2022 American Control Conference (ACC)
Atlanta, USA, June 8-10, 2022

On the noise amplification of primal-dual gradient flow dynamics based on proximal augmented Lagrangian

Hesameddin Mohammadi and Mihailo R. Jovanović

***Abstract*—In this paper, we examine amplification of additive stochastic disturbances to primal-dual gradient flow dynamics based on proximal augmented Lagrangian. These dynamics can be used to solve a class of non-smooth composite optimization problems and are convenient for distributed implementation. We utilize the theory of integral quadratic constraints to show that the upper bound on noise amplification is inversely proportional to the strong-convexity module of the smooth part of the objective function. Furthermore, to demonstrate tightness of these upper bounds, we exploit the structure of quadratic optimization problems and derive analytical expressions in terms of the eigenvalues of the corresponding dynamical generators. We further specialize our results to a distributed optimization framework and discuss the impact of network topology on the noise amplification.**

***Index Terms*—Convex optimization, distributed computation, integral quadratic constraints, linear matrix inequalities, noise amplification, primal-dual dynamics, proximal augmented Lagrangian, saddle-point dynamics.**

I. Introduction

We consider a class of primal-dual gradient flow dynamics based on proximal augmented Lagrangian [1] that can be used for solving large-scale non-smooth constrained optimization problems in continuous time. These problems arise in many areas including signal processing [2], statistical estimation [3], and control [4]. In addition, primal-dual methods have received renewed attention due to their prevalent application in distributed optimization [5] and their convergence and stability properties have been greatly studied [6]–[12].

While gradient-based methods are not readily applicable to non-smooth optimization, we can utilize their proximal counterparts to address such problems [13]. In the context of non-smooth constrained optimization, proximal-based extensions of primal-dual methods can also be obtained using the augmented Lagrangian [1], which preserve structural separability and remain suitable for distributed optimization.

Employing primal-dual algorithms in real-world distributed settings motivates the robustness analysis of such methods as uncertainty can potentially enter the dynamics due to noisy communication channels [14]. Moreover, uncertainties can also arise in applications where the exact value of the gradient is not fully available, e.g., when the objective function is obtained via costly simulations or its computation relies on noisy measurements e.g., real-time and embedded applications.

In this paper, we consider the scenario in which the dynamics of the primal-dual flow are perturbed by additive white noise. We examine the mean-squared error of the primal optimization variable as a measure of how noise gets amplified by the dynamics – we refer to this quantity as *noise* (or *variance*) *amplification.* For convex quadratic optimization problems, the primal-dual flow becomes a linear time invariant system, for which the noise amplification can be characterized using Lyapunov equations. For non-quadratic problems, the flow is no longer linear, however, tools from robust control theory can be utilized to quantify upper bounds on the noise amplification. In particular, we use the theory of Integral Quadratic Constraints (IQC) [15], [16] to characterize upper bounds on the noise amplification of the primal-dual flow based on proximal augmented Lagrangian using solutions to a certain linear matrix inequality. Our results establish tight upper-upper bounds on the noise amplification that are inversely proportional to the strong-convexity module of the corresponding objective function.

The approach taken in this paper is similar to those in [17]–[22], wherein IQCs have been used to analyze convergence and robustness of first-order optimization algorithms and their accelerated variants. The noise amplification of primal-dual methods has also been studied in [14] where the authors have focused on quadratic problems and considered the average error in the objective function. In contrast, we consider the average error in the optimization variable and extend the noise amplification analysis to the case of strongly convex and non-smooth optimization problems. For smooth strongly convex problems, an input-output analysis with a focus on the induced $\mathcal{L}_2$ norm using the passivity theory has been provided in [23]. In contrast, we study stochastic performance of primal-dual algorithms that can be utilized to solve non-smooth composite optimization problems.

The rest of the paper is structured as follows. We describe the proximal-augmented Lagrangian and the noisy primal-dual gradient flow dynamics in Section II. We next study the variance amplification for quadratic problems in Section III. We present our IQC-based approach for general strongly convex but non-smooth optimization problems in Section IV. We study the noise amplification in a distributed optimization setting in Section V, and provide concluding remarks in Section VI.

II. Proximal Augmented Lagrangian

We study a nonsmooth composite 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 $f\colon \mathbb{R}^n \to \mathbb{R}$ is a convex, continuously differentiable function, $g\colon \mathbb{R}^k \to \mathbb{R}$ is a convex, but possibly non-differentiable

This work is supported in part by the National Science Foundation under Awards ECCS-1708906 and ECCS-1809833.

The authors are with the Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA 90089. Emails: {hesamedm, mihailo}@usc.edu.

function, and $T \in \mathbb{R}^{k \times n}$ is a given matrix. The augmented Lagrangian associated with (1) is given by

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

where $\mu > 0$ is a parameter and ν is the Lagrange multiplier. The infimum of the augmented Lagrangian $\mathcal{L}_\mu$ with respect to z is given by the proximal augmented Lagrangian [1]

$$\begin{aligned} \mathcal{L}_\mu(x; \nu) \; &:= \; \inf_z \; \mathcal{L}_\mu(x, z; \nu) \\ &= \; f(x) \, + \, \mathrm{M}_{\mu g}(Tx + \mu\nu) \, - \, \tfrac{\mu}{2}\|\nu\|_2^2 \end{aligned} \tag{2}$$

where $\mathrm{M}_{\mu g}(\xi) := g(\mathbf{prox}_{\mu g}(\xi)) + \frac{1}{2\mu}\|\mathbf{prox}_{\mu g}(\xi) - \xi\|_2^2$ is the Moreau envelope of the function g and

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

is the corresponding proximal operator. In addition, the Moreau envelope is continuously differentiable and its gradient is determined by $\mu \nabla \mathrm{M}_{\mu g}(\xi) = \xi - \mathbf{prox}_{\mu g}(\xi)$.

For convex problems, solving (1) amounts to finding the saddle points of $\mathcal{L}_\mu(x; \nu)$. To this end, continuous differentiability of $\mathcal{L}_\mu(x; \nu)$ was utilized in [1] to introduce associated Arrow-Hurwicz-Uzawa gradient flow dynamics

$$\begin{aligned} \dot{x} \; &= \; -\nabla_x \mathcal{L}_\mu(x; \nu) \\ \dot{\nu} \; &= \; \nabla_\nu \mathcal{L}_\mu(x; \nu) \end{aligned} \tag{3}$$

which is a continuous-time algorithm that performs gradient primal-descent and dual-ascent on the proximal augmented Lagrangian. For $\mathcal{L}_\mu(x; \nu)$ given by (2), gradient flow dynamics (3) take the following form,

$$\begin{aligned} \dot{x} \; &= \; -\nabla f(x) - \tfrac{1}{\mu} T^T(Tx + \mu\nu - \mathbf{prox}_{\mu g}(Tx + \mu\nu)) \\ \dot{\nu} \; &= \; Tx \, - \, \mathbf{prox}_{\mu g}(Tx + \mu\nu). \end{aligned} \tag{4}$$

A. Stability properties

When f is convex with a Lipschitz continuous gradient, and g is proper, closed, and convex, the set of equilibrium points of (4) is characterized by minimizers of problem (1) and is globally asymptotically stable [1, Theorem 2]. Furthermore, when f is strongly convex and T is full-row-rank, there is a unique equilibrium point $(x^\star, \nu^\star)$ which is globally exponentially stable and $(x^\star, z^\star = \mathbf{prox}_{\mu g}(Tx^\star + \mu\nu^\star))$ is the unique optimal solution of problem (1) [8, Theorem 6].

B. Noise amplification

We examine the impact of additive stochastic uncertainties on performance of the primal-dual gradient flow dynamics. In particular, we consider the noisy version of (4),

$$\begin{aligned} \mathrm{d}x \; &= \; -\left(\nabla f(x) \, + \, T^T \nabla \mathrm{M}_{\mu g}(Tx + \mu\nu)\right) \mathrm{d}t \, + \, \mathrm{d}w_1 \\ \mathrm{d}\nu \; &= \; \left(Tx \, - \, \mathbf{prox}_{\mu g}(Tx + \mu\nu)\right) \mathrm{d}t \, + \, \mathrm{d}w_2 \end{aligned} \tag{5}$$

where $\mathrm{d}w_i(t)$ are the increments of independent Wiener processes with covariance $\mathbb{E}[w_i(t) w_i^T(t)] = s_i I t$ and $s_i > 0$ for $i \in \{1, 2\}$. We quantify the noise amplification using [16]

$$J \; = \; \limsup_{T \to \infty} \frac{1}{T} \int_0^T \mathbb{E}[\|x(t) \, - \, x^\star\|_2^2] \, \mathrm{d}t. \tag{6}$$

For quadratic objective functions $f(x) := \frac{1}{2} x^T Q x$, if we let g be the indicator function of the set $\{b\}$ with $b \in R^k$, (5) is a linear time-invariant system and J quantifies the steady-state variance of the error in the optimization variable $x(t) - x^\star$,

$$J \; = \; \lim_{t \to \infty} \mathbb{E}[\|x(t) \, - \, x^\star\|_2^2]. \tag{7}$$

In the next section, we examine this class of problems.

III. Quadratic optimization problems

To provide insight into the noise amplification of the primal-dual gradient flow dynamics, we first examine the special case in which the quadratic objective function $f(x) = \frac{1}{2} x^T Q x$ is strongly convex with $Q = Q^T \succ 0$ and $g(z) = \mathrm{I}_{\{b\}}(z)$, where $\mathrm{I}_\mathcal{S}$ is the indicator function of the set $\mathcal{S}$, i.e., $\mathrm{I}_\mathcal{S}(z) := 0$ for $z \in \mathcal{S}$ and $\mathrm{I}_\mathcal{S}(z) := \infty$ for $z \notin \mathcal{S}$. For this choice of g, optimization problem (1) simplifies to

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & Tx \; = \; b \end{array} \tag{8}$$

and the nonlinear terms in (5) are determined by

$$\nabla f(x) \; = \; Qx, \; \mathbf{prox}_{\mu g}(\xi) \; = \; b, \; \nabla \mathrm{M}_{\mu g}(\xi) \; = \; \tfrac{1}{\mu}(\xi \, - \, b).$$

Hence, (5) simplifies to

$$\begin{aligned} \mathrm{d}x \; &= \; -\left((Q \, + \, \tfrac{1}{\mu} T^T T) x + T^T \nu - \tfrac{1}{\mu} T b\right) \mathrm{d}t \, + \, \mathrm{d}w_1 \\ \mathrm{d}\nu \; &= \; (Tx \, - b) \, \mathrm{d}t \, + \, \mathrm{d}w_2 \end{aligned} \tag{9}$$

In what follows, without loss of generality, we set $b = 0$. In this case, noisy dynamics (5) are described by an LTI system

$$\mathrm{d}\psi \; = \; A \, \psi \, \mathrm{d}t \, + \, \mathrm{d}w \tag{10}$$

where $w := \left[\begin{array}{cc} w_1^T & w_2^T \end{array}\right]^T$ and

$$\psi \; := \; \left[\begin{array}{c} x \, - \, x^\star \\ \nu \, - \, \nu^\star \end{array}\right], \; A \; = \; \left[\begin{array}{cc} -(Q + \frac{1}{\mu} T^T T) & -T^T \\ T & 0 \end{array}\right].$$

For $Q \succ 0$ and a full-row-rank T, A is a Hurwitz matrix and LTI system (10) is stable. Moreover from linearity, it follows that the variance amplification, $J = \lim_{t \to \infty} \mathbb{E}[\|x(t) - x^\star\|_2^2]$, can be computed as

$$J \; = \; \operatorname{trace}(X C^T C) \; = \; \operatorname{trace}(X_1) \tag{11}$$

where $X := \lim_{t \to \infty} \mathbb{E}[\psi(t) \psi^T(t)] = \left[\begin{array}{cc} X_1 & X_2 \\ X_2^T & X_3 \end{array}\right]$ is the steady-state covariance matrix of the state $\psi(t)$ which can be obtained by solving the algebraic Lyapunov equation

$$AX \, + \, XA^T \; = \; -\operatorname{diag}(s_1 I, s_2 I) \tag{12}$$

and $C := \left[\begin{array}{cc} I & 0 \end{array}\right]$. Theorem 1 addresses the special case with $Q = m_f I$ and provides an analytical expression for the variance amplification of the corresponding primal-dual gradient flow dynamics. This result is obtained by computing the steady-state covariance matrix of the state ψ.

Theorem 1: Let $f(x) = \frac{m_f}{2}\|x\|^2$, $g(z) = \mathrm{I}_{\{0\}}(z)$, and T be a full-row-rank matrix in (1). Then, the steady-state variance of the primal optimization variable in (5) with $\mathrm{d}w_i(t)$ being the increments of independent Wiener processes with covariance $\mathbb{E}[w_i(t)w_i^T(t)] = s_i I t$ is determined by

$$J \;=\; \frac{(n-k)s_1}{2m_f} \;+\; \sum_{i\,=\,1}^{k} \frac{s_1 + s_2}{2(m_f + (1/\mu)\sigma_i^2(T))}$$

where $\sigma_i(T)$ is the ith singular values of the matrix T.

Proof: Let $T = U\Sigma V^T$ be the singular value decomposition of the matrix T, with unitary matrices $U \in \mathbb{R}^{k\times k}$ and $V \in \mathbb{R}^{n\times n}$ and $\Sigma = \left[\begin{array}{cc} \Sigma_0 & 0_{k\times(n-k)} \end{array}\right] \in \mathbb{R}^{k\times n}$, with

$$\Sigma_0 \;:=\; \mathrm{diag}\,(\sigma_1,\ldots,\sigma_k) \;\in\; \mathbb{R}^{k\times k}.$$

Multiplication of the Lyapunov equation (12) by $M = \mathrm{diag}\,(V,U)$ and M^T from right and left, respectively, yields

$$\hat{A}\,\hat{X} \;+\; \hat{X}\,\hat{A}^T \;=\; -\mathrm{diag}\,(s_1 I, s_2 I) \tag{13}$$

where

$$\hat{A} \;=\; \left[\begin{array}{cc} -m_f I - \frac{1}{\mu}\Sigma^T\Sigma & -\Sigma^T \\ \Sigma & 0 \end{array}\right]$$

$$\hat{X} \;=\; \left[\begin{array}{cc} \hat{X}_1 & \hat{X}_2 \\ \hat{X}_2^T & \hat{X}_3 \end{array}\right] \;:=\; M^T X M.$$

Finally, it is straightforward to verify that

$$\hat{X}_1 \;=\; \left[\begin{array}{cc} \frac{s_1+s_2}{2}\left(m_f I + \frac{1}{\mu}\,\Sigma_0\Sigma_0\right)^{-1} & 0 \\ 0 & \frac{s_1}{2m_f} I \end{array}\right]$$

$$\hat{X}_2 \;=\; \left[\begin{array}{c} -\frac{s_2}{2}\,\Sigma_0^{-1} \\ 0_{(n-k)\times k} \end{array}\right] \;\in\; \mathbb{R}^{n\times k}$$

$$\hat{X}_3 \;=\; \mathrm{diag}\,(a_1,\ldots,a_k) \;\in\; \mathbb{R}^{k\times k}$$

where $a_i \;=\; \dfrac{s_1 + s_2}{2(m_f + \sigma_i^2/\mu)} + \dfrac{s_2(m_f + \sigma_i^2/\mu)}{2\sigma_i^2}$. The result follows from $J = \mathrm{trace}\,(X_1) = \mathrm{trace}\,(\hat{X}_1)$. ∎

The following corollary is immediate from Theorem 1.

Corollary 1: Under the conditions of Theorem 1, the steady-state variance of the primal optimization variable in (5) is upper bounded by $J \leq (ns_1 + ks_2)/(2m_f)$.

Corollary 1 establishes that, for $\mu > 0$ and a full-row-rank matrix T, the variance of the primal optimization variable in (5) satisfies an upper bound that is independent of T and μ. In addition, using the explicit expression for J provided in Theorem 1, it follows that for any fixed $\mu > 0$, in the limit of $\sigma_{\max}(T) \to 0$ and/or $n/k \to \infty$, the upper bound on the variance amplification J in Corollary 1 becomes exact.

It is also noteworthy that, as demonstrated in the proof of Theorem 1, the dual variable ν may experience an unbounded steady-state variance for $s_2 > 0$ if $\sigma_{\min}(T) \to 0$.

Even though it is challenging to derive an analytical expression for the covariance matrix X for a general strongly convex quadratic objective function f, we next demonstrate that the upper bound in Corollary 1 remains valid.

Theorem 2: Let $f(x) = \frac{1}{2}x^T Q x$ with $Q \succeq m_f I$, $g(z) = \mathrm{I}_{\{0\}}(z)$, and T be a full-row-rank matrix in (1). Then, the steady-state variance of the primal optimization variable in (5) with $\mathrm{d}w_i(t)$ being the increments of independent Wiener processes with covariance $\mathbb{E}[w_i(t)w_i^T(t)] = s_i I t$ satisfies

$$J \;\leq\; \frac{ns_1 + ks_2}{2m_f}. \tag{14}$$

Proof: To quantify J, an alternative method to using the state covariance matrix is to write

$$J \;=\; \mathrm{trace}\,(P\,\mathrm{diag}\,(s_1 I, s_2 I))$$

where P is the observability gramian of system (10)

$$A^T P \,+\, PA \;=\; -C^T C \tag{15}$$

with $C = \left[\begin{array}{cc} I & 0 \end{array}\right]$. Thus, any matrix $P' \succeq P$ satisfies

$$J \;\leq\; \mathrm{trace}\,(P'\mathrm{diag}\,(s_1 I, s_2 I)).$$

To find such a P', we note that A satisfies

$$A^T I + IA \;=\; -2\mathrm{diag}\,(Q + \tfrac{1}{\mu}T^T T, 0) \;\preceq\; -2\lambda_{\min}(Q) C^T C.$$

Dividing this inequality by $2\lambda_{\min}(Q)$ and subtracting from (15) yields

$$A^T(\tfrac{1}{2\lambda_{\min}(Q)}\,I \,-\, P) \,+\, (\tfrac{1}{2\lambda_{\min}(Q)}\,I \,-\, P)A \;\preceq\; 0.$$

Since A is Hurwitz, it follows that $P \preceq \frac{1}{2\lambda_{\min}(Q)}I$, and hence

$$\begin{aligned} J \;&=\; \mathrm{trace}\,(P\,\mathrm{diag}\,(s_1 I, s_2 I)) \\ &\leq\; \frac{1}{2\lambda_{\min}(Q)}\,\mathrm{trace}\,(\mathrm{diag}\,(s_1 I, s_2 I)) \;\leq\; \frac{ns_1 + ks_2}{2m_f}. \end{aligned}$$

∎

IV. Beyond quadratic problems

In this section, we extend our upper bounds on the noise amplification of the primal-dual gradient flow dynamics to problems with a general strongly convex function f, a convex but possibly non-differentiable function g, and a matrix T of an arbitrary rank. Our approach is based on Integral Quadratic Constraints (IQCs) which provide a convex control-theoretic framework for stability and robustness analysis of systems with structured nonlinear components [15]. This framework has been recently used to analyze convergence and robustness of first-order optimization methods [17]–[20]. In what follows, we first demonstrate how IQCs can be combined with quadratic storage functions to characterize upper bounds on the noise amplification of continuous-time dynamical systems via solutions to a certain linear matrix inequality (LMI). We then specialize this result to the primal-dual gradient flow dynamics and establish tight upper bounds on the noise amplification by finding feasible solutions to the associated LMI.

A. An IQC-based approach

As demonstrated in Section IV-B, noisy primal-dual gradient flow dynamics can be represented via feedback interconnection of an LTI system with a static nonlinear component

$$\begin{aligned} \mathrm{d}\psi \;&=\; A\,\psi\,\mathrm{d}t \;+\; B\,u\,\mathrm{d}t \;+\; \mathrm{d}w \\ \begin{bmatrix} z \\ y \end{bmatrix} \;&=\; \begin{bmatrix} C_z \\ C_y \end{bmatrix}\psi, \quad u(t) \;=\; \Delta(y(t)). \end{aligned} \tag{16}$$

Here, $\psi(t)$ is the state, $\mathrm{d}w(t)$ is the increment of a Wiener process with covariance $\mathbb{E}[w(t)w^T(t)] = Wt$, where W is a positive semidefinite matrix, $z(t)$ is the performance output, and $u(t)$ is the output of the nonlinear term $\Delta\colon \mathbb{R}^n \to \mathbb{R}^n$ that satisfies the quadratic inequalities

$$\begin{bmatrix} y \\ \Delta(y) \end{bmatrix}^T \Pi_i \begin{bmatrix} y \\ \Delta(y) \end{bmatrix} \;\geq\; 0 \tag{17}$$

for some matrices Π_i and all $y \in \mathbb{R}^n$.

Lemma 1 utilizes property (17) of the nonlinear mapping Δ and provides an upper bound on the average energy [16]

$$J \;=\; \limsup_{T\to\infty}\, \frac{1}{T}\int_0^T \mathbb{E}\left[\|z(t)\|\right]_2^2 \mathrm{d}t.$$

Lemma 1: Let the nonlinear function $u = \Delta(y)$ satisfy

$$\begin{bmatrix} y \\ u \end{bmatrix}^T \Pi_i \begin{bmatrix} y \\ u \end{bmatrix} \;\geq\; 0 \tag{18}$$

for some matrices Π_i, let P be a positive semidefinite matrix, and let λ_i be nonnegative scalars such that system (16) satisfies

$$\begin{bmatrix} A^TP + PA + C_z^TC_z & PB \\ B^TP & 0 \end{bmatrix} + \sum_i \lambda_i \begin{bmatrix} C_y^T & 0 \\ 0 & I \end{bmatrix} \Pi_i \begin{bmatrix} C_y & 0 \\ 0 & I \end{bmatrix} \;\preceq\; 0. \tag{19}$$

Then the average energy of the performance output in statisticaly steady-state is bounded by $J \leq \mathrm{trace}(PW)$.

The proof of Lemma 1 follows from similar arguments as in [16, Theorem 7.2] and is omitted due to space limitation. Lemma 1 introduces a quadratic storage function, $\psi^TP\psi$, for continuous-time primal-dual gradient flow dynamics. We note that discrete-time variants of this result were used to quantify noise amplification of accelerated first-order optimization algorithms [19, Lemmas 1, 2], [22].

B. State-space representation

We next demonstrate how noisy primal-dual gradient flow dynamics (5) can be brought into the standard state-space form (16). In particular, choosing the state variable

$$\psi \;=\; \begin{bmatrix} x^T & \nu^T \end{bmatrix}^T$$

along with $z := x$ and

$$\begin{aligned} y \;&=\; \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \;:=\; \begin{bmatrix} x \\ Tx + \mu\nu \end{bmatrix} \\ u \;&=\; \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \;:=\; \begin{bmatrix} \nabla f(x) - m_f x \\ \mathbf{prox}_{\mu g}(Tx + \mu\nu) \end{bmatrix} \end{aligned}$$

brings system (5) into the state-space form (16) with

$$A \;=\; \begin{bmatrix} -(m_f I + \frac{1}{\mu}T^TT) & -T^T \\ T & 0 \end{bmatrix} \tag{20a}$$

$$B \;=\; \begin{bmatrix} -I & \frac{1}{\mu}T^T \\ 0 & -I \end{bmatrix}, \quad C_y \;=\; \begin{bmatrix} I & 0 \\ T & \mu I \end{bmatrix}. \tag{20b}$$

and $C_z = \begin{bmatrix} I & 0 \end{bmatrix}$, where m_f is the strong-convexity module of f. We note that the input-output pair (u, y) satisfies the pointwise nonlinear equation $u = \Delta(y)$ with $\Delta = \mathrm{diag}\,(\Delta_1, \Delta_2)$, where

$$\begin{aligned} u_1 \;&=\; \Delta_1(y_1) \;:=\; \nabla f(y_1) - m_f y_1 \\ u_2 \;&=\; \Delta_2(y_2) \;:=\; \mathbf{prox}_{\mu g}(y_2). \end{aligned}$$

It is worth mentioning that for the special case $g(z) = \mathrm{I}_{\{0\}}(z)$, which we considered in our analysis of quadratic problems in Section III, the nonlinear term u_2 vanishes and the primal-dual gradient flow dynamics simplify to

$$\begin{aligned} \mathrm{d}x \;&=\; -\Big(\nabla f(x) \,+\, \tfrac{1}{\mu}T^TTx \,+\, T^T\nu\Big)\,\mathrm{d}t \,+\, \mathrm{d}w_1 \\ \mathrm{d}\nu \;&=\; Tx\,\mathrm{d}t \,+\, \mathrm{d}w_2. \end{aligned} \tag{21}$$

C. Characterizing the structural properties via IQCs

The input-output pairs (y_i, u_i) associated with nonlinear mappings Δ_i satisfy

$$\begin{bmatrix} y_i - y_i' \\ u_i - u_i' \end{bmatrix}^T \pi_i \begin{bmatrix} y_i - y_i' \\ u_i - u_i' \end{bmatrix} \geq 0 \tag{22}$$

where

$$\begin{aligned} \pi_1 \;&:=\; \begin{bmatrix} 0 & (L_f - m_f)I \\ (L_f - m_f)I & -2I \end{bmatrix} \\ \pi_2 \;&:=\; \begin{bmatrix} 0 & I \\ I & -2I \end{bmatrix}. \end{aligned}$$

The above inequalities follow from the facts that Δ_1 is the gradient of the $(L_f - m_f)$-smooth convex function $f(\cdot) - (m_f/2)\|\cdot\|^2$ and that $\Delta_2 = \mathbf{prox}_{\mu g}$ is firmly non-expansive.

To make the above IQCs conform to the required format in Lemma 1, we can employ a suitable permutation combined with a change of variables that utilizes deviations from the optimal solution to obtain the inequalities in (18) with

$$\Pi_1 \;=\; \begin{bmatrix} 0 & 0 & (L_f - m_f)\,I & 0 \\ 0 & 0 & 0 & 0 \\ (L_f - m_f)\,I & 0 & -2I & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{23a}$$

$$\Pi_2 \;=\; \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I \\ 0 & 0 & 0 & 0 \\ 0 & I & 0 & -2I \end{bmatrix}. \tag{23b}$$

D. General convex g

The main result of the paper is presented in Theorem 3. It demonstrates that proximal primal-dual gradient flow dynamics enjoys the same upper bound on noise amplification as the primal-dual gradient flow dynamics for smooth problems.

Theorem 3: Let the function f be m_f-strongly convex and let g be closed, proper, convex. Then, the noise amplification of noisy primal-dual gradient flow dynamics satisfies (14).

Proof: It is easy to verify that $P = pI$, $\lambda_1 = 1/(L_f - m_f)$, $\lambda_2 = 1/\mu$ with $p \geq 1/(2m_f)$ provides a feasible solution to the LMI in Lemma 1 for the system matrices in (20) and matrices Π_1, Π_2 in (23). Thus, the result follows from Lemma 1. ∎

For general strongly convex problems, Theorem 3 establishes the same upper bound on the noise amplification as what we obtained using Lyapunov equations for quadratic problems in Theorem 2. In addition, as we discussed in Section III, this upper bound is tight in the sense that the noise amplification for the quadratic problem in Theorem 1 converges to this upper bound in the limit $\sigma_{\max}(T) \to 0$ and/or as $n/k \to \infty$. Another advantage of the IQC framework is that it does not require the matrix A to be Hurwitz. Therefore, the upper bound established in Theorem 3 holds for any matrix T independent of its rank.

V. Application to distributed optimization

The primal-dual gradient flow dynamics provide a distributed strategy for solving the optimization problem

$$\underset{\theta}{\text{minimize}} \quad \sum_{i=1}^{n} f_i(\theta) \tag{24}$$

where f_i are convex functions [5]. Assuming without loss of generality that $\theta \in \mathbb{R}$, given a connected network with an incidence matrix $E = T^T$, we can assign a different scalar variable x_i to each agent and define the equivalent problem

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \displaystyle\sum_{i=1}^{n} f_i(x_i) \\ \text{subject to} & T x \;=\; 0 \end{array} \tag{25}$$

where the constraint enforces that

$$x \;:=\; \left[\begin{array}{ccc} x_1 & \cdots & x_n \end{array}\right]^T \;\in\; \text{Null}\,(T) \;=\; \{c\mathbb{1} \,|\, c \in \mathbb{R}\}$$

where $\mathbb{1} := [1 \;\cdots\; 1]^T$. Letting $f(x) := \sum_i f_i(x_i)$, the primal-dual gradient flow for solving problem (25) is determined by (21) and, in the absence of noise, it converges to $x = \theta^\star \mathbb{1}$, where $\theta^\star$ is an optimal solution of problem (24). In this formulation, the primal and dual variables x_i and ν_i correspond to the nodes and the edges of the network, respectively.

Theorem 3 provides an upper bound on noise amplification of a distributed primal-dual algorithm

$$J \;\leq\; \frac{n s_1 + k s_2}{2 m_f}$$

for strongly convex problems. Here, k denotes the number of edges in the network and m_f is the strong convexity module of the function f. However, if f lacks strong convexity, then an additive white noise with a full-rank covariance matrix can result in unbounded variance of $x(t)$ as $t \to \infty$.

To see one such example, we can let f_i be constants, in which case the primal-dual gradient flow simplifies to a consensus-type algorithm. In this case, the average mode $a(t) := \frac{1}{n}(\mathbb{1}^T x(t))\mathbb{1}$ experiences a random walk, and its variance

$$J_a \;:=\; \lim_{t \to \infty} \mathbb{E}\left(\|a(t) - \theta^\star \mathbb{1}\|^2\right) \tag{26a}$$

is unbounded. However, the mean-square deviation from the network average

$$\bar{J} \;:=\; \lim_{t \to \infty} \mathbb{E}\left(\|x(t) - a(t)\|^2\right) \tag{26b}$$

becomes a relevant quantity and it can be used in lieu of J to quantify stochastic performance as it remains bounded [24].

Using the fact that $\langle x(t) - a(t), \mathbb{1}\rangle = 0$, this idea can be generalized to the distributed optimization framework by noting that the variance amplification can be split into two terms,

$$J \;=\; J_a \,+\, \bar{J}.$$

To provide insight, let us examine the special case with $f_i(\theta) = \frac{1}{2}m_f(\theta - c_i)^2$, where the agents aim to compute the average of c_i. Although the underlying dynamics are linear in this case, the results of Theorem 1 are not applicable because the matrix T is full row-rank only when the corresponding graph is a tree. However, by eliminating modes from the dual-variable that are not stable, a similar argument as in the proof of Theorem 1 can be used to establish an expression for the noise amplification in the distributed setting in terms of the non-zero eigenvalues λ_i of the Laplacian matrix $\mathbf{L} = T^T T$.

Proposition 1: The noisy primal-dual gradient flow dynamics (9) for solving distributed optimization problem (25) with $f_i(x_i) = \frac{1}{2}m_f(x_i - c_i)^2$ satisfies $J = J_a + \bar{J}$, where

$$J_a \;=\; \frac{s_1}{2m_f}, \quad \bar{J} \;=\; \sum_{i=1}^{n-1} \frac{s_1 + s_2}{2(m_f + \lambda_i(\mathbf{L})/\mu)}$$

and λ_i are the non-zero eigenvalues of the Laplacian matrix $\mathbf{L} = T^T T$ of connected undirected network.

Proof: See Appendix A. ∎

For quadratic optimization problems, Proposition 1 demonstrates that, in addition to the strong-convexity module of the function f, the topology of the network also impacts the variance amplification. In the limit $m_f \to 0$, while the variance of the average mode J_a becomes unbounded, the mean-square deviation from the average mode remains bounded and is captured by the sum of reciprocals of the eigenvalues of the graph Laplacian. This dependence of variance amplification on the spectral properties of $\mathbf{L}$ is identical to the one observed in standard consensus algorithms [19], [24].

VI. Concluding remarks

We have examined the noise amplification of proximal primal-dual gradient flow dynamics that can be used to solve non-smooth composite optimization problems. For quadratic problems, we have employed algebraic Lyapunov equations to establish analytical expressions for the noise amplification. We have also utilized the theory of IQCs to characterize tight upper bounds in terms of a solution to an LMI. Our results show that stochastic performance of the primal-dual dynamics is inversely proportional to the strong-convexity module of the smooth part of the objective function. The ongoing work focuses on examining the impact of network topology on the noise amplification in distributed settings and on extension of our results to discrete-time versions of primal-dual algorithms.

APPENDIX

A. Proof of Proposition 1

Let us without loss of generality assume that $c_i = 0$; using the change of variables $y := T^T \nu$, we obtain that the noisy primal-dual flow satisfies

$$\begin{bmatrix} \mathrm{d}x \\ \mathrm{d}y \end{bmatrix} = \begin{bmatrix} -m_f I - \frac{1}{\mu}\mathbf{L} & -I \\ \mathbf{L} & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \mathrm{d}t + \begin{bmatrix} \mathrm{d}w_1 \\ T^T \mathrm{d}w_2 \end{bmatrix}.$$

Noting that $\mathbf{L}\mathbb{1} = 0$, we can let $\mathbf{L} = V \Lambda V^T$, where $\Lambda = \mathrm{diag}(0, \hat{\Lambda})$ is the diagonal matrix of eigenvalues and the columns of the unitary matrix $V = \begin{bmatrix} \mathbb{1}/\sqrt{n} & U \end{bmatrix}$ are the corresponding eigenvectors. Using the change of variables

$$\hat{x} := U^T x, \quad \hat{y} := U^T y, \quad \hat{\psi}^T = \begin{bmatrix} \hat{x}^T & \hat{y}^T \end{bmatrix}$$

it is easy to verify that

$$\mathrm{d}\hat{\psi} = \begin{bmatrix} -m_f I - \frac{1}{\mu}\hat{\Lambda} & -I \\ \hat{\Lambda} & 0 \end{bmatrix} \hat{\psi} \mathrm{d}t + \begin{bmatrix} \mathrm{d}\hat{w}_1 \\ \mathrm{d}\hat{w}_2 \end{bmatrix}$$

where $\mathrm{d}\hat{w}_1$ and $\mathrm{d}\hat{w}_2$ are the increments of independent Wiener process with covariance $s_1 I t$ and $s_2 \hat{\Lambda} t$, respectively. In addition, the average modes associated with the primal and dual variables $a = (x^T \mathbb{1})\mathbb{1}/n$ and $b = (y^T \mathbb{1})\mathbb{1}/n$ satisfy

$$\mathrm{d}a = -m_f \, a \, \mathrm{d}t + \mathrm{d}w_a, \quad b = 0$$

and the variance amplification is determined by

$$\begin{aligned} J \;=\; J_a + \bar{J} \;&=\; \lim_{t \to \infty} \mathbb{E}[\|\hat{x}\|^2] + \mathbb{E}[a^2] \\ &=\; \mathrm{trace}(X_1) + \frac{s_1}{2m_f} \end{aligned}$$

where $X = \begin{bmatrix} X_1 & X_2 \\ X_2^T & X_3 \end{bmatrix}$ is the corresponding state covariance matrix at the steady state

$$\begin{aligned} &\begin{bmatrix} -m_f I - \frac{1}{\mu}\hat{\Lambda} & -I \\ \hat{\Lambda} & 0 \end{bmatrix} X + X \begin{bmatrix} -m_f I - \frac{1}{\mu}\hat{\Lambda} & \hat{\Lambda} \\ -I & 0 \end{bmatrix} \\ &\qquad = \begin{bmatrix} -s_1 I & 0 \\ 0 & -s_2 \hat{\Lambda} \end{bmatrix} \end{aligned}$$

The result follows from noting that X_1, X_2, and X_3 are all diagonal and

$$X_1 = \frac{s_1 + s_2}{2}(m_f I + \hat{\Lambda})^{-1}, \quad X_2 = \frac{-s_2}{2} I.$$

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*, vol. 64, no. 7, pp. 2861–2868, July 2019.

[2] E. J. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure and Applied Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[3] P. J. Bickel and E. Levina, "Regularized estimation of large covariance matrices," *The Annals of Statistics*, vol. 36, no. 1, pp. 199–227, 2008.

[4] 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, September 2013.

[5] 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.

[6] 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, 2018.

[7] 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.

[8] 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*, Philadelphia, PA, 2019, pp. 3414–3419.

[9] Y. Tang, G. Qu, and N. Li, "Semi-global exponential stability of augmented primal–dual gradient dynamics for constrained convex optimization," *Systems & Control Letters*, vol. 144, p. 104754, 2020.

[10] D. Ding and M. R. Jovanović, "Global exponential stability of primal-dual gradient flow dynamics based on the proximal augmented Lagrangian: A Lyapunov-based approach," in *Proceedings of the 59th IEEE Conference on Decision and Control*, Jeju Island, Republic of Korea, 2020, pp. 4836–4841.

[11] D. Jakovetić, D. Bajović, J. Xavier, and J. M. Moura, "Primal–dual methods for large-scale and distributed convex optimization and data analytics," *Proceedings of the IEEE*, vol. 108, no. 11, pp. 1923–1938, 2020.

[12] P. You and E. Mallada, "Saddle flow dynamics: Observable certificates and separable regularization," in *Proceedings of the 2021 American Control Conference*, 2021, pp. 4817–4823.

[13] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 123–231, 2013.

[14] J. W. Simpson-Porco, B. K. Poolla, N. Monshizadeh, and F. Dörfler, "Input-output performance of linear–quadratic saddle-point algorithms with application to distributed resource allocation problems," *IEEE Trans. Automat. Control*, vol. 65, no. 5, pp. 2032–2045, 2019.

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

[16] F. Paganini and E. Feron, "Linear matrix inequality methods for robust H_2 analysis: A survey with comparisons," in *Advances in linear matrix inequality methods in control*. SIAM, 2000, pp. 129–151.

[17] B. Hu and L. Lessard, "Dissipativity theory for Nesterov's accelerated method," in *Proc. ICML*, vol. 70, 2017, pp. 1549–1557.

[18] M. Fazlyab, A. Ribeiro, M. Morari, and V. M. Preciado, "Analysis of optimization algorithms via integral quadratic constraints: Nonstrongly convex problems," *SIAM J. Optim.*, vol. 28, no. 3, pp. 2654–2689, 2018.

[19] H. Mohammadi, M. Razaviyayn, and M. R. Jovanović, "Robustness of accelerated first-order algorithms for strongly convex optimization problems," *IEEE Trans. Automat. Control*, vol. 66, no. 6, pp. 2480–2495, June 2021.

[20] S. Michalowsky, C. Scherer, and C. Ebenbauer, "Robust and structure exploiting optimization algorithms: An integral quadratic constraint approach," *International Journal of Control*, pp. 1–24, 2020.

[21] H. Mohammadi, S. Samuelson, and M. R. Jovanović, "Transient growth of accelerated optimization algorithms," *IEEE Trans. Automat. Control*, 2022, in press; also arXiv:2103.08017.

[22] N. S. Aybat, A. Fallah, M. M. Gürbüzbalaban, and A. Ozdaglar, "Robust accelerated gradient methods for smooth strongly convex functions," *SIAM J. Opt.*, vol. 30, no. 1, pp. 717–751, 2020.

[23] J. W. Simpson-Porco, "Input/output analysis of primal-dual gradient algorithms," in *Proceedings of the 54th Annual Allerton Conference on Communication, Control, and Computing*, 2016, pp. 219–224.

[24] B. Bamieh, M. R. Jovanović, P. Mitra, and S. Patterson, "Coherence in large-scale networks: dimension dependent limitations of local feedback," *IEEE Trans. Automat. Control*, vol. 57, no. 9, pp. 2235–2249, September 2012.