

Tight lower bounds on the convergence rate of primal-dual dynamics for equality constrained convex problems

Ibrahim K. Ozaslan, *Graduate Student Member, IEEE*, and Mihailo R. Jovanović, *Fellow, IEEE*

Abstract—We study the exponential stability of continuous-time primal-dual gradient flow dynamics for convex optimization problems with linear equality constraints. Without making any assumptions on the rank of the constraint matrix, we obtain a tight lower bound on the worst-case convergence rate for smooth strongly convex objective functions. Our analysis identifies two different convergence regimes depending on the ratio between primal and dual time scales. When the primal time scale is inversely proportional to the Lipschitz parameter of the objective function and the dual time scale is large enough, the convergence rate is inversely proportional to the condition number of the problem. In contrast to the existing results, our lower bound on the convergence rate does not depend on the condition number of the constraint matrix.

Index Terms—Convex optimization, exponential stability, gradient flow dynamics, primal-dual methods.

I. INTRODUCTION

We study the constrained convex problems of the form

$$\begin{aligned} & \underset{x}{\text{minimize}} && f(x) \\ & \text{subject to} && Ex - q = 0 \end{aligned} \quad (1)$$

where $x \in \mathbb{R}^n$ is the optimization variable, $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is the objective function, $E \in \mathbb{R}^{d \times n}$ is the constraint matrix, and $q \in \mathbb{R}^d$ is a given vector. The Lagrangian associated with problem (1) is given by

$$\mathcal{L}(x; y) = f(x) + y^T(Ex - q) \quad (2)$$

where $y \in \mathbb{R}^d$ is the vector of dual variables associated with the linear equality constraint in (1). We are interested in characterizing a tight lower bound on the worst-case convergence rate of Primal-Dual (PD) gradient flow dynamics

$$\begin{aligned} \dot{x} &= -\alpha_1 \nabla_x \mathcal{L}(x; y) \\ \dot{y} &= +\alpha_2 \nabla_y \mathcal{L}(x; y) \end{aligned} \quad (3)$$

where $\alpha_1, \alpha_2 > 0$ are primal and dual time scale parameters.

Stability and convergence properties of PD gradient flow dynamics resulting from possibly different saddle functions have been studied since the seminal paper [1]. Early results [2], [3] provided conditions for the global asymptotic stability of the projected PD dynamics based on the standard Lagrangian for convex problems with inequality constraint. In [4]–[7], asymptotic stability of the projected PD dynamics was established for general saddle functions under different assumptions. More recently, the focus shifted toward

establishing conditions for the exponential stability of PD dynamics. In [8]–[14], global exponential stability of the PD dynamics resulting from various types of Lagrangian functions was proved for possibly non-differentiable strongly convex optimization problems.

Analyses of gradient flow dynamics have proved useful in studying the behavior of iterative optimization algorithms. For example, the continuous-time dynamical system viewpoint has recently provided insights into the acceleration phenomenon of the gradient methods for unconstrained optimization problems [15]–[17]. Similarly, study of PD gradient flow dynamics represents an important initial step toward understanding the behavior of PD gradient methods. These primal-descent dual-ascent algorithms are the workhorse for solving large-scale min-max problems that arise, for example, in training of generative adversarial networks [18]. However, in contrast to first-order methods for unconstrained minimization, the theory of PD gradient methods is still in its early stages. While lower complexity bounds have been established for gradient methods used in unconstrained minimization almost a half-century ago [19], equivalent results are still scarce for their PD equivalents. Such results are absent even when saddle functions have convex-linear forms that arise in convex problems with equality constraints.

In [8], by assuming that E is a full-row rank matrix and f is a strongly convex smooth function, it was shown that dynamics (3) are exponentially stable with conservative convergence rate estimate. In [14], we showed that the rank assumption on E is not necessary for the exponential stability and a conservative rate estimate was provided. The main motivation in these papers, including [9]–[11], was to prove the exponential stability of the dynamics rather than obtaining tight convergence rates, whose knowledge is essential for designing accelerated algorithms. Recently [20]–[23] proposed accelerated versions of (3) and established improved sub-linear convergence rate of $O(1/t^2)$ in the absence of strong convexity. However, similar improvements have yet to be seen for strongly convex functions for which we do not have tight bounds on the linear convergence rate.

In this paper, we obtain a tight lower bound on the linear convergence rate for (3) under the assumption that f is a strongly convex function with a Lipschitz continuous gradient. Our analysis shows how the time scale parameters impact the convergence rate without making any assumptions on the rank of the constraint matrix E . These parameters have been used as an effective heuristics for avoiding limit

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.

cycles and divergence of primal-dual algorithms [7], [24]. Theoretical understanding of their role for the case study that we examine serves as a motivation for a careful analysis of more intricate convex-concave problems.

The rest of the paper is organized as follows. In Section II, we introduce problem setup and provide background material. We present our main result in Section III and relegate the proofs to Section IV. We demonstrate the validity of our findings through computational experiments in Section V and conclude the paper in Section VI with remarks.

Notation: The range space and the null space of matrix A are denoted by $\mathcal{R}(A)$ and $\mathcal{N}(A)$, respectively.

II. BACKGROUND

If q belongs to the range space of E , $q \in \mathcal{R}(E)$, problem (1) is feasible and the Karush-Kuhn-Tucker conditions

$$\nabla f(x^*) = -E^T y^*, \quad E x^* = q \quad (4)$$

are necessary and sufficient for (x^*, y^*) to be a primal-dual solution pair to (1) [25]. When f is strongly convex, problem (1) has a unique solution x^* and the optimal dual variable y^* belongs to an affine set,

$$\begin{aligned} \mathcal{Y}^* &= \{y^* \in \mathbb{R}^d \mid \nabla f(x^*) = -E^T y^*\} \\ &= \{y^* \in \mathbb{R}^d \mid y^* = y_0^* + U_2 w, w \in \mathbb{R}^{d-r}\}. \end{aligned}$$

Here, r is the rank of E , $y_0^* \in \mathcal{R}(E)$ is the unique vector satisfying $\nabla f(x^*) = -E^T y_0^*$, the columns of U_2 form an orthonormal basis for $\mathcal{N}(E^T)$, and E has the singular value decomposition

$$E = [U_1 \quad U_2] \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}$$

where $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r)$ with $\sigma_1 \geq \dots \geq \sigma_r > 0$. In what follows, we use $\bar{\sigma}$ and $\underline{\sigma}$ to denote the largest and the smallest nonzero singular values σ_1 and σ_r of E .

To compute a solution pair (x^*, y^*) for some $y^* \in \mathcal{Y}^*$, we cannot directly solve the system of equations (4) because of the nonlinear term ∇f . Instead, we use an alternative characterization of optimal primal-dual pairs: every solution (x^*, y^*) of (4) is a saddle point of the Lagrangian satisfying

$$\mathcal{L}(x^*; y) \leq \mathcal{L}(x^*; y^*) \leq \mathcal{L}(x; y^*), \quad \forall x, y.$$

Based on this characterization, a solution to problem (1) can be computed by simultaneous minimization and maximization of the Lagrangian over primal variable x and dual variable y , respectively. To this end, we deploy primal-dual gradient flow dynamics (3) which simplify to

$$\dot{x} = -\alpha_1 \nabla_x \mathcal{L}(x; y) = -\alpha_1 (\nabla f(x) + E^T y) \quad (5a)$$

$$\dot{y} = +\alpha_2 \nabla_y \mathcal{L}(x; y) = +\alpha_2 (E x - q) \quad (5b)$$

The equilibrium points of these primal-descent dual-ascent gradient flow dynamics are obtained by setting $\dot{x} = 0$ and $\dot{y} = 0$ and they correspond to the saddle points of the Lagrangian characterized by conditions (4). Thus, we can use these dynamics to compute a solution to problem (1).

III. MAIN RESULTS

Our main result characterizes a tight lower bound on the worst-case convergence rate of PD gradient flow dynamics (5) under the following assumption.

Assumption 1: Problem (1) is feasible, i.e., $q \in \mathcal{R}(E)$, and the objective function f is m -strongly convex with an L -Lipschitz continuous gradient.

Theorem 2 in [14] shows that any solution to (5) converges exponentially to a limit point with rate ρ , i.e.,

$$\begin{aligned} \|(x(t), y(t)) - (x^*, \bar{y}^*)\|_2 \\ \leq c e^{-\rho t} \|(x(0), y(0)) - (x^*, \bar{y}^*)\|_2, \quad \forall t \geq 0 \end{aligned}$$

where c is a positive constant and

$$\bar{y}^* = y_0^* + U_2 U_2^T y(0).$$

Theorem 1: Let Assumption 1 hold. The primal-dual gradient flow dynamics (5) are globally exponentially stable with the convergence rate $\rho \geq \bar{\rho}$ where $\bar{\rho}$ satisfies,

$$\frac{\alpha_1 m}{4} \leq \bar{\rho} \leq \frac{\alpha_1 m}{2}, \quad \frac{\alpha_2}{\alpha_1} \geq \frac{mL}{4\underline{\sigma}^2} \quad (6a)$$

$$\bar{\rho} = \frac{\alpha_1 L}{2} \left(1 - \sqrt{1 - \frac{4\sigma^2 \alpha_2}{L^2 \alpha_1}} \right), \quad \frac{\alpha_2}{\alpha_1} \leq \frac{mL}{4\underline{\sigma}^2}. \quad (6b)$$

Furthermore, in both cases, the upper bounds on $\bar{\rho}$ cannot be improved for any ratio α_2/α_1 .

Proof of Theorem 1 is given in Section IV. As discussed in [26], for a continuous-time optimization algorithm any convergence rate can be achieved by increasing algorithmic parameters (in our case, α_1 and α_2). This ambiguity can be avoided by fixing α_1 .

If in addition to Assumption 1 the matrix E is full row rank and f is twice continuously differentiable, reference [8] showed that $\bar{\rho} = 0.25 \min(\alpha_2 \underline{\sigma}^2/L, m \underline{\sigma}^2/\bar{\sigma}^2)$ for $\alpha_1 = 1$. This lower bound suggests that, when α_1 is fixed, the convergence rate of (5) is inversely proportional to the square of the condition number $\bar{\sigma}^2/\underline{\sigma}^2$ of the constraint matrix E . In contrast, our result shows that the lower bound on ρ does not depend on the condition number of E . Theorem 1 also identifies two convergence regimes for dynamics (5); these respectively occur for large and small values of α_2/α_1 . If α_2/α_1 is smaller than $\gamma := mL/(4\underline{\sigma}^2)$, the lower bound $\bar{\rho}$ determined by (6b) is bounded from below by $\alpha_2 \underline{\sigma}^2/L$. Thus, the convergence rate does not depend on α_1 in this regime. On the other hand, when $\alpha_2/\alpha_1 \geq \gamma$ the lower bound given by (6a) demonstrates that the convergence rate is independent of α_2 , i.e., $\bar{\rho} \sim O(\alpha_1 m)$. Similar observation was made using computational experiments in Figure 1 of [8] but theoretical explanation was not provided.

IV. PROOF OF MAIN THEOREM

The coordinate transformations

$$\begin{aligned} \tilde{x}_1 &:= V_1^T (x - x^*) & \tilde{y}_1 &:= U_1^T (y - y_0^*) \\ \tilde{x}_2 &:= V_2^T (x - x^*) & \tilde{y}_2 &:= U_2^T (y - y(0)) \end{aligned} \quad (7)$$

brings dynamics (5) to

$$\begin{aligned} \begin{bmatrix} \dot{\tilde{x}}_1 \\ \dot{\tilde{y}}_1 \end{bmatrix} &= \begin{bmatrix} -\alpha_1 m I & -\alpha_1 \Sigma \\ \alpha_2 \Sigma & 0 \end{bmatrix} \begin{bmatrix} \tilde{x}_1 \\ \tilde{y}_1 \end{bmatrix} - \begin{bmatrix} \alpha_1 I \\ 0 \end{bmatrix} \tilde{u}_1 \\ \dot{\tilde{x}}_2 &= -\alpha_1 m \tilde{x}_2 - \alpha_1 \tilde{u}_2 \end{aligned} \quad (8)$$

where $\dot{\tilde{y}}_2 = 0$ and

$$\begin{bmatrix} \tilde{u}_1 \\ \tilde{u}_2 \end{bmatrix} := \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} (\nabla f(x) - mx - (\nabla f(x^*) - mx^*)).$$

Since trajectories of (5) are perpendicular to $\mathcal{N}(E^T)$, for any $t \geq 0$ we have $\tilde{y}_2(t) = U_2^T y(0)$, and only dynamics (8) have to be analyzed; see [14] for additional details.

System (8) can be viewed as a linear dynamical system in feedback with a nonlinear block Δ ,

$$\dot{\psi} = A\psi + B\tilde{u}, \quad \tilde{x} = C\psi, \quad \tilde{u} = \Delta(\tilde{x}) \quad (9)$$

where $\tilde{x} = [\tilde{x}_1^T \tilde{x}_2^T]^T$, $\tilde{u} = [\tilde{u}_1^T \tilde{u}_2^T]^T$, $\psi = [\tilde{x}_1^T \tilde{y}_1^T \tilde{x}_2^T]^T$, and

$$A = \begin{bmatrix} -\alpha_1 m I & -\alpha_1 \Sigma & 0 \\ \alpha_2 \Sigma & 0 & 0 \\ 0 & 0 & -\alpha_1 m I \end{bmatrix}, B = \begin{bmatrix} -\alpha_1 I & 0 \\ 0 & 0 \\ 0 & -\alpha_1 I \end{bmatrix}$$

$$C = \begin{bmatrix} I & 0 & 0 \\ 0 & 0 & I \end{bmatrix}.$$

Since affine transformation with a unitary matrix does not change parameters of strong convexity and Lipschitz continuity, $f(V\tilde{x} + x^*)$ is an m -strongly convex function with an L -Lipschitz continuous gradient. This implies Lipschitz continuity with modulus $\tilde{L} := L - m$ and monotonicity [10] of the nonlinear term Δ which is defined as

$$\begin{aligned} \Delta(\tilde{x}) &:= V^T (\nabla f(x) - mx - (\nabla f(x^*) - mx^*)) \\ &= V^T (\nabla f(V\tilde{x} + x^*) - mV\tilde{x} - \nabla f(x^*)). \end{aligned} \quad (10)$$

This equation along with the aforementioned properties of Δ can be used to characterize it in terms of the following sector bound condition,

$$\begin{bmatrix} \tilde{x} - \bar{x} \\ \tilde{u} - \bar{u} \end{bmatrix}^T \Pi \begin{bmatrix} \tilde{x} - \bar{x} \\ \tilde{u} - \bar{u} \end{bmatrix} \geq 0 \quad (11a)$$

where (\bar{x}, \bar{u}) is any reference point satisfying $\bar{u} = \Delta(\bar{x})$ and Π is given by [27],

$$\Pi = \begin{bmatrix} 0 & \tilde{L}I \\ \tilde{L}I & -2I \end{bmatrix} \quad (11b)$$

The existence of a quadratic Lyapunov function $V(\psi) = \psi^T P \psi$ whose derivative along the solutions of (9) satisfies $\dot{V} \leq -2\rho V$ provides a sufficient condition for the ρ -exponential stability of (9). This condition is equivalent to the existence of matrix $P \succ 0$ that satisfies [28, Theorem 3],

$$\begin{bmatrix} A_\rho^T P + P A_\rho & P B \\ B^T P & 0 \end{bmatrix} + \begin{bmatrix} C^T & 0 \\ 0 & I \end{bmatrix} \Pi \begin{bmatrix} C & 0 \\ 0 & I \end{bmatrix} \preceq 0. \quad (12)$$

where $A_\rho := A + \rho I$.

Since (A, B) is a controllable pair, if matrix A_ρ is Hurwitz, then an equivalent characterization of condition (12) in

the frequency domain is obtained using KYP Lemma [29, Theorem 1],

$$\begin{bmatrix} G_\rho^*(j\omega) & I \end{bmatrix} \Pi \begin{bmatrix} G_\rho(j\omega) \\ I \end{bmatrix} \preceq 0 \quad \forall \omega \in \mathbb{R} \quad (13)$$

where $G_\rho(j\omega) := C(j\omega I - A_\rho)^{-1} B$ and $G_\rho^*(j\omega)$ is the complex conjugate transpose of $G_\rho(j\omega)$. For the matrix Π in (11), condition (13) simplifies to

$$\text{Re}[G_\rho(j\omega)] \preceq (1/\tilde{L})I, \quad \forall \omega \in \mathbb{R}. \quad (14)$$

The transfer function of system (9) is given by

$$G_\rho(s) = \begin{bmatrix} G_1(s - \rho) & 0 \\ 0 & G_2(s - \rho) \end{bmatrix}$$

where G_1 and G_2 are diagonal matrices that determine transfer functions from \tilde{u}_k to \tilde{x}_k , for $k = 1, 2$. The nonzero entries of these diagonal matrices are given by

$$\begin{aligned} [G_1(s)]_{ii} &= \frac{-\alpha_1 s}{s^2 + \alpha_1 m s + \alpha_1 \alpha_2 \sigma_i^2} \quad i = 1, \dots, m \\ [G_2(s)]_{ii} &= -\alpha_1 / (s + \alpha_1 m) \quad i = 1, \dots, n - m. \end{aligned}$$

The rest of the proof is organized as follows. We first determine the conditions for the matrix A_ρ to be Hurwitz; since system (9) is controllable and observable, we can equivalently study stability of the transfer function $G_\rho(s)$. The $(\tilde{x}_1, \tilde{y}_1)$ - and \tilde{x}_2 -dynamics are decoupled and we examine their stability properties in Section IV-A and Section IV-B, respectively. Then, in Section IV-C and Section IV-D, we identify conditions that algorithmic parameters α_1 and α_2 as well as the convergence rate ρ have to satisfy in order for the KYP lemma to hold. In Section IV-D, we obtain the lower bound identified in (6) on the convergence rate. Finally, in Section IV-E, we demonstrate the tightness of the bounds given in (6) by providing an example where the best convergence rates match those established in Theorem 1.

A. Stability of the \tilde{x}_2 -dynamics

Transfer function $G_2(s - \rho)$ is stable if and only if all its poles have negative real parts. Since the denominator of $G_2(s - \rho)$ is given by $\Phi_2(s) := s - \rho + \alpha_1 m$, the convergence rate ρ must satisfy $\rho < \alpha_1 m$.

B. Stability of the \tilde{x}_1 -dynamics

Transfer function $G_1(s - \rho)$ is stable if and only if the roots of its denominator $\Phi_1(s) = (s - \rho)^2 + \alpha_1 m(s - \rho) + \alpha_1 \alpha_2 \sigma_i^2$ have negative real parts for all $i = 1, \dots, m$. The coefficients of Φ_1 are polynomials in ρ , i.e.,

$$\Phi_1(s) = s^2 + a_1(\rho)s + a_0(\rho)$$

where

$$\begin{aligned} a_0(\rho) &= \rho^2 - \alpha_1 m \rho + \alpha_1 \alpha_2 \sigma_i^2 \\ a_1(\rho) &= \alpha_1 m - 2\rho. \end{aligned}$$

The Routh-Hurwitz stability criterion

$$a_0(\rho) > 0, \quad a_1(\rho) > 0 \quad (15)$$

is a necessary and sufficient condition for the stability of Φ_1 . Due to the positivity of a_0 , the convergence rate must satisfy

$$\rho < \alpha_1 m / 2. \quad (16)$$

We skip positivity of $a_1(\rho)$ for now since it is guaranteed later by a stronger condition in Section IV-D.

C. KYP lemma for the \tilde{x}_2 -dynamics

Condition (14) implies that G_2 has to satisfy

$$\operatorname{Re}[G_2(j\omega - \rho)] = \operatorname{Re} \left[\frac{-\alpha_1}{j\omega + \alpha_1 m - \rho} \right] \leq 1/\tilde{L}$$

for all $\omega \in \mathbb{R}$, which is true under stability condition (16) since $G_2(j\omega - \rho)$ is a first-order transfer function with negative real part for any $\omega \in \mathbb{R}$.

D. KYP Lemma for the \tilde{x}_1 -dynamics

Condition (14) implies that G_1 has to satisfy

$$\begin{aligned} & \operatorname{Re}[G_1(j\omega - \rho)] \\ &= \operatorname{Re} \left[\frac{-\alpha_1(j\omega - \rho)}{a_1(\rho) - \omega^2 + j\omega a_0(\rho)} \right] \\ &= \frac{\alpha_1 \rho (a_1(\rho) - \omega^2) - \alpha_1 \omega^2 a_0(\rho)}{(a_1(\rho) - \omega^2)^2 + a_0(\rho)^2 \omega^2} \leq 1/\tilde{L} \\ &\iff \Gamma(\omega) := \omega^4 + b_2(\rho)\omega^2 + b_0(\rho) \geq 0 \end{aligned}$$

for all $\omega \in \mathbb{R}$ where

$$\begin{aligned} b_0(\rho) &= a_1(\rho)(a_1(\rho) - \alpha_1 \tilde{L}\rho) \\ b_2(\rho) &= a_0^2(\rho) - 2a_1(\rho) + \alpha_1 \tilde{L}(\rho + a_0(\rho)). \end{aligned}$$

A necessary and sufficient condition for $\Gamma(\omega) \geq 0, \forall \omega$ is

$$\omega^2 + b_2(\rho)\omega + b_0(\rho) \geq 0, \quad \forall \omega \geq 0$$

which is equivalent to showing that

$$b_0(\rho) \geq 0, \quad b_2(\rho) + 2\sqrt{b_0(\rho)} \geq 0. \quad (17)$$

It is worth noting that verifying only sufficient conditions for nonnegativity of Γ , i.e.,

$$b_0(\rho) \geq 0, \quad b_2(\rho) \geq 0$$

results in highly conservative bounds on ρ .

First inequality in (17). Fourth order polynomial $b_0(\rho)$ can be decomposed into the multiplication of two second order polynomials

$$\begin{aligned} b_0(\rho) &= (\rho^2 - \alpha_1 m \rho + \alpha_1 \alpha_2 \sigma_i^2)(\rho^2 - \alpha_1 L \rho + \alpha_1 \alpha_2 \sigma_i^2) \\ &= a_1(\rho) a_2(\rho) \end{aligned}$$

where $a_2(\rho) = \rho^2 - \alpha_1 L \rho + \alpha_1 \alpha_2 \sigma_i^2$.

Due to stability condition (15), $a_1(\rho)$ is positive, which implies that for nonnegativity of $b_0(\rho)$, $a_2(\rho)$ must be nonnegative since $a_1(\rho) \geq a_2(\rho)$ for any ρ . Hence, it suffices to check whether

$$\rho^2 - \alpha_1 L \rho + \alpha_1 \alpha_2 \sigma_i^2 \geq 0 \quad (18)$$

which is true in either of the following cases:

- (i) $\alpha_2/\alpha_1 \geq L^2/(4\sigma^2)$, i.e., the discriminant is non-positive. In this case, there is no restriction on the convergence rate except stability condition (16).
- (ii) $\alpha_2/\alpha_1 \leq L^2/(4\sigma^2)$, i.e., the discriminant is non-negative. Considering also (16), the convergence rate has to satisfy

$$\rho \leq \min \left\{ \frac{\alpha_1 m}{2}, \frac{\alpha_1 L}{2} \left(1 - \sqrt{1 - \frac{4\sigma^2 \alpha_2}{L^2 \alpha_1}} \right) \right\}.$$

In this case, the upper bound on the convergence rate is determined by the second argument in the min operator if and only if $\alpha_2/\alpha_1 \leq (2mL - m^2)/(4\sigma^2)$. Moreover, since $1 - \sqrt{1 - 2/z} \geq 1/z$ for any $z \geq 2$, if $\alpha_2/\alpha_1 \geq mL/(4\sigma^2)$, we have

$$\frac{\alpha_1 L}{2} \left(1 - \sqrt{1 - \frac{4\sigma^2 \alpha_2}{L^2 \alpha_1}} \right) \geq \frac{\alpha_1 m}{4}. \quad (19)$$

Second inequality in (17). The left hand side of second inequality can be expressed in terms of $a_1(\rho)$ and $a_2(\rho)$ as

$$\begin{aligned} \tilde{b}(\rho) &:= b_2(\rho) + 2\sqrt{b_0(\rho)} \\ &= 2\rho^2 - \alpha_1(m + L)\rho - 2\alpha_1\alpha_2\sigma_i^2 + \alpha_1^2 Lm \\ &\quad + 2\sqrt{a_1(\rho)a_2(\rho)} \\ &= a_1(\rho) + a_2(\rho) + 2\sqrt{a_1(\rho)a_2(\rho)} + \alpha_1^2 Lm \\ &\quad - 4\alpha_1\alpha_2\sigma_i^2 \\ &= \left(\sqrt{a_1(\rho)} + \sqrt{a_2(\rho)} \right)^2 + \alpha_1^2 Lm - 4\alpha_1\alpha_2\sigma_i^2 \end{aligned} \quad (20)$$

Since $\tilde{b}(\rho)$ is decreasing in α_2 and σ_i , it suffices to check whether $\tilde{b}(\rho) \geq 0$ for the m^{th} -mode, i.e., $\sigma_i = \sigma$. Unfortunately we do not have analytical expressions for the roots of $\tilde{b}(\rho)$ since it is signomial, i.e., its powers are not natural numbers. Nonetheless, in what follows, we show that

- (a) $\tilde{b}(\alpha_1 m/2) < 0$ for finite values of α_2/α_1 ratio, but the smallest root of $\tilde{b}(\rho)$ converges to $\alpha_1 m/2$ as ratio $\alpha_2/\alpha_1 \rightarrow \infty$.
- (b) $\tilde{b}(\alpha_1 m/4) > 0$ if $\alpha_2/\alpha_1 \geq (4mL - m^2)/(16\sigma^2)$.

Combining results (a) and (b) with (19) yields (6a). Moreover, if $\alpha_2/\alpha_1 \leq mL/4\sigma^2$, then $\tilde{b}(\rho) \geq a_2(\rho)$, which implies that (18) is the only limiting factor on the convergence rate, thus (6b) is obtained. These two results complete the first part of the proof.

To obtain results (a) and (b) above, let

$$v := \alpha_2 \sigma^2 / (\alpha_1 L^2), \quad w := m / (kL)$$

for some $k \geq 2$. Then, $\tilde{b}(\rho)$ at $\rho = \alpha_1 m/k$ becomes

$$\begin{aligned} & \tilde{b}(\alpha_1 m/k) / (\alpha_1^2 L^2) \\ &= (\sqrt{w^2 - w + v} + \sqrt{v - (k-1)w^2})^2 - 4v + kw \\ &= (2 - k)w^2 + (k - 1)w - 2v \\ &\quad + 2\sqrt{(w^2 - w + v)(v - (k-1)w^2)} \end{aligned} \quad (21)$$

Setting $k = 2$ in (21) yields

$$\begin{aligned} \frac{\tilde{b}(\alpha_1 m/2)}{2\alpha_1^2 L^2} &= w/2 - v + \sqrt{(w^2 - w + v)(v - w^2)} \\ &= \sqrt{(v - w/2)^2 + w^3 - w^4 - w^2/4} - (v - w/2) \end{aligned}$$

which converges to zero as $v \rightarrow \infty$ since

$$\lim_{z \rightarrow \infty} \sqrt{z^2 - a} - z = 0$$

for any $a < \infty$. For finite values of v on the other hand,

$$a_2(\alpha_1 m/2) \geq 0 \iff \alpha_2/\alpha_1 \geq (2mL - m^2)/(4\sigma^2)$$

which implies that the square-root term in (21) is positive and $\tilde{b}(\alpha_1 m/2) \geq 0$ if and only if

$$\begin{aligned} -wv + v^2 + w^3 - w^4 &\geq v^2 - wv + w^2/4 \\ \iff w^2(1/2 - w)^2 &\leq 0 \iff w = 1/2 \iff m = L. \end{aligned}$$

Thus, except for the least-norm problems with linear equality constraints, the convergence rate of (5) is strictly less than $\alpha_1 m/2$ for finite values of α_2 and it converges to $\alpha_1 m/2$ as $\alpha_2 \rightarrow \infty$. This proves result (a).

For result (b), we have

$$a_2(\alpha_1 m/4) \geq 0 \iff \alpha_2/\alpha_1 \geq (4mL - m^2)/(16\sigma^2).$$

Also, setting $k = 4$ in (21) yields

$$\frac{\tilde{b}(\alpha_1 m/4)}{2\alpha_1^2 L^2} = -w^2 + \frac{3}{2}w - v + \sqrt{(w^2 - w + v)(v - 3w^2)}$$

where the square-root term is positive if $\alpha_2/\alpha_1 \geq (4mL - m^2)/(16\sigma^2)$ and the other term is negative if and only if $v \geq (6mL - m^2)/(16L^2)$. Therefore, for the positivity of $\tilde{b}(\alpha_1 m/4)$, it suffices to check if

$$\begin{aligned} (w^2 - w + v)(v - 3w^2) &\geq (w^2 - \frac{3}{2}w + v)^2 \\ \iff v &\geq w(4w - 3)^2/(8(1 - 2w)) \end{aligned}$$

for $v \geq (6mL - m^2)/(16L^2)$. This condition is equivalent to verifying

$$\begin{aligned} (6mL - m^2)/(16L^2) &\geq w(4w - 3)^2/(8(1 - 2w)) \\ \iff 12 - 2/\kappa &\geq (1/\kappa - 3)^2/(1 - 1/(2\kappa)) \end{aligned}$$

which holds for any $\kappa = L/m \geq 1$. Thus, (b) is proved.

E. Tightness of the upper bound in (6)

To prove the second part of the theorem, we consider the least norm problem with linear equality constraints, i.e., we set $f(x) = \frac{m}{2}\|x\|^2$ in problem (1). After the coordinate transformation (7) is applied, dynamics (5) read

$$\begin{aligned} \begin{bmatrix} \dot{\tilde{x}}_1 \\ \dot{\tilde{y}}_1 \end{bmatrix} &= \begin{bmatrix} -\alpha_1 m I & -\alpha_1 \Sigma \\ \alpha_2 \Sigma & 0 \end{bmatrix} \begin{bmatrix} \tilde{x}_1 \\ \tilde{y}_1 \end{bmatrix} \\ \dot{\tilde{x}}_2 &= -\alpha_1 m \tilde{x}_2. \end{aligned} \quad (22)$$

The convergence rate of linear system (22) is determined by the eigenvalues of the state matrix. For \tilde{x}_1 -dynamics and $i = 1, \dots, m$, the eigenvalues are

$$\lambda_i = -\frac{\alpha_1 m}{2} \left(1 \pm \sqrt{1 - (4\alpha_2 \sigma_i^2)/(\alpha_1 m^2)} \right)$$

thus, the convergence rate of (22) is given by

$$\rho = \frac{\alpha_1 m}{2} \min_i \operatorname{Re} \left[1 - \sqrt{1 - (4\alpha_2 \sigma_i^2)/(\alpha_1 m^2)} \right]$$

which implies that the best lower bound on the worst-case convergence rate of (22) is $\rho = \alpha_1 m/2$ if $\alpha_2/\alpha_1 \geq m^2/(4\sigma^2)$; $\rho = \alpha_1 m(1 - \sqrt{1 - (4\sigma^2/m^2)(\alpha_2/\alpha_1)})/2$ otherwise. Since $m = L$ in this problem, bounds in (6) are recovered.

V. COMPUTATIONAL EXPERIMENTS

We demonstrate the validity of our analysis on problems where f has quadratic form $f(x) = 0.5x^T W x$. In this case (5) becomes a linear dynamical system whose worst-case convergence rate can be exactly determined by computing the eigenvalues of the state matrix. We use two different weight matrices W_1 and W_2 . To obtain these weight matrices, we first compute $Q^T Q$, where the entries of Q are sampled from standard normal distribution. Then, we alter the singular values in two different ways: for W_1 , the singular values are uniformly sampled from interval $[m, L]$ in the logarithmic scale; for W_2 , the largest singular value is set to L , and the rest is set to m . Then, two different constraint matrices $E_1, E_2 \in \mathbb{R}^{d \times n}$ are generated in the same way as Q . Singular values of E_1 and E_2 are scaled in the same way as W_1 . For both E_1 and E_2 , the smallest non-zero singular value is set to $\sigma = 0.1$, whereas the largest singular values are set to 1 and 10^4 , respectively.

In Figure 1, we plot the worst-case convergence rate of (5) and the lower bounds given in (6) for three different problems with parameters $(d, n, r, m, L) = (10, 20, 5, 1, 10)$. For (W_2, E_1) pair, the problem is very similar to finding the least-norm solution, hence as explained in Section IV-E, the convergence rate is saturated when α_2/α_1 exceeds $m^2/(4\sigma^2)$, then it is lower bounded by $\alpha_1 m/2$. This example illustrates the tightness of the lower bound. The other two pairs (W_1, E_1) and (W_1, E_2) demonstrate that the condition number of the constraint matrices, 10 and 10^5 respectively, does not have much impact on the worst-case convergence rate of (5).

In Figure 2, we plot the numerical convergence of (5) on (W_1, E_1) pair for different values of α_1 and α_2 . The ratio α_2/α_1 in the first six cases is smaller than the threshold $\gamma = mL/(4\sigma^2)$; hence, the convergence rate is substantially smaller than $\alpha_1 m/2$ as anticipated in Theorem 1. Moreover, in that region, $\bar{\rho}$ is well approximated by $\alpha_2 \sigma^2/L$ suggesting that the convergence is insensitive to α_1 , which can be observed in the figure as well. The ratio of α_2/α_1 is larger than γ in the remaining three cases where the convergence rate is lower bounded by $\alpha_1 m/2$. In that region, the convergence rate is proportional to α_1 but independent of α_2 . As seen in the figure, even if α_2 is multiplied by hundred, the change in the rate is insignificant.

VI. CONCLUSION

We studied the convergence of primal-dual gradient flow dynamics applied to the equality constrained convex prob-

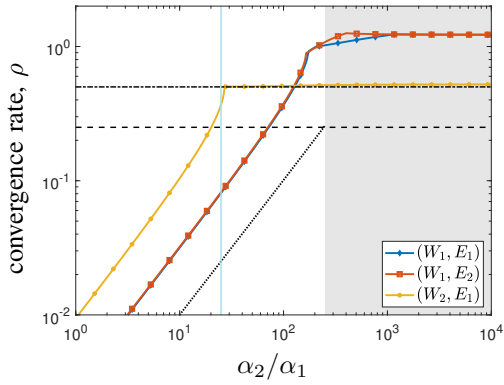


Fig. 1. Worst-case convergence rate and proposed lower bounds. The dot-dashed line is $\rho = \alpha_1 m/2$, the dashed line is $\rho = \alpha_1 m/4$ and the dotted line is $\bar{\rho}$ given in (6b). The boundary of the white and grey region is $\alpha_2/\alpha_1 = mL/(4\sigma^2)$, whereas the thin vertical line is $\alpha_2/\alpha_1 = m^2/(4\sigma^2)$. Thick lines with different markers show the worst-case convergence rate of (5) when the specified (weight matrix, constraint matrix) pairs in the legend are used in (1). In all cases, $\alpha_1 = 1$.

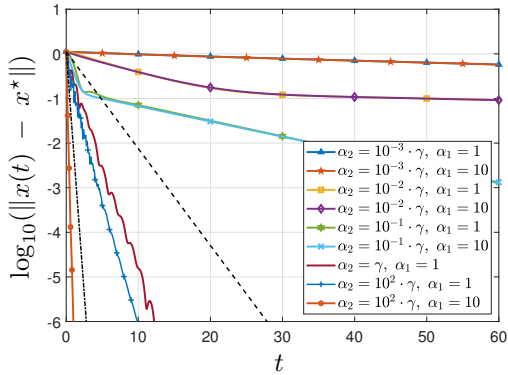


Fig. 2. Effect of α_1 and α_2 on the numerical convergence. The lines given in the legend show the convergence of (5) on the problem with pair (W_1, E_1) where $\gamma = mL/(4\sigma^2)$. The black dashed line is $\|x^*\|e^{-tm/2}$ and the black dot-dashed line is $\|x^*\|e^{-5tm}$.

lems with smooth strongly convex objective functions. Without making any assumptions on the constraint matrix, we obtained a tight lower bound on the linear convergence rate. Our analysis also reveals the effect of primal and dual time scales on convergence. Our ongoing effort aims to design accelerated dynamics for strongly convex problems with constraints using constant parameters. Also, it would be interesting to study the discrepancy between the convergence of continuous and discrete time dynamics.

REFERENCES

- [1] K. Arrow, L. Hurwicz, and U. H., *Studies in Linear and Non-linear Programming*. Palo Alto CA: Stanford Univ. Press, 1958.
- [2] D. Feijer and F. Paganini, "Stability of primal-dual gradient dynamics and applications to network optimization," *Automatica*, vol. 46, no. 12, pp. 1974–1981, Dec. 2010.
- [3] A. Cherukuri, E. Mallada, and J. Cortés, "Asymptotic convergence of constrained primal-dual dynamics," *Systems Control Lett.*, vol. 87, pp. 10–15, Jan. 2016.
- [4] R. Goebel, "Stability and robustness for saddle-point dynamics through monotone mappings," *Systems Control Lett.*, vol. 108, pp. 16–22, Oct. 2017.
- [5] A. Cherukuri, B. Ghahserifard, 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.

- [6] 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.
- [7] 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, Aug. 2021.
- [8] 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, June 2018.
- [9] 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.
- [10] N. K. Dhingra, S. Z. Khong, and M. R. Jovanović, "The proximal augmented Lagrangian method for nonsmooth composite optimization," *IEEE Trans. Automat. Control*, vol. 64, no. 7, pp. 2861–2868, July 2019.
- [11] X. Chen and N. Li, "Exponential stability of primal-dual gradient dynamics with non-strong convexity," in *Amer. Control Conf.*, 2020, pp. 1612–1618.
- [12] 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 *Proc. IEEE Conf. Decis. Control*, 2020, pp. 4836–4841.
- [13] I. K. Ozaslan and M. R. Jovanović, "Exponential convergence of primal-dual dynamics for multi-block problems under local error bound condition," in *Proc. IEEE Conf. Decis. Control*, 2022, pp. 7579–7584.
- [14] I. K. Ozaslan and M. R. Jovanović, "On the global exponential stability of primal-dual dynamics for convex problems with linear equality constraints," in *Proc. Am. Control Conf.*, 2023, pp. 210–215.
- [15] W. Su, S. Boyd, and E. Candes, "A differential equation for modeling nesterov's accelerated gradient method: theory and insights," in *Adv. Neural Inf. Process. Syst.*, 2014, pp. 5312–5354.
- [16] A. Wibisono, A. C. Wilson, and M. I. Jordan, "A variational perspective on accelerated methods in optimization," *Proc. Natl. Acad. Sci.*, vol. 113, no. 47, pp. E7351–E7358, Nov. 2016.
- [17] M. Muehlebach and M. I. Jordan, "Optimization with momentum: Dynamical, control-theoretic, and symplectic perspectives," *J. Mach. Learn. Res.*, vol. 22, no. 1, pp. 3407–3456, Jan. 2021.
- [18] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, Nov. 2020.
- [19] Y. E. Nesterov, "A method of solving a convex programming problem with convergence rate $O(1/k^2)$," *Dokl. Akad. Nauk SSSR*, vol. 269, no. 3, pp. 543–547, 1983.
- [20] X. Zeng, J. Lei, and J. Chen, "Dynamical primal-dual accelerated method with applications to network optimization," *IEEE Trans. Automat. Control*, vol. 68, no. 3, Feb. 2022.
- [21] H. Attouch, Z. Chbani, J. Fadili, and H. Riahi, "Fast convergence of dynamical ADMM via time scaling of damped inertial dynamics," *J. Optim. Theory Appl.*, vol. 193, no. 1-3, pp. 704–736, June 2022.
- [22] X. He, R. Hu, and Y. P. Fang, "Convergence rates of inertial primal-dual dynamical methods for separable convex optimization problems," *SIAM J. Control Optim.*, vol. 59, no. 5, pp. 3278–3301, June 2021.
- [23] X. He, R. Hu, and Y.-P. Fang, "Fast primal-dual algorithm via dynamical system for a linearly constrained convex optimization problem," *Automatica*, vol. 146, p. 110547, Dec. 2022.
- [24] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," in *Adv. Neural Inf. Process. Syst.*, 2017, pp. 6629–6640.
- [25] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge U.K.: Cambridge Univ. Press, 2004.
- [26] M. Muehlebach and M. Jordan, "Continuous-time lower bounds for gradient-based algorithms," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 7088–7096.
- [27] L. Lessard, B. Recht, and A. Packard, "Analysis and design of optimization algorithms via integral quadratic constraints," *SIAM J. Optim.*, vol. 26, no. 1, pp. 57–95, Jan. 2016.
- [28] B. Hu and P. Seiler, "Exponential decay rate conditions for uncertain linear systems using integral quadratic constraints," *IEEE Trans. Automat. Control*, vol. 61, no. 11, pp. 3631–3637, Jan. 2016.
- [29] A. Rantzer, "On the Kalman—Yakubovich—Popov lemma," *Systems Control Lett.*, vol. 28, no. 1, pp. 7–10, June 1996.