Semi-Global Exponential Stability of Primal-Dual Gradient Dynamics for Constrained Convex Optimization

Yujie Tang, Guannan Qu and Na Li

Abstract—Primal-dual gradient dynamics that find saddle points of a Lagrangian have been widely employed for handling constrained optimization problems. Building on existing methods, we extend the augmented primal-dual gradient dynamics to incorporate general convex and nonlinear inequality constraints, and we establish its semi-global exponential stability when the objective function has a quadratic gradient growth. Numerical simulation also suggests that the exponential convergence rate could depend on the initial distance to the KKT point.

I. INTRODUCTION

This paper introduces and analyzes a version of primal-dual gradient dynamics applied to the following convex optimization problem:

\[
\begin{align*}
\min_{x} & \quad f(x) \\
\text{s.t.} & \quad g(x) \leq 0, \quad Ax = b.
\end{align*}
\]

Here \( f : \mathbb{R}^n \to \mathbb{R} \) is continuously differentiable and convex, \( g : \mathbb{R}^n \to \mathbb{R}^{m_f} \) is continuously differentiable and convex, and \( A \in \mathbb{R}^{m_E \times n} \), \( b \in \mathbb{R}^{m_E} \). We consider the augmented primal-dual gradient dynamics (Aug-PDGD) given by

\[
\begin{align*}
\dot{x}(t) &= -\nabla_x L_\rho(x(t), \lambda(t), \nu(t)), \\
\dot{\lambda}(t) &= \nabla_\lambda L_\rho(x(t), \lambda(t), \nu(t)), \\
\dot{\nu}(t) &= \nabla_\nu L_\rho(x(t), \lambda(t), \nu(t)),
\end{align*}
\]

where \( L_\rho(x, \lambda, \nu) \) is the augmented Lagrangian [1]

\[
L_\rho(x, \lambda, \nu) = f(x) + \nu^T(Ax - b) + \sum_{i=1}^{m_f} \max\{\rho g_i(x) + \lambda_i, 0\}^2 - \lambda_i^2 \over 2\rho
\]

and \( \rho \) is a positive constant.

Since its first introduction in [2], [3], primal-dual gradient dynamics have been applied for solving optimization problems in various applications, including power system operation [4], [5], communication networks [6], distributed optimization [7], [8], etc. Theoretical studies of the primal-dual gradient dynamics, particularly its stability analysis, have been of continuing interest to researchers [2], [3], [7]–[15]. Several existing works have established global asymptotic stability of the primal-dual gradient dynamics under different settings. For instance, in [9], the authors proved global asymptotic stability of the projected primal-dual gradient dynamics for constrained convex problems with strictly convex objectives by using a version of the LaSalle invariance principle for hybrid systems. In [12], the authors proposed some general conditions for global asymptotic stability of the primal-dual gradient dynamics. In [13], the authors considered projected primal-dual gradient dynamics that can be applied for convex optimization with inequality constraints, and showed global asymptotic stability for locally strong convex-concave Lagrangian.

Exponential stability is a desirable property both theoretically and in practice. Particularly, given a continuous-time dynamics that is exponentially stable, one can obtain a discrete-time iterative algorithm through explicit Euler discretization that achieves linear convergence for sufficiently small step sizes under appropriate conditions [16], [17]. It is well-known that for unconstrained convex optimization, when the objective function is smooth and strongly convex, the projected gradient dynamics achieve global exponential stability, and as the discrete-time counterpart, the projected gradient descent algorithm achieves global linear convergence [18], and [19] suggests that the condition of strong convexity could be relaxed. In the context of primal-dual gradient dynamics and constrained convex optimization, it is known that local exponential stability can be established by resorting to spectral bounds of saddle matrices [20], [21]. Regarding global exponential stability, [11] and [8] studied saddle-point-like dynamics and proved global exponential stability when applying such dynamics to equality constrained convex optimization problems. [14] proposed proximal gradient flow that can be applied to convex programs with affine inequality constraints \( Ax \leq b \), and showed global exponential stability when \( A \) has full row rank by modeling the dynamics as a linear system with nonlinear feedback and employing the theory of integral quadratic constraints.

This work is an extension of the results in [15], which introduced the augmented primal-dual gradient dynamics for convex optimization with affine inequality constraints. We extend the algorithm to a very general setting of convex optimization, where the constraint functions can be convex and nonlinear. The use of augmented Lagrangian for handling inequality constraints results in a continuous dynamical system, which is different from projection-based primal-dual gradient dynamics. We generalize and improve on the approach employed in [15] to show that the Aug-PDGD achieves semi-global exponential stability [22] when the objective function \( f \) has a quadratic gradient growth [19]. The proof is based on a quadratic Lyapunov function that
has non-zero off-diagonal terms.

**Notation**

For any \( x \in \mathbb{R} \), we denote \( [x]_+ := \max\{x, 0\} \). For any \( x \in \mathbb{R}^p \), we use \( x \geq 0 \) to mean that all the entries of \( x \) are nonnegative. For any real symmetric matrices \( P \) and \( Q \), \( P \succeq Q \) and \( Q \preceq P \) mean that \( P - Q \) is positive semidefinite; similarly \( P \succ Q \) and \( Q \prec P \) mean that \( P - Q \) is positive definite. For any \( x \in \mathbb{R}^p \), we use \( \|x\| \) to denote the \( \ell_2 \) norm of \( x \), and denote \( \|x\|_Q := \sqrt{x^T Q x} \) when \( Q \) is a positive definite matrix. For any matrix \( Q \), we use \( \|Q\| \) to denote the spectral norm of \( Q \). The identity matrix will be denoted by \( I \). For a finite set \( S \), we use \( |S| \) to denote the number of elements in \( S \).

II. AUGMENTED PRIMAL-DUAL GRADIENT DYNAMICS

In this section we present a more detailed description of the augmented primal-dual gradient dynamics and introduce some preliminary results regarding its equilibrium point and trajectory behavior.

We introduce the following auxiliary function

\[
\Theta_\rho(x, \lambda) := \sum_{i=1}^{m_1} \left[ p g_i(x) + \lambda_i \right]^2 - \lambda_i^2 \frac{2 \rho}{\lambda_i^2}.
\]

It can be checked that \( \Theta_\rho(x, \lambda) \) is convex in \( x \) and concave in \( \lambda \), and that \( \Theta_\rho(x, \lambda) \) is continuously differentiable. The augmented Lagrangian of (1) can then be formulated as

\[
L_\rho(x, \lambda, \nu) := f(x) + \Theta_\rho(x, \lambda) + \nu^T (Ax - b).
\]

The augmented primal-dual gradient dynamics are then given by

\[
\begin{align*}
\dot{x}(t) &= -\nabla f(x(t)) - \sum_{i=1}^{m_1} [p g_i(x(t)) + \lambda_i(t)]_+ \nabla g_i(x(t)) \\
&\quad - A^T \nu(t), \quad (2a) \\
\dot{\lambda}(t) &= \sum_{i=1}^{m_1} \left[ p g_i(x(t)) + \lambda_i(t) \right]_+ - \lambda_i(t) e_i, \quad (2b) \\
\dot{\nu}(t) &= A x(t) - b, \quad (2c)
\end{align*}
\]

where \( \{e_i\}_{i=1}^{m_1} \) denotes the standard basis of \( \mathbb{R}^{m_1} \).

**Remark 1.** In [15], an additional parameter \( \eta > 0 \) that scales the dual gradients was introduced in the Aug-PDGD

\[
\begin{align*}
\dot{x} &= -\nabla_x L_\rho(x, \lambda, \nu), \\
\dot{\lambda} &= \eta \nabla_\lambda L_\rho(x, \lambda, \nu), \\
\dot{\nu} &= \eta \nabla_\nu L_\rho(x, \lambda, \nu).
\end{align*}
\]

In this paper we neglect this parameter, as one can scale the constraint functions and dual variables by

\[
\begin{align*}
g &\to \sqrt{\eta} g, \\
A &\to \sqrt{\eta} A, \\
b &\to \sqrt{\eta} b, \\
\lambda &\to \frac{\lambda}{\sqrt{\eta}}, \\
\nu &\to \frac{\nu}{\sqrt{\eta}}, \\
\rho &\to \frac{\rho}{\eta}
\end{align*}
\]

to recover (3).

Now suppose \( (x(t), \lambda(t), \nu(t)), t \geq 0 \) is a differentiable trajectory that satisfies the Aug-PDGD (2) for all \( t \geq 0 \). The following proposition summarizes some preliminary results on the equilibrium and trajectory behavior of Aug-PDGD.

**Proposition 1.**

1. A primal-dual pair is an equilibrium point of the Aug-PDGD (2) if and only if it is a KKT point of (1).

2. Suppose \( \lambda(0) \geq 0 \). Then \( \lambda(t) \geq 0 \) for all \( t \in [0, T] \).

**Proof.** 1) Let \((x_e, \lambda_e, \nu_e)\) denote any equilibrium point of (2). Obviously \( \nabla_\nu L_\rho(x_e, \lambda_e, \nu_e) = 0 \) implies \( Ax_e = b \). Then \( \nabla_\lambda L_\rho(x_e, \lambda_e, \nu_e) = 0 \) shows \( \lambda_{e,i} = [\rho g_i(x_e) + \lambda_{e,i}]_+ \), which further implies \( \lambda_{e,i} \geq 0 \) and

\[
g_i(x_e) \begin{cases} = 0, & \lambda_{i,e} > 0, \\ \leq 0, & \lambda_{i,e} = 0, \end{cases}
\]

for each \( i \). Finally, \( \nabla_{x} L_\rho(x_e, \lambda_e, \nu_e) = 0 \) leads to

\[
\nabla f(x_e) + \sum_{i=1}^{m_1} \rho g_i(x_e) + \lambda_{e,i} \nabla g_i(x_e) + A^T \nu_e = 0.
\]

Conversely, if \((x^*, \lambda^*, \nu^*)\) is a KKT point of (1), it can be checked by direct calculation that

\[
\begin{bmatrix}
-\nabla_\lambda L_\rho(x^*, \lambda^*, \nu^*) \\
\nabla_\nu L_\rho(x^*, \lambda^*, \nu^*)
\end{bmatrix} = 0.
\]

2) Suppose we have \( \lambda_i(t_1) < 0 \) for some \( t_1 > 0 \) and \( i \in \{1, \ldots, m_1\} \). By the continuity of \( \lambda_i(t) \), there exists some \( t_0 \in [0, t_1) \) such that \( \lambda_i(t_0) = 0 \) and \( \lambda_i(t) < 0 \) for \( t \in (t_0, t_1] \). Therefore for all \( t \in [t_0, t_1] \) we have

\[
\dot{\lambda}_i(t) = \frac{[\rho g_i(x(t)) + \lambda_i(t)]_+ - \lambda_i(t)}{\rho} \geq 0,
\]

which contradicts

\[
\int_{t_0}^{t_1} \dot{\lambda}_i(t) \, dt = \lambda_i(t_1) - \lambda_i(t_0) < 0.
\]

**Remark 2.** The augmented primal-dual gradient dynamics differ from the standard projected primal-dual gradient dynamics by employing \( \Theta_\rho(x, \lambda) \) instead of \( \lambda^T g(x) \) in constructing the Lagrangian and by the lack of projection of \( \lambda \) onto the nonnegative orthant. However, as Proposition 1 shows, the KKT point of (1) coincides with the equilibrium of (2), and as long as \( \lambda(0) \geq 0 \), \( \lambda(t) \) will remain nonnegative even if there is no explicit projection onto the nonnegative orthant. One advantage of (2) is that its right-hand sides are all continuous in \( (x, \lambda, \nu) \), unlike the standard projected primal-dual gradient dynamics in which projection introduces discontinuity.

Since (2) can be viewed as primal-dual gradient dynamics on a differentiable convex-concave function, we have the following result that guarantees the boundedness of the trajectory.

Lemma 1. Suppose \((x^*, \lambda^*, \nu^*)\) is a KKT point of (1). We have
\[
\|x(t) - x^*\|^2 + \|\lambda(t) - \lambda^*\|^2 + \|\nu(t) - \nu^*\|^2 \\
\leq \|x(0) - x^*\|^2 + \|\lambda(0) - \lambda^*\|^2 + \|\nu(0) - \nu^*\|^2
\]
for all \(t \geq 0\).

Proof. This follows from that the Aug-PDGD (2) are continuous convex-concave saddle point dynamics. See [9].

III. Main Results

The following additional assumptions will be employed for the remaining part of the paper:

Assumption 1. The problem (1) is feasible and has a unique solution \(x^*\), and linear independence constraint qualification (LICQ) [23] holds at \(x^*\).

As a result, there exist unique optimal Lagrange multipliers \(\lambda^*, \nu^*\) such that \((x^*, \lambda^*, \nu^*)\) satisfies the KKT conditions.

We denote the active set at \(x^*\) by
\[
\mathcal{I} := \{i : g_i(x^*) = 0\},
\]
and denote \(\mathcal{I}^c = \{1, \ldots, m_1\} \setminus \mathcal{I}\). The Jacobian of \(g(x)\) at \(x^*\) will be denoted by \(J\), and we define
\[
\kappa := \lambda_{\min} \left( \begin{bmatrix} J_{\mathcal{I}} & J_{\mathcal{I}} \\ A & A \end{bmatrix}^T \right),
\]
where \(J_{\mathcal{I}}\) is formed by the rows of \(J\) whose indices are in \(\mathcal{I}\). The assumption that LICQ holds at \(x^*\) then implies \(\kappa > 0\).

We also define
\[
d_0 := \left( \|x(0) - x^*\|^2 + \|\lambda(0) - \lambda^*\|^2 + \|\nu(0) - \nu^*\|^2 \right)^{1/2},
\]
i.e., \(d_0\) is the distance from the initial primal-dual pair \((x(0), \lambda(0), \nu(0))\) to the KKT point \((x^*, \lambda^*, \nu^*)\).

Assumption 2. \(f(x)\) has a quadratic gradient growth with parameter \(\mu > 0\), i.e.,
\[
(x - x^*)^T (\nabla f(x) - \nabla f(x^*)) \geq \mu \|x - x^*\|^2
\]
for all \(x \in \mathbb{R}^n\).

The notion of quadratic gradient growth, together with other notions such as quasi-strong convexity and quadratic under-approximation, has been introduced in [19] as a relaxation of the strong convexity condition for linear convergence of gradient-based optimization algorithms. It can be shown that the class of convex functions with quadratic gradient growth is a proper superset of the class of strongly convex functions.

Assumption 3. \(\nabla f(x)\) is \(\ell\)-Lipschitz, and for each \(i = 1, \ldots, m_1\), \(\|\nabla g_i(x)\| \leq L_{g,i}\) and \(\nabla g_i(x)\) is \(M_{g,i}\)-Lipschitz continuous over \(x \in \{y \in \mathbb{R}^n : \|y - x^*\| \leq d_0\}\).

We also denote
\[
L_g := \sqrt{\sum_{i=1}^{m_1} L_{g,i}^2}, \quad M_g := \sqrt{\sum_{i=1}^{m_1} M_{g,i}^2}.
\]

We mention that \(L_g\) can be viewed as an upper bound on the Frobenius norm (and consequently the spectral norm) of the Jacobian matrix of \(g(x)\).

Lemma 2. Let \(\lambda \geq 0\) satisfy \(\|\lambda - \lambda^*\| \leq d_0\). Then for any \(x_1, x_2\) such that \(\|x_1 - x^*\| \leq d_0\) and \(\|x_2 - x^*\| \leq d_0\),
\[
\|\nabla_x \Theta_\rho(x_1, \lambda) - \nabla_x \Theta_\rho(x_2, \lambda)\| \leq M_\Theta \|x_1 - x_2\|,
\]
where
\[
M_\Theta := \rho L_g^2 + (\rho L_g d_0 + d_0 + \|\lambda^*\|) M_g.
\]

Proof. By a direct calculation, we have
\[
\|\nabla_x \Theta_\rho(x_1, \lambda) - \nabla_x \Theta_\rho(x_2, \lambda)\| = \sum_{i=1}^{m_1} \left( (\|[pg_i(x_1) + \lambda_i]_+ - [pg_i(x_2) + \lambda_i]_+\| \|\nabla g_i(x_1)\| + [pg_i(x_1) + \lambda_i]_+ \|\nabla g_i(x_1) - \nabla g_i(x_2)\|) \right)
\]
\[
\leq \sum_{i=1}^{m_1} \left( (\|[pg_i(x_1) - pg_i(x_2)]\| L_{g,i} + \rho L_g d_0 + \|\lambda^*\| M_g) \|x_1 - x_2\| \right)
\]
\[
\leq (\rho L_g^2 + (\rho L_g d_0 + d_0 + \|\lambda^*\|) M_g) \|x_1 - x_2\|.
\]

This lemma shows that, \(M_\Theta\) can be viewed as the Lipschitz constant of \(\nabla \Theta_\rho(x, \lambda)\) with respect to \(x\) in the region \(\{x \in \mathbb{R}^n : \|x - x^*\| \leq d_0\}\).

The main result of this paper is summarized as follows.

Theorem 1. Suppose \(\lambda(0) \geq 0\). Under Assumptions 1, 2 and 3, the trajectory \((x(t), \lambda(t), \nu(t))\) of the augmented primal-dual gradient dynamics (2) satisfies
\[
\|x(t) - x^*\|, \|\lambda(t) - \lambda^*\|, \|\nu(t) - \nu^*\| = O(e^{-\beta t}).
\]

Here \(\beta > 0\) is any constant satisfying
\[
\beta \leq \frac{\kappa \delta_{\min}}{23 \rho (L_g^2 + \|A\|^2)}
\]
and
\[
\|A\|^2 + L_g^2 + \frac{\kappa}{4} + (\ell + M_\Theta) \left( \mu + M_\Theta + \frac{1}{\rho} \right) + \frac{1}{2 \rho^2} \leq \frac{\kappa \mu}{2} - \beta^2,
\]
where
\[
\delta_{\min} := 1 - \left[ 1 + \rho \cdot \max_{i \in \mathcal{I}^c} g_i(x^*) \right] d_0
\]
(5).
In other words, the augmented primal-dual gradient dynamics achieve semi-global exponential stability.

It can be seen that a constant $\beta > 0$ satisfying (4) will always exist, as the right-hand side of (4b) is a decreasing function of $\beta$ that goes to $+\infty$ as $\beta \to 0^+$, and $\delta_{\min}$ is strictly positive.

We note that, while for any initial point $(x(0), \lambda(0), \nu(0))$, there exists a constant $\beta > 0$ such that the distance to the KKT point decays exponentially with rate $\beta$, the upper bound on $\beta$ given by Theorem 1 depends on the initial distance $d_0$ and decreases to zero as $d_0 \to +\infty$; in other words, Theorem 1 does not guarantee the existence of a universal exponential convergence rate, and consequently only implies semi-global exponential stability [22]. This is different from the cases where we only have equality constraints or where $g$ is an affine function $Ax - b$ with $A$ having full row rank [8], [11], [14], [15]. The numerical example in Section V suggests that the semi-global exponential stability might be the nature of the Aug-PDGD rather than an artifact of the proof, but further investigation is needed for concrete conclusions.

The quantity $\delta_{\min}$ characterizes how close the inactive constraints are to being active at $x^*$, and it can be seen that $\beta \to 0$ when $\delta_{\min} \to 0$. As a consequence, the bound on the exponential convergence rate $\beta$ is not robust to perturbations of the problem (1). On the other hand, we suspect that it is possible to generalize the results if we make the stronger assumption that the set $\{\nabla g_i(x^*) : g_i(x^*) \geq -\epsilon\}$ is linearly independent for some given $\epsilon > 0$, and $\beta$ can then be robust to small perturbations. Detailed analysis will be left for future work.

Note that by definition we have $\kappa \leq L_g^2 + \|A\|^2$, and we can interpret the ratio $(L_g^2 + \|A\|^2)/\kappa$ as the “condition number” of the constraints. Then it can be inferred from (3) that better conditioned constraints can facilitate faster convergence. It can also be seen that when $M_g$ is smaller and the function $g(x)$ is closer to being affine, the upper bound on $\beta$ given by (4b) will also be larger.

Compared to the results in [15] where the function $f$ needs to be twice differentiable and strongly convex, we relax the assumption and require $f$ to have Lipschitz continuous gradient and have a quadratic gradient growth.

IV. PROOF OF SEMI-GLOBAL EXPONENTIAL STABILITY

For notational simplicity we suppress the dependence on $t$ and use $(x, \lambda, \nu)$ to denote the trajectory that satisfies (2).

We define

$$V_c := \frac{1}{2} \begin{bmatrix} x - x^* \\ \lambda - \lambda^* \\ \nu - \nu^* \end{bmatrix}^T \begin{bmatrix} I & cJ^T & cA^T \\ cJ & I & 0 \\ cA & 0 & I \end{bmatrix} \begin{bmatrix} x - x^* \\ \lambda - \lambda^* \\ \nu - \nu^* \end{bmatrix},$$

where $c = 2\kappa^{-1}\beta$ and $\beta$ satisfies the conditions (4). The goal is to prove that the matrix

$$\begin{bmatrix} I & cJ^T & cA^T \\ cJ & I & 0 \\ cA & 0 & I \end{bmatrix}$$

is positive definite, and $\dot{V}_c \leq -\beta V_c$, which then lead to the conclusion in Theorem 1 by the equivalence of norms in Euclidean spaces.

Step 1: Prove that (6) is positive definite. By the condition (4b) and the fact that $\mu \leq \ell$, we have

$$\frac{\kappa \ell}{2\beta} > \frac{\kappa \mu}{2\beta} - \beta^2 > (\ell + M_{\Theta})(\mu + M_{\Theta} + \rho^{-1}) > \rho^{-1} \ell,$$

which implies $\rho^{-1} < \kappa/(2\beta)$. Then by (4a) and the fact that $\delta_{\min} \leq 1$, we have

$$\beta \leq \frac{\kappa}{23\rho(L_g^2 + \|A\|^2)} < \frac{\kappa^2}{46\rho(L_g^2 + \|A\|^2)},$$

and consequently

$$c^2 = 4\kappa^{-2}\beta^2 < \frac{2}{23(L_g^2 + \|A\|^2)}.$$

We then have

$$\left\| c^2 \begin{bmatrix} J^T \\ A \end{bmatrix} \right\| = c^2 \| J^T J + A^T A \|
\leq c^2 \left( \|J\|^2 + \|A\|^2 \right) \leq c^2 \left( L_g^2 + \|A\|^2 \right) < 1,$$

and by the Schur complement condition, we see that (6) is positive definite.

Step 2: Prove $\dot{V}_c \leq -\beta V_c$. Firstly, we can write $V_c$ as the following:

$$V_c := \begin{bmatrix} x - x^* \\ \lambda - \lambda^* \\ \nu - \nu^* \end{bmatrix}^T \begin{bmatrix} I & cJ^T & cA^T \\ cJ & I & 0 \\ cA & 0 & I \end{bmatrix} \begin{bmatrix} x - x^* \\ \lambda - \lambda^* \\ \nu - \nu^* \end{bmatrix}.$$

Since $\Theta_\rho$ is convex in $x$ and concave in $\lambda$, we have

$$(x - x^*)^T \nabla x \Theta_\rho(x, \lambda) \leq \Theta_\rho(x^*, \lambda) - \Theta_\rho(x, \lambda),$$

$$(x - x^*)^T \nabla x \Theta_\rho(x^*, \lambda^*) \leq \Theta_\rho(x, \lambda^*) - \Theta_\rho(x^*, \lambda^*),$$

$$(\lambda^* - \lambda)^T \nabla \lambda \Theta_\rho(x, \lambda) \geq \Theta_\rho(x, \lambda^*) - \Theta_\rho(x, \lambda),$$

and so the diagonal terms in (7) can be bounded by

$$\begin{bmatrix} x - x^* \\ \lambda - \lambda^* \\ \nu - \nu^* \end{bmatrix}^T \begin{bmatrix} -\nabla x L_{\rho}(x, \lambda, \nu) \\ \nabla \lambda L_{\rho}(x, \lambda, \nu) \\ \nabla \nu L_{\rho}(x, \lambda, \nu) \end{bmatrix}.$$

(7)
We define
\[ \tilde{\gamma}_{\lambda, i} := \begin{cases} 1, & i \in \mathcal{I} \text{ or } \lambda_i = 0, \\ [1 + \rho g_i(x^*)/\lambda_i]^2, & i \notin \mathcal{I} \text{ and } \lambda_i > 0, \end{cases} \]
and \( \tilde{\Gamma}_\lambda := \text{diag}(\tilde{\gamma}_{\lambda,1}, \ldots, \tilde{\gamma}_{\lambda,m}) \). It can be seen that \( 0 \preceq \tilde{\Gamma}_\lambda \preceq I \)
and
\[ \Theta_\rho(x^*, \lambda) - \Theta_\rho(x^*, \lambda^*) = -\frac{1}{2\rho} (\lambda - \lambda^*)^T (I - \tilde{\Gamma}_\lambda)(\lambda - \lambda^*). \]

Then we consider the off-diagonal terms of \( \mathcal{Q} \). For the term \( \nabla_x L_\rho(x, \lambda, \nu) \), we have
\[
\nabla f(x) - \nabla f(x^*) + \nabla_x \Theta_\rho(x, \lambda) - \nabla_x \Theta_\rho(x^*, \lambda) + A^T \nabla \nu^\top + A^T \nu - A^T \nabla \nu^\top
\]
where we define
\[ \gamma_{\lambda, i} := \begin{cases} [\rho g_i(x^*) + \lambda_i] + -[\rho g_i(x^*) + \lambda_i^*] +, & \lambda_i \neq \lambda_i^*, \\ \lambda_i - \lambda_i^*, & \lambda_i = \lambda_i^*, \end{cases} \]
and \( \Gamma_\lambda = \text{diag}(\gamma_{\lambda,1}, \ldots, \gamma_{\lambda,m}) \).

Now, if the \( i \)th inequality constraint is active at the optimal, then \( g_i(x^*) = 0 \) and \( \lambda_i^* \geq 0 \), which leads to
\[ \gamma_{\lambda, i} = \tilde{\gamma}_{\lambda, i} = 1. \]

If the \( i \)th inequality constraint is inactive at the optimal, then \( g_i(x^*) < 0 \) and \( \lambda_i^* = 0 \), implying that
\[ 1 - \tilde{\gamma}_{\lambda, i} = 1 - \frac{1 + \rho g_i(x^*)/\lambda_i}{1 + \rho g_i(x^*)/\lambda_i^*} \leq \inf_{u \in (0,1)} \frac{1}{u^2} = 1 \]
when \( \lambda_i \neq \lambda_i^* \), and trivially \( \tilde{\gamma}_{\lambda, i} = \gamma_{\lambda, i} = 1 \) when \( \lambda_i = \lambda_i^* = 0 \). Thus we can see that
\[
I - \tilde{\Gamma}_\lambda \succeq I - \Gamma_\lambda.
\]

Next, we can show that
\[
\nabla_x L_\rho(x, \lambda, \nu) = \nabla_x L_\rho(x, \lambda, \nu) - \nabla_x \Theta_\rho(x^*, \lambda, \nu^*)
\]
where we denote
\[
\hat{x}_{\lambda, i} := \begin{cases} [\rho g_i(x) + \lambda_i] + -[\rho g_i(x^*) + \lambda_i^*], & g_i(x) \neq g_i(x^*), \\ \rho(g_i(x) - g_i(x^*)), & g_i(x) = g_i(x^*), \end{cases} \]
and
\[ \hat{\Gamma}_{x, \lambda} := \text{diag}(\hat{\gamma}_{x, \lambda,1}, \ldots, \hat{\gamma}_{x, \lambda,m}). \]

Summarizing the above derivations, we get
\[ V_\rho \leq - (x - x^*)^T (\nabla f(x) - \nabla f(x^*)) + \frac{1}{2\rho} (\lambda - \lambda^*)^T (I - \tilde{\Gamma}_\lambda)(\lambda - \lambda^*) \]
where \( J_\rho(x) \) is the Jacobian matrix of \( g \) evaluated at \( x \).

It can be checked that \( \mathcal{Q} \) can be equivalently written as
\[ V_\rho \leq -\beta V_\rho - Z, \]

where we denote
\[
Z = b(x) - \ddot{x}^T Q_1 \ddot{x} + \dot{y}^T Q_2 \dot{y}
\]
and
\[ \dot{\hat{x}} = x - x^*, \quad \ddot{y} = \begin{bmatrix} \nu - \nu^* \end{bmatrix} \lambda - \lambda^* \]
where
\[ Q_1 = cA^T A + c \left( J^T \hat{\Gamma}_{x, \lambda} J + J^T \hat{\Gamma}_{x, \lambda} J \right) + \frac{\beta}{2} I, \]
\[ \dot{\hat{y}} = \frac{1}{\rho}(I - \tilde{\Gamma}_\lambda) + \frac{c}{2}(J^T \Gamma_\lambda + \Gamma_\lambda J^T) - \frac{\beta}{2} I, \]
\[ Q_2 = \begin{bmatrix} cA^T - \frac{\beta}{2} I \\ cA^T \end{bmatrix} \begin{bmatrix} \hat{Q}_2 \\ \frac{\gamma}{2} (I + \Gamma_\lambda) J A \end{bmatrix}, \]
and
\[ Q_3 = \frac{1}{2\rho} \left[ (\Gamma_\lambda - I) J \right]. \]

We now need to show \( Z \geq 0 \), which will then imply \( V_\rho \leq -\beta V_\rho \) by \( 10 \). Without loss of generality we assume that \( \mathcal{I} = \{1, 2, \ldots, |\mathcal{I}|\} \). We first present the following lemma to give a positive definite lower bound of \( Q_2 \), whose proof is postponed to Appendix C.

**Lemma 3.** We have
\[ Q_2 \succeq \frac{c}{2} \begin{bmatrix} A^T A & A^T J \\ J^T A & J^T J \end{bmatrix} L^2 \]
when \( 4a \) is satisfied.
Lemma 3 implies that $Q_2$ is positive definite as well as $Q_2^{-1}$. This allows us to rewrite $Z$ in (11) to the following form

$$Z = b(x) - \hat{x}^T Q_1 \hat{x} - c^2 \|Q_3 \hat{x} + Kw(x)\|^2_{Q_2^{-1}} + \|\hat{y} + c Q_2^{-1} Q_3 \hat{x} + c Q_2^{-1} Kw(x)\|^2_{Q_2}.$$ 

Now $Z \geq 0$ will follow directly from the following lemma:

**Lemma 4.** We have

$$b(x) - \hat{x}^T Q_1 \hat{x} - c^2 \|Q_3 \hat{x} + Kw(x)\|^2_{Q_2^{-1}} \geq 0$$

when the conditions 4 are satisfied.

The proof of Lemma 4 is presented in Appendix II.

We consider a convex program that models the optimal curtailment of generations by $n$ solar panels in a distribution feeder of $m$ nodes. The problem is formulated as

$$\begin{align*}
\min_{p,q \in \mathbb{R}^n} & \quad \sum_{i=1}^{n} c_p (p_i - p_i^{PV})^2 + c_q q_i^2 \\
\text{s.t.} & \quad p_i^2 + q_i^2 \leq S_{\text{max},i}, \quad i = 1, \ldots, n, \\
& \quad 0 \leq p \leq p^{PV}, \\
& \quad v_{\text{min}} \leq Mp + Nq + r \leq v_{\text{max}}.
\end{align*}$$

Here $p, q \in \mathbb{R}^n$ model the real and reactive power injections of the inverters connected to solar panels, $S_{\text{max}} \in \mathbb{R}^n$ represents the rated apparent power of the inverters, $p^{PV} \in \mathbb{R}^n$ represents the real power generated by solar panels; the map $(p,q) \mapsto Mp + Nq + r$ is a linear model derived from the DistFlow equations [24] that maps power injections to voltage magnitudes, and $v_{\text{min}}, v_{\text{max}} \in \mathbb{R}^m$ represent bounds on voltage magnitudes. $c_p$ and $c_q$ are real positive constants.

The distribution feeder is based on the IEEE 37-node test feeder [25], where we adopt its topology, line parameters and loads, and modify it to be a single-phase network. Figure 1 shows the network topology and the locations where the solar panels and inverters are installed, and Table I gives the rated apparent power $S_{\text{max},i}$ for each inverter.

![Fig. 1. Topology of the distribution feeder. Solar panels and inverters are installed at buses marked by green hollow circles.](image)

| Inverter ID | 1 | 2 | 3 | 4 | 5 | 6 |
|-------------|---|---|---|---|---|---|
| Bus No.     | 2 | 4 | 5 | 6 | 7 | 10 |
| $S_{\text{max},i}$ (p.u.) | 2.7 | 1.35 | 2.7 | 1.35 | 2.025 | 2.025 |
| Inverter ID | 7 | 8 | 9 | 10 | 11 | 12 |
| Bus No.     | 13 | 15 | 16 | 20 | 21 | 27 |
| $S_{\text{max},i}$ (p.u.) | 2.7 | 2.7 | 1.35 | 2.025 | 2.025 | 2.025 |
| Inverter ID | 13 | 14 | 15 | 16 | 17 | 18 |
| Bus No.     | 28 | 29 | 31 | 32 | 33 | 34 |
| $S_{\text{max},i}$ (p.u.) | 2.7 | 2.7 | 1.35 | 2.7 | 2.025 | 1.35 |

As a function of time $t$ for 20 instances of randomly selected initial points $(x(0), \lambda(0))$. We see that while $\|(x(t) - x^*, \lambda(t) - \lambda^*)\|$ decreases exponentially on the whole, the exponential decay rates can change as $(x(t), \lambda(t))$ approaches the KKT point. Especially, we observe that the exponential decay rates for $t \in (0, 10)$ are smaller than those for $t > 40$ where $(x(t), \lambda(t))$ finally achieves the same stable decay rate for all 20 instances. This observation suggests that the semi-global exponential stability for nonlinearly constrained problems is not an artifact of the analysis, but further investigations are necessary for solid conclusions.

![Fig. 2. Illustration of the relative distances to $(x^*, \lambda^*)$ with respect to time $t$ for 20 random instances.](image)
VI. CONCLUSION

This paper introduced the augmented primal-dual gradient dynamics for convex optimization, and analyzed its stability behavior. Specifically, it was shown that the augmented primal-dual gradient dynamics achieve semi-global exponential stability when the objective has a quadratic gradient growth. This work extended the results in [15] to more general settings where the constraint functions can be convex and nonlinear.

Out theoretical analysis only justifies semi-global exponential stability of the augmented primal-dual gradient dynamics, and the bound on the convergence rate \( \beta \) is not robust to small perturbations of the problem; these observations and issues are worth further investigations. We are also interested in investigating the performance of the augmented primal-dual gradient dynamics in time-varying settings.

APPENDIX I

PROOF OF LEMMA 3

Obviously (44) implies

\[
\delta \leq \frac{2\delta_{\min}}{23(4L_g^2 + \|A\|^2)}.
\]

Noting that \( \gamma_{\lambda,i} = \tilde{\gamma}_{\lambda,i} = 1 \) for \( i \in I \), we can partition the matrix \( Q_2 \) as

\[
Q_2 = \begin{bmatrix}
Q_{2,I^c} & Q_{2,I^c} \\
Q_{2,I^c}^T & Q_{2,I^c}
\end{bmatrix},
\]

where we denote

\[
Q_{2,I^c}, Q_{2,I^c} = \begin{bmatrix}
\frac{c}{2} A \, A^T & \frac{c}{2} \tilde{J}_J^T \\
\frac{c}{2} \tilde{J}_J & \tilde{J}_J^T (I + \Gamma_{\lambda,I^c})
\end{bmatrix},
\]

\[
Q_{2,I^c}^T = \begin{bmatrix}
\frac{1}{2} (I - \Gamma_{\lambda,I^c}) \\
\frac{1}{2} (I - \Gamma_{\lambda,I^c})^T
\end{bmatrix},
\]

\[
\frac{c}{2} \Gamma_{\lambda,I^c} = \text{diag}(\gamma_{\lambda,i})_{i \in I}, \quad \tilde{\Gamma}_{\lambda,I^c} = \text{diag}(\tilde{\gamma}_{\lambda,i})_{i \in I},
\]

and \( J_{I^c} \) is formed by the rows of \( J \) whose indices are in \( I^c \).

By the definition of \( \delta_{\min} \), we have \( 1 - \Gamma_{\lambda,I^c} \geq \delta_{\min} I \) for all \( t \geq 0 \). Together with (8), it can be shown that

\[
Q_{2,I^c} \geq \begin{bmatrix}
\frac{1}{2k} I - cL_g^2 \delta_{\min} I + cL_g^2 (I - \Gamma_{\lambda,I^c}) \\
\frac{1}{2k} (I - \Gamma_{\lambda,I^c})^T
\end{bmatrix},
\]

\[
\frac{c}{2k} \Gamma_{\lambda,I^c} = \text{diag}(\gamma_{\lambda,i})_{i \in I}, \quad \tilde{\Gamma}_{\lambda,I^c} = \text{diag}(\tilde{\gamma}_{\lambda,i})_{i \in I},
\]

and \( J_{I^c} \) is formed by the rows of \( J \) whose indices are in \( I^c \).

By [15, Lemma 6], we have \( L_g^2 (I - \Gamma_{\lambda,I^c}) + \frac{1}{2} (J_{I^c}, J_{I^c}^T \Gamma_{\lambda,I^c} + \tilde{\Gamma}_{\lambda,I^c} J_{I^c}, J_{I^c}^T) \geq 0 \), and so

\[
Q_{2,I^c} \geq \begin{bmatrix}
\frac{1}{2k} I - cL_g^2 \delta_{\min} I - \frac{c}{2k} I
\end{bmatrix},
\]

then we have

\[
Q_{2,I^c} = \frac{c}{2k} \begin{bmatrix}
A^T J_{I^c}^T & A J_{I^c}^T \\
J_{I^c} A^T & J_{I^c} J_{I^c}^T
\end{bmatrix},
\]

by the definition of \( \kappa \), and

\[
Q_{2,I^c} = \frac{cL_g^2}{2} I - Q_{2,I^c} - \left( 2 - \frac{c}{2k} J_{I^c} A^T J_{I^c}^T \right) \left( J_{I^c} J_{I^c}^T \right)^{-1} Q_{2,I^c} \geq c \left( \frac{1}{2k} - L_g^2 \right) \delta_{\min} I - \frac{L_g^2}{2} I.
\]

We have w(x)^T K x = (A^T J_{I^c}^T A J_{I^c}^T I^T (J_{I^c})^T I^T I) \left( J_{I^c} \right) w(x) \leq w(x)^T \left( I + L_g^2 J_{I^c}^T J_{I^c} \right) w(x) \leq 2 \|w(x)\|^2, and

\[
\frac{1}{4L_g^2} J_{I^c}^T (\Gamma_{\lambda,I^c} - I) J_{I^c} \leq \frac{1}{4L_g^2} I,
\]

and

\[
w(x)^T K x = \frac{1}{2L_g^2} \|w(x)\|^2 \|J_{I^c} (\Gamma_{\lambda,I^c} - I) J_{I^c} \| \leq \frac{1}{2L_g^2} \|w(x)\|^2 \|\tilde{x}\|.
\]
Therefore
\[
\begin{aligned}
    b(x) - \hat{x}^T Q_1 \hat{x} - c^2 \|Q_3 \hat{x} + K w(x)\|^2_{Q_2^{-2}} \\
    \geq b(x) - \hat{x}^T Q_1 \hat{x} \\
    - 2c \left( 2\|w(x)\|^2 + \frac{1}{4\rho^2} \|\hat{x}\|^2 + \frac{1}{\rho} \|w(x)\| \|\hat{x}\| \right).
\end{aligned}
\]

Since \( f(x) + \Theta_\rho(x, \lambda) \) is convex in \( x \) and its gradient with respect to \( x \) is \((\ell + M_\Theta)\)-Lipschitz in \( x \), we have (see [18, Theorem 2.1.5])
\[
(\nabla f(x) - \nabla f(x^*) + \nabla x \Theta_\rho(x, \lambda) - \nabla x \Theta_\rho(x^*, \lambda))^T (x - x^*) 
\geq \frac{1}{\ell + M_\Theta} \|\nabla f(x) - \nabla f(x^*) + \nabla x \Theta_\rho(x, \lambda) - \nabla x \Theta_\rho(x^*, \lambda)\|^2.
\]

In addition, (4b) implies that
\[
\frac{\kappa \mu}{2 \beta} > \|A\|^2 + L_g^2 \\
\implies \beta < \frac{\kappa \mu}{2\|A\|^2 + \|A\|^2} \leq \frac{\ell}{\ell + M_\Theta}.
\]

where we have used \( \kappa \leq \ell \Delta_{L^2} + \|A\|^2 \) and \( \beta \leq \ell \). Therefore we can bound \( \|w(x)\| \) by
\[
\begin{aligned}
    4\|w(x)\|^2 \\
    = \|\nabla f(x) - \nabla f(x^*) + \nabla x \Theta_\rho(x, \lambda) - \nabla x \Theta_\rho(x^*, \lambda)\|^2 \\
    - 2\beta \hat{x}^T (\nabla f(x) - \nabla f(x^*) + \nabla x \Theta_\rho(x, \lambda) - \nabla x \Theta_\rho(x^*, \lambda)) \\
    + \beta^2 \|\hat{x}\|^2 \\
    \leq \|\nabla f(x) - \nabla f(x^*) + \nabla x \Theta_\rho(x, \lambda) - \nabla x \Theta_\rho(x^*, \lambda)\|^2 + \beta^2 \|\hat{x}\|^2 \\
    \leq \ell (\ell + M_\Theta) \|\nabla f(x) - \nabla f(x^*)\|^2 + \beta^2 \|\hat{x}\|^2 + (\ell + M_\Theta) M_\Theta \|\hat{x}\|^2 \\
    + \beta^2 \|\hat{x}\|^2,
\end{aligned}
\]

The first inequality in (12) can also be relaxed by
\[
\begin{aligned}
    (\ell + M_\Theta - 2\beta) \hat{x}^T \left[ \nabla f(x) - \nabla f(x^*) \right. \\
    + \nabla x \Theta_\rho(x, \lambda) - \nabla x \Theta_\rho(x^*, \lambda) \left]. \right] + \beta^2 \|\hat{x}\|^2 \\
    \leq \ell (\ell + M_\Theta - 2\beta) (\ell + M_\Theta) \|\hat{x}\|^2 + \beta^2 \|\hat{x}\|^2 \\
    = \ell (\ell + M_\Theta)^2 \|\hat{x}\|^2,
\end{aligned}
\]

which shows that \( \|w(x)\| \leq (\ell + M_\Theta) \|\hat{x}\|/2 \). It’s not hard to see that \( Q_1 \preceq c(\|A\|^2 + L_g^2 + 4/\rho) \). Now we can get
\[
\begin{aligned}
    b(x) - \hat{x}^T Q_1 \hat{x} - c^2 \|Q_3 \hat{x} + K w(x)\|^2_{Q_2^{-2}} \\
    \geq (1 - c(\ell + M_\Theta)) \|\nabla f(x) - \nabla f(x^*)\|^2 \hat{x} \\
    - c \left( \|A\|^2 + L_g^2 + \frac{\kappa}{4} \right) \|\hat{x}\|^2 \\
    - c \left( (\ell + M_\Theta) M_\Theta + \beta^2 \right) \|\hat{x}\|^2 \\
    - \frac{c}{2\rho^2} \|\hat{x}\|^2 - \frac{c}{\rho} (\ell + M_\Theta) \|\hat{x}\|^2 \\
    \geq \|\hat{x}\|^2 \left( \mu (1 - c(\ell + M_\Theta)) \\
    - c \left( \|A\|^2 + L_g^2 + \frac{\kappa}{4} + \beta^2 + (\ell + M_\Theta)(M_\Theta + 1/\rho) + \frac{1}{2\rho^2} \right) \right).
\end{aligned}
\]

By the condition (4b) and that \( c = 2\kappa^{-1}\beta \), we get the desired inequality.