gamma_1 & \Gamma_0^T \\ \Gamma_0 & \Gamma_2 \end{bmatrix}, \Gamma_1 \in \mathbb{R}^{k \times k}$$

we have

$$\bar{Q}_3(R) \;=\; \begin{bmatrix} 2\mu\alpha I + 2\mu\beta\Gamma_1 & (\mu\beta+1)\Gamma_0^T \\ (\mu\beta+1)\Gamma_0 & 2\Gamma_2 \end{bmatrix}.$$

For any $\mu > 0$, if we choose $0 \le \beta \le \frac{1}{\mu}$ and $\alpha \ge \frac{u_0}{\mu}(1 - \mu\beta)$, we have

$$\bar{Q}_3(R) \;\succeq\; \begin{bmatrix} 2\Gamma_1 & (\mu\beta+1)\Gamma_0^T \\ (\mu\beta+1)\Gamma_0 & 2\Gamma_2 \end{bmatrix} \tag{22a}$$

$$\succeq \quad \tfrac{\mu\beta+3}{2}\,\Gamma \tag{22b}$$

where the inequality (22a) follows from $\|\Gamma_1\| \le \|\Gamma\| \le u_0$ and $\alpha \ge \frac{u_0}{\mu}(1 - \mu\beta)$, and the inequality (22b) follows from $\Gamma \succ 0$ and $0 \le \beta \le \frac{1}{\mu}$. This completes the proof.

B. Functions h_0 and h_0' in Lemma 5 and Theorem 6

$$\begin{aligned}
h_0(\beta,\mu) &= \tfrac{L_f^2}{2m_f} + \tfrac{1}{2m_f}\textstyle\sum_{i=1}^{4} h_i(\alpha;\beta,\mu)\Big|_{\alpha=\alpha_m} \\
h_1(\alpha;\beta,\mu) &= 2L_f\big(\tfrac{u_0}{\mu}(1-\mu\beta) + \tfrac{l_0(1+\mu\beta)}{2(\alpha+\beta u_0)}\big) \\
&\quad + \big(\tfrac{u_0}{\mu}(1-\mu\beta) + \tfrac{l_0(1+\mu\beta)}{2(\alpha+\beta u_0)}\big)^2 \\
h_2(\alpha;\beta,\mu) &= \tfrac{2\mu u_0}{l_0}\big(L_f + \tfrac{u_0}{\mu}(1-\mu\beta) + \tfrac{l_0(1+\mu\beta)}{2(\alpha+\beta u_0)}\big) \\
h_3(\alpha;\beta,\mu) &= \tfrac{\mu^2 u_0}{l_0}
\end{aligned}$$

$$h_4(\alpha;\beta,\mu) = 2u_0 + \tfrac{l_0(1+\mu\beta)}{2}$$

$$h'_0(\beta,\mu) = \tfrac{L_f^2}{2m_f} + \tfrac{1}{2m_f}\textstyle\sum_{i=1,4} h'_i(\alpha;\beta,\mu)\Big|_{\alpha=\sqrt{u_0}}$$

$$h'_1(\alpha;\beta,\mu) = 2L_f(\tfrac{u_0}{\mu}|1-\mu\beta| + \tfrac{l_0}{\alpha+\beta u_0}) + (\tfrac{u_0}{\mu}|1-\mu\beta| + \tfrac{l_0}{\alpha+\beta u_0})^2$$

$$h'_4(\alpha;\beta,\mu) = 2u_0 + l_0$$

where $\mu > 0$ is the augmented Lagrangian parameter, m_f and L_f are defined in Assumption 1, l_0 and u_0 are defined in Assumption 2, and $\alpha_m = \min(u_0(1-\mu\beta)/\mu, \sqrt{u_0})$.

C. Proof of Lemma 5

We use the Schur complement condition [18] to show positive semidefiniteness of Q by checking: (i) $Q_2 \succ 0$; (ii) $Q_1 - Q_0^T Q_2^{-1} Q_0 \succeq 0$.

(i) Using Lemma 4 and Assumption 2, we have

$$\begin{aligned} Q_2 - TT^T &\succeq \tfrac{\mu\beta+1}{2}TT^T - \rho(\alpha I + \beta TT^T) \\ &\succeq (\tfrac{\mu\beta+1}{2}l_0 - \rho(\alpha+\beta u_0))I. \end{aligned}$$

With the choice of ρ in (11), we have $Q_2 \succeq TT^T \succ 0$.

(ii) Using the above lower bound for Q_2, we have

$$\begin{aligned} Q_0^T Q_2^{-1} Q_0 &\preceq Q_0^T (TT^T)^{-1} Q_0 \\ &\preceq Q_4^T Q_4 + 2\|Q_4^T T^T (TT^T)^{-1} Q_5\| I \\ &\quad + \|Q_5^T (TT^T)^{-1} Q_5\| I \end{aligned} \tag{23}$$

where $T^T(TT^T)^{-1}T \preceq I$ and the triangle inequality yields the second inequality.

Next, we further bound the three terms in the upper bound of $Q_0^T Q_2^{-1} Q_0$ in (23). We apply the triangle inequality, sub-multiplicativity of spectral norm, and bounds in Lemma 3 and Assumption 2 on each of the three terms to get the following results:

$$\begin{aligned} Q_4^T Q_4 &\preceq L_f B + h_1(\alpha;\beta,\mu) I \\ 2\|Q_4^T T^T (TT^T)^{-1} Q_5\| I &\preceq h_2(\alpha;\beta,\mu) I \\ \|Q_5^T (TT^T)^{-1} Q_5\| I &\preceq h_3(\alpha;\beta,\mu) I. \end{aligned} \tag{24}$$

Here, $h_1(\alpha;\beta,\mu), h_2(\alpha;\beta,\mu)$, and $h_3(\alpha;\beta,\mu)$ are given in Appendix B.

On the other hand, notice that $0 \preceq T(I-D)T^T \preceq u_0 I$, we put a lower bound on Q_1 as follows

$$\begin{aligned} Q_1 &\succeq 2\alpha B - 2u_0 I - \alpha\rho I \\ &= 2\alpha B - h_4(\alpha;\beta,\mu). \end{aligned} \tag{25}$$

Then, we use upper bounds in (24) for $Q_0^T Q_2^{-1} Q_0$ and lower bound in (25) for Q_1 to bound the Schur complement

$$Q_1 - Q_0^T Q_2^{-1} Q_0 \succeq (2\alpha - L_f)B - \textstyle\sum_{i=1}^4 h_i(\alpha;\beta,\mu) I \tag{26}$$

When $0 \le \beta \le \frac{1}{\mu}$, the last term $\sum_{i=1}^4 h_i(\alpha;\beta,\mu)$ is a strictly decreasing function of α and it is positive. To lower bound (26) above zero, we further require

$$\begin{aligned} Q_1 - Q_0^T Q_2^{-1} Q_0 &\succeq 2\alpha m_f I - L_f^2 I \\ &\quad - \textstyle\sum_{i=1}^4 h_i(\alpha;\beta,\mu) I\Big|_{\alpha=\alpha_m} \end{aligned} \tag{27}$$

where the last evaluation is because of the lower bounds of α from Lemma 4 and (6). The positive semi-definiteness of the lower bound in (27) gives $\alpha \ge h_0(\beta,\mu)$. This completes the verification of condition (ii).

Finally, due to Assumption 2, we can combine results in Lemma 4 to get the condition (10).

References

[1] 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.

[2] 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.

[3] 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*, Milwaukee, WI, 2018, pp. 5316–5320.

[4] 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.

[5] K. J. Arrow, L. Hurwicz, and H. Uzawa, *Studies in linear and nonlinear programming*. Stanford University Press, 1958.

[6] A. Cherukuri, B. Gharesifard, 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] J. Cortés and S. K. Niederländer, "Distributed coordination for nonsmooth convex optimization via saddle-point dynamics," *J. Nonlinear Sci.*, pp. 1–26, 2018.

[10] 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.

[11] H. D. Nguyen, T. L. Vu, K. Turitsyn, and J. Slotine, "Contraction and robustness of continuous time primal-dual dynamics," *IEEE Control Syst. Lett.*, vol. 2, no. 4, pp. 755–760, 2018.

[12] 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.

[13] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[14] N. K. Dhingra, S. Z. Khong, and M. R. Jovanović, "A second order primal-dual method for nonsmooth convex composite optimization," 2017, arXiv:1709.01610.

[15] 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.

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

[17] 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.

[18] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.