Exponential convergence of primal-dual dynamics for multi-block problems under local error bound condition

Ibrahim K. Ozaslan and Mihailo R. Jovanović

***Abstract*— We study composite optimization problems in which the objective function is given by the sum of a smooth strongly convex term and multiple, potentially non-differentiable, convex regularizers. For a class of problems that satisfy a structural property expressed in terms of a local error bound condition, we establish the existence of a finite time after which a primal-dual method based on the proximal augmented Lagrangian converges exponentially fast to the set of optimal primal-dual variables.**

***Index Terms*— Gradient flow dynamics, error bound condition, Lyapunov functions, operator splitting, proximal augmented Lagrangian.**

I. Introduction

We consider composite optimization problems of the form

$$\underset{x}{\text{minimize}} \;\; f(x) \, + \, \sum_{i\,=\,1}^{r} g_i(T_i x) \qquad (1)$$

where $x \in \mathbb{R}^n$ is the optimization variable, f: $\mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable convex function, each g_i: $\mathbb{R}^{m_i} \to \mathbb{R}$ is a possibly non-differentiable convex regularization function, and each $T_i \in \mathbb{R}^{m_i \times n}$ is a matrix that imposes regularization in the desired coordinates. Since g_i's are allowed to be non-differentiable, constrained problems can be brought into the form of (1) using indicator functions associated with the convex sets. Hence, (1) has a broad usage in applications that range from machine learning [1], to statistics [2], and control theory [3]–[5].

Primal-dual methods provide an effective means for solving (1). These methods decouple regularizers and treat them separately and they are convenient for distributed computing and parallelization [2]. Stability and convergence of primal-dual algorithms have been studied in various scenarios since their introduction in the seminal paper [6] as a continuous-time dynamical system. Early results [7], [8] focused on global asymptotic stability of the primal-dual dynamics for strictly-convex differentiable problems with inequality constraints. In [9]–[11], the asymptotic stability results were extended to general convex-concave saddle functions and, in [12], global exponential stability of the primal-dual dynamics was established for differentiable convex problems with equality constraints. Primal-dual dynamics based on the generalized augmented Lagrangian [13] were examined in [14] and [15]; these references proved global exponential stability (for smooth problems with linear inequality constraints) and semi-global exponential stability (for differentiable problems with smooth inequality constraints), respectively.

In [16], the augmented Lagrangian was brought into a continuously differentiable form by exploiting the structure of the proximal operator associated with a non-differentiable regularizer in (1). This approach was utilized to obtain the Proximal Augmented Lagrangian (PAL), which is determined by the sum of the smooth part of the objective function and the Moreau envelope associated with the non-differentiable regularizer. In contrast to the augmented Lagrangian, PAL is a continuously differentiable function of primal and dual variables and the primal-dual dynamics can be used to compute its saddle points. For the two-block case, i.e., for $r = 1$ in (1), with strongly convex f and full-row rank T, the primal-dual dynamics based on PAL are globally exponentially stable in both continuous [16], [17] and discrete [18] time and similar properties are enjoyed by the second-order method based on PAL [19].

Although the aforementioned results have direct extensions to the multi-block setup, the full-row rank assumption on $T := [T_1^T \cdots T_r^T]^T$ does not hold in applications that arise in multi-block optimization (e.g., convex formulations of neural networks [1], distributed averaging [2], [20], and empirical risk minimization via support vector machines and logistic regression [21]).

In the absence of the full-row rank assumption, even if f is strongly convex, the primal-dual dynamics based on PAL may have a continuum of equilibria because of non-differentiable terms. This is the main challenge for proving the global exponential stability. In this paper, we assume that the problem satisfies a structural property expressed in terms of a local error bound condition [22]. Similar assumption was made in [23] to establish convergence of the Alternating Direction Method of Multipliers (ADMM) for the multi-block problem (1). Without making any rank assumptions on T, we prove that, after a finite time, the primal-dual dynamics based on PAL converges exponentially fast to the equilibrium set. This together with the global asymptotic stability result established in [24] imply the semi-global exponential stability of the primal-dual dynamics based on PAL (see [15] for the definition). The class of problems (1) satisfies the local error bound condition for a broad class of regularization functions including group lasso penalization and indicator functions of polyhedral sets [23], [25]. Thus, our results are applicable to convex problems with affine constraints without requiring any constraint qualifications.

The rest of the paper is organized as follows. In Sec-

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

I. K. Ozaslan 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: ozaslan@usc.edu, mihailo@usc.edu.

tion II, we provide the background material and introduce the primal-dual gradient flow dynamics based on PAL. In Section III, we discuss the local error bound condition and utilize a Lyapunov-based approach to characterize stability and convergence properties of the primal-dual dynamics. We conclude the paper in Section IV with remarks.

II. Problem formulation and background

To streamline our analysis, we define a separable function $g \colon \mathbb{R}^m \to \mathbb{R}$ as

$$g(Tx) \;:=\; \sum_{i=1}^{r} g_i(T_i x) \tag{2}$$

where $m := m_1 + \cdots + m_r$ and cast (1) as a two-block convex composite optimization problem

$$\underset{x}{\text{minimize}} \;\; f(x) \,+\, g(Tx). \tag{3}$$

By introducing auxiliary variables $z := [z_1^T \cdots z_r^T]^T \in \mathbb{R}^m$, we can rewrite problem (3) as

$$\begin{array}{ll} \underset{x,\, z}{\text{minimize}} & f(x) \,+\, g(z) \\ \text{subject to} & Tx \,-\, z \,=\, 0. \end{array} \tag{4}$$

The Lagrangian associated with (4) is given by

$$\mathcal{L}(x,z;y) \;:=\; f(x) \,+\, g(z) \,+\, \langle y,\, Tx - z\rangle$$

where $y := [y_1^T \ldots y_r^T]^T \in \mathbb{R}^m$ is the vector of dual variables associated with equality constraint $Tx = z$. For a given positive parameter μ, the associated augmented Lagrangian is

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

where $\|\cdot\|$ denotes Euclidean norm. The first order optimality conditions for (4) are given by

$$0 \;=\; \nabla f(x^\star) \,+\, T^T y^\star \tag{5a}$$

$$0 \;\in\; \partial g(z^\star) \,-\, y^\star \tag{5b}$$

$$0 \;=\; Tx^\star \,-\, z^\star \tag{5c}$$

where ∂g denotes the subdifferential of g. Substitution of (5c) into (5b) and the resulting inclusion into (5a) yield the optimality condition for (3).

The minimizer of $\mathcal{L}_\mu(x,z;y)$ with respect to z is given by the proximal operator associated with the non-differentiable components

$$z^\star(x,y) \;=\; \underset{z}{\text{argmin}}\; \mathcal{L}_\mu(x,z;y) \;=\; \mathbf{prox}_{\mu g}(Tx + \mu y). \tag{6}$$

Here, the proximal operator of g for a positive parameter μ is the one-to-one mapping defined as [26]

$$\mathbf{prox}_{\mu g} \;:=\; (I \,+\, \mu \partial g)^{-1}. \tag{7}$$

Alternatively, the proximal operator can be obtained as the minimizer of the following optimization problem

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

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

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

which is a continuously differentiable function, even for a non-differentiable g, with $1/\mu$-Lipschitz gradient [27, Proposition 12.30]

$$\mu \nabla M_{\mu g}(v) \;=\; v \,-\, \mathbf{prox}_{\mu g}(v). \tag{8}$$

The proximal augmented Lagrangian is obtained by evaluating $\mathcal{L}_\mu$ along the manifold determined by $z^\star(x,y)$

$$\begin{aligned} \mathcal{L}_\mu(x;y) \;&:=\; \inf_z \mathcal{L}_\mu(x,z;y) \;=\; \mathcal{L}_\mu(x, z^\star(x,y);y) \\ &=\; f(x) \,+\, M_{\mu g}(Tx \,+\, \mu y) \,-\, \tfrac{\mu}{2}\|y\|^2. \end{aligned} \tag{9}$$

Unlike the augmented Lagrangian, PAL is a continuously differentiable saddle function. Minimizing PAL over primal variable x yields the Lagrange dual function associated with problem (4)

$$d(y) \;:=\; \underset{x}{\text{minimize}}\; \mathcal{L}_\mu(x;y) \;=\; \mathcal{L}_\mu(x^\star(y);y) \tag{10}$$

where $x^\star(y) := \text{argmin}_x \; \mathcal{L}_\mu(x;y)$. The set of optimal dual variables is denoted by $\mathcal{Y}^\star := \text{argmax}\;\, d(y)$ and the set of optimal primal variables is denoted by $\mathcal{X}^\star$. Since (4) is a convex optimization problem with linear equality constraints, the strong duality holds [21], hence, we have $x^\star(y^\star) \in \mathcal{X}^\star$ for any $y^\star \in \mathcal{Y}^\star$. Moreover, the optimal value of problem (4) equals to the dual optimal value $d^\star := \max\;\, d(y)$.

The proximal augmented Lagrangian (9) is a convex-concave saddle function and its saddle points that satisfy the following chain of inequalities

$$\mathcal{L}_\mu(x^\star;y) \;\le\; \mathcal{L}_\mu(x^\star;y^\star) \;\le\; \mathcal{L}_\mu(x;y^\star), \quad \forall x, y$$

are characterized by the following system of equations

$$0 \;=\; \nabla f(\bar{x}) \,+\, T^T \nabla M_{\mu g}(T\bar{x} \,+\, \mu\bar{y}) \tag{11a}$$

$$0 \;=\; \mu \nabla M_{\mu g}(T\bar{x} \,+\, \mu\bar{y}) \,-\, \mu\bar{y}. \tag{11b}$$

Moreover, the set of saddle points characterized by (11) is equivalent to the set of optimal primal-dual pairs given by (5). To see this, substitute (11b) into (11a) which gives (5a). Then, substitute expression (8) into (11b) to get

$$T\bar{x} \;=\; \mathbf{prox}_{\mu g}(T\bar{x} \,+\, \mu\bar{y}) \tag{12}$$

which means that $T\bar{x} = z^\star(\bar{x},\bar{y})$, hence (5c) is satisfied. Lastly, one-to-one correspondence in (7) together with (12) gives that $\bar{y} \in \partial g(T\bar{x})$ and (5b) is satisfied. The converse direction can also be shown by following similar steps. Consequently, a solution to problem (1) can be found by computing a saddle point characterized by (11). To this end, we deploy the scaled Arrow-Hurwitz-Uzawa gradient flow dynamics

$$\begin{aligned}\dot{x} \;&=\; -\nabla_x \mathcal{L}_\mu(x;y) \;=\; -\nabla f(x) \,-\, T^T \nabla_x M_{\mu g}(Tx \,+\, \mu y)\\ \dot{y} \;&=\; \alpha \nabla_y \mathcal{L}_\mu(x;y) \;=\; \alpha\left(Tx \,-\, \mathbf{prox}_{\mu g}(Tx \,+\, \mu y)\right)\end{aligned} \tag{13}$$

where x: $[t_0,\infty) \to \mathbb{R}^n$, y: $[t_0,\infty) \to \mathbb{R}^m$, and t_0 is the initial time. The equilibrium points of (13) are clearly the same as the saddle points given by (11). Here, $\alpha \in (0,1]$ is the scaling parameter that enforcers time separation between primal and dual flows [28]; it determines the rate of slowing down in the dual dynamics. In discrete time, $\alpha < 1$ corresponds to having smaller step sizes for dual updates. As $\alpha \to 0$, the trajectories resulting from (13) become similar to the primal-dual sequence generated by the method of multipliers.

Owing to the separable structure in (2), minimizer $z^\star(x,y)$ in (6) takes the following form

$$z^\star(x,y) \;=\; \begin{bmatrix} z_1^\star(x,y_1)\\ \vdots \\ z_r^\star(x,y_r)\end{bmatrix} \;=\; \begin{bmatrix} \mathbf{prox}_{\mu g_1}(T_1 x \,+\, \mu y_1)\\ \vdots \\ \mathbf{prox}_{\mu g_r}(T_r x \,+\, \mu y_r)\end{bmatrix}$$

and the associated Moreau envelope can be written as the sum of individual envelopes associated with each regularizer

$$M_{\mu g}(v) \;=\; \sum_{i\,=\,1}^{r} M_{\mu g_i}(T_i x \,+\, \mu y_i).$$

As a result, primal-dual dynamics (13) can be implemented by using multiple blocks as

$$\begin{aligned}\dot{x} \;&=\; -\nabla f(x) \;-\; \sum_{i\,=\,1}^{r} T_i^T\Big(y_i \,+\, \tfrac{1}{\mu\alpha}\,\dot{y}_i\Big)\\ \dot{y} \;&=\; \alpha\left(T_i x \,-\, \mathbf{prox}_{\mu g_i}(T_i x \,+\, \mu y_i)\right), \quad i \,=\, 1,\ldots,r\end{aligned}$$

where y_i–blocks can be run in parallel since they are independent from each other.

We next introduce the local error bound condition and utilize a Lyapunov-based approach to analyze stability and convergence rate of the primal-dual dynamics.

III. Exponential convergence of primal-dual dynamics

Throughout our analysis, we have the following main assumption on the differentiable part of problem (4).

Assumption 1: Let function f be m_f-strongly convex and its gradient ∇f be L_f-Lipschitz continuous.

Under Assumption 1, the proximal augmented Lagrangian $\mathcal{L}_\mu$ is strongly convex in x and the set of optimal primal variables $\mathcal{X}^\star$ is a singleton. However, even if problem (1) under Assumption 1 is strongly convex, the set of optimal dual variables $\mathcal{Y}^\star$ may not be a singleton. Using optimality conditions (5), it can be easily shown that $\mathcal{Y}^\star$ is given by the intersection of $\partial g(Tx^\star)$ and an affine set determined by the null space of T^T; hence $\mathcal{Y}^\star$ can be even unbounded.

If in addition to the strong convexity assumption the matrix T is full row rank, $\mathcal{Y}^\star$ also becomes a singleton. In that case, dynamics (13) with $\alpha = 1$ are globally exponentially stable for problem (3) with a single non-differentiable regularizer [16], [17]. This immediately implies the global exponential stability of the multi-block problem (1) due to the equivalence between (1) and (3). However, since the full-row rank assumption on T is restrictive for problems with several regularizers, we do not made it in this paper.

When T is not full-row rank[1], the set of equilibrium points becomes a continuum which may not even be bounded; yet, the classical notion of stability is mainly defined for compact sets [30, Section 4.7-4.9]. In our work, we focus on the convergence properties of dynamics (13). We assume that the problem satisfies a structural property that gives a local bound on the distance to $\mathcal{Y}^\star$ and we utilize this bound to prove the exponential convergence rate. This structural property allows us to avoid making any assumptions on the rank of T and g_i's are allowed to be indicator functions of polyhedral sets. On the other hand, the local error bound condition allows us to establish exponential decay of trajectories to the set of equilibrium points of primal-dual dynamics only after a certain finite time.

A. *Local (dual) error bound condition*

The local error bound condition holds if the inequality

$$\mathbf{dist}(y,\mathcal{Y}^\star) \;\le\; \gamma\|\nabla d(y)\| \tag{14}$$

is valid when $d(y) \ge \eta$ and $\|\nabla d(y)\| \le \delta$, where

$$\mathbf{dist}(y,\mathcal{Y}^\star) \;:=\; \underset{\nu\,\in\,\mathcal{Y}^\star}{\operatorname{argmin}}\,\|\nu \,-\, y\|$$

and η, γ, and δ are real parameters. Strong concavity of the dual function immediately yields an upper bound on the distance to $\mathcal{Y}^\star$ and the error bound condition provides a local substitute for it. In the absence of strong convexity of the objective function, similar error bounds were utilized to prove linear and even superlinear convergence for different methods, including the interior point, proximal gradient, and coordinate descent algorithms [31].

The next lemma discusses classes of functions for which local error bound condition (14) holds for problem (4).

Lemma 1: Let each g_i be given by $g_i(x) = h_{i1}(E_i x) + h_{i2}(x)$ where

1) h_{i1} is a strictly convex differentiable function and $E_i \in \mathbb{R}^{m_i \times n}$.
2) h_{i2} is either
   a) a polyhedral function or
   b) group lasso penalization, i.e.,

$$h_{i2}(x) \;=\; \lambda\|x\|_1 \;+\; \sum_{\mathcal{I}} \omega_{\mathcal{I}}\|x_{\mathcal{I}}\|_2$$

   where $x = [\,\cdots\; x_{\mathcal{I}}^T \;\cdots]^T$ is a partition of x with $\omega_{\mathcal{I}} \ge 0$ and $\mathcal{I}$ is a partition index.

Then, for any scalar η, there exist δ and γ such that the local error bound condition (14) holds for (4). Moreover, γ is independent of x and y.

[1] In this case, Mangasarian-Fromovitz constraint qualification, that guarantees the boundedness of $\mathcal{Y}^\star$ for differentiable functions [29], does not hold.

Proof: Since dynamics (13) are globally asymptotically stable [16], [24], its trajectories remain inside a compact set for all times. Then, together with the strong convexity assumption, all conditions in [23, Lemma 2.3] are satisfied and Lemma 1 becomes a corollary of [23, Lemma 2.3]; see [32, Theorem 4.1] and [31] for additional details. ■

Remark 1: Since E_i can be rank-deficient or possibly zero, Lemma 1 does not require g_i to be strictly convex. Thus, h_{i1}'s are also allowed to be equal to zero.

Epigraph of a polyhedral function can be expressed as the intersection of finitely many halfspaces. Examples of such functions include, but are not limited to, piece-wise affine functions, ℓ_1 and ℓ_∞ norms, and indicator functions of polyhedral sets. Among other applications, the classes of functions described in Lemma 1 arise in empirical risk minimization and sparse signal recovery [31]. Additional details about the relation between the error bound, quadratic growth, and proximal Polyak-Lojasiewicz conditions can be found in [33], [34].

B. Lyapunov-based analysis

We start our analysis by proposing a Lyapunov function candidate

$$V(x,y) \;:=\; \mathcal{L}_\mu(x;y) \,-\, d(y) \,+\, d^\star \,-\, d(y) \tag{15}$$

which represents the sum of the primal gap $\mathcal{L}_\mu(x;y) - d(y)$ and the dual gap $d^\star - d(y)$. In [23] it was observed that V given by (15) decreases along the trajectories of ADMM but, to the best of our knowledge, $V(x,y)$ was not previously utilized for a Lyapunov-based analysis of the primal-dual gradient flow dynamics.

We next show that $V(x,y)$ given by (15) is a *strict Lyapunov function* for (13), which implies the global asymptotic stability of the corresponding equilibria.

Lemma 2 (Strict Lyapunov Function): The time derivative of $V(x,y)$ in (15) along the trajectories of primal-dual dynamics (13) with parameter $\alpha \in (0, m_f^2 \|T\|_2^{-2}/4)$ satisfies

$$\dot V \;\le\; -c_0\left(\|\nabla_x \mathcal{L}(x;y)\|^2 \,+\, \|\nabla d(y)\|^2\right)$$

where $c_0 = \min(\alpha, 1 - 4\alpha\|T\|^2/m_f^2)$.

Proof: See Appendix B. ■

Since $\mathcal{L}_\mu$ is convex in x and d is a concave function of y, $\dot V(x,y) = 0$ if and only if $(x,y) \in \mathcal{X}^\star \times \mathcal{Y}^\star$. This implies that V in (15) is a monotonically decreasing function of time along the trajectories of (13). Moreover, Lemma 3 shows that the norms of $\nabla_x \mathcal{L}_\mu$ and ∇d are upper bounded by V and, thus, they decay to zero at the same rate as V. In Corollary 6, we use this fact to prove that the distance to the equilibria decreases exponentially fast. We note that in the rest of the paper L_x and L_y denote the Lipschitz parameters of $\nabla_x \mathcal{L}_\mu(x;y)$ and ∇d, respectively. The explicit expressions for L_x and L_y are provided in Lemma 7 of Appendix A.

Lemma 3 (Decaying Gradient): The time derivative of V along the solutions of (13) satisfies $c_1 V \ge -\dot V$, where $c_1 = 2\max(L_x, L_y)$. Thus, $\|\nabla_x \mathcal{L}_\mu(x;y)\| \to 0$ and $\|\nabla d(y)\| \to 0$ as $t \to \infty$.

Proof: See Appendix C. ■

In Lemma 4, we establish a quadratic upper bound on V in terms of the distance to $\mathcal{Y}^\star$ and, in Thereom 5, we utilize this quadratic upper bound in conjunction with the error bound condition to derive an upper bound on $\dot V$ in terms of V.

Lemma 4 (Quadratic Upper Bound): Along the trajectories of primal-dual dynamics (13), the Lyapunov function $V(x,y)$ in (15) is upper bounded by

$$V(x,y) \;\le\; c_2\left(\|x \,-\, x^\star(y)\|^2 \,+\, \mathbf{dist}^2(y, \mathcal{Y}^\star)\right)$$

where $c_2 = (1/2)\max(L_x, \mu)$.

Proof: See Appendix D. ■

We next prove our main result, which implies that, after a finite time, the Lyapunov function decreases at an exponential rate.

Theorem 5: Let g satisfy Lemma 1. Then, there exists $t_1 \in (t_0, \infty)$ such that the Lyapunov function V in (15) for primal-dual gradient flow dynamics (13) with parameter $\alpha \in (0, m_f^2\|T\|_2^{-2}/4)$ satisfies

$$\dot V(t) \;\le\; -\rho V(t) \text{ for all } t \,\ge\, t_1$$

where $\rho = (c_0/c_2)\min(m_f^2, \gamma^{-2})$ and γ is the error bound parameter.

Proof: Let $V(t)$ denote Lyapunov function (15) along the trajectories of (13). Lemma 2 implies that $V(t)$ is monotonically decreasing outside the equilibria; thus, the primal and dual gaps are bounded from above since both are non-negative. Hence, dual function $d(y)$ is bounded below by $d^\star - V(0)$. Let $\eta := d^\star - V(0)$; by Lemma 1, there exist positive constants δ and γ such that inequality (14) holds when $\|\nabla d(y)\| \le \delta$. Furthermore, Lemma 3 guarantees the existence of a finite $t_1 > t_0$ such that $\|\nabla d(y)\| \le \delta$ for $t \ge t_1$. Consequently, error bound (14) holds when $t \ge t_1$, which allows us to derive the following upper bound on $V(t)$

$$\begin{aligned} V \;&\le\; c_2(\|x \,-\, x^\star(y)\|^2 \,+\, \mathbf{dist}(y,\, \mathcal{Y}^\star)^2) \\ &\le\; c_2(\|x \,-\, x^\star(y)\|^2 \,+\, \gamma^2\|\nabla d(y)\|^2) \\ &\le\; c_2(m_f^{-2}\|\nabla_x \mathcal{L}_\mu(x;y)\|^2 \,+\, \gamma^2\|\nabla d(y)\|^2) \\ &\le\; c_2\max(m_f^{-2},\; \gamma^2)\left(\|\nabla_x \mathcal{L}_\mu(x;y)\|^2 \,+\, \|\nabla d(y)\|^2\right) \\ &\le\; -\,(c_2/c_0)\max(m_f^{-2},\; \gamma^2)\dot V \end{aligned}$$

where the first line follows from Lemma 4, the second line is obtained using (14), the third line is a consequence of the strong convexity of $\mathcal{L}_\mu$ with respect to x, and the last inequality follows from Lemma 2. ■

An immediate consequence of Theorem 5 is that the distance of the solution to (13) to the equilibrium set decreases exponentially fast after a finite time.

Corollary 6: For $t \,\ge\, t_1$, the distance of the solution to (13) to $\mathcal{X}^\star$ and $\mathcal{Y}^\star$ obey the following exponentially decaying upper bound

$$\|x(t) \;-\; x^\star\| \;\leq\; c_3(2/\rho)\mathrm{e}^{-\rho(t \,-\, t_1)/2}$$
$$\mathbf{dist}(y(t),\; \mathcal{Y}^\star) \;\leq\; c_3\gamma\, \mathrm{e}^{-\rho(t \,-\, t_1)/2}$$

where $c_3 = \sqrt{(c_1/c_0)V(t_1)}$.

Proof: See Appendix E. ■

Remark 2: Although the exponential convergence rate in Corollary 6 holds only for $t > t_1$, a global algebraic convergence rate can be established using Lyapunov function (15). In conjunction with Corollary 6, this implies the existence of a global exponential upper bound on the distance to the set of equilibrium points. Since this bound depends on the initial condition, primal-dual gradient flow dynamics (13) are semi-globally exponentially stable.

The regularizers in Lemma 1 are allowed to be indicator functions of polyhedral sets. Hence, the exponential rate given in Corollary 6 is also valid for convex problems with affine constraints and, in contrast to [14], we do not have any constraint qualifications or any assumptions on the matrices T_i. Furthermore, our results hold for a class of non-differentiable regularizers that are widely encountered in machine learning and statistics applications.

IV. Concluding remarks

We have considered a class of composite optimization problems where the objective function can be expressed as the sum of a smooth convex term and multiple possibly non-differentiable convex regularizers. For primal-dual dynamics based on the proximal augmented Lagrangian, we show that under a local bound condition on the distance to the dual solutions the dynamics converge exponentially fast to the set of equilibrium points. Since this structural property is satisfied by a broad class of functions, including indicator functions of polyhedral sets, our analysis covers convex problems with linear equality and inequality constraints.

Appendix

In what follows, we omit the arguments of functions whenever it is easy to infer them from the context.

A. Lipschitz Continuity of $\nabla_x \mathcal{L}_\mu$ and ∇d

Lemma 7: The gradients $\nabla_x \mathcal{L}_\mu$ and ∇d are Lipschitz continuous with moduli $L_x = L_f + \|T\|^2/\mu$ and $L_y = \mu$, respectively.

Proof: The Lipschitz continuity of $\nabla_x \mathcal{L}_\mu$ over x follows from the Lipschitz continuity of ∇f and $\nabla M_{\mu g}$ [27, Proposition 12.30].

We now show the Lipschitz continuity of ∇d. For arbitrary y and y', let $\tilde{y} = y - y'$, $\tilde{x} = x^\star(y) - x^\star(y')$, $\widetilde{\nabla f} = \nabla f(x^\star(y)) - \nabla f(x^\star(y'))$, and $\tilde{z} = \mathbf{prox}_{\mu g}(Tx^\star(y) + \mu y) - \mathbf{prox}_{\mu g}(Tx^\star(y') + \mu y')$. Since $x^\star(y)$ and $x^\star(y')$ are minimizers of $\mathcal{L}_\mu(x;y)$ over x evaluated at y and y', respectively, $\nabla_x \mathcal{L}_\mu$ vanishes at these points. Hence, we have

$$\begin{aligned} 0 \;&=\; \langle \nabla_x \mathcal{L}_\mu(x^\star(y);y) \;-\; \nabla_x \mathcal{L}_\mu(x^\star(y');y'),\; \tilde{x}\rangle \\ &=\; \langle \widetilde{\nabla f} \;+\; T^T\tilde{y} \;+\; \tfrac{1}{\mu}T^T T\tilde{x} \;-\; \tfrac{1}{\mu}T^T\tilde{z},\; \tilde{x}\rangle \end{aligned}$$

and monotonicity of ∇f implies

$$\langle T^T\tilde{y},\tilde{x}\rangle \;+\; \tfrac{1}{\mu}\,\|T\tilde{x}\|^2 \;\leq\; \tfrac{1}{\mu}\,\langle \tilde{z}, T\tilde{x}\rangle.$$

Completion of squares along with the nonexpansiveness of proximal operator yields

$$\begin{aligned} \|T\tilde{x} - \tilde{z}\|^2 \;&\leq\; -\langle \mu\tilde{y}, T\tilde{x}\rangle \;-\; \langle \tilde{z}, T\tilde{x}\rangle \;+\; \|\tilde{z}\|^2 \\ &\leq\; -\langle \mu\tilde{y}, T\tilde{x}\rangle \;+\; \langle \mu\tilde{y}, \tilde{z}\rangle \\ &=\; \langle \mu\tilde{y}, \tilde{z} \;-\; T\tilde{x}\rangle \end{aligned}$$

and Cauchy-Schwarz inequality completes the proof. ■

B. Proof of Lemma 2

From (10), $V(x,y) = 0$ if $(x,y) \in \mathcal{X}^\star \times \mathcal{Y}^\star$ and $V(x,y) > 0$ otherwise. Thus, V is strictly positive outside the equilibria. The time derivative of V along the solutions of (13) satisfies

$$\begin{aligned} \dot{V} \;&=\; \langle \nabla_x V, \dot{x}\rangle \;+\; \langle \nabla_y V, \dot{y}\rangle \\ &=\; \langle \nabla_x \mathcal{L}_\mu, \dot{x}\rangle \;+\; \langle \nabla_y \mathcal{L}_\mu \;-\; 2\nabla d, \dot{y}\rangle \\ &=\; -\,\|\nabla_x \mathcal{L}_\mu\|^2 \;+\; \alpha\|\nabla_y \mathcal{L}_\mu \;-\; \nabla d\|^2 \;-\; \alpha\|\nabla d\|^2 \end{aligned}$$

where the third equality is obtained by completing squares in the second term in the second line. Since $\mathcal{L}_\mu$ is strongly convex in x, both $\mathcal{X}^\star$ and $x^\star(y)$ are singletons for any y. Hence, we can use Danskin's Theorem [35, Prop. B.25] to show that $\nabla d(y) = Tx^\star(y) - z^\star(x^\star(y), y)$. For a given y, let $\bar{x} = x^\star(y)$ and $\bar{z} = z^\star(x^\star(y), y)$. The second term of $\dot{V}$ can be upper bounded as

$$\begin{aligned} \|\nabla_y \mathcal{L}_\mu \;-\; \nabla d\|^2 \;&=\; \|T(x - \bar{x}) \;-\; (z^\star(x,y) \;-\; \bar{z})\|^2 \\ &\leq\; (\|T(x \;-\; \bar{x})\| \;+\; \|z^\star(x,y) \;-\; \bar{z}\|)^2 \\ &\leq\; 4\|T(x \;-\; \bar{x})\|^2 \\ &\leq\; 4\|T\|^2\|x \;-\; \bar{x}\|^2 \\ &\leq\; 4m_f^{-2}\|T\|^2\|\nabla_x \mathcal{L}_\mu\|^2. \end{aligned}$$

Here, the first equality follows from (6), the second inequality follows from nonexpansiveness of the proximal operator, and the forth inequality follows from the strong convexity of $\mathcal{L}_\mu$ in x. Substitution of this bound to $\dot{V}$ and rearrangement of terms complete the proof.

C. Proof of Lemma 3

The time derivative of V can be lower bounded as

$$\begin{aligned} -\,\dot{V} \;&=\; \|\nabla_x \mathcal{L}_\mu\|^2 \;-\; \alpha\|\nabla_y \mathcal{L}_\mu \;-\; \nabla d\|^2 \;+\; \alpha\|\nabla d\|^2 \\ &\leq\; \|\nabla_x \mathcal{L}_\mu\|^2 \;+\; \alpha\|\nabla d\|^2 \\ &\leq\; \|\nabla_x \mathcal{L}_\mu\|^2 \;+\; \|\nabla d\|^2 \end{aligned}$$

where the last line follows from $\alpha \in (0,1]$. By Lemma 7, $\nabla_x \mathcal{L}_\mu$ and ∇d are Lipschitz continuous, which implies

$$\|\nabla_x \mathcal{L}_\mu(x;y)\|^2 \;\leq\; 2L_x\,(\mathcal{L}_\mu(x;y) \;-\; d(y))$$
$$\|\nabla d(y)\|^2 \;\leq\; 2L_y\,(d^\star \;-\; d(y))\,.$$

Substitution of these two inequalities into the lower bound of $\dot{V}$ gives the desired inequality. Furthermore, Lemma 2 implies that V is also an upper on the norm of $\nabla_x \mathcal{L}_\mu$ and

∇d. Since V is monotonically decreasing and non-negative, we conclude that the gradients are decaying.

D. Proof of Lemma 4

Since $\nabla_x \mathcal{L}_\mu$ and ∇d are Lipschitz continuous, we have

$$d^\star \,-\, d(y) \;\leq\; (L_y/2)\|y \,-\, y^\star\|^2$$
$$\mathcal{L}_\mu(x;y) \,-\, \min_x \mathcal{L}_\mu(x;y) \;\leq\; (L_x/2)\|x \,-\, x^\star(y)\|^2$$

where the first inequality holds for any $y^\star \in \mathcal{Y}^\star$. Summing these two upper bounds completes the proof.

E. Proof of Corollary 6

Separation of variables along with Theorem 5 implies

$$V(t) \;\leq\; V(t_1)\,\mathrm{e}^{-\,\rho(t\,-\,t_1)} \text{ for all } t \,\geq\, t_1$$

which together with Lemmas 2 and 3 for $t \geq t_1$ yield

$$\begin{aligned}\mathbf{dist}(y(t),\, \mathcal{Y}^\star) \;&\leq\; \gamma\|\nabla d(y(t))\|\\ &\leq\; \gamma\sqrt{(c_1/c_0)V(t)}\\ &\leq\; \gamma\sqrt{(c_1/c_0)V(t_1)}\,\mathrm{e}^{-\,\rho(t\,-\,t_1)/2}.\end{aligned}$$

Furthermore, the strong duality and convergence of dual variable imply that $x^\star(y(t)) \to x^\star$ as $t \to \infty$ and application of Lemmas 2 and 3 completes the proof,

$$\begin{aligned}\|x^\star \,-\, x(t)\| \;&\leq\; \int_t^\infty \|\dot{x}(\tau)\|\,\mathrm{d}\tau\\ &\leq\; \int_t^\infty \sqrt{(c_1/c_0)V(t_1)}\,\mathrm{e}^{-\,\rho(\tau\,-\,t_1)/2}\,\mathrm{d}\tau\\ &=\; (2/\rho)\sqrt{(c_1/c_0)V(t_1)}\,\mathrm{e}^{-\,\rho(t\,-\,t_1)/2}.\end{aligned}$$

References

[1] M. Pilanci and T. Ergen, "Neural networks are convex regularizers: Exact polynomial-time convex optimization formulations for two-layer networks," in *Int. Conf. Mach. Learn.*, 2020, pp. 7695–7705.

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–124, Jan. 2011.

[3] F. Lin, M. Fardad, and M. R. Jovanovic, "Augmented Lagrangian approach to design of structured optimal state feedback gains," *IEEE Tran. Autom. Control*, vol. 56, no. 12, pp. 2923–2929, Dec. 2011.

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

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

[6] K. Arrow, L. Hurwicz, and U. H., *Studies in Linear and Non-linear Programming*. Palo Alto CA: Stanford Univ. Press, 1958.

[7] D. Feijer and F. Paganini, "Stability of primal-dual gradient dynamics and applications to network optimization," *Automatica*, vol. 46, no. 12, pp. 1974–1981, Dec. 2010.

[8] A. Cherukuri, E. Mallada, and J. Cortés, "Asymptotic convergence of constrained primal-dual dynamics," *Systems Control Lett.*, vol. 87, pp. 10–15, Jan. 2016.

[9] 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, Feb. 2017.

[10] A. Cherukuri, E. Mallada, S. Low, and J. Cortes, "The role of convexity on saddle-point dynamics: Lyapunov function and robustness," *IEEE Trans. Automat. Control*, vol. 63, no. 8, pp. 2449–2464, Aug. 2018.

[11] T. Holding and I. Lestas, "Stability and instability in saddle point dynamics—part i," *IEEE Trans. Automat. Control*, vol. 66, no. 7, pp. 2933–2944, July 2021.

[12] J. Cortés and S. K. Niederländer, "Distributed coordination for nonsmooth convex optimization via saddle-point dynamics," *J. Nonlinear Sci.*, vol. 29, no. 4, pp. 1247–1272, Aug. 2019.

[13] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1970.

[14] G. Qu and N. Li, "On the exponential stability of primal-dual gradient dynamics," *IEEE Contr. Syst. Lett.*, vol. 3, no. 1, pp. 43–48, 2019.

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

[16] N. K. Dhingra, S. Z. Khong, and M. R. Jovanović, "The proximal augmented Lagrangian method for nonsmooth composite optimization," *IEEE Tran. Autom. Control*, vol. 64, no. 7, pp. 2861–2868, July 2019.

[17] 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 *IEEE Conf. Decis. Control*, 2020, pp. 4836–4841.

[18] D. Ding, B. Hu, N. K. Dhingra, and M. R. Jovanović, "An exponentially convergent primal-dual algorithm for nonsmooth composite minimization," in *IEEE Conf. Decis. Control*, 2018, pp. 4927–4932.

[19] N. K. Dhingra, S. Z. Khong, and M. R. Jovanović, "A second order primal-dual method for nonsmooth convex composite optimization," *IEEE Trans. Automat. Control*, vol. 67, no. 8, pp. 4061–4076, 2022.

[20] S. Hassan-Moghaddam and M. R. Jovanović, "Distributed proximal augmented Lagrangian method for nonsmooth composite optimization," in *Amer. Control Conf.*, 2018, pp. 2047–2052.

[21] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge U.K.: Cambridge Univ. Press, 2004.

[22] Z.-Q. Luo and P. Tseng, "On the linear convergence of descent methods for convex essentially smooth minimization," *SIAM J. Control Optim.*, vol. 30, no. 2, pp. 408–425, Mar. 1992.

[23] M. Hong and Z.-Q. Luo, "On the linear convergence of the alternating direction method of multipliers," *Math. Program*, vol. 162, no. 1, pp. 165–199, Mar. 2017.

[24] I. K. Ozaslan, S. Hassan-Moghaddam, and M. R. Jovanović, "On the asymptotic stability of proximal algorithms for convex optimization problems with multiple non-smooth regularizers," in *Amer. Control Conf.*, 2022, pp. 132–137.

[25] Z. Zhou and A. M.-C. So, "A unified approach to error bounds for structured convex optimization problems," *Math. Program.*, vol. 165, no. 2, pp. 689–728, Oct. 2017.

[26] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Opt.*, vol. 1, no. 3, pp. 123–231, Jan. 2014.

[27] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. New York NY: Springer, 2011.

[28] M. Picallo, S. Bolognani, and F. Dorfler, "Predictive-sensitivity: Beyond singular perturbation for control design on multiple time scales," *arXiv preprint arXiv:2101.04367*, 2021.

[29] J. Gauvin, "A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming," *Math. Program.*, vol. 12, no. 1, pp. 136–138, Dec. 1977.

[30] W. M. Haddad and V. Chellaboina, *Nonlinear Dynamical Systems and Control*. Princeton NJ: Princeton Univ. Press, 2011.

[31] P. Tseng, "Approximation accuracy, gradient methods, and error bound for structured convex optimization," *Math. Program.*, vol. 125, no. 2, pp. 263–295, Oct. 2010.

[32] Z.-Q. Luo and P. Tseng, "On the convergence rate of dual ascent methods for linearly constrained convex minimization," *Math. Oper. Res.*, vol. 18, no. 4, pp. 846–867, Nov. 1993.

[33] H. Karimi, J. Nutini, and M. Schmidt, "Linear convergence of gradient and proximal-gradient methods under the Polyak-Lojasiewicz condition," in *Joint Eur. Conf. Mach. Learn. Knowl. Discov. Data.*, 2016, pp. 795–811.

[34] D. Drusvyatskiy and A. S. Lewis, "Error bounds, quadratic growth, and linear convergence of proximal methods," *Math. Oper. Res.*, vol. 43, no. 3, pp. 919–948, Mar. 2018.

[35] D. Bertsekas, *Nonlinear Programming*. Belmont, Mass: Athena Scientific, 1999.