Accelerated Saddle Flow Dynamics for Bilinearly Coupled Minimax Problems

Yingzhu Liu, Enrique Mallada, Zhongkui Li, and Pengcheng You

***Abstract*— Minimax problems have attracted much attention due to various applications in constrained optimization problems and zero-sum games. Identifying saddle points within these problems is crucial, and saddle flow dynamics offer a straightforward yet useful approach. This study focuses on a class of bilinearly coupled minimax problems and designs an accelerated algorithm based on saddle flow dynamics that achieves a convergence rate beyond stereotype limits. The algorithm is derived based on a sequential two-step transformation of the objective function. First, a change of variable is aimed at a better-conditioned saddle function. Second, a proximal regularization, when staggered with the first step, guarantees strong convexity-strong concavity of the objective function that can be tuned for accelerated exponential convergence. Besides, such an approach can be extended to a class of weakly convex-weakly concave functions and still achieve exponential convergence to one stationary point. The theory is verified by a numerical test on an affine equality-constrained convex optimization problem.**

I. INTRODUCTION

The study of saddle flow dynamics traces back to the seminal work [1] and contributes to the fundamentals of mathematical optimization from a dynamical system perspective. Saddle flow dynamics aim to search for minimax saddle points by combining gradient descent with gradient ascent on two respective subsets of variables (that form a partition). This approach is the basis of primal-dual methods used for solving constrained optimization problems [2] and best-response dynamics used for locating Nash equilibria in zero-sum games [3], [4]. As a result, saddle flow dynamics have been widely used for resources allocation and stabilizing controller design in a variety of areas, including power systems [5], [6], communication networks [7], [8], and cloud computing [9].

To fully understand the dynamic behavior of saddle flows (and its variants), the convergence properties have been extensively studied. Preliminary results [10]–[13] are centered on asymptotic stability, primarily using advanced analytical tools to provide insights into the important special case of primal-dual dynamics. More recent efforts [14]–[16] have been made to study the exponential convergence of saddle flows for not only theory development but also practical uses. For example, [14], [15] have shown particular exponential convergence in the absence of strong convexity. [2] has further extended the study to discrete-time problems, showing linear convergence–the counterpart of exponential convergence in continuous time.

A fundamental challenge in studying these exponentially convergent dynamics is to estimate their decay rates using several algorithmic variants [17]–[23]. A vast amount of literature resorts to proximal regularization, especially when handling non-smooth functions. In particular, [17] only shows the existence of exponential convergence rates that depend on the strong convexity constant but fail to provide an estimate. Lower bounds on the convergence rate have been developed in [18], [19] and [20], respectively, using saddle flow variants and frequency-domain Integral Quadratic Constraint (IQC) approaches. [21] further establishes a best-case upper-bound rate with a Lyapunov method in the time domain. Another set of results [22], [23] develop a novel projection on the standard Lagrangian and derive lower bounds on the decay rates when proving (semi-)global exponential stability for the augmented primal-dual dynamics. By and large, the rate estimates provided in the literature depend on strong convexity, regularization coefficients, and singular values of any coupling matrices. Notably, the constant of strong convexity seems to be a bottleneck inherent in the convergence rate of saddle flow dynamics.

In this paper, we focus particularly on a class of bilinearly coupled minimax problems and design an algorithm based on carefully designed saddle flow dynamics that exploits the problem structure and achieves an exponential convergence rate beyond the given strong convexity constant. The algorithm involves two sequential steps - change-of-variable conditioning and proximal regularization - that jointly enhance the strong convexity-strong concavity of the objective function. Building upon our recent results [24], this property immediately results in a lower-bound estimate for the convergence rate that breaks the stereotype limit. In addition, such a performance guarantee makes the algorithm suited for a class of weakly convex-weakly concave functions with exponential convergence to a stationary point. Extensive simulations are run to compare our algorithm with existing methods and the actual rate with its lower-bound rate estimate.

Our contributions are summarized as follows.

(i) For bilinearly coupled minimax problems, we design a saddle flow-based algorithm with an exponential convergence rate beyond the given strong convexity constant.

(ii) We provide an explicit lower-bound estimate for the

This work was supported by the National Natural Science Foundation of China under grants 72201007, T2121002, 72131001, U2241214, and 62373008.

Y. Liu, Z. Li, and P. You are with the College of Engineering, Peking University, Beijing, China. Y. Liu and Z. Li are also with the State Key Laboratory for Turbulence and Complex Systems, Peking University, Beijing, China. {yzliucoe, zhongkli, pcyou}@pku.edu.cn

E. Mallada is with the Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA. mallada@jhu.edu

exponential convergence rate of the algorithm, which offers a guideline for the choice of parameters to optimize the rate further.

(iii) The accelerated rate guarantee accommodates weakly convex-weakly concave functions, thus showing exponential convergence in nonconvex-nonconcave scenarios.

(iv) Numerical results validate the proposed algorithm's superiority to many existing methods and further show that the lower-bound rate estimate is almost tight.

Notation: Let $\mathbb{R}$ and $\mathbb{R}_+$ be the set of real numbers and positive real numbers, respectively. $I_n \in \mathbb{R}^{n\times n}$ denotes the identity matrix of size n. Given a twice differentiable function $L(x,y) \in \mathcal{C}^2$ with $L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, we use $\frac{\partial}{\partial x}L(x,y) \in \mathbb{R}^{1\times n}$ and $\frac{\partial}{\partial y}L(x,y) \in \mathbb{R}^{1\times m}$ to denote the partial derivatives with respect to x and y, respectively. We further define $\nabla_x L(x,y) := \left[\frac{\partial}{\partial x}L(x,y)\right]^T$. Meanwhile, $\frac{\partial^2}{\partial x^2}L(x,y) \in \mathbb{R}^{n\times n}$ and $\frac{\partial^2}{\partial y^2}L(x,y) \in \mathbb{R}^{m\times m}$ represent the second-order partial derivatives of $L(x,y)$ with respect to x and y, respectively.

II. Problem and Result

A. Problem Statement

In this paper, we aim to solve bilinearly coupled minimax problems of the form:

$$\min_{x\in\mathbb{R}^n}\max_{y\in\mathbb{R}^m} L(x,y) := f(x) + \eta y^T Ax - \eta y^T b. \quad (1)$$

Here $\eta \in \mathbb{R}_+$ denotes a constant, $b \in \mathbb{R}^m$ is a constant vector, and $A \in \mathbb{R}^{m\times n}$ is the coupling matrix. $f(x) : \mathbb{R}^n \to \mathbb{R}$ is assumed to be twice differentiable. We assume a saddle point exists. The minimax problem (1) is equivalent to an affine equality-constrained optimization problem and $L(x,y)$ is its Lagrangian:

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & Ax - b = 0 : \ y \in \mathbb{R}^m \end{aligned} \quad (2)$$

where $y \in \mathbb{R}^m$ represents the dual variable.

The primary goal is to locate a saddle (min-max) point of $L(x,y)$, as indicated in (1). We will particularly center our focus around saddle flow dynamics that run a continuous-time version of gradient descent ascent on $L(x,y)$:

Definition 1 (Saddle Flow Dynamics)**.** *The saddle flow dynamics on $L(x,y)$ are defined as the following dynamic law:*

$$\nabla_x L(x,y) = -\nabla_x L(x,y), \quad (3a)$$
$$\nabla_y L(x,y) = +\nabla_y L(x,y), \quad (3b)$$

In the following subsections, we will develop an algorithm based on saddle flow dynamics that exponentially converges to a saddle point of $L(x,y)$.

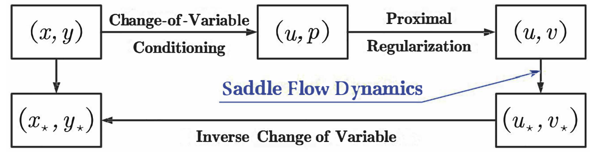


Fig. 1: Algorithm development.

B. Algorithm

Our algorithm is defined on a set of new auxiliary variables. Three auxiliary variables $u \in \mathbb{R}^n$, $v,p \in \mathbb{R}^m$ are introduced, and their connection with the original variables x and y is briefly illustrated in Figure 1. First, we propose a change of variable to transform (x,y) into (u,p). This conditioning step exploits the linearity in y. Second, we introduce a proximal regularizer on the variable p to obtain a Moreau envelope function defined (u,v). The saddle flow dynamics on the Moreau envelope are expected to reach an equilibrium point $(u_\star,v_\star)$, which immediately implies a saddle point $(x_\star,y_\star)$ of the original Lagrangian $L(x,y)$ via the inverse change of variable.

More specifically, we define our algorithm as the following dynamic law on (u,v):

$$\dot{u} = -\nabla f(u - \alpha A^T B^{-1}p^\star) - \eta A^T B^{-1} p^\star, \quad (4a)$$
$$\dot{v} = \rho(p^\star - v). \quad (4b)$$

Here $\alpha,\rho \in \mathbb{R}_+$ are positive constants, and $B \in \mathbb{R}^{m\times m}$ is a given full-rank transformation matrix, with $\bar{\sigma}_B I \succeq BB^T \succeq \underline{\sigma}_B I$ for some $\bar{\sigma}_B, \underline{\sigma}_B \in \mathbb{R}_+$. Note that $p^\star$ is a shorthand for the mapping $p^\star(u,v)$ that represents the unique solution (uniqueness to be shown later) subject to

$$\begin{aligned} &-\alpha B^{-T}A\nabla f(u - \alpha A^T B^{-1}p^\star(u,v)) + \eta B^{-T}(Au-b) \\ &\quad -2\eta\alpha B^{-T}AA^T B^{-1}p^\star(u,v) - \rho(p^\star(u,v) - v) = 0. \end{aligned} \quad (5)$$

We will provide the conditions under which the algorithm (5) is guaranteed to achieve the goal, and analyze its performance in the next subsection.

C. Main Result

Before stating the main result, we outline the assumptions required for the algorithm (5) that basically characterize the objective function $f(x)$ and the matrix A of the affine equality constraint in (2).

Assumption 1. *$f(x)$ is m_f-strongly convex with a l_f-Lipschitz gradient $\nabla f(x)$, i.e., $l_f I \succeq \nabla^2 f(x) \succeq m_f I$.*

Given Assumption 1, $L(x,y)$ is a strongly convex-linear function.

Assumption 2. *The coupling matrix A is full row rank with $\bar{\sigma}_A I \succeq AA^T \succeq \underline{\sigma}_A I$ for some $\bar{\sigma}_A, \underline{\sigma}_A \in \mathbb{R}_+$.*

Given Assumptions 1 and 2, the saddle point $(x_\star,y_\star)$ of $L(x,y)$ is unique since the problem (2) meets the linear independence constraint qualification and admits a unique primal-dual solution [25], [26]. Note that the saddle point

$(x_\star, y_\star)$ of $L(x,y)$ is also the equilibrium point of the corresponding saddle flow dynamics (3).

Our first argument is that an equilibrium point $(u_\star, v_\star)$ of the algorithm (4) is immediately related to the target saddle point $(x_\star, y_\star)$, as the following theorem suggests.

Theorem 1 (Characterization of Saddle Point)**.** *Let Assumptions 1 and 2 hold. Given an equilibrium point $(u_\star, v_\star)$ of the dynamic law* (4)*, the unique saddle point $(x_\star, y_\star)$ of $L(x,y)$ can be attained by*

$$x_\star = u_\star - \alpha A^T B^{-1} v_\star, \tag{6a}$$
$$y_\star = B^{-1} v_\star. \tag{6b}$$

The proof of Theorem 1 is provided in Appendix A.

Given Theorem 1, it remains to show the convergence of the algorithm 4 to an equilibrium point. We establish its exponential convergence and provide a lower-bound rate estimate in the following theorem.

Theorem 2 (Convergence)**.** *Let Assumptions 1 and 2 hold. Given any $\eta, \alpha > 0$ that satisfy $2\eta > \alpha l_f$, the dynamic law* (4) *globally exponentially converges to a unique equilibrium point $(u_\star, v_\star)$. More precisely, with $w := (u, v)$,*

$$\|w(t) - w_\star\| \leq \|w(0) - w_\star\| e^{-ct}$$

holds with the lower-bound rate

$$c := \min\left\{ m_f + \frac{\underline{\sigma}_B}{\bar{\sigma}_B} \frac{\underline{\sigma}_A \gamma_1}{(2\eta\alpha - \alpha^2 m_f)\bar{\sigma}_A + \rho \underline{\sigma}_B}, \right. \\ \left. \frac{\rho \underline{\sigma}_A (2\eta\alpha - \alpha^2 l_f)}{\underline{\sigma}_A (2\eta\alpha - \alpha^2 l_f) + \bar{\sigma}_B \rho} \right\},$$

where $\gamma_1 = \min\{(\eta - \alpha m_f)^2, (\eta - \alpha l_f)^2\}$.

The proof will become clear as we develop the algorithm (4) in detail in Section III.

Theorem 2 not only guarantees the exponential convergence of the algorithm (4), but also implies the following corollary that suggests an accelerated rate beyond the constant of the strong convexity of $f(x)$.

Corollary 3. *Given any $\rho, \bar{\sigma}_B, \underline{\sigma}_B$ that satisfy*

1) $\bar{\sigma}_B m_f < \underline{\sigma}_A (2\eta\alpha - \alpha^2 l_f)$,
2) $\rho \geq \frac{\underline{\sigma}_A (2\eta\alpha - \alpha^2 l_f)}{\underline{\sigma}_A (2\eta\alpha - \alpha^2 l_f) - \bar{\sigma}_B m_f} m_f$,

the dynamical law (4) *achieves a convergence rate lower bounded by*

$$c \geq m_f.$$

Remark 1. *The parameters that satisfy all the conditions always exist. A straightforward way is to pick a sufficiently small $\bar{\sigma}_B$ and a sufficiently large ρ. For example, the following set of parameters*

$$\eta = 8,\ \alpha = \frac{1}{l_f},\ \bar{\sigma}_B = \frac{\underline{\sigma}_A}{m_f l_f},\ \rho = \frac{3}{2} m_f \tag{7}$$

satisfy all the conditions.

III. Algorithm Development and Analysis

In this section, we explain the development of the algorithm, which naturally provides insights into the accelerated convergence performance the algorithm (4). Inspired by [27] and [28], we aim to enhance the convexity-concavity of the original Lagrangian in (1) by exploiting the bilinear couplings.

A. Change-of-Variable Conditioning

In the first step, consider the following change of variable:

$$\begin{bmatrix} u \\ p \end{bmatrix} = \begin{bmatrix} I_n & \alpha A^T \\ 0 & B \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} =: \Theta \begin{bmatrix} x \\ y \end{bmatrix}, \tag{8}$$

where we use Θ to denote the transformation. According to (8), we obtain the expression for the Lagrangian $L(x,y)$ in the (u,p) space:

$$\begin{aligned} \bar{L}(u,p) = {} & f(u - \alpha A^T B^{-1} p) + \eta p^T B^{-T}(Au - b) \\ & - \eta\alpha \|A^T B^{-1} p\|_2^2. \end{aligned} \tag{9}$$

The role of the change-of-var conditioning (8) is captured below.

Lemma 4. *Let Assumptions 1 and 2 hold. Given any $\eta, \alpha > 0$ that satisfy $2\eta > \alpha l_f$, then $\bar{L}(u,p)$ is m_f-strongly convex in u and $\frac{\underline{\sigma}_A(2\eta\alpha - \alpha^2 l_f)}{\bar{\sigma}_B}$-strongly concave in p.*

Proof. Given Assumptions 1 and 2, Lemma 4 follows immediately from the calculation of the respective second-order partial derivatives:

$$\begin{aligned} \frac{\partial^2}{\partial u^2} \bar{L}(u,p) = {} & \frac{\partial^2}{\partial x^2} f(u - \alpha A^T B^{-1} p) \succeq m_f I_n \\ \frac{\partial^2}{\partial p^2} \bar{L}(u,p) = {} & \alpha^2 B^{-T} A \frac{\partial^2}{\partial x^2} f(u - \alpha A^T B^{-1} p) A^T B^{-1} \\ & - 2\eta\alpha B^{-T} A A^T B^{-1} \\ = {} & -B^{-T} A \left(2\eta\alpha I_n - \alpha^2 \frac{\partial^2}{\partial x^2} f(u - \alpha A^T B^{-1} p) \right) A^T B^{-1} \\ \preceq {} & -B^T A (2\eta\alpha - \alpha^2 l_f) I_n A^T B^{-1} \preceq -\frac{\underline{\sigma}_A}{\bar{\sigma}_B} (2\eta\alpha - \alpha^2 l_f) I_m. \end{aligned}$$

■

Lemma 4 suggests that the change-of-variable conditioning enhances the concavity (from linearity) while maintaining the strong convexity for the Lagrangian. In fact, the constant of strong concavity of $\bar{L}(u,p)$ in p could be further optimized with proper choices of parameters. See an example below.

Remark 2. *When first fixing η and $\bar{\sigma}_B$, setting $\alpha = \eta / l_f$ gives the largest constant of strong concavity $\frac{\eta^2}{\bar{\sigma}_B} \frac{\underline{\sigma}_A}{l_f}$. In this case, $\frac{\eta^2}{\bar{\sigma}_B} \geq \frac{m_f l_f}{\underline{\sigma}_A}$ suffices to guarantee*

$$\frac{\partial^2}{\partial u^2} \bar{L}(u,p) \succeq m_f I_n, \quad \frac{\partial^2}{\partial p^2} \bar{L}(u,p) \preceq -m_f I_m.$$

Furthermore, $\bar{L}(u,p)$ is relevant since its unique saddle point $(u_\star, p_\star)$ is directly connected with the saddle point

$(x_\star, y_\star)$ of the original Lagrangian $L(x,y)$, as the corollary states.

Corollary 5. *$(u_\star, p_\star)$ is the unique saddle point of $\bar{L}(u,p)$ if and only if $(x_\star, y_\star)$ is the unique saddle point of $L(x,y)$ with*

$$x_\star = u_\star - \alpha A^T B^{-1} p_\star, \tag{10a}$$
$$y_\star = B^{-1} p_\star. \tag{10b}$$

B. Proximal Regularization

This subsection builds on $\bar{L}(u,p)$ and aims to further improve its strong convexity-strong concavity. As the second step, we apply proximal regularization on p to obtain the proximal Lagrangian:

$$\begin{aligned}\tilde{L}(u,v) =& \max_{p\in\mathbb{R}^m} \left\{ \bar{L}(u,p) - \frac{\rho}{2}\|p-v\|_2^2 \right\} \\ =& \max_{p\in\mathbb{R}^m} \left\{ f(u-\alpha A^T B^{-1}p) + \eta p^T B^{-T}(Au-b) \right. \\ & \left. -\eta\alpha\|A^T B^{-1} p\|_2^2 - \frac{\rho}{2}\|p-v\|_2^2 \right\} \\ =& f(u-\alpha A^T B^{-1}p^\star) + \eta (p^\star)^T B^{-T}(Au-b) \\ & -\eta\alpha\|A^T B^{-1}p^\star\|_2^2 - \frac{\rho}{2}\|p^\star - v\|_2^2,\end{aligned} \tag{11}$$

where $p^\star$ is the solution to (5). Note that $p^\star$ is unique due to the strong concavity of $\bar{L}(u,p)$ in p. It can be efficiently computed, either using its closed form when it is available, i.e., the inverse function of $\nabla f(\cdot)$ has an explicit expression, or using numerical methods since $\nabla f(\cdot)$ is strictly monotone.

The following lemma characterizes the impact of proximal regularization on the convexity-concavity property.

Lemma 6. *Let Assumptions 1 and 2 hold. Given any $\eta, \alpha > 0$ such that $2\eta > \alpha l_f$, $\tilde{L}(u,v)$ is c-strongly convex-strongly concave.*

Proof. Sice $L(x,y)$ is twice differentiable, we are allowed to derive the partial derivatives of $\tilde{L}(u,v)$ as

$$\frac{\partial}{\partial u}\tilde{L}(u,v) = \frac{\partial}{\partial u} f(u-\alpha A^T B^{-1} p^\star(u,v)) \tag{12a}$$
$$+ \eta p^\star(u,v)^T B^{-T} A, \tag{12b}$$
$$\frac{\partial}{\partial v}\tilde{L}(u,v) = \rho\left(p^\star(u,v) - v\right). \tag{12c}$$

Using Danskin's Theorem, we can further compute the second-order partial derivatives. Define $H := \frac{\partial^2}{\partial x^2} f(u - \alpha A^T B^{-1} p^\star(u,v))$, which is a symmetric matrix. Then we obtain

$$\frac{\partial^2}{\partial u^2}\tilde{L}(u,v) = H + \left(\eta I_n - \alpha H\right) A^T B^{-1} \mathbf{J}_{p^\star}^u, \tag{13a}$$
$$\frac{\partial^2}{\partial v^2}\tilde{L}(u,v) = \rho\left(\mathbf{J}_{p^\star}^v - I_m\right), \tag{13b}$$

where $\mathbf{J}_{p^\star}^u$ and $\mathbf{J}_{p^\star}^v$ are the Jacobian matrices of $p^\star(u,v)$ with respect to u and v, respectively, given by

$$\mathbf{J}_{p^\star}^u = \mathbf{J}_c^{-1} B^T A \left(\eta I_n - \alpha H\right), \tag{14a}$$
$$\mathbf{J}_{p^\star}^v = \rho \mathbf{J}_c^{-1}, \tag{14b}$$

with

$$\mathbf{J}_c := B^{-T} A \left(2\eta\alpha I_n - \alpha^2 H\right) A^T B^{-1} + \rho I_m.$$

The explicit second-order partial derivatives above allow us to characterize the strong convexity-strong concavity:

$$\begin{aligned}\frac{\partial^2}{\partial u^2}\tilde{L}(u,v) &= H + \left(\eta I_n - \alpha H\right) A^T B^{-1} \mathbf{J}_c^{-1} B^T A \left(\eta I_n - \alpha H\right) \\ &\succeq m_f I + \gamma_1 A^T B^{-1} \mathbf{J}_c^{-1} B^T A \\ &\succeq \left(m_f + \frac{\underline{\sigma}_B}{\bar{\sigma}_B} \frac{\underline{\sigma}_A \gamma_1}{(2\eta\alpha - \alpha^2 m_f)\bar{\sigma}_A + \rho\underline{\sigma}_B} \right) I_n \\ &=: c_1 I_n \succeq 0,\end{aligned} \tag{15a}$$

$$\begin{aligned}\frac{\partial^2}{\partial v^2}\tilde{L}(u,v) &= \rho\left(\mathbf{J}_{p^\star}^v - I_m\right) \\ &= \rho(\rho \mathbf{J}_c^{-1} - I_m) \\ &\preceq -\rho \frac{\underline{\sigma}_A (2\eta\alpha - \alpha^2 l_f)}{\underline{\sigma}_A (2\eta\alpha - \alpha^2 l_f) + \bar{\sigma}_B \rho} I_m \\ &=: -c_2 I_m \preceq -0.\end{aligned} \tag{15b}$$

Obviously c-strong convexity-strong concavity holds for $\tilde{L}(u,v)$ with $c = \min\{c_1, c_2\}$. ■

Compared with the strong convexity-strong concavity of $\bar{L}(u,p)$ identified in Lemma 4, Lemma 6 reveals an important tradeoff role of proximal regularization here: it increases the constant of strong convexity while decreasing the constant of strong concavity.

Similarly, we can establish the immediate connection between the saddle points of $\bar{L}(u,p)$ and $\tilde{L}(u,v)$ as follows.

Corollary 7. *$(u_\star, v_\star)$ is the unique saddle point of $\tilde{L}(u,p)$ if and only if $(u_\star, p_\star)$ is the unique saddle point of $\bar{L}(u,p)$ with $p_\star = v_\star$.*

Note that given Lemmas 4, 6 and Corollaries 5, 7, Theorem 1 becomes straightforward. We summarize the brief proof in Appendix A.

C. Convergence Analysis

Now it becomes clear that the algorithm (4) is essentially the saddle flow dynamics on $\tilde{L}(u,v)$. The exponential convergence of the algorithm established in Theorem 2 could be built upon our recent result [24], which we briefly present here for completeness.

Lemma 8. *Given a twice differentiable function $S(x,y)$ with m-strong convexity in x and q-strong concavity in y, the saddle flow dynamics on $S(x,y)$:*

$$\dot{z} = F(z) = \begin{bmatrix} -\nabla_x S(x,y) \\ \nabla_y S(x,y) \end{bmatrix}, \tag{16}$$

with $z := (x,y)$, are globally exponentially stable. More precisely,

$$\|z(t) - z_\star\| \le \|z(0) - z_\star\| e^{-rt}$$

holds with the lower-bound rate

$$r := \min\{m, q\}.$$

The proof of Lemma 8 is available in [24].

In virtue of Lemma 8, the strong saddle property of $\tilde{L}(u,v)$ in Lemma 6 immediately implies Theorem 2.

Rate Optimization: According to the parameter conditions in Corollary 3, there exist two constants $0 < \zeta < \beta < 1$ subject to

$$\bar{\sigma}_B = \beta \underline{\sigma}_A \frac{(2\eta\alpha - \alpha^2 l_f)}{m_f}, \quad \underline{\sigma}_B = \zeta \underline{\sigma}_A \frac{(2\eta\alpha - \alpha^2 l_f)}{m_f}.$$

Then the strong convexity and strong concavity constants of $\tilde{L}(u,v)$ can be rewritten as

$$c_1 = \left(1 + \frac{\zeta}{\beta} \frac{\gamma_1}{(2\eta\alpha - \alpha^2 m_f)\kappa_A m_f + (2\eta\alpha - \alpha^2 l_f)\rho\zeta}\right) m_f, \tag{17}$$

$$c_2 = \frac{\rho m_f}{\rho\beta + m_f} = \left(1 + \frac{(1-\beta)\rho - m_f}{\rho\beta + m_f}\right) m_f, \tag{18}$$

where $\kappa_A := \bar{\sigma}_A / \underline{\sigma}_A$ is the condition number of AA^T. Since c_1 decreases while c_2 increases, both monotonically in ρ, and there is bound to be one crossing point (given the respective ranges), we can obtain the fastest convergence rate in terms of ρ by setting $c_1 = c_2$. It will be achieved at $\rho^\star$ that satisfies

$$\begin{aligned}(\beta - 1)\beta\zeta L \rho_\star^2 + m_f(\zeta\gamma_1 + M\beta m_f \kappa_A)\\ + \rho_\star \beta\left(\zeta\gamma_1 + m_f\left(L\zeta + M(\beta-1)\kappa_A\right)\right) = 0,\end{aligned}$$

with $L =: 2\eta\alpha - \alpha^2 l_f$ and $M := 2\eta\alpha - \alpha^2 m_f$. Further optimization over the other parameters is intricate and will be left for future studies.

IV. Extension to Nonconvex-Nonconcave Minimax Problem

In this section, we study a broader class of the bilinearly coupled minimax problem:

$$\min_{x\in\mathbb{R}^n} \max_{y\in\mathbb{R}^m} L(x,y) = f(x) + y^T A x - g(y), \tag{19}$$

where $A \in \mathbb{R}^{m\times n}$ is the coupling matrix while $f(x) : \mathbb{R}^n \to \mathbb{R}$ and $g(y) : \mathbb{R}^m \to \mathbb{R}$ are assumed to be twice differentiable, but not necessarily convex and concave, respectively. We assume a stationary point exists.

Searching for saddle/stationary points of nonconvex-nonconcave functions is a popular goal in the machine learning community where such objective functions are common [29], [30]. While in general saddle points may not exist, certain special structures of the objective function, e.g., bilinear couplings, make this problem still tractable [28], [31], [32].

In this section, we can extend the results of this paper to solve the problem (19), in the case where $L(x,y)$ is weakly convex-weakly concave and smooth.

Assumption 3. *Assume that $f(x) \in \mathcal{C}^2$ is m_f-weakly convex and l_f-smooth, i.e., $l_f I \succeq \nabla^2 f(x) \succeq -m_f I$ with $l_f, m_f \in \mathbb{R}_+$, and $g(y) \in \mathcal{C}^2$ is m_g-weakly convex and l_g-smooth, i.e., $l_g I \succeq \nabla^2 g(y) \succeq -m_g I$ with $l_g, m_g \in \mathbb{R}_+$.*

Remark 3. *Note that, the objective function $L(x,y)$ is convex-concave when $m_f = m_g = 0$, transitioning to strong convexity and strong concavity if $m_f, m_g < 0$. This study specifically addresses scenarios in which $m_f, m_g > 0$, thereby quantifying how nonconvex-noncave $L(x,y)$ is.*

In what follows we follow a similar development as in Section II and Section III highlighting mostly the main changes due to Assumption 3. As before, applying the change-of-variable (8), we obtain corresponding objective function $\bar{L}(u,p)$

$$\begin{aligned}\bar{L}(u,p) = f(u - \alpha A^T B^{-1} p) + p^T B^{-T} A u\\ - \alpha\|A^T B^{-1} p\|_2^2 - g(B^{-1} p).\end{aligned} \tag{20}$$

Even though, a priori $\bar{L}(u,p)$ is not necessarily concave in p, following a procedure similar to Lemma 4, one can show that $\bar{L}(u,p)$ is strongly concave in p with the constant $\frac{\underline{\sigma}_A(2\eta\alpha - \alpha l_f) - m_g}{\bar{\sigma}_B}$ whenever $\underline{\sigma}_A(2\eta\alpha - \alpha l_f) - m_g > 0$.

This guarantees the existence and uniqueness of the minimizer of the proimal step (11), i.e.,

$$\begin{aligned}\tilde{L}(u,v) &= \max_{p\in\mathbb{R}^m}\left\{\bar{L}(u,p) - \frac{\rho}{2}\|p - v\|_2^2\right\}\\ &= f(u - \alpha A^T B^{-1}\tilde{p}^\star) + (\tilde{p}^\star)^T B^{-T} A u\\ &\quad - \alpha\|A^T B^{-1}\tilde{p}^\star\|_2^2 - g(B^{-1}\tilde{p}^\star) - \frac{\rho}{2}\|\tilde{p}^\star - v\|_2^2,\end{aligned} \tag{21}$$

where $\tilde{p}^\star = \tilde{p}^\star(u,v)$ is a unique minimizer in (21).

We will show soon that under mild conditions (21) is strongly convex-strongly concave, thus guaranteeing convergence to a unique stationary point $(u_\star, v_\star)$. The following theorem relates such a point, with a stationary point of the original problem.

Theorem 9. *If $(u_\star, v_\star)$ is a stationary point of $\tilde{L}$, then $(u_\star, p_\star = v_\star)$ is a stationary point of $\bar{L}$. Moreover, $(x_\star, y_\star)$ as described by* (6) *is a stationary point of $L(x,y)$.*

The proof is provided in Appendix B.

Theorem 9 thus guarantees that if the saddle flow dynamics of $\tilde{L}(u,v)$, i.e.,

$$\dot{u} = -\nabla f(u - \alpha A^T B^{-1}\tilde{p}^\star) - A^T B^{-1}\tilde{p}^\star, \tag{22a}$$

$$\dot{v} = \rho(\tilde{p}^\star - v), \tag{22b}$$

converges, then the corresponding $(x_\star, y_\star)$ will be a stationary point of the origianl problem.

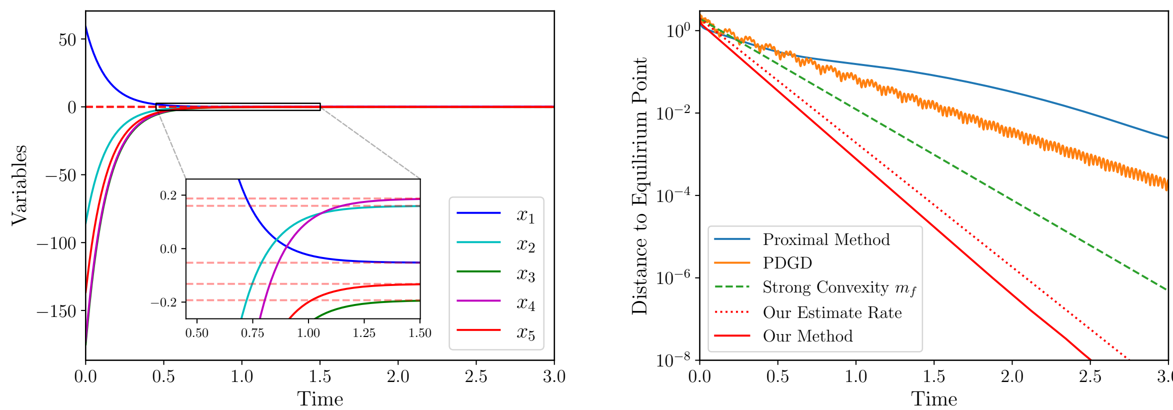
We finalize this section by providing a convergence guarantee for (22).

Theorem 10. *Let Assumption 2 and 3 hold. Given any $\alpha, \rho, \bar{\sigma}_B, \underline{\sigma}_B$ that satisfy*

1) $\underline{\sigma}_A(2\alpha - \alpha^2 l_f) - m_g > 0$ *and* $\alpha l_f \neq 1$,
2) $\kappa_A := \frac{\bar{\sigma}_A}{\underline{\sigma}_A} < \frac{\gamma_2}{2\alpha + \alpha^2 m_f}$,
3) $\kappa_B := \frac{\bar{\sigma}_B}{\underline{\sigma}_B} < \frac{\underline{\sigma}_A\gamma_2}{l_g} - \frac{\bar{\sigma}_A(2\alpha + \alpha^2 m_f)}{lg}$,
4) $0 < \rho < \frac{\underline{\sigma}_A\gamma_2}{\bar{\sigma}_B m_f} - \frac{\bar{\sigma}_A(2\alpha + \alpha^2 m_f)}{\bar{\sigma}_B} - \frac{lg}{\underline{\sigma}_B}$,

where $\gamma_2 = \min\{(1 + \alpha m_f)^2, (1 - \alpha l_f)^2\} > 0$, then the dynamic law (22) *exponentially converges to the unique stationary point $(u_\star, v_\star)$. More precisely, given $w := (u,v)$,*

$$\|w(t) - w_\star\| \le \|w(0) - w_\star\| e^{-ct}$$



(a) Primal Variables.

(b) Convergence rate.

Fig. 2: The solution for the quadratic problem.

holds with the rate

$$c := \min\left\{-m_f + \frac{\underline{\sigma}_B}{\bar{\sigma}_B}\frac{\underline{\sigma}_A\gamma_2}{\bar{\sigma}_A(2\alpha+\alpha^2 m_f)+\rho\underline{\sigma}_B + l_g},\right.$$
$$\left.\rho\frac{\underline{\sigma}_A(2\alpha-\alpha^2 l_f)-m_g}{\underline{\sigma}_A(2\alpha-\alpha^2 l_f)+\rho\bar{\sigma}_B - m_g}\right\} > 0.$$

The proof is similar to that of Theorem 2, omitted here due to the space limit.

Remark 4. *Since the original objective function $L(x, y)$ is nonconvex-nonconcave, there might be multiple stationary points. However, the adjusted Lagrangian becomes strongly convex-strongly concave which implies the uniqueness of the saddle point. As a result, our algorithm only converges to one of these stationary points.*

V. Simulation Results

In this section, we numerically validate the results of this paper and compare them with existing algorithms in the literature. To that end, we validate the effectiveness on algorithm (4), for solving an equality-constrained optimization problem as in (2).

Precisely, we let $n = 5$, $m = 4$, with (quadratic) objective given by $f(x) = \frac{1}{2}x^T Q x$, $Q \in \mathbb{R}^{5\times 5}$ given by $Q = 5I + Q_0^T Q_0$, and $Q_0 \in \mathbb{R}^{10\times 10}$ is a Gaussian random matrix. Not that by definition f is strongly convex. The constraints parameters A and b are also Gaussian Random matrices and vectors. We numerically establish the strong convexity constant m_f at 0.5 and the smoothness constant l_f at 0.693. Finally, we set the rest of the parameters using (7) and choose the initial point arbitrarily.

The convergence results are shown in Fig. 2. It can be observed that all primal variables converge to the optimal solution in Fig. 2a. The distance from the variable pair $w = (u, v)$ to the equilibrium point $w_\star = (u_\star, v_\star)$ exponentially decays to zero with an almost tight lower-bound rate estimate. Moreover, our method is superior to many existing methods which are limited by the given strong convexity constant, such as the proximal method [20] and PDGD in [23].

VI. Conclusion

This paper studies bilinearly coupled minimax problems and proposes an accelerated algorithm based on saddle flow dynamics to locate the saddle points. When the function is strongly convex-linear with bilinear couplings, we demonstrate that after change-of-variable conditioning and proximal regularization, the saddle flow dynamics of the transformed proximal Lagrangian exhibit exponential convergence with proper choices of parameters. The same design strategy can also be applied to construct a stationary point searching algorithm when facing a general class of weakly convex-weakly concave functions.

References

[1] K. J. Arrow, L. Hurwicz, and H. Uzawa, *Studies in Linear and Non-Linear Programming*, ser. Stanford Mathematical Studies in the Social Sciences. Stanford University Press, 1982, no. 2.

[2] D. Ding, B. Hu, N. K. Dhingra, and M. R. Jovanovic, "An exponentially convergent primal-dual algorithm for nonsmooth composite minimization," in *IEEE Conference on Decision and Control (CDC)*. IEEE, 2018, pp. 4927–4932.

[3] L. Baudin and R. Laraki, "Fictitious play and best-response dynamics in identical interest and zero-sum stochastic games," in *International Conference on Machine Learning*. PMLR, 2022, pp. 1664–1690.

[4] B. Gharesifard and J. Cortés, "Distributed convergence to Nash equilibria in two-network zero-sum games," *Automatica*, vol. 49, no. 6, pp. 1683–1692, 2013.

[5] C. Zhao, U. Topcu, N. Li, and S. Low, "Design and stability of load-side primary frequency control in power systems," *IEEE Transactions on Automatic Control*, vol. 59, no. 5, pp. 1177–1189, 2014.

[6] X. Zhang, N. Li, and A. Papachristodoulou, "Achieving real-time economic dispatch in power networks via a saddle point design approach," in *IEEE Power & Energy Society General Meeting*. IEEE, 2015, pp. 1–5.

[7] F. Paganini and E. Mallada, "A unified approach to congestion control and node-based multipath routing," *IEEE/ACM Transactions on Networking*, vol. 17, no. 5, pp. 1413–1426, 2009.

[8] T. Holding and I. Lestas, "Stability and instability in primal-dual algorithms for multi-path routing," in *IEEE Conference on Decision and Control (CDC)*. IEEE, 2015, pp. 1651–1656.

[9] D. Goldsztajn and F. Paganini, "Proximal regularization for the saddle point gradient dynamics," *IEEE Transactions on Automatic Control*, vol. 66, no. 9, pp. 4385–4392, 2021.

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

[11] A. Cherukuri, B. Gharesifard, and J. Cortés, "Saddle-point dynamics: Conditions for asymptotic stability of saddle points," *SIAM Journal on Control and Optimization*, 2017.

[12] A. Cherukuri, E. Mallada, S. Low, and J. Cortés, "The role of convexity in saddle-point dynamics: Lyapunov function and robustness," *IEEE Transactions on Automatic Control*, vol. 63, no. 8, pp. 2449–2464, 2017.

[13] P. You and E. Mallada, "Saddle flow dynamics: Observable certificates and separable regularization," in *American Control Conference (ACC)*. IEEE, 2021, pp. 4817–4823.

[14] X. Chen and N. Li, "Exponential stability of primal-dual gradient dynamics with non-strong convexity," in *American Control Conference (ACC)*. IEEE, 2020, pp. 1612–1618.

[15] S. Liang, L. Y. Wang, and G. Yin, "Exponential convergence of distributed primal–dual convex optimization algorithm without strong convexity," *Automatica*, vol. 105, pp. 298–306, 2019.

[16] P. A. Bansode, V. Chinde, S. R. Wagh, R. Pasumarthy, and N. M. Singh, "On the exponential stability of projected primal-dual dynamics on a riemannian manifold," *arXiv preprint arXiv:1905.04521*, 2019.

[17] S. K. Niederländer, F. Allgower, and J. Cortés, "Exponentially fast distributed coordination for nonsmooth convex optimization," in *IEEE Conference on Decision and Control (CDC)*, 2016, pp. 1036–1041.

[18] Z. Wang, W. Wei, C. Zhao, Z. Ma, Z. Zheng, Y. Zhang, and F. Liu, "Exponential stability of partial primal–dual gradient dynamics with nonsmooth objective functions," *Automatica*, vol. 129, p. 109585, 2021.

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

[20] N. K. Dhingra, S. Z. Khong, and M. R. Jovanovic, "The proximal augmented lagrangian method for nonsmooth composite optimization," *IEEE Transactions on Automatic Control*, vol. 64, no. 7, pp. 2861–2868, 2019.

[21] D. Ding and M. R. Jovanovic, "Global exponential stability of primal-dual gradient flow dynamics based on the proximal augmented lagrangian: A lyapunov-based approach," in *IEEE Conference on Decision and Control (CDC)*. IEEE, 2020, pp. 4836–4841.

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

[23] G. Qu and N. Li, "On the exponential stability of primal-dual gradient dynamics," *IEEE Control Systems Letters*, vol. 3, no. 1, pp. 43–48, 2019.

[24] P. You, Y. Liu, and E. Mallada, "A first principles analysis of saddle flow dynamics: Stability, convergence, and novel algorithms," 2024, https://pengcheng-you.github.io/desires-lab/papers/YLM_TAC.pdf, Last accessed on 2024-03-22.

[25] G. Wachsmuth, "On licq and the uniqueness of lagrange multipliers," *Operations Research Letters*, vol. 41, no. 1, pp. 78–80, 2013.

[26] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

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

[28] B. Grimmer, H. Lu, P. Worah, and V. Mirrokni, "The landscape of the proximal point method for nonconvex–nonconcave minimax optimization," *Mathematical Programming*, vol. 201, no. 1-2, pp. 373–407, 2023.

[29] F. Farnia and A. Ozdaglar, "Do GANs always have Nash equilibria?" in *Proceedings of the 37th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 119. PMLR, 2020, pp. 3029–3039.

[30] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[31] J. Diakonikolas, C. Daskalakis, and M. I. Jordan, "Efficient methods for structured nonconvex-nonconcave min-max optimization," in *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*. PMLR, 2021, pp. 2746–2754.

[32] J. Yang, N. Kiyavash, and N. He, "Global convergence and variance reduction for a class of nonconvex-nonconcave minimax problems," in *Advances in Neural Information Processing Systems*, vol. 33. Curran Associates, Inc., 2020, pp. 1153–1165.

APPENDIX

A. Proof of Theorem 1

In Section III, we already know that $\tilde{L}(u,v)$ is strongly convex-strongly concave. As a result, the equilibrium of the dynamic law (4) is unique, and is also the saddle point of $\tilde{L}(u,v)$, denoted as $(u_\star, v_\star)$. According to Corollary 7, the equilibrium $(u_\star, v_\star)$ is related to the saddle point $(u_\star, p_\star)$ with $v_\star = p_\star$. Besides, Corollary 5 implies that given the saddle point $(u_\star, p_\star)$, we can get the unique saddle point $(x_\star, y_\star)$ by (6). This proves Theorem 1.

B. Proof of Theorem 9

We only need to demonstrate

$$\begin{array}{c} \nabla_u \tilde{L}(u_\star, v_\star) = 0 \\ \nabla_v \tilde{L}(u_\star, v_\star) = 0 \end{array} \Rightarrow \begin{array}{c} \nabla_x L(x_\star, y_\star) = 0 \\ \nabla_y L(x_\star, y_\star) = 0 \end{array} \tag{23}$$

where the corresponding stationary point $(x_\star, y_\star)$ satisfies (6). To process the proof, we move towards the calculation of stationary points of $\tilde{L}(u,v)$:

$$\begin{cases} \nabla_u \tilde{L}(u_\star, v_\star) = \nabla f(u_\star - \alpha A^T B^{-1} \tilde{p}^\star) + A^T B^{-1} \tilde{p}^\star = 0 \\ \nabla_v \tilde{L}(u_\star, v_\star) = \rho(\tilde{p}^\star - v_\star) = 0 \end{cases} \tag{24}$$

where $\tilde{p}^\star$ is the unique solution to

$$\begin{array}{c} -\alpha B^{-T} A \nabla f(u - \alpha A^T B^{-1} \tilde{p}^\star) - B^{-T} \nabla g(B^{-1} \tilde{p}^\star) \\ -2\alpha B^{-T} A A^T B^{-1} \tilde{p}^\star + B^{-T} A u - \rho(\tilde{p}^\star - v) = 0. \end{array} \tag{25}$$

Substituting equations (6), (24) and (25) into the calculations of the stationary point of $L(x,y)$, we can show that

$$\begin{cases} \nabla_x L(x_\star, y_\star) = \nabla f(x_\star) + A^T y_\star = 0 \\ \nabla_y L(x_\star, y_\star) = A x_\star - b = 0 \end{cases} \tag{26}$$

which evidently implies (23).