

# A Unified Analysis of Saddle Flow Dynamics: Stability and Algorithm Design

Pengcheng You, *Member, IEEE*, Yingzhu Liu, *Student Member, IEEE*, and Enrique Mallada, *Senior Member, IEEE*

**Abstract**—This work examines the conditions for asymptotic and exponential convergence of saddle flow dynamics of convex-concave functions. First, we propose an observability-based certificate for asymptotic convergence, directly bridging the gap between the invariant set in a LaSalle argument and the equilibrium set of saddle flows. This certificate generalizes conventional conditions for convergence, e.g., strict convexity-concavity, and leads to a novel state-augmentation method that requires minimal assumptions for asymptotic convergence. We also show that global exponential stability follows from strong convexity-strong concavity, providing a lower-bound estimate of the convergence rate. This insight also explains the convergence of proximal saddle flows for strongly convex-concave objective functions. Our results generalize to dynamics with projections on the vector field and have applications in solving constrained convex optimization via primal-dual methods. Based on these insights, we study four algorithms built upon different Lagrangian function transformations. We validate our work by applying these methods to solve a network flow optimization and a Lasso regression problem.

**Index Terms**—Saddle flow dynamics, stability of nonlinear systems, optimization algorithms, nonlinear systems, optimization

## I. INTRODUCTION

Studying optimization algorithms from a continuous time dynamical-systems viewpoint has become an insightful technique in the analysis of algorithms, providing alternative means to understand their stability [2], [3], rate of convergence [4]–[6], and robustness [6]–[9]. For example, in the basic case of gradient descent dynamics for unconstrained convex optimization, the objective function monotonically decreases along trajectories towards the optimum, naturally rendering a Lyapunov function [10]. Such realization leads to multiple extensions, including finite-time convergence [11], [12], acceleration [5], [6], and time-varying optimization [13]–[15], and in many cases it can be used to help understand the convergence of the discrete-time counterparts [16].

One prominent problem in this field is the study of saddle flow dynamics, i.e., dynamics that move in the gradient descent direction for a sub-set of variables and the gradient ascent direction for the complement. Designed for locating min-max saddle points, saddle flow dynamics, or saddle flows,

are particularly suited for solving constrained optimization problems via primal-dual methods [17] and finding Nash equilibria of zero-sum games via gradient play [18], which have led to a broad application spectrum, including power systems [19], [20], communication networks [21], [22], and cloud computing [23]. As a result, understanding the properties that affect the behavior of these dynamics has been a major focus of study.

The seminal work [24] first explored the asymptotic convergence of primal-dual dynamics using first principles. Since then, emerging advanced analytical tools have been used to re-validate conventional conditions or unveil new, weaker ones. A major set of conditions are related to the convexity properties of objective functions, including [24]. For instance, [17] revisits the strict convexity-concavity condition in the case of discontinuous vector fields, using LaSalle’s invariance principle for discontinuous Caratheodory systems. Besides, weaker conditions have been discovered, such as local strong convexity-concavity [9], convexity-linearity, or strong quasiconvexity-quasiconcavity [25]. Regularization serves as an alternative to circumventing the above conditions. This includes various forms of penalty terms on equality/inequality constraints to handle the Lagrangian of constrained optimization [8], [26], as well as the proximal methods [23], [27]. Despite the merit of regularization that often relaxes conditions for convergence, the extra penalty terms may introduce couplings that require additional computation and communication overheads in distributed implementation.

More recently, the research focus has shifted toward the exponential stability of saddle flows and characterizing the convergence rate. This has been made possible through diverse conditions and techniques. For instance, the saddle flow dynamics of Lagrangian functions augmented with specific projection are proven exponentially stable in [28]. However, the convergence rate estimate is not provided therein. Instead, [29] proposes a projection-free design of dynamics for the Lagrangian of affine inequality-constrained convex programs. It is followed up by [30] that further accounts for convex inequality constraints with semi-global exponential stability guarantee – the convergence rate is estimated to be dependent on initial points. Besides, [31]–[33] use the Moreau envelope of Lagrangian functions as a proxy and validate the exponential stability of the resulting proximal primal-dual gradient flows using integral quadratic constraints. However, with this method, accurately estimating the convergence rate is difficult, thus leading to other alternatives, including contraction [34], [35], Riemannian geometric frameworks [36], and two-timescale approaches [37]. It shall be noted that

P. You and Y. Liu are with the College of Engineering, Peking University, Beijing, China (email: {pyou,yzliucoe}@pku.edu.cn).

E. Mallada is with the Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218, USA (email: mallada@jhu.edu).

A preliminary version of this work was presented in [1].

most existing results are restricted to primal-dual dynamics of Lagrangian functions, and remain highly case-specific. Thus, the conditions required for the exponential convergence of general saddle flow dynamics and any convenient methods to estimate the convergence rate are still not fully understood.

In this paper, we develop a unified analysis for saddle flow dynamics that renders general conditions for asymptotic and exponential convergence. Our results hold for, but are not limited to, Lagrangian functions derived from optimization problems and can readily be extended for scenarios where the saddle flow is projected onto a closed convex polyhedron. The novel analysis provides a unified framework to explain the convergence of several existing algorithms in the literature and offers valuable insights for developing new ones. We showcase several new algorithms that converge under weaker conditions and/or exhibit superior performance. Specifically, the paper makes the following contributions to the literature:

- *Asymptotic Convergence Condition:* We provide a general certificate for the asymptotic convergence of saddle flow dynamics. The proposed certificate, which is based on an observability condition, directly connects the equilibrium set with the invariant set that is derived from applying LaSalle’s invariance principle and can be shown to generalize existing asymptotic convergence results.
- *Universal Algorithm for Convex-Concave Saddle Problems:* We further leverage our condition for asymptotic convergence to design a novel saddle flow algorithm, based on state augmentation, that is able to guarantee convergence to *some* saddle point, depending on the initial condition, under minimal assumptions on convexity-concavity. In particular, our algorithm is able to converge even for bilinear functions.
- *Exponential Convergence Condition:* We show that a strongly convex-strongly concave objective function (with respective constants  $\mu$  and  $q$ ) is sufficient for saddle flows’ exponential convergence and that a lower bound  $c = \min\{\mu, q\}$  on the rate of convergence can be readily established. We further show how proximal regularization on a strongly convex-concave function can achieve exponential convergence by “shifting” some of the strong convexity, making the resulting function also strongly concave.
- *Projected Saddle Flows:* Our analysis on both asymptotic and exponential convergence can be readily extended to the case where saddle flow dynamics are projected so as to constrain the trajectories within a closed convex polyhedron. This is achieved without requiring any strengthening of the conditions for convergence of the unconstrained case.
- *Applications to Constrained Convex Optimization:* Finally, we bring together all our theoretical development and analyze a wide variety of primal-dual algorithms for constrained convex optimization, new and existing ones, that lead to a novel augmented primal-dual algorithm to solve linear programs, a novel bound on the convergence rate of proximal primal-dual algorithms, a novel preconditioned primal-dual algorithm that outperforms the proximal methods for strongly convex problems with affine constraints, and a novel reduced primal-dual algorithm that further exploits separable problem structures and exponentially converges

with a potentially even faster rate than the preconditioned case.

The versatility of the analysis and new algorithms are validated with numerical illustrations, showcasing the solution to a network flow optimization problem using our novel primal-dual LP algorithm and the solution to a Lasso regression problem via a combination of our novel preconditioning method with proximal regularization.

*Organization:* The remainder of the paper is organized as follows. Section II introduces the problem formulation with basic definitions and assumptions, followed by the main results on the asymptotic convergence of saddle flow dynamics in Section III. The exponential stability analysis of saddle flow dynamics is demonstrated in Section IV. We further generalize the results to the projected version in Section V, with various applications in algorithm designs for constrained convex optimization in Section VI. Section VII provides simulation validations and Section VIII concludes.

*Contributions w.r.t. [1]:* A preliminary version of this work appeared in [1]. This paper extends those results in many ways. First, our asymptotic convergence of the proximal regularization method in Section III is extended to general convex-concave saddle functions. The exponential convergence analysis in Section IV, and its extension to projected saddle flows in Section V-B, is entirely novel, as well as its application to (a) characterizing conditions for the exponential convergence of proximal primal-dual algorithms (Section VI-B), and (b) deriving two new algorithms, the preconditioned primal-dual (Section VI-C) and the reduced primal-dual (Section VI-D), with guaranteed exponential convergence. Finally, the numerical validation in Section VII is also a new contribution.

*Notation:* Let  $\mathbb{R}$  and  $\mathbb{R}_{\geq 0}$  be the sets of real and non-negative real numbers, respectively.  $I_n \in \mathbb{R}^{n \times n}$  denotes the identity matrix of size  $n$ . Given two vectors  $x, y \in \mathbb{R}^n$ ,  $x_i$  and  $y_i$  denote their  $i^{\text{th}}$  entries, respectively; and  $x \leq y$  holds if  $x_i \leq y_i$  holds for  $\forall i$ . Given a continuously differentiable function  $S(x, y) \in \mathcal{C}^1$  with  $S : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}$ , we use  $\partial_x S(x, y) \in \mathbb{R}^{1 \times n}$  and  $\partial_y S(x, y) \in \mathbb{R}^{1 \times m}$  to denote the partial derivatives with respect to  $x$  and  $y$ , respectively, and define  $\nabla_x S(x, y) := [\partial_x S(x, y)]^T$ . We further use  $\partial_{xy}^2 S(x, y) := \partial_x (\nabla_y S(x, y))$  to denote the corresponding second-order partial derivative.

## II. PROBLEM FORMULATION

We consider a function  $S : \mathcal{D} \mapsto \mathbb{R}$  with  $\mathcal{D} = \mathcal{X} \times \mathcal{Y}$ , where both  $\mathcal{X} \subseteq \mathbb{R}^n$  and  $\mathcal{Y} \subseteq \mathbb{R}^m$  are convex sets. Our goal is to study different dynamic laws that seek to converge to some saddle point  $(x_*, y_*)$  of  $S(x, y)$ . While in general, such questions could be asked in a setting without any further restrictions, neither the existence of saddle points nor convergence towards them is easy to guarantee. For this paper, we focus our attention on functions  $S(x, y)$  that are *convex-concave*.

**Definition 1** (Convex-Concave Functions).  $S(x, y)$  is *convex-concave*, if  $S(\cdot, y)$  is convex for  $\forall y \in \mathcal{Y}$  and  $S(x, \cdot)$  is concave for  $\forall x \in \mathcal{X}$ .  $S(x, y)$  is *strictly convex-concave*, if  $S(x, y)$  is

convex-concave, and further either  $S(\cdot, y)$  is strictly convex for  $\forall y \in \mathcal{Y}$  or  $S(x, \cdot)$  is strictly concave for  $\forall x \in \mathcal{X}$ .

In this case, a general definition of a saddle point of  $S(x, y)$  is given as follows:

**Definition 2** (Saddle Point). *A point  $(x_*, y_*) \in \mathcal{D}$  is a saddle point of a convex-concave function  $S(x, y)$  if*

$$S(x_*, y) \leq S(x_*, y_*) \leq S(x, y_*) \quad (1)$$

holds for  $\forall x \in \mathcal{X}$  and  $\forall y \in \mathcal{Y}$ .

Due to the convexity-concavity of  $S(x, y)$ , we are specifically interested in minimizing  $S(x, y)$  over  $x$  and meanwhile maximizing  $S(x, y)$  over  $y$ . Throughout this work, we will assume that  $S(x, y)$  is continuously differentiable, i.e.,  $S(x, y) \in \mathcal{C}^1$ , as formally summarized below.

**Assumption 1.**  *$S(x, y)$  is convex-concave, continuously differentiable, and there exists at least one saddle point  $(x_*, y_*) \in \mathcal{D}$  satisfying (1).*

The continuous differentiability in Assumption 1 is introduced to simplify the exposition. It does not significantly limit the scope of the results as one can always derive (albeit possibly increased computational complexity) a continuously differentiable surrogate of a continuous convex-concave function by means of the Moreau Envelope [38].

Given a convex-concave function  $S(x, y)$  satisfying Assumption 1, we refer to the following dynamic law

$$\dot{x} = -\nabla_x S(x, y), \quad (2a)$$

$$\dot{y} = +\nabla_y S(x, y), \quad (2b)$$

as the saddle flow dynamics of  $S(x, y)$ . Basically the dynamic law (2) drives the system in gradient descent and ascent directions for  $x$  and  $y$ , respectively.

Suppose the saddle flow (2) does converge to an equilibrium point  $(x_*, y_*) \in \mathcal{D}$ . It has to be a stationary point of  $S(x, y)$ , defined by

**Definition 3** (Stationary Point). *A point  $(x_*, y_*) \in \mathcal{D}$  is a stationary point of a function  $S(x, y)$  if*

$$\begin{cases} \nabla_x S(x_*, y_*) = 0 \\ \nabla_y S(x_*, y_*) = 0 \end{cases} \quad (3)$$

holds.

**Remark 1.** *In the case of a convex-concave function  $S(x, y)$ , any stationary point must be a saddle point:*

$$\nabla_x S(x_*, y_*) = 0 \Rightarrow S(x_*, y_*) \leq S(x, y_*), \quad \forall x \in \mathcal{X},$$

$$\nabla_y S(x_*, y_*) = 0 \Rightarrow S(x_*, y) \leq S(x_*, y_*), \quad \forall y \in \mathcal{Y}.$$

However, the converse does not necessarily hold.

We will mainly work with the standard form (2) of saddle flow dynamics to locate a saddle point of  $S(x, y)$ . In the following Sections III and IV, we first consider a special case where the feasible domain of  $S(x, y)$  is full space, i.e.,  $\mathcal{D} = \mathbb{R}^n \times \mathbb{R}^m$ , to derive our main results. In this case Assumption 1 implies that any saddle point  $(x_*, y_*) \in \mathbb{R}^n \times \mathbb{R}^m$  is a stationary point and satisfies (3). Then we move on to relax

the restriction and introduce a solution using a projection on the vector field to handle more general cases in Section V.

### III. ASYMPTOTIC CONVERGENCE

This section presents an observable certificate for a convex-concave function  $S(x, y)$  that, if exists, ensures the asymptotic convergence of the saddle flow dynamics (2) towards a saddle point. We show that such certificates exist in the cases of two conventional sufficient conditions – strict convexity-concavity and proximal regularization – and are thus weaker. We further build on this certificate to develop a state-augmentation method that entails minimal convexity-concavity requirements on  $S(x, y)$  for saddle flow dynamics to converge to a saddle point asymptotically.

#### A. Observable Certificates

We first define the proposed observable certificate for  $S(x, y)$ .

**Definition 4** (Observable Certificate). *A function  $h(x, y)$  with  $h : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}_{\geq 0}^2$  is an observable certificate of  $S(x, y)$ , if there exists a saddle point  $(x_*, y_*)$  such that*

$$\begin{bmatrix} S(x_*, y_*) - S(x_*, y) \\ S(x, y_*) - S(x_*, y_*) \end{bmatrix} \geq h(x, y) \geq 0 \quad (4)$$

holds and for any trajectory  $(x(t), y(t))$  of (2) that satisfies  $h(x(t), y(t)) \equiv 0$ , we have  $\dot{x} \equiv 0$ ,  $\dot{y} \equiv 0$ .

We call  $h(x, y)$  an observable certificate [10], due to the second property of Definition 4, which is akin to (2) having  $h(x, y)$  as an observable output. Our first main result is that the existence of an observable certificate suffices to guarantee the asymptotic convergence of the saddle flow dynamics (2), as we formally state below.

**Assumption 2.**  *$S(x, y)$  has an observable certificate  $h(x, y)$  as given by Definition 4.*

**Theorem 1.** *Let Assumptions 1 and 2 hold. Then the saddle flow dynamics (2) globally asymptotically converge to some saddle point  $(x_*, y_*)$  of  $S(x, y)$ .*

*Proof.* The proof follows from applying LaSalle's invariance principle [10] to the following candidate Lyapunov function

$$V(x, y) = \frac{1}{2} \|x - x_*\|^2 + \frac{1}{2} \|y - y_*\|^2, \quad (5)$$

where  $(x_*, y_*)$  is the saddle point identified in Definition 4. Taking the Lie derivative of (5) along the trajectory  $(x(t), y(t))$  of (2) gives

$$\begin{aligned} \dot{V} &= (x - x_*)^T \dot{x} + (y - y_*)^T \dot{y} \\ &= (x - x_*)^T [-\nabla_x S(x, y)] + (y - y_*)^T [+ \nabla_y S(x, y)] \\ &= (x_* - x)^T \nabla_x S(x, y) - (y_* - y)^T \nabla_y S(x, y) \\ &\leq S(x_*, y) - S(x, y) - (S(x, y_*) - S(x, y)) \\ &= S(x_*, y) - S(x, y_*) \\ &= \underbrace{S(x_*, y) - S(x_*, y_*)}_{\leq 0} + \underbrace{S(x_*, y_*) - S(x, y_*)}_{\leq 0}, \end{aligned}$$

where the second equality follows from the dynamic law (2), the first inequality follows from the convexity-concavity of  $S(x, y)$ , and the last inequality follows from the saddle property (1) of  $(x_*, y_*)$ .

Since (5) is radially unbounded, all its sub-level sets are compact. From above, it follows that the trajectories of (2) are bounded and contained in an invariant domain

$$D_0(x(0), y(0)) := \{(x, y) \mid V(x, y) \leq V(x(0), y(0))\}, \quad (6)$$

where  $(x(0), y(0))$  is any given initial point. LaSalle's invariance principle then implies that any trajectory of (2) should converge to the largest invariant set

$$\mathbb{S} := D_0(x(0), y(0)) \cap \{(x, y) \mid \dot{V}(x(t), y(t)) \equiv 0\}. \quad (7)$$

Given Assumption 2, (4) implies that  $\mathbb{S}$  is indeed a subset of

$$\mathbb{H} = \{(x, y) \mid h(x(t), y(t)) \equiv 0\}, \quad (8)$$

which is further a subset of the equilibrium set of (2), denoted as

$$\mathbb{E} := \{(x, y) \mid \dot{x}(t), \dot{y}(t) \equiv 0\}, \quad (9)$$

i.e.,  $\mathbb{S} \subset \mathbb{H} \subset \mathbb{E}$ .

Therefore, the invariant set  $\mathbb{S}$  contains only equilibrium points. If  $\mathbb{S}$  were to be composed of isolated points – only possible when there is a unique saddle point – this would be sufficient to prove convergence to the (unique) saddle point. However, in general, LaSalle's invariance principle only shows asymptotic convergence to the invariant set, without guaranteeing convergence to a point within it, even in the case where the set is composed of equilibrium points.

This issue is circumvented by the fact that all the equilibria within  $\mathbb{S}$  are stable. See, e.g., [39, Corollary 5.2]. Alternatively, notice that  $\mathbb{S}$  is compact, and as a result any trajectory within the  $\Omega$  limit set of (2) has a convergent sub-sequence. Let  $(\bar{x}, \bar{y})$  be the limit point of such a sequence. Due to  $(\bar{x}, \bar{y}) \in \mathbb{S}$ , it is also a saddle point. By changing  $(x_*, y_*)$  in the definition of  $V(x, y)$  specifically to  $(\bar{x}, \bar{y})$ , it follows that  $0 \leq V(x(t), y(t)) \rightarrow 0$  holds, which implies  $(x(t), y(t)) \rightarrow (\bar{x}, \bar{y})$ . ■

Checking whether Assumption 2 holds basically requires hunting for a qualified observable certificate  $h(x, y)$  of  $S(x, y)$ . The existence and characterization of such observable certificates are still vague from Definition 4. We next discuss how they can be identified through concrete examples. We show that the observable certificate is indeed a weaker condition underneath some of the conventional ones required for the asymptotic convergence of the saddle flow dynamics (2).

1) *Strict Convexity-Concavity*: The most common condition is arguably the strict convexity-concavity of  $S(x, y)$  [17]. We formalize its connection with our observable certificate as below.

**Assumption 3.**  $S(x, y)$  is strictly convex-concave.

**Proposition 2.** Let Assumptions 1 and 3 hold. Then the function

$$h(x, y) := \begin{bmatrix} S(x_*, y_*) - S(x_*, y) \\ S(x, y_*) - S(x_*, y_*) \end{bmatrix}, \quad (10)$$

with  $(x_*, y_*)$  being an arbitrary saddle point of  $S(x, y)$ , is an observable certificate of  $S(x, y)$ .

The fact that (10) meets the second requirement

$$h(x(t), y(t)) \equiv 0 \Rightarrow \dot{x} \equiv 0, \dot{y} \equiv 0$$

to be an observable certificate can be readily verified using the strict convexity-concavity of  $S(x, y)$ . See [1] for a full analysis.

The asymptotic convergence of the saddle flow dynamics (2) then immediately follows from Theorem 1.

**Corollary 3.** Let Assumptions 1 and 3 hold. Then the saddle flow dynamics (2) globally asymptotically converge to some saddle point  $(x_*, y_*)$  of  $S(x, y)$ .

2) *Proximal Regularization*: Proximal regularization has been widely recognized as an alternative method to guarantee the asymptotic convergence of proximal primal-dual dynamics – an exemplar of saddle flows [23], [27]. More generally, a surrogate function

$$\tilde{S}(u, y) := \min_{x \in \mathbb{R}^n} \left\{ S(x, y) + \frac{\rho}{2} \|x - u\|^2 \right\} \quad (11)$$

can be defined on  $S(x, y)$ , where  $\rho > 0$  is a constant regularization coefficient (throughout the paper). An important fact is that the surrogate function  $\tilde{S}(u, y)$  is also convex-concave and always maintains the same set of saddle points as the original function  $S(x, y)$ , as stated in the following lemma.

**Lemma 4.** Let Assumption 1 hold. Then  $\tilde{S}(u, y)$  is convex in  $u$ , concave in  $y$  and continuously differentiable with the gradient:

$$\nabla_u \tilde{S}(u, y) = \rho u - \rho \tilde{x}(u, y), \quad (12a)$$

$$\nabla_y \tilde{S}(u, y) = \nabla_y S(\tilde{x}(u, y), y), \quad (12b)$$

where  $\tilde{x}(u, y)$  is the unique minimizer in (11) satisfying

$$\nabla_x S(\tilde{x}(u, y), y) + \rho(\tilde{x}(u, y) - u) = 0. \quad (13)$$

Moreover, a point  $(x_*, y_*)$  is a saddle point of  $S(x, y)$  if and only if  $(u_*, y_*)$  is a saddle point of  $\tilde{S}(u, y)$  with  $u_* = x_*$ .

The proof essentially follows [27, Theorem 2], with a general convex-concave function replacing a convex-linear Lagrangian function. Lemma 4 also implies  $\tilde{x}(u_*, y_*) = x_*$ . It basically allows us to focus on the proximal saddle flow dynamics of  $S(u, y)$ , i.e.,

$$\dot{u} = -\nabla_u \tilde{S}(u, y) = -(\rho u - \rho \tilde{x}(u, y)), \quad (14a)$$

$$\dot{y} = +\nabla_y \tilde{S}(u, y) = \nabla_y S(\tilde{x}(u, y), y), \quad (14b)$$

which suffice to locate a saddle point of the original function  $S(x, y)$ . We formalize the connection of this method with our observable certificate as follows.

**Proposition 5.** *Let Assumption 1 hold. Then the function*

$$\tilde{h}(u, y) := \left[ \begin{array}{c} \tilde{S}(u_*, y_*) - \tilde{S}(u_*, y) \\ \frac{\rho}{2} \|\tilde{x}(u, y_*) - u\|^2 \end{array} \right], \quad (15)$$

with  $(u_*, y_*)$  being an arbitrary saddle point of  $\tilde{S}(u, y)$ , is an observable certificate of  $\tilde{S}(u, y)$ .

*Proof.* (15) is a qualified observable certificate since it has the following two properties. First, to show

$$\left[ \begin{array}{c} \tilde{S}(u_*, y_*) - \tilde{S}(u_*, y) \\ \tilde{S}(u, y_*) - \tilde{S}(u_*, y_*) \end{array} \right] \geq \tilde{h}(u, y) \geq 0,$$

the main task boils down to showing the following inequality:

$$\begin{aligned} & \tilde{S}(u, y_*) - \tilde{S}(u_*, y_*) \\ &= S(\tilde{x}(u, y_*), y_*) + \frac{\rho}{2} \|\tilde{x}(u, y_*) - u\|^2 - S(\tilde{x}(u_*, y_*), y_*) \\ & \quad - \frac{\rho}{2} \|\tilde{x}(u_*, y_*) - u_*\|^2 \\ &= S(\tilde{x}(u, y_*), y_*) - S(x_*, y_*) + \frac{\rho}{2} \|\tilde{x}(u, y_*) - u\|^2 \\ & \geq \frac{\rho}{2} \|\tilde{x}(u, y_*) - u\|^2, \end{aligned}$$

where the second equality follows from  $\tilde{x}(u_*, y_*) = x_* = u_*$  and the inequality follows from the convexity of  $S(x, y)$  in  $x$ .

Second, to show

$$\tilde{h}(u(t), y(t)) \equiv 0 \Rightarrow \dot{u} \equiv 0, \dot{y} \equiv 0,$$

we refer readers to the proof of [27, Proposition 10]. ■

Similarly, Theorem 1 guarantees the asymptotic convergence of the proximal saddle flow dynamics (14).

**Corollary 6.** *Let Assumption 1 hold. Then the proximal saddle flow dynamics (14) globally asymptotically converge to some saddle point  $(u_*, y_*)$  of  $\tilde{S}(u, y)$ , with  $(x_* = u_*, y_*)$  being a saddle point of  $S(x, y)$ .*

In fact, even the differentiability in Assumption 1 is not required since the surrogate  $\tilde{S}(u, y)$  can be continuously differentiable regardless.

### B. Augmented Saddle Flow Dynamics

We further design a novel state-augmentation method that exploits our observable certificate and only requires Assumption 1 for augmented saddle flow dynamics to asymptotically converge to a saddle point. The key of this method is to augment the domain of  $S(x, y)$  and introduce regularization terms that provide a convenient observable certificate without altering the positions of the original saddle points. In particular, we propose a surrogate for  $S(x, y)$  via the following augmentation

$$\hat{S}(x, \hat{x}, y, \hat{y}) := \frac{\rho}{2} \|x - \hat{x}\|^2 + S(x, y) - \frac{\rho}{2} \|y - \hat{y}\|^2, \quad (16)$$

where  $\hat{x} \in \mathbb{R}^n$  and  $\hat{y} \in \mathbb{R}^m$  serve as two new sets of virtual variables. It can be readily verified that the saddle points remain invariant for  $\hat{S}(x, \hat{x}, y, \hat{y})$ .

**Lemma 7.** *Let Assumption 1 hold. Then a point  $(x_*, y_*)$  is a saddle point of  $S(x, y)$  if and only if  $(x_*, \hat{x}_*, y_*, \hat{y}_*)$  is a saddle point of  $\hat{S}(x, \hat{x}, y, \hat{y})$ , with*

$$x_* = \hat{x}_* \quad \text{and} \quad y_* = \hat{y}_*. \quad (17)$$

*Proof.* Recall the saddle property (1) of a saddle point, this theorem follows immediately from the equivalence of the following three arguments:

- $\hat{S}(x_*, \hat{x}_*, y, \hat{y}) \leq \hat{S}(x_*, x_*, y_*, y_*) \leq \hat{S}(x, \hat{x}, y_*, \hat{y}_*)$  holds for  $\forall(x, \hat{x}, y, \hat{y})$ ;
- $S(x_*, y) - \frac{\rho}{2} \|y - \hat{y}\|^2 \leq S(x_*, y_*) \leq S(x, y_*) + \frac{\rho}{2} \|x - \hat{x}\|^2$  holds for  $\forall(x, \hat{x}, y, \hat{y})$ ;
- $S(x_*, y) \leq S(x_*, y_*) \leq S(x, y_*)$  holds for  $\forall(x, y)$ .

Here the first and second arguments are equivalent due to the definition (16) of  $\hat{S}(x, \hat{x}, y, \hat{y})$ , while the second and third arguments are equivalent since the regularization terms attain zero at the minimum. ■

Under Assumption 1 for  $S(x, y)$ , the augmented function  $\hat{S}(x, \hat{x}, y, \hat{y})$  is convex in  $(x, \hat{x})$ , concave in  $(y, \hat{y})$ , and continuously differentiable with at least one saddle point, by its definition (16) and Lemma 7. Therefore, Assumption 1 also holds for  $\hat{S}(x, \hat{x}, y, \hat{y})$ . Lemma 7 further ensures that whenever we locate a saddle point of  $\hat{S}(x, \hat{x}, y, \hat{y})$ , a saddle point of  $S(x, y)$  is attained simultaneously. This motivates us to instead look at the saddle flow dynamics of  $\hat{S}(x, \hat{x}, y, \hat{y})$ .

Following (2), this augmented version of saddle flow dynamics is given by

$$\dot{x} = -\nabla_x S(x, y) - \rho(x - \hat{x}), \quad (18a)$$

$$\dot{\hat{x}} = \rho(x - \hat{x}), \quad (18b)$$

$$\dot{y} = +\nabla_y S(x, y) - \rho(y - \hat{y}), \quad (18c)$$

$$\dot{\hat{y}} = \rho(y - \hat{y}). \quad (18d)$$

Although this dynamic law has twice as many state variables as its prototype (2), it is important to notice that, unlike proximal regularization [23], [27], [38] and quadratic regularizers on equality constraints [8], [26], the state augmentation does not introduce couplings and (18) preserves any distributed structure that the system may originally have.

We now provide the key result on the asymptotic convergence of the augmented saddle flow dynamics (18).

**Proposition 8.** *Let Assumption 1 hold. Then the function*

$$\hat{h}(x, \hat{x}, y, \hat{y}) := \left[ \begin{array}{c} \frac{\rho}{2} \|y - \hat{y}\|^2 \\ \frac{\rho}{2} \|x - \hat{x}\|^2 \end{array} \right] \quad (19)$$

is an observable certificate of  $\hat{S}(x, \hat{x}, y, \hat{y})$ .

*Proof.*  $\hat{h}(x, \hat{x}, y, \hat{y})$  satisfies (4) as follows:

$$\begin{aligned} & \left[ \begin{array}{c} \hat{S}(x_*, \hat{x}_*, y_*, \hat{y}_*) - \hat{S}(x_*, \hat{x}_*, y, \hat{y}) \\ \hat{S}(x, \hat{x}, y_*, \hat{y}_*) - \hat{S}(x_*, \hat{x}_*, y_*, \hat{y}_*) \end{array} \right] \\ &= \left[ \begin{array}{c} \underbrace{S(x_*, y_*) - S(x_*, y)}_{\geq 0} + \frac{\rho}{2} \|y - \hat{y}\|^2 \\ \underbrace{S(x, y_*) - S(x_*, y_*)}_{\geq 0} + \frac{\rho}{2} \|x - \hat{x}\|^2 \end{array} \right] \end{aligned}$$

$$\begin{aligned} &\geq \left[ \begin{array}{c} \frac{\rho}{2} \|y - \hat{y}\|^2 \\ \frac{\rho}{2} \|x - \hat{x}\|^2 \end{array} \right] \\ &\geq 0. \end{aligned}$$

Moreover,  $\hat{h}(x, \hat{x}, y, \hat{y}) \equiv 0$  implies  $x(t) \equiv \hat{x}(t)$  and  $y(t) \equiv \hat{y}(t)$ , which, by (18d) and (18b), enforce  $\dot{\hat{x}} \equiv 0$ ,  $\dot{\hat{y}} \equiv 0$ . Therefore,  $\hat{x}(t)$  and  $\hat{y}(t)$ , and thus  $x(t)$  and  $y(t)$ , all remain constant and have reached an equilibrium point. ■

Assumption 2 thus holds for the augmented function  $\hat{S}(x, \hat{x}, y, \hat{y})$  and the asymptotic convergence of the augmented saddle flow dynamics (18) follows immediately from Theorem 1. We summarize this result next.

**Corollary 9.** *Let Assumption 1 hold. Then the augmented saddle flow dynamics (18) globally asymptotically converge to some saddle point  $(x_*, \hat{x}_*, y_*, \hat{y}_*)$  of  $\hat{S}(x, \hat{x}, y, \hat{y})$ , with  $(x_*, y_*)$  being a saddle point of  $S(x, y)$ .*

Corollary 9 indicates that only the convexity-concavity of  $S(x, y)$  is required to asymptotically arrive at a saddle point through the augmented saddle flow dynamics (18). This condition applies to general convex-concave functions, including bilinear functions, and is milder than many existing conditions in the literature. The fact that the state-augmented saddle flow (18) asymptotically converges for any convex-concave saddle function  $S(x, y)$  (even in cases where the saddle flow of  $S(x, y)$  does not converge) should come as a surprise, given that the augmented function  $\hat{S}(x, \hat{x}, y, \hat{y})$  is identical to  $S(x, y)$  on the hyperplane defined by  $x = \hat{x}$  and  $y = \hat{y}$ . One explanation for this phenomenon is rooted in the dissipative theory. That is, (18d) and (18b) act as a dynamic feedback aimed at dissipating the “energy” that prevents the regular saddle flow (2) from converging. We refer readers to [40] for further discussion on this subject.

#### IV. EXPONENTIAL CONVERGENCE

This section explores the conditions under which the saddle flow dynamics (2) are globally exponentially stable. We establish exponential stability as a direct consequence of strong convexity-strong concavity of  $S(x, y)$ . This insight can be used to explain the convergence behavior of the proximal saddle flow dynamics (14), which generalizes existing results for the proximal gradient algorithms [33], [41].

##### A. Strong Convexity-Strong Concavity

We first rewrite the saddle flow dynamics (2) in a more compact form

$$\dot{z} = F(z) \quad (20)$$

with  $z := (x, y)$  and

$$F(z) = \begin{bmatrix} -\nabla_x S(x, y) \\ +\nabla_y S(x, y) \end{bmatrix}. \quad (21)$$

We further assume absolute continuity for  $F(z)$ .

**Assumption 4.** *The gradient of  $S(x, y)$ , i.e.,  $\nabla S(x, y) := [\partial_x S(x, y), \partial_y S(x, y)]^T$ , is absolutely continuous [42].*

**Remark 2.** *Assumption 4 is slightly weaker than Lipschitz continuity - the common assumption used in the studies of global exponential stability of saddle flows [29], [32], [43].*

Assumption 4 basically enables

$$\partial_z F(z) = \begin{bmatrix} -\partial_{xx}^2 S(x, y) & -\partial_{xy}^2 S(x, y) \\ \partial_{yx}^2 S(x, y) & \partial_{yy}^2 S(x, y) \end{bmatrix}, \quad (22)$$

and

$$\frac{1}{2}(\partial_z F(z) + \partial_z F(z)^T) = \begin{bmatrix} -\partial_{xx}^2 S(x, y) & 0 \\ 0 & \partial_{yy}^2 S(x, y) \end{bmatrix}, \quad (23)$$

wherever the second-order partial derivatives are defined [44].

We now show that strong convexity-strong concavity of  $S(x, y)$  is conducive to the saddle flow’s exponential convergence.

**Assumption 5.**  *$S(x, y)$  is  $\mu$ -strongly convex in  $x$  and  $q$ -strongly concave in  $y$ .*

Note that  $\mu > 0$  and  $q > 0$  are given constants. Throughout the paper, such constants are always positive unless specified.

**Remark 3.** *One consequence of Assumption 5 is that there is a unique saddle point  $(x_*, y_*)$ . Further,*

$$\hat{S}(x, y) := S(x, y) - \frac{\mu}{2} \|x - x_*\|^2 + \frac{q}{2} \|y - y_*\|^2$$

*is convex in  $x$ , concave in  $y$ , and  $(x_*, y_*)$  is also a saddle point of  $\hat{S}(x, y)$ .*

**Theorem 10.** *Let Assumptions 1, 4 and 5 hold. Then the saddle flow dynamics (2) are globally exponentially stable. More precisely, given the (unique) saddle point  $z_*$  and any initial point  $z(0)$  with  $z = (x, y)$ ,*

$$\|z(t) - z_*\| \leq \|z(0) - z_*\| e^{-ct}$$

*holds with the rate*

$$c = \min\{\mu, q\} > 0.$$

*Proof.* The proof features a reformulation of the Lie derivative of the Lyapunov function (5) based on the fundamental theorem of calculus for absolutely continuous functions [45].

We consider again the Lyapunov function

$$V(z) = \frac{1}{2} \|z - z_*\|^2 = \frac{1}{2} \|x - x_*\|^2 + \frac{1}{2} \|y - y_*\|^2.$$

Now taking the Lie derivative with respect to (20) gives

$$\begin{aligned} \dot{V}(z) &= (z - z_*)^T F(z) \\ &= \frac{1}{2} ((z - z_*)^T F(z) + F(z)^T (z - z_*)). \end{aligned} \quad (24)$$

Assumption 4 allows us to write  $F(z)$  as

$$F(z) = \int_0^1 \underbrace{\partial_z F(z(s))}_{=0} (z - z_*) ds + \underbrace{F(z_*)}_{=0}, \quad (25)$$

with  $z(s) = (z - z_*)s + z_*$ , where we introduce a scalar  $s \in \mathbb{R}$  and use the fact  $dz(s) = (z - z_*)ds$ .

Now substituting (25) into (24) gives

$$\begin{aligned}\dot{V}(z) &= (z - z_*)^T \int_0^1 \frac{1}{2} (\partial_z F(z) + \partial_z F(z)^T) ds (z - z_*) \\ &= (z - z_*)^T \int_0^1 \begin{bmatrix} -\partial_{xx}^2 S(z(s)) & 0 \\ 0 & \partial_{yy}^2 S(z(s)) \end{bmatrix} ds (z - z_*).\end{aligned}\quad (26)$$

Note that up to this point,  $\dot{V}(z)$  is exactly given by (26). With Assumption 5, (26) can be relaxed as

$$\begin{aligned}\dot{V}(z) &\leq -\mu \|x - x_*\|^2 - q \|y - y_*\|^2 \\ &\leq -c \|z - z_*\|^2 = -2cV(z).\end{aligned}$$

Then the Comparison Lemma allows us to derive the exponential convergence [10]:

$$\begin{aligned}V(z(t)) &\leq e^{-2ct} V(z(0)) \\ \iff \|z(t) - z_*\|^2 &\leq e^{-2ct} \|z(0) - z_*\|^2 \\ \iff \|z(t) - z_*\| &\leq e^{-ct} \|z(0) - z_*\|.\end{aligned}$$

The exponential convergence of the saddle flow dynamics (2) established in Theorem 10 is tightly connected with the notion of contractivity in contraction theory [46]. It can be shown that, under the same condition, the vector field  $F(z)$  in (21) is infinitesimally contracting with the rate  $\min\{\mu, q\}$ . Our proof however highlights that, while contraction may be sufficient, it is not necessary for exponential convergence as long as the integrated matrix in (26) can be uniformly bounded, thus opening the path for future extensions on the conditions provided in this paper.

## B. Proximal Saddle Flow Dynamics

In this subsection, we use Theorem 10 to understand how the proximal methods in general enable the exponential convergence of saddle flows. The proximal method has been commonly used to handle non-smooth optimization problems [38]. This points to the most relevant applications in primal-dual dynamics [27], [31], [47], among which [31], [47] have shown that running primal-dual dynamics on a proximal augmented Lagrangian function is exponentially convergent. Our analysis based on Theorem 10 aims to complement these results and provide more insights into the exponential stability of proximal saddle flows.

Consider again the surrogate convex-concave function  $\tilde{S}(u, y)$  in (11). It can be re-written as

$$\tilde{S}(u, y) = S(\tilde{x}(u, y), y) + \frac{\rho}{2} \|\tilde{x}(u, y) - u\|^2, \quad (27)$$

using the unique minimizer  $\tilde{x}(u, y)$  that satisfies the optimality condition (13), given  $(u, y)$ . On this basis, the proximal saddle flow dynamics are given by (14). We make the following assumptions on the original function  $S(x, y)$  that are sufficient to guarantee the exponential convergence of the proximal saddle flow dynamics (14) to a unique saddle point.

**Assumption 6.** The function  $S(x, y)$  is  $\mu$ -strongly convex with  $l$ -Lipschitz gradient in  $x$ , i.e.,  $lI \succeq \partial_{xx}^2 S(x, y) \succeq \mu I$  for  $\forall(x, y)$  wherever  $\partial_{xx}^2 S(x, y)$  is defined.

**Assumption 7.** The matrix  $\partial_{yx}^2 S(\tilde{x}(u, y), y)$  is full row rank with  $\sigma I \succeq \partial_{yx}^2 S(\tilde{x}(u, y), y) \partial_{xy}^2 S(\tilde{x}(u, y), y) \succeq \kappa I$  for  $\forall(u, y)$  wherever the second-order partial derivatives are defined.

**Remark 4.** Assumption 7 is not that restrictive in practice. For instance, consider  $S(x, y) = f(x) + y^T A x$ , which could be a common Lagrangian function. Assumption 7 basically requires that  $A$  should be full row rank with  $\sigma I \succeq A A^T \succeq \kappa I$ , a common assumption for the exponential convergence of primal-dual dynamics [29], [36]. Assumption 7 could thus be regarded as a more general version.

**Proposition 11.** Let Assumptions 1, 4, 6 and 7 hold. Then the proximal saddle flow dynamics (14) are globally exponentially stable. More precisely, given the (unique) saddle point  $w_*$  and any initial point  $w(0)$  with  $w := (u, y)$ ,

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

holds with the rate

$$c = \min \left\{ \frac{\mu\rho}{\mu + \rho}, \frac{\kappa}{l + \rho} \right\} > 0.$$

*Proof.* For ease of presentation, we simply use  $\tilde{x}$  to represent  $\tilde{x}(u, y)$  in the proof. The main task is to characterize the following second-order partial derivatives of  $\tilde{S}(u, y)$ :

$$\partial_{uu}^2 \tilde{S}(u, y) = \rho I - \rho \partial_u \tilde{x}, \quad (28a)$$

$$\partial_{yy}^2 \tilde{S}(u, y) = [\partial_{xy}^2 S(\tilde{x}, y)]^T \partial_y \tilde{x} + \partial_{yy}^2 S(\tilde{x}, y), \quad (28b)$$

which can be readily derived based on the gradient in (12). To further show the properties of the second-order partial derivatives in (28), we utilize the optimality condition (13) and take partial derivatives of both sides with respect to  $x$  and  $y$ , respectively, leading to

$$\partial_{xx}^2 S(\tilde{x}, y) \partial_u \tilde{x} + \rho \partial_u \tilde{x} - \rho I = 0, \quad (29a)$$

$$\partial_{xx}^2 S(\tilde{x}, y) \partial_y \tilde{x} + [\partial_{yx}^2 S(\tilde{x}, y)]^T + \rho \partial_y \tilde{x} = 0. \quad (29b)$$

Since  $S(x, y)$  is  $\mu$ -strongly convex in  $x$ ,  $\partial_{xx}^2 S(x, y)$  is positive definite for  $\forall(x, y)$ . Therefore, we can re-organize (29) into

$$\partial_u \tilde{x} = \rho (\partial_{xx}^2 S(\tilde{x}, y) + \rho I)^{-1}, \quad (30a)$$

$$\partial_y \tilde{x} = -(\partial_{xx}^2 S(\tilde{x}, y) + \rho I)^{-1} [\partial_{yx}^2 S(\tilde{x}, y)]^T. \quad (30b)$$

Combining (28) and (30) yields

$$\partial_{uu}^2 \tilde{S}(u, y) = \rho I - \rho^2 (\partial_{xx}^2 S(\tilde{x}, y) + \rho I)^{-1}, \quad (31a)$$

$$\begin{aligned}\partial_{yy}^2 \tilde{S}(u, y) &= \partial_{yy}^2 S(\tilde{x}, y) - [\partial_{yx}^2 S(\tilde{x}, y)]^T \\ &\quad \cdot (\partial_{xx}^2 S(\tilde{x}, y) + \rho I)^{-1} [\partial_{yx}^2 S(\tilde{x}, y)]^T.\end{aligned}\quad (31b)$$

Recall  $\partial_{yy}^2 S(\tilde{x}, y) \preceq 0$  due to the concavity of  $S(x, y)$  in  $y$ . Then Assumption 7 implies the strong convexity-strong concavity of  $\tilde{S}(u, y)$ :

$$\partial_{uu}^2 \tilde{S}(u, y) \succeq \rho I - \frac{\rho^2}{\mu + \rho} I \succeq \frac{\mu\rho}{\mu + \rho} I \succ 0, \quad (32a)$$

$$\begin{aligned}
\partial_{yy}^2 \tilde{S}(u, y) &\preceq -\partial_{yx}^2 S(\tilde{x}, y)(\partial_{xx}^2 S(\tilde{x}, y) + \rho I)^{-1} \partial_{xy}^2 S(\tilde{x}, y) \\
&\preceq -\frac{1}{l + \rho} \partial_{yx}^2 S(\tilde{x}, y) \partial_{xy}^2 S(\tilde{x}, y) \\
&\preceq -\frac{\kappa}{l + \rho} I \prec 0,
\end{aligned} \tag{32b}$$

where  $\partial_{yx}^2 S(\tilde{x}, y) = [\partial_{xy}^2 S(\tilde{x}, y)]^T$  has been employed. Then the exponential stability of the proximal saddle flow dynamics (14) follows immediately from Theorem 10. ■

**Remark 5.** According to Assumption 7, the second term on the right-hand side of (31b) is positive definite and reflects the cross-term interaction between  $x$  and  $y$  in  $S(x, y)$ . This analysis following Theorem 10 reveals the impact of such interaction on the convergence rate of the proximal saddle flow dynamics (14). This insight is consistent with the results from [48] that focuses on discrete-time proximal algorithms.

Proposition 11 highlights the strong convexity-strong concavity of the surrogate function  $\tilde{S}(u, y)$  (with respective constants  $\frac{\mu\rho}{\mu+\rho}$  and  $\frac{\kappa}{l+\rho}$ ), compared with the strong convexity-concavity of the original function  $S(x, y)$  (with respective constants  $\mu$  and 0). The fact of

$$\frac{\mu\rho}{\mu+\rho} < \mu \text{ and } \frac{\kappa}{l+\rho} > 0$$

unveils the rationale behind the proximal regularization that enables exponential convergence of saddle flows. Further note that a larger  $\rho$  increases the strong convexity constant while decreasing the strong concavity constant. This tradeoff suggests that the convergence rate estimate - the smaller between the two constants - could be optimized with a proper choice of  $\rho$ .

**Corollary 12.** The fastest convergence rate estimate is attained at

$$c_* = \frac{2\mu\kappa}{\sqrt{(\mu l - \kappa)^2 + 4\mu^2\kappa} + \mu l + \kappa} < \mu \tag{33}$$

by optimizing  $\rho$  to be  $\rho_* > 0$  that satisfies

$$\frac{\mu\rho_*}{\mu + \rho_*} = \frac{\kappa}{l + \rho_*}. \tag{34}$$

The uniqueness of  $\rho_* > 0$  is straightforward from the fact that the two sides of (34) are both strictly monotone in  $\rho$  and are guaranteed to intersect. However, Corollary 12 points out a bottleneck for the proximal saddle flow's convergence rate estimate, i.e., it is always limited by the strong convexity constant  $\mu$  of  $S(x, y)$  in  $x$ .

## V. PROJECTED SADDLE FLOW DYNAMICS

In this section, we generalize the convergence results in Sections III and IV to account for a projection defined on the vector field of the saddle flow dynamics (2), which is commonly introduced to constrain solution trajectories within a feasible region.

Specifically, we look at a projected version of saddle flow dynamics of a convex-concave function  $S(x, y)$  as

$$\dot{z} = \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} \Pi_{\mathcal{X}} [x, -\nabla_x S(x, y)] \\ \Pi_{\mathcal{Y}} [y, +\nabla_y S(x, y)] \end{bmatrix}. \tag{35}$$

Here we consider the case where  $\mathcal{X}$  and  $\mathcal{Y}$  are both closed convex polyhedra feasible for  $x$  and  $y$ , respectively. The projection is explicitly defined as follows.

**Definition 5** (Vector Field Projection [22]). Given  $p \in \mathcal{P} \subseteq \mathbb{R}^n$  where  $\mathcal{P}$  is a closed convex set and  $s \in \mathbb{R}^n$ , the vector field projection  $\Pi_{\mathcal{P}}[p, s]$  of  $s$  at  $p$  with respect to  $\mathcal{P}$  is defined as

$$\Pi_{\mathcal{P}}[p, s] := \lim_{\delta \rightarrow 0^+} \frac{\Psi_{\mathcal{P}}[p + \delta s] - p}{\delta}, \tag{36}$$

with  $\delta \in \mathbb{R}$  and  $\Psi_{\mathcal{P}}[r] := \arg \min_{\hat{p} \in \mathcal{P}} \|\hat{p} - r\|$  denoting the point in  $\mathcal{P}$  closest to  $r \in \mathbb{R}^n$ .

With the projection, any trajectory of (35) will remain in the feasible set  $\mathcal{D} = \mathcal{X} \times \mathcal{Y}$  as long as it starts with a feasible initial point. Meanwhile, the existence and uniqueness of the solutions of (35) are guaranteed by [49, Theorem 2.5] when the vector field of (35) is Lipschitz continuous. Therefore, to work with the projected saddle flow dynamics (35), we enhance Assumptions 1 and 4 as follows.

**Assumption 8.** The gradient of  $S(x, y)$ , i.e.,  $\nabla S(x, y) := [\partial_x S(x, y), \partial_y S(x, y)]^T$ , is Lipschitz continuous.

On this basis, a key enabler for generalizing the convergence results to such projected saddle flow dynamics is the following property of the projection.

**Lemma 13.** Given any closed convex set  $\mathcal{P} \subset \mathbb{R}^n$ , and  $p \in \mathcal{P}$ ,  $p_* \in \mathcal{P}$ ,  $s \in \mathbb{R}^n$ , the following inner product inequality always holds:

$$\langle p_* - p, s - \Pi_{\mathcal{P}}[p, s] \rangle \leq 0.$$

*Proof.* We start with a variational inequality from [50, Chapter 0.6, Corollary 1]:

$$\langle p_* - \Psi_{\mathcal{P}}[r], r - \Psi_{\mathcal{P}}[r] \rangle \leq 0, \quad \forall r \in \mathbb{R}^n,$$

which is intuitive, considering the projection of a point onto a convex set.

Apply the inequality to  $r = p + \delta s$ , and further divide both sides by  $\delta > 0$ , then we arrive at

$$\left\langle p_* - \Psi_{\mathcal{P}}[p + \delta s], \frac{p + \delta s - \Psi_{\mathcal{P}}[p + \delta s]}{\delta} \right\rangle \leq 0.$$

By taking the limit  $\delta \rightarrow 0^+$ , the first term of the inner product converges to  $p_* - p$  while the second term converges to  $s - \Pi_{\mathcal{P}}[p, s]$ . This completes the proof of the lemma. ■

Following Lemma 13, it can be readily verified that if any trajectory of (38) does converges to an equilibrium point, it is also a saddle point of  $S(x, y)$ .

**Lemma 14.** Any equilibrium point of the projected saddle flow dynamics (38) is a saddle point  $(x_*, y_*) \in \mathcal{D}$  of  $S(x, y)$ , satisfying Definition 2.

*Proof.* Consider an equilibrium point  $(x_*, y_*)$  subject to

$$\begin{bmatrix} \Pi_{\mathcal{X}} [x_*, -\nabla_x S(x_*, y_*)] \\ \Pi_{\mathcal{Y}} [y_*, +\nabla_y S(x_*, y_*)] \end{bmatrix} = 0.$$

When either of the element-wise projection is inactive, it boils down to the case without the projection discussed in Section II.



Therefore, we focus on active projections. Without loss of generality, suppose  $\Pi_{\mathcal{X}}[x_*, -\nabla_x S(x_*, y_*)] = 0$  holds. Based on Lemma 13, we can readily attain the saddle property from the following:

$$\begin{aligned} & \langle x - x_*, -\nabla_x S(x_*, y_*) - \Pi_{\mathcal{X}}[x_*, -\nabla_x S(x_*, y_*)] \rangle \leq 0, \\ \Rightarrow & \langle x - x_*, -\nabla_x S(x_*, y_*) \rangle \leq 0, \\ \Rightarrow & S(x_*, y_*) - S(x, y_*) \leq 0, \end{aligned}$$

which holds for  $x_* \in \mathcal{X}$  and  $\forall x \in \mathcal{X}$ .  $\blacksquare$

Next, we formally generalize the sufficiency of observable certificates (Theorem 1) in Section III-A and the sufficiency of strong convexity-strong concavity (Theorem 10) in Section IV-A to prove the asymptotic and exponential convergence, respectively, of the projected saddle flow dynamics (35).

#### A. Asymptotic Convergence of Projected Saddle Flows

The sufficiency of observable certificates to guarantee the asymptotic convergence of the projected saddle flow dynamics (35) to a saddle point of  $S(x, y)$  in the feasible domain  $\mathcal{D}$  is summarized as follows.

**Theorem 15.** *Let Assumptions 1, 2 and 8 hold. Then starting from an arbitrary initial point  $(x(0), y(0)) \in \mathcal{D}$ , the projected saddle flow dynamics (35) asymptotically converge to some saddle point  $(x_*, y_*) \in \mathcal{D}$  of  $S(x, y)$ .*

*Proof.* Consider the same quadratic Lyapunov function (5). Taking its Lie derivative along the trajectory  $(x(t), y(t))$  of (35) yields

$$\begin{aligned} \dot{V} &= (x - x_*)^T \dot{x} + (y - y_*)^T \dot{y} \\ &= (x - x_*)^T \Pi_{\mathcal{X}}[x, -\nabla_x S(x, y)] \\ &\quad + (y - y_*)^T \Pi_{\mathcal{Y}}[y, +\nabla_y S(x, y)] \\ &= (x_* - x)^T \nabla_x S(x, y) - (y_* - y)^T \nabla_y S(x, y) \\ &\quad + \underbrace{(x_* - x)^T (-\nabla_x S(x, y) - \Pi_{\mathcal{X}}[x, -\nabla_x S(x, y)])}_{\leq 0} \\ &\quad + \underbrace{(y_* - y)^T (\nabla_y S(x, y) - \Pi_{\mathcal{Y}}[y, \nabla_y S(x, y)])}_{\leq 0} \\ &\leq (x_* - x)^T \nabla_x S(x, y) - (y_* - y)^T \nabla_y S(x, y) \\ &\leq S(x_*, y) - S(x, y) - (S(x, y_*) - S(x, y)) \\ &= S(x_*, y) - S(x, y_*) \\ &= \underbrace{S(x_*, y) - S(x_*, y_*)}_{\leq 0} + \underbrace{S(x_*, y_*) - S(x, y_*)}_{\leq 0}, \end{aligned}$$

where the key step is to use Lemma 13 in the first inequality.

We then define the largest invariant set between the on-off switches of the projection as

$$\mathbb{S} := D_0(x(0), y(0)) \cap \left\{ (x, y) \mid \dot{V}(x(t), y(t)) \equiv 0, t \in \mathbb{R}_{\geq 0} \setminus \mathbb{T} \right\}, \quad (37)$$

where  $\mathbb{T}$  consists of all the time epochs when the projection switches on and off. According to the invariance principle for Caratheodory systems [51], any trajectory of (35) converges to  $\mathbb{S}$ . Then Assumption 2 basically still implies  $\mathbb{S} \subset \mathbb{H} \subset \mathbb{E}$ , i.e., the invariant set contains only equilibrium points. Following the same discussion in the proof of Theorem 1, any trajectory

of (35) indeed asymptotically converges to an equilibrium point, i.e., a saddle point of  $S(x, y)$ .  $\blacksquare$

#### B. Exponential Convergence of Projected Saddle Flows

We then generalize the role of a strongly convex-strongly concave objective function in the exponential convergence of the projected saddle flow dynamics (35).

**Theorem 16.** *Let Assumptions 1, 5 and 8 hold. Then the projected saddle flow dynamics (35) are globally exponentially stable. More precisely, given the (unique) saddle point  $z_*$  and any initial point  $z(0) \in \mathcal{D}$  with  $z = (x, y)$ ,*

$$\|z(t) - z_*\| \leq \|z(0) - z_*\| e^{-ct}$$

*holds with the rate*

$$c = \min\{\mu, q\} > 0.$$

*Proof.* Lemma 13 allows the proof pipeline of Theorem 10 to still apply here. In particular, given  $z = (x, y)$ , consider again the quadratic Lyapunov function

$$V(z) = \frac{1}{2} \|z - z_*\|^2 = \frac{1}{2} \|x - x_*\|^2 + \frac{1}{2} \|y - y_*\|^2.$$

From the proof of Theorem 15, we already have its Lie derivative with respect to (35) that satisfies

$$\begin{aligned} \dot{V}(z) &\leq (x_* - x)^T \nabla_x S(x, y) - (y_* - y)^T \nabla_y S(x, y) \\ &\leq (z - z_*)^T F(z). \end{aligned}$$

Following the analysis in the proof of Theorem 10, we can directly obtain

$$\dot{V}(z) \leq -\mu \|x - x_*\|^2 - q \|y - y_*\|^2 \leq -c \|z - z_*\|^2 = -2cV(z).$$

Using the Comparison Lemma defined on the Dini derivative of non-differentiable functions [10], we can still derive the exponential convergence:

$$\begin{aligned} V(z(t)) &\leq e^{-2ct} V(z(0)) \\ \Leftrightarrow \|z(t) - z_*\|^2 &\leq e^{-2ct} \|z(0) - z_*\|^2 \\ \Leftrightarrow \|z(t) - z_*\| &\leq e^{-ct} \|z(0) - z_*\|. \end{aligned}$$

$\blacksquare$

Theorems 15 and 16 significantly broaden the use of the proposed conditions. In the next section we will demonstrate how they could be applied to algorithm analysis and design for solving constrained convex programs.

## VI. APPLICATIONS TO CONSTRAINED CONVEX OPTIMIZATION

We present four algorithms for constrained convex optimization. Three of them are novel and inspired by the proposed convergence results. Their asymptotic/exponential stability either requires weaker conditions and/or exhibits superior performance, i.e., faster convergence rate estimates. For an existing algorithm based on proximal primal-dual dynamics, we use our convergence results to develop a novel bound on its convergence rate. All the algorithms are rooted in different transformations of Lagrangian functions that aim to meet our

convergence conditions while preserving saddle points. This provides a systematic approach to algorithm design.

The Lagrangian of an inequality-constrained convex program is, in general, convex in  $x \in \mathbb{R}^n$  and linear in  $y \in \mathbb{R}_{\geq 0}^m$ . Due to non-negative dual variables, we will, in particular, focus on the vector field projection of the form  $[s]_p^+ := \Pi_{\mathbb{R}_{\geq 0}^m}[p, s]$ . In this case, this projection can be implemented element-wise, i.e.,

$$[s_j]_{p_j}^+ = \begin{cases} s_j, & \text{if } p_j > 0, \\ \max\{s_j, 0\}, & \text{otherwise,} \end{cases}$$

for  $j = 1, 2, \dots, m$ , given  $p \in \mathbb{R}_{\geq 0}^m$  and  $s \in \mathbb{R}^m$ . Therefore, in this section we mainly look at the particular projected saddle flow dynamics:

$$\dot{x} = -\nabla_x S(x, y), \quad (38a)$$

$$\dot{y} = [+ \nabla_y S(x, y)]_y^+. \quad (38b)$$

Along any solution trajectories,  $y(t)$  will be constrained to be non-negative as long as it starts with a non-negative initial point.

### A. Augmented Primal-Dual Dynamics

Theorem 15 allows the state-augmentation method in Section III-B to work for the projected saddle flow dynamics (38), given the same observable certificate

$$\hat{h}(x, \hat{x}, y, \hat{y}) = \begin{bmatrix} \frac{\rho}{2} \|y - \hat{y}\|^2 \\ \frac{\rho}{2} \|x - \hat{x}\|^2 \end{bmatrix},$$

which consists of only the regularization terms, to satisfy Assumption 2. One interesting application is a novel augmented primal-dual algorithm to solve linear programs.

Consider the following problem:

$$\min_{x \in \mathbb{R}^n} c^T x \quad (39a)$$

$$\text{s.t.} \quad Ax - b \leq 0 : y \in \mathbb{R}_{\geq 0}^m \quad (39b)$$

where  $c \in \mathbb{R}^n$ ,  $A \in \mathbb{R}^{m \times n}$ , and  $b \in \mathbb{R}^m$  are given. The problem (39) corresponds to a bi-linear Lagrangian

$$L(x, y) := c^T x + y^T (Ax - b).$$

We introduce virtual variables  $\hat{x} \in \mathbb{R}^n$ ,  $\hat{y} \in \mathbb{R}^m$  to define the following augmented Lagrangian:

$$\hat{L}(x, \hat{x}, y, \hat{y}) := \frac{\rho}{2} \|x - \hat{x}\|^2 + c^T x + y^T (Ax - b) - \frac{\rho}{2} \|y - \hat{y}\|^2.$$

From Proposition 8 and Theorem 15, the projected saddle flow dynamics (38) of the augmented Lagrangian suggest an algorithm that globally asymptotically converges to one of its saddle points:

$$\dot{x} = -c - A^T y - \rho(x - \hat{x}), \quad (40a)$$

$$\dot{\hat{x}} = \rho(x - \hat{x}), \quad (40b)$$

$$\dot{y} = [Ax - b - \rho(y - \hat{y})]_y^+, \quad (40c)$$

$$\dot{\hat{y}} = \rho(y - \hat{y}). \quad (40d)$$

The augmented primal-dual dynamics (40) maintain the distributed structure where each agent  $i = 1, 2, \dots, n$  may locally manage

$$\dot{x}_i = -c_i - A_i^T y - \rho(x_i - \hat{x}_i), \quad (41a)$$

$$\dot{\hat{x}}_i = \rho(x_i - \hat{x}_i), \quad (41b)$$

and/or each dual agent  $j = 1, 2, \dots, m$  may locally manage

$$\dot{y}_j = [A_j x - b_j - \rho(y_j - \hat{y}_j)]_{y_j}^+, \quad (41c)$$

$$\dot{\hat{y}}_j = \rho(y_j - \hat{y}_j), \quad (41d)$$

with  $A_i$  and  $A_j$  denoting the  $i^{\text{th}}$  column and the  $j^{\text{th}}$  row of  $A$ , respectively. Lemma 7 implies that any saddle point of  $\hat{L}(x, \hat{x}, y, \hat{y})$  corresponds to a saddle point of  $L(x, y)$ , i.e., an optimal primal-dual solution to the linear program (39). Therefore, the augmented primal-dual dynamics (40) can be used as a distributed linear programming solver.

### B. Proximal Primal-Dual Dynamics

Likewise, Theorem 16 suggests that the proximal saddle flow dynamics in Section IV-B can be extended to handle inequality-constrained convex programs, leading to the proximal primal-dual algorithm proposed in [27]. However, rather than the asymptotic stability established therein, our theories enable a novel analysis of the algorithm that provides a bound on its convergence rate.

Consider the following problem:

$$\min_{x \in \mathbb{R}^n} f(x) \quad (42a)$$

$$\text{s.t.} \quad g(x) \leq 0 : y \in \mathbb{R}_{\geq 0}^m \quad (42b)$$

Here  $f : \mathbb{R}^n \mapsto \mathbb{R}$  is continuously differentiable and convex.  $g : \mathbb{R}^n \mapsto \mathbb{R}^m$  is a vector-valued function, and  $g_j(x)$ ,  $j = 1, 2, \dots, m$ , is convex with locally Lipschitz gradient. We define the standard Lagrangian:

$$L(x, y) := f(x) + y^T g(x),$$

and its surrogate function using proximal regularization:

$$\begin{aligned} \tilde{L}(u, y) &:= \min_{x \in \mathbb{R}^n} \left\{ f(x) + y^T g(x) + \frac{\rho}{2} \|x - u\|^2 \right\} \\ &= f(\tilde{x}(u, y)) + y^T g(\tilde{x}(u, y)) + \frac{\rho}{2} \|\tilde{x}(u, y) - u\|^2, \end{aligned} \quad (43)$$

where  $\tilde{x}(u, y)$  is the unique minimizer or the unique solution to the following first-order optimality condition:

$$H(x, u, y) = 0, \quad (44)$$

with

$$H(x, u, y) := \nabla f(x) + [\partial_x g(x)]^T y + \rho(x - u). \quad (45)$$

Note that here  $\partial_x g(x)$  is the Jacobian matrix of  $g(x)$ .

The proximal primal-dual dynamics, i.e., the projected saddle flow dynamics (38) of  $\tilde{L}(u, y)$ ,

$$\dot{u} = \rho u - \rho \tilde{x}(u, y), \quad (46a)$$

$$\dot{y} = [g(\tilde{x}(u, y))]_y^+, \quad (46b)$$

enjoy exponential convergence to a (unique) saddle point if the following assumption holds.

**Assumption 9.** *The function  $f(x)$  is  $\mu$ -strongly convex with  $l$ -Lipschitz gradient, i.e.,  $lI \succeq \nabla^2 f(x) \succeq \mu I$  wherever  $\nabla^2 f(x)$  is defined. The Jacobian matrix  $\partial_x g(x)$  is full row rank*

with  $\sigma I \succeq \partial_x g(x)[\partial_x g(x)]^T \succeq \kappa I$ , and locally Lipschitz continuous row-wise.

**Remark 6.**  $\partial_x g(x)$  being full row rank implies that the linear independence constraint qualification is satisfied, and the saddle point is unique.

Given Assumption 9, we can derive the partial derivatives of  $\tilde{x}(u, y)$  as

$$\begin{aligned} \partial_u \tilde{x} &= -(\partial_x H)^{-1} \partial_u H \\ &= \rho \left( \nabla^2 f(x) + \rho I + \sum_{j=1}^m y_j \nabla^2 g_j(x) \right)^{-1}, \end{aligned} \quad (47a)$$

$$\begin{aligned} \partial_y \tilde{x} &= -(\partial_x H)^{-1} \partial_y H \\ &= - \left( \nabla^2 f(x) + \rho I + \sum_{j=1}^m y_j \nabla^2 g_j(x) \right)^{-1} [\partial_x g(x)]^T, \end{aligned} \quad (47b)$$

which imply the second-order partial derivatives of  $\tilde{L}(u, y)$ :

$$\partial_{uu}^2 \tilde{L} = \rho I - \rho \partial_u \tilde{x} \succeq \frac{\mu \rho}{\mu + \rho} I \succ 0, \quad (48a)$$

$$\begin{aligned} \partial_{yy}^2 \tilde{L} &= \partial_x g(x) \partial_y \tilde{x} \\ &\preceq -\partial_x g(x) \left( \rho I + \sum_{j=1}^m y_j \nabla^2 g_j(x) \right)^{-1} [\partial_x g(x)]^T \\ &< 0. \end{aligned} \quad (48b)$$

Here we have used the fact of  $y_j \geq 0$  and  $\nabla^2 g_j(x) \succeq 0, \forall j = 1, 2, \dots, m$ .

Recall the analysis in (26) used to establish saddle flows' exponential convergence. Given the saddle point  $(u_*, y_*)$  of  $\tilde{L}(u, y)$ , (48) suffices to guarantee that the following quadratic Lyapunov function

$$\tilde{V}(u, y) := \frac{1}{2} \|u - u_*\|^2 + \frac{1}{2} \|y - y_*\|^2$$

is non-increasing along any trajectories of the proximal primal-dual dynamics (46). As a result, given an arbitrary initial point  $(u(0), y(0))$ , the trajectory of (46) is bounded and contained in an invariant domain

$$\tilde{D}_0(u(0), y(0)) := \left\{ (u, y) \mid \tilde{V}(u, y) \leq \tilde{V}(u(0), y(0)) \right\},$$

which is compact and convex since  $\tilde{V}(u, y)$  is radially unbounded. Suppose

$$\|y\| \leq \zeta(u(0), y(0)),$$

$$\nabla^2 g_j(x) \preceq \gamma(u(0), y(0)) I, \quad \forall j = 1, 2, \dots, m,$$

hold for any point  $(u, y) \in \tilde{D}_0(u(0), y(0))$ . Then along the trajectory starting from  $(u(0), y(0))$ ,

$$\partial_{yy}^2 \tilde{L} \preceq - \frac{\kappa}{l + \rho + m\zeta(u(0), y(0))\gamma(u(0), y(0))} I \prec 0 \quad (49)$$

always holds.

Akin to Theorem 16, it can thus be shown that the proximal primal-dual dynamics (46) are semi-globally exponentially stable. In other words, given the unique saddle point  $w_*$  and any arbitrary initial point  $w(0) \in \mathbb{R}^n \times \mathbb{R}_{\geq 0}^m$  with  $w = (u, y)$ ,

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

holds with the rate

$$c = \min \left\{ \frac{\mu \rho}{\mu + \rho}, \frac{\kappa}{l + \rho + m\zeta(w(0))\gamma(w(0))} \right\} > 0, \quad (50)$$

which is initial point-dependent. Meanwhile, according to [27, Theorem 5], the unique saddle point of  $\tilde{L}(u, y)$  is also the saddle point of  $L(x, y)$ , i.e., the unique optimal primal-dual solution to the convex program (42).

### C. Preconditioned Primal-Dual Dynamics

In light of the strong convexity-strong concavity condition in Theorems 10 and 16, we have discussed the use of the proximal methods in Sections IV-B and VI-B to achieve the exponential convergence of saddle flows. As observed in both cases, there is a bottleneck for the estimated convergence rate

$$c \leq \frac{\mu \rho}{\mu + \rho} < \mu,$$

i.e., the strong convexity constant  $\mu$ . We propose in this subsection a new algorithm that employs a change of variables to alter the condition number of the integrand (matrix) in (26) such that an accelerated convergence rate as fast as  $\mu$  can be achieved.

Consider the following problem:

$$\min_{x \in \mathbb{R}^n} f(x) \quad (51a)$$

$$\text{s.t.} \quad Ax - b \leq 0 : y \in \mathbb{R}_{\geq 0}^m \quad (51b)$$

where  $f: \mathbb{R}^n \mapsto \mathbb{R}$  is continuously differentiable and convex, and  $A \in \mathbb{R}^{m \times n}$  and  $b \in \mathbb{R}^m$  are given in the affine constraints. Define its (weighted) Lagrangian as

$$L(x, y) := f(x) + \eta y^T (Ax - b), \quad (52)$$

where  $\eta > 0$  is a constant. We propose the following change of variables (with abuse of the notation  $u$ )

$$u := x + \alpha A^T y \quad (53)$$

to transform  $L(x, y)$  into

$$\tilde{L}(u, y) := f(u - \alpha A^T y) + \eta y^T (Au - b) - \eta \alpha \|A^T y\|^2, \quad (54)$$

where  $\alpha > 0$  is also a constant. This creates an extra quadratic term in  $y$  and obviously,  $\tilde{L}(u, y)$  is still convex-concave. In fact,  $(x_*, y_*)$  is a saddle point of  $L(x, y)$  if and only if  $(u_*, y_*)$  is a saddle point of  $\tilde{L}(u, y)$  with  $u_* = x_* + \alpha A^T y_*$ , which follows from the equivalence below:

$$\begin{cases} \nabla_x L(x_*, y_*) = 0 \\ [\nabla_y L(x_*, y_*)]_y^+ = 0 \end{cases} \iff \begin{cases} \nabla_u \tilde{L}(u_*, y_*) = 0 \\ [\nabla_y \tilde{L}(u_*, y_*)]_y^+ = 0 \end{cases}$$

Since the problem (51) is a special case of (42), we show that given the same assumption below (Assumption 9 for the problem (51)), the projected saddle flow dynamics (38) of  $\tilde{L}(u, y)$ ,

$$\dot{u} = -\nabla f(u - \alpha A^T y) - \eta A^T y, \quad (55a)$$

$$\dot{y} = [-\alpha A \nabla f(u - \alpha A^T y) + \eta (Au - b) - 2\eta \alpha A A^T y]_y^+, \quad (55b)$$

converge exponentially to a unique saddle point with a rate as fast as  $\mu$ .

**Assumption 10.** *The function  $f(x)$  is  $\mu$ -strongly convex with  $l$ -Lipschitz gradient, i.e.,  $lI \succeq \nabla^2 f(x) \succeq \mu I$  whenever  $\nabla^2 f(x)$  is defined. The matrix  $A$  is full row rank with  $\sigma I \succeq AA^T \succeq \kappa I$ .*

Given Assumption 10, we can readily verify the strong convexity-strong concavity of  $\tilde{L}(u, y)$ , as along as the tunable constants  $\eta$  and  $\alpha$  are properly chosen subject to  $2\eta > l\alpha$ :

$$\partial_{uu}^2 \tilde{L} = \nabla^2 f(u - \alpha A^T y) \succeq \mu I \succ 0, \quad (56a)$$

$$\begin{aligned} \partial_{yy}^2 \tilde{L} &= \alpha^2 A \nabla^2 f(u - \alpha A^T y) A^T - 2\eta \alpha A A^T \\ &= A (\alpha^2 \nabla^2 f(u - \alpha A^T y) - 2\eta \alpha I) A^T \\ &\preceq A (l\alpha^2 - 2\eta \alpha) I A^T \\ &\preceq -(2\eta \alpha - l\alpha^2) \kappa I \\ &\prec 0. \end{aligned} \quad (56b)$$

Following Theorem 16, (56) implies that the preconditioned primal-dual dynamics (55) are globally exponentially stable, i.e., given the unique saddle point  $w_*$  and any arbitrary initial point  $w(0) \in \mathbb{R}^n \times \mathbb{R}_{\geq 0}^m$  with  $w = (u, y)$ ,

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

holds with the rate

$$c = \min \{ \mu, (2\eta \alpha - l\alpha^2) \kappa \} > 0.$$

**Remark 7.** *We can always pick  $\eta$  and  $\alpha$  to satisfy  $2\eta > l\alpha + \frac{\mu}{\kappa \alpha}$ , i.e.,  $(2\eta \alpha - l\alpha^2) \kappa > \mu$ , such that*

$$c = \min \{ \mu, (2\eta \alpha - l\alpha^2) \kappa \} = \mu$$

*holds. In this way, the preconditioned primal-dual dynamics (55) are guaranteed to converge with a faster rate than the one obtained for the proximal primal-dual dynamics (46).*

Due to the linear coordinate transformation, we can also directly work with the original primal-dual variables  $(x, y)$  by converting the preconditioned primal-dual dynamics (55) to the following:

$$\dot{x} = -\alpha A^T [-\alpha A (\nabla f(x) + \eta A^T y) + \eta (Ax - b)]_y^+ \quad (57a)$$

$$-\nabla f(x) - \eta A^T y, \quad (57b)$$

$$\dot{y} = [-\alpha A (\nabla f(x) + \eta A^T y) + \eta (Ax - b)]_y^+. \quad (57c)$$

Essentially (57), inspired by

$$\dot{x} := \dot{u} - \alpha A^T \dot{y},$$

yields a trajectory  $(x(t), y(t))$  that satisfies

$$x(t) \equiv u(t) - \alpha A^T y(t)$$

for any trajectory  $(u(t), y(t))$  of the preconditioned primal-dual dynamics (55). Thus, given Assumption 10, (57) exponentially converges to the unique saddle point  $(x_*, y_*)$  of  $L(x, y)$  with  $x_* = u_* - \alpha A^T y_*$ . Note that  $(x_*, \eta y_*)$  is the unique optimal primal-dual solution to the convex program (51).

**Proposition 17.** *Let Assumption 10 hold. Then the dynamics (57) are globally exponentially stable, given any  $\eta > 0$  and*

*$\alpha > 0$  that satisfy  $2\eta > l\alpha + \frac{\mu}{\kappa \alpha}$ . More precisely, given the (unique) saddle point  $z_*$  and any initial point  $z(0) \in \mathbb{R}^n \times \mathbb{R}_{\geq 0}^m$  with  $z = (x, y)$ ,*

$$\|z(t) - z_*\| \leq K \|z(0) - z_*\| e^{-\mu t}$$

*holds with  $K := \max\{2, 2\sigma\alpha^2 + 1\}$ .*

*Proof.* We start with the following inequalities:

$$\begin{aligned} &\|x(t) - x_*\|^2 \\ &= \|(u(t) - u_*) - \alpha A^T (y(t) - y_*)\|^2 \\ &\leq 2\|u(t) - u_*\|^2 + 2\alpha^2 \|A^T (y(t) - y_*)\|^2 \\ &\leq 2\|u(t) - u_*\|^2 + 2\sigma\alpha^2 \|y(t) - y_*\|^2 \end{aligned} \quad (58)$$

and

$$\begin{aligned} &\|u(t) - u_*\|^2 \\ &= \|(x(t) - x_*) + \alpha A^T (y(t) - y_*)\|^2 \\ &\leq 2\|x(t) - x_*\|^2 + 2\alpha^2 \|A^T (y(t) - y_*)\|^2 \\ &\leq 2\|x(t) - x_*\|^2 + 2\sigma\alpha^2 \|y(t) - y_*\|^2. \end{aligned} \quad (59)$$

Recall that here  $\sigma$  is the largest eigenvalue of  $AA^T$  from Assumption 10. Then given  $w = (u, y)$ , we are able to derive

$$\begin{aligned} \|z(t) - z_*\| &\leq \sqrt{2\|u(t) - u_*\|^2 + (2\sigma\alpha^2 + 1)\|y(t) - y_*\|^2} \\ &\leq \sqrt{K} \|w(t) - w_*\| \\ &\leq \sqrt{K} \|w(0) - w_*\| e^{-\mu t} \\ &\leq \sqrt{K} \sqrt{2\|x(0) - x_*\|^2 + (2\sigma\alpha^2 + 1)\|y(0) - y_*\|^2} e^{-\mu t} \\ &\leq K \|z(0) - z_*\| e^{-\mu t}, \end{aligned}$$

where the first inequality uses (58), the third inequality uses the exponential stability of the preconditioned primal-dual dynamics (55), and the fourth inequality uses (59). ■

**Remark 8.** *Despite the strongly convex-linear Lagrangian  $L(x, y)$ , the coordinate transformation*

$$\begin{bmatrix} u \\ y \end{bmatrix} = \begin{bmatrix} I & \alpha A^T \\ 0 & I \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

*renders strong convexity-strong concavity of the function in the  $(u, y)$  space, and thus establishes the exponential stability of the corresponding saddle flow dynamics. When the trajectories are converted back to the original  $(x, y)$  space, they are still exponentially convergent with the same rate, but the distance to equilibrium is not monotonically decreasing due to the constant  $K$ .*

## D. Reduced Primal-Dual Dynamics

To show the versatility of our condition for saddle flows' exponential convergence, we further propose a reduced primal-dual algorithm to solve the convex program (51) in Section VI-C if it admits the following separable structure:

$$\min_{x \in \mathbb{R}^n} f(x) := f_s(x_s) + f_c(x_c) \quad (60a)$$

$$\text{s.t. } A_s x_s + A_c x_c - b \leq 0 : y \in \mathbb{R}_{\geq 0}^m \quad (60b)$$

where  $f : \mathbb{R}^n \mapsto \mathbb{R}$  is additively separable into  $f_s : \mathbb{R}^{n_s} \mapsto \mathbb{R}$  and  $f_c : \mathbb{R}^{n_c} \mapsto \mathbb{R}$ . Accordingly,  $x_s \in \mathbb{R}^{n_s}$  and  $x_c \in \mathbb{R}^{n_c}$  are

the two separated subsets of variables with  $x = [x_s^T, x_c^T]^T \in \mathbb{R}^n$  and  $n = n_s + n_c$ . Both  $f_s(x_s)$  and  $f_c(x_c)$  are assumed to be continuously differentiable and convex.  $A \in \mathbb{R}^{m \times n}$  and  $b \in \mathbb{R}^m$  are given.  $A_s \in \mathbb{R}^{m \times n_s}$  and  $A_c \in \mathbb{R}^{m \times n_c}$  are the submatrices of  $A = [A_s, A_c]$  that consist of the columns corresponding to  $x_s$  and  $x_c$ , respectively.

We make the following assumption to proceed.

**Assumption 11.** *The function  $f_s(x)$  is  $\mu_s$ -strongly convex with  $l_s$ -Lipschitz gradient, i.e.,  $l_s I \succeq \nabla^2 f_s(x) \succeq \mu_s I$  wherever  $\nabla^2 f_s(x)$  is defined. Similarly,  $l_c I \succeq \nabla^2 f_c(x) \succeq \mu_c I$  holds wherever  $\nabla^2 f_c(x)$  is defined. The matrix  $A_s$  is full row rank with  $\sigma_s I \succeq A_s A_s^T \succeq \kappa_s I$ .*

Instead of introducing regularization, we use the standard Lagrangian

$$L(x, y) := f_s(x_s) + f_c(x_c) + y^T (A_s x_s + A_c x_c - b), \quad (61)$$

but minimize it over  $x_s$  to attain a reduced Lagrangian

$$\begin{aligned} \bar{L}(x_c, y) &:= \min_{x_s \in \mathbb{R}^{n_s}} L(x, y) \\ &= f_s(\bar{x}_s(y)) + f_c(x_c) + y^T (A_s \bar{x}_s(y) + A_c x_c - b) \end{aligned} \quad (62)$$

with  $\bar{x}_s(y)$  being the unique minimizer given  $y$  such that

$$\nabla f_s(\bar{x}_s(y)) + A_s^T y = 0 \quad (63)$$

holds.

The reduced Lagrangian  $\bar{L}(x_c, y)$  can be exploited due to the properties below.

**Lemma 18.** *Let Assumption 11 hold. Then  $\bar{L}(x_c, y)$  is strongly convex-strongly concave with the Lipschitz gradient:*

$$\nabla_{x_c} \bar{L}(x_c, y) = \nabla f_c(x_c) + A_c^T y, \quad (64a)$$

$$\nabla_y \bar{L}(x_c, y) = A_s \bar{x}_s(y) + A_c x_c - b. \quad (64b)$$

Moreover, a point  $(x_*, y_*)$  is a saddle point of  $L(x, y)$  if and only if  $(x_{c*}, y_*)$  is a saddle point of  $\bar{L}(x_c, y)$  with  $[\bar{x}_s^T(y_*), x_{c*}^T]^T = x_*$ .

*Proof.* It follows from (63) and the strict convexity of  $f_s$  that

$$\bar{x}_s(y) = (\nabla f_s)^{-1}(-A_s^T y) \quad (65)$$

is Lipschitz. Denote the Jacobian matrix of  $(\nabla f_s)^{-1}$  as  $\mathbf{J}_s$  wherever it is defined. Taking the derivative with respect to  $y$  on both sides of (63) leads to the following:

$$(-\nabla^2 f_s(\bar{x}_s(y)) \mathbf{J}_s + I) A_s^T \equiv 0 \iff \nabla^2 f_s(\bar{x}_s(y)) \mathbf{J}_s \equiv I \quad (66)$$

Then the gradient of  $\bar{L}(x_c, y)$  can be obtained using the chain rule:

$$\nabla_{x_c} \bar{L}(x_c, y) = \nabla f_c(x_c) + A_c^T y, \quad (67a)$$

$$\begin{aligned} \nabla_y \bar{L}(x_c, y) &= -A_s \mathbf{J}_s^T \nabla f_s(\bar{x}_s(y)) - A_s \mathbf{J}_s^T A_s^T y \\ &\quad + A_s \bar{x}_s(y) + A_c x_c - b \\ &= A_s \bar{x}_s(y) + A_c x_c - b. \end{aligned} \quad (67b)$$

This allows us to further compute the second-order partial derivatives as

$$\partial_{x_c x_c}^2 \bar{L}(x_c, y) = \nabla^2 f_c(x_c) \succeq \mu_c, \quad (68a)$$

$$\partial_{yy}^2 \bar{L}(x_c, y) = -A_s \mathbf{J}_s^T A_s^T \preceq -\frac{1}{l_s} A_s A_s^T \preceq -\frac{\kappa_s}{l_s}, \quad (68b)$$

where  $\mathbf{J}_s^T \succeq \frac{1}{l_s}$  follows from (66) with  $\nabla_{x_s}^2 f_s(x_s) \succeq l_s$ .

Given the strong convexity-strong concavity of  $\bar{L}(x_c, y)$  above, the correspondence between saddle points of  $L(x, y)$  and  $\bar{L}(x_c, y)$  is straightforward from

$$\begin{cases} \nabla_x L(x_*, y_*) = 0 \\ [\nabla_y L(x_*, y_*)]_y^+ = 0 \end{cases} \iff \begin{cases} \nabla_{x_c} \bar{L}(x_{c*}, y_*) = 0 \\ \nabla_y \bar{L}[(x_{c*}, y_*)]_y^+ = 0 \end{cases} \quad \blacksquare$$

Lemma 18 allows us to focus on the projected saddle flow dynamics (38) of  $\bar{L}(x_c, y)$ :

$$\dot{x}_c = -\nabla_{x_c} \bar{L}(x_c, y), \quad (69a)$$

$$\dot{y} = [+ \nabla_y \bar{L}(x_c, y)]_y^+, \quad (69b)$$

since such reduced primal-dual dynamics (69) converge exponentially to its unique saddle point according to Theorem 16. In other words, given the unique saddle point  $\bar{z}_*$  and any initial point  $\bar{z}(0) \in \mathbb{R}^{n_c} \times \mathbb{R}_{\geq 0}^m$  with  $\bar{z} := (x_c, y)$ ,

$$\|\bar{z}(t) - \bar{z}_*\| \leq \|\bar{z}(0) - \bar{z}_*\| e^{-ct}$$

holds with the rate

$$c = \min \left\{ \mu_c, \frac{\kappa_s}{l_s} \right\} > 0.$$

**Remark 9.** *The key enabler for the reduced primal-dual dynamics (69) is the separability of  $x$  into  $x_s$  and  $x_c$  such that all the necessary assumptions hold. It is possible that multiple ways of separation exist, then the convergence rate could be further optimized based on how the variables are split. In this case, comparing the reduced primal-dual dynamics (69) with the preconditioned primal-dual dynamics (55) might also suggest a better option between the two since in general neither of them dominates the other in terms of the convergence rate.*

## VII. SIMULATION RESULTS

### A. Network Flow Optimization

We apply the augmented primal-dual dynamics (40) to a network flow optimization problem, which is commonly formulated as a linear program. Consider a minimum-cost flow problem with a network specified in Fig. 1(a). Let  $\mathcal{V}$  and  $\mathcal{E}$  denote the sets of the nodes and the directed edges, respectively.  $x := (x_{ij}, (i, j) \in \mathcal{E}) \in \mathbb{R}^{|\mathcal{E}|}$  is the variable defined as the amount of flow on each edge  $(i, j)$  from node  $i$  to node  $j$ .  $c \in \mathbb{R}^{|\mathcal{E}|}$  and  $b \in \mathbb{R}^{2|\mathcal{E}|}$  denote the unit cost of flow transmission and the capacity for all the edges, respectively. Note that  $b$  is the concatenation of upper bounds (given in Fig. 1(a)) and lower bounds (set to 0).  $d \in \mathbb{R}^{|\mathcal{V}|}$  denotes the nodal (net) injection. Then given the incidence matrix  $B \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{E}|}$  and the matrix  $A := \text{diag}(I_{|\mathcal{E}|}, -I_{|\mathcal{E}|})$ , the minimum-cost flow problem is formulated as follows:

$$\min_{x \in \mathbb{R}^{|\mathcal{E}|}} c^T x \quad (70a)$$

$$\text{s.t.} \quad Bx - d = 0 : y^E \in \mathbb{R}^{|\mathcal{V}|} \quad (70b)$$

$$Ax - b \leq 0 : y^I \in \mathbb{R}_{\geq 0}^{2|\mathcal{E}|} \quad (70c)$$

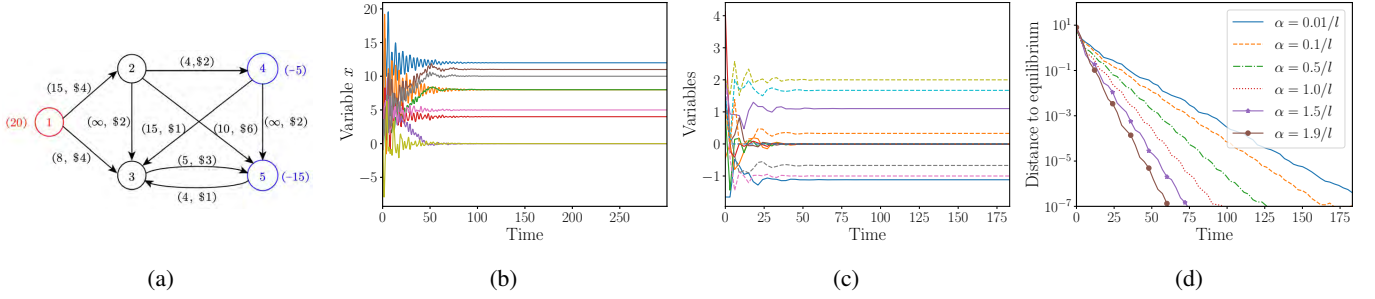


Fig. 1: (a) Network specification. The nodes are parameterized by injection (positive) or withdrawal (negative), and the edges are parameterized by capacity and unit cost. (b) Asymptotic convergence of trajectories for primal variable  $x$  with  $\rho = 0.05$ . (c) Exponential convergence of trajectories for variables  $u$  (solid lines) and  $v$  (dashed lines) with  $\alpha = \frac{0.01}{l}$ . (d) Exponential decay of distance to equilibrium  $\sqrt{\|u(t) - u_*\|^2 + \|v(t) - v_*\|^2}$  with varying  $\alpha$ .

where the equality constraints enforce the flow conservation while the inequality constraints enforce the flow capacity.

Given the linear programming formulation (70), the augmented primal-dual dynamics (40) can be applied to attain an optimal solution with asymptotic convergence guarantee. Note the projection is only defined on  $y^l$ . Fig. 1(b) shows three stages of the convergence process. Initially, the trajectories exhibit large-amplitude oscillations in the first 10 seconds, as the system searches around for the region of optimal solutions. The oscillations decay over the following 40 seconds. After approximately 50 seconds, the trajectories start to center around the equilibrium, and finally asymptotically converge to an optimal solution of the minimum flow cost problem (70).

### B. Lasso Regression

We use a classical Lasso regression problem to test the performance of an algorithm that combines the preconditioning method in Section VI-C with proximal regularization. Consider the following problem

$$\min_{\hat{x} \in \mathbb{R}^n} \hat{f}(\hat{x}) + \lambda \|\hat{x}\|_1 \quad (71)$$

where  $\lambda$  is a constant.  $\hat{f}: \mathbb{R}^n \mapsto \mathbb{R}$  is convex with  $l$ -Lipschitz gradient, i.e.,  $lI \succeq \nabla^2 \hat{f}(\hat{x}) \succeq 0$  wherever  $\nabla^2 \hat{f}(\hat{x})$  is defined.

To handle the non-smooth term  $\|\hat{x}\|_1$ , we define  $\bar{x}_i^+ := \max\{\hat{x}_i, 0\}$  and  $\bar{x}_i^- = \max\{-\hat{x}_i, 0\}$  for  $\forall i = 1, 2, \dots, n$ , such that

$$\hat{x}_i = \bar{x}_i^+ - \bar{x}_i^-, \quad (72a)$$

$$|\hat{x}_i| = \bar{x}_i^+ + \bar{x}_i^-, \quad (72b)$$

hold. Denote  $\bar{x}^+ := (\bar{x}_i^+, i = 1, 2, \dots, n) \in \mathbb{R}^n$ ,  $\bar{x}^- := (\bar{x}_i^-, i = 1, 2, \dots, n) \in \mathbb{R}^n$ , and  $\bar{x} := [(\bar{x}^+)^T, (\bar{x}^-)^T]^T$ . The Lasso regression problem (71) can be transformed equivalently into

$$\min_{\hat{x} \in \mathbb{R}^n, \bar{x} \in \mathbb{R}^{2n}} \hat{f}(\hat{x}) + \lambda \mathbf{1}^T \bar{x} \quad (73a)$$

$$\text{s.t.} \quad \hat{x} + C\bar{x} = 0 : \hat{y} \in \mathbb{R}^n \quad (73b)$$

$$\bar{x} \geq 0 : \bar{y} \in \mathbb{R}_{\geq 0}^{2n} \quad (73c)$$

with  $C = [-I_n, I_n]$ . Given  $x := [\hat{x}^T, \bar{x}^T]^T$  and  $y := [\hat{y}^T, \bar{y}^T]^T$ , we can define its Lagrangian as

$$\begin{aligned} L(x, y) &:= \hat{f}(\hat{x}) + \lambda \mathbf{1}^T \bar{x} + \hat{y}^T (\hat{x} + C\bar{x}) - \bar{y}^T \bar{x} \\ &= \underbrace{\hat{f}(\hat{x}) + \lambda \mathbf{1}^T \bar{x}}_{=: f(x)} + y^T \underbrace{\begin{bmatrix} I_n & C \\ 0 & -I_{2n} \end{bmatrix}}_A x. \end{aligned} \quad (74)$$

Note that the matrix  $A$  is full row rank. Thus we bound the eigenvalues of  $AA^T$  by  $\sigma I \succeq AA^T \succeq \kappa I$ .

The Lagrangian (74) is only convex-linear. We propose an algorithm that implements the primal-dual dynamics of a transformed function. The transformation involves two steps: first use the change of variables (53) to attain a convex-strongly concave function and then apply a more general version of proximal regularization on the dual variable  $y$  to attain the final function (recall our analysis of the proximal method in Section IV-B showing the “shift” of strong convexity).

More specifically, with the change of variables (53), the Lagrangian (74) is transformed into

$$L^C(u, y) := f(u - \alpha A^T y) + y^T A u - \alpha \|A^T y\|^2. \quad (75)$$

By properly choosing the constant  $\alpha > 0$  such that  $\alpha < \frac{2}{l}$ , we can verify the convexity-strong concavity of  $L^C(u, y)$  with

$$\partial_{uu}^2 L^C(u, y) = \nabla^2 f(u - \alpha A^T y) \succeq 0, \quad (76a)$$

$$\begin{aligned} \partial_{yy}^2 L^C(u, y) &= \alpha^2 A \nabla^2 f(u - \alpha A^T y) A^T - 2\alpha A A^T \\ &= A (\alpha^2 \nabla^2 f(u - \alpha A^T y) - 2\alpha I) A^T \\ &\preceq A (l\alpha^2 - 2\alpha) A^T \\ &\preceq -(2\alpha - l\alpha^2) \kappa I \\ &\prec 0. \end{aligned} \quad (76b)$$

In terms of  $L^C(u, y)$ , we apply the following proximal method on  $y$  to attain the final transformed function:

$$\begin{aligned} L^P(u, v) &:= \max_{y \in \mathbb{R}^n \times \mathbb{R}_{\geq 0}^{2n}} \{L^C(u, y) - \frac{\rho}{2} \|y - v\|^2\} \\ &= f(u - \alpha A^T \tilde{y}) + \tilde{y}^T A u - \alpha \|A^T \tilde{y}\|^2 - \frac{\rho}{2} \|\tilde{y} - v\|^2, \end{aligned} \quad (77)$$

where  $\tilde{y}$  is a shorthand of  $\tilde{y}(u, v)$ , denoting the unique maximizer given  $(u, v)$ .

We thus employ the saddle flow dynamics of  $L^P(u, v)$ , i.e.,

$$\dot{u} = -\nabla f(u - \alpha A^T \tilde{y}(u, v)) - A^T \tilde{y}(u, v), \quad (78a)$$

$$\dot{v} = \rho \tilde{y}(u, v) - \rho v. \quad (78b)$$

to solve the Lasso regression problem (71), since whenever (78) converges to an equilibrium point  $(u_*, v_*)$ , we can easily recover  $(x_* = u_* - \alpha A^T v_*, y_* = v_*)$  as an optimal solution to (71). Note that the maximizer  $\tilde{y}(u, v)$  can be efficiently computed using numerical methods since  $\nabla f(\cdot)$  is strictly monotone.

In particular, we specify  $\hat{f}(\hat{x}) := \frac{1}{2} \|A\hat{x} - b\|^2$ , where  $A \in \mathbb{R}^{m \times n}$  and  $b \in \mathbb{R}^m$  could represent a collection of known input and output data of size  $m$ , respectively. Fig. 1(c) shows the trajectories of (78) that converge rapidly in approximately 40 seconds. Fig. 1(d) further shows the exponentially decaying distance to the equilibrium as we implement the dynamics (78). Moreover, the decaying rate increases in the tunable constant  $\alpha$ . While the more general version of proximal regularization (77) goes beyond our analysis in Section IV-B due to the constrained feasible region  $\mathbb{R}^n \times \mathbb{R}_{\geq 0}^{2n}$  for  $y$ , the numerical results suggest that the desirable exponential stability of the saddle flow dynamics (78) may still hold, thus opening the path for future extensions to further relax our conditions.

### VIII. CONCLUSION

This paper focuses on the convergence behavior of saddle flow dynamics and provides a unified analysis that leads to two novel conditions for asymptotic and exponential convergence. The first condition is an observable certificate which, if identified for a convex-concave function, guarantees the saddle flow dynamics to converge asymptotically. It generalizes some existing conditions and further inspires the design of a novel augmented algorithm that only requires minimal assumptions on convexity-concavity for asymptotic convergence and even works for bilinear functions. The second condition is the strong convexity-strong concavity of an objective function which suffices to guarantee saddle flows' global exponential stability and provides a convenient lower-bound estimate of the rate of convergence. The insight also reveals how proximal regularization on a strongly convex-concave function leads to exponential convergence by "shifting" some of the strong convexity to make the resulting function also strongly concave. Our analysis and results can be extended to a projected version of saddle flow dynamics, the trajectories of which are constrained within a closed convex polyhedron. Such an extension allows the immediate application of our theory to develop and analyze primal-dual algorithms for constrained convex optimization. This leads to three novel algorithm designs and one novel bound on the convergence rate of proximal primal-dual algorithms, highlighting weaker convergence conditions and/or faster convergence rates. Finally, we run extensive numerical experiments with our new algorithms applied to solve network flow optimization and Lasso regression. The simulations validate our theoretical development and also suggest potential extensions.

### REFERENCES

- [1] P. You and E. Mallada, "Saddle flow dynamics: Observable certificates and separable regularization," in *IEEE American Control Conference (ACC)*, 2021, pp. 4817–4823.
- [2] T. Holding and I. Lestas, "Stability and instability in saddle point dynamics—Part I," *IEEE Transactions on Automatic Control*, vol. 66, no. 7, pp. 2933–2944, 2021.
- [3] T. Holding and I. Lestas, "Stability and instability in saddle point dynamics Part II: The subgradient method," *IEEE Transactions on Automatic Control*, vol. 66, no. 7, pp. 2945–2960, 2021.
- [4] A. Wibisono, A. C. Wilson, and M. I. Jordan, "A variational perspective on accelerated methods in optimization," *Proceedings of the National Academy of Sciences*, vol. 113, no. 47, pp. E7351–E7358, 2016.
- [5] G. França, D. P. Robinson, and R. Vidal, "Gradient flows and proximal splitting methods: A unified view on accelerated and stochastic optimization," *Physical Review E*, vol. 103, no. 5, p. 053304, 2021.
- [6] H. Mohammadi, M. Razaviyayn, and M. R. Jovanović, "Robustness of accelerated first-order algorithms for strongly convex optimization problems," *IEEE Transactions on Automatic Control*, vol. 66, no. 6, pp. 2480–2495, 2021.
- [7] L. Lessard, B. Recht, and A. Packard, "Analysis and design of optimization algorithms via integral quadratic constraints," *SIAM Journal on Optimization*, vol. 26, no. 1, pp. 57–95, 2016.
- [8] D. Richert and J. Cortés, "Robust distributed linear programming," *IEEE Transactions on Automatic Control*, vol. 60, no. 10, pp. 2567–2582, 2015.
- [9] 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.
- [10] H. K. Khalil, *Nonlinear systems*, 3rd ed. Prentice Hall, 2002.
- [11] J. Cortés, "Finite-time convergent gradient flows with applications to network consensus," *Automatica*, vol. 42, no. 11, pp. 1993–2000, 2006.
- [12] K. Garg and D. Panagou, "Fixed-time stable gradient flows: Applications to continuous-time optimization," *IEEE Transactions on Automatic Control*, vol. 66, no. 5, pp. 2002–2015, 2021.
- [13] S. Rahili and W. Ren, "Distributed continuous-time convex optimization with time-varying cost functions," *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 1590–1605, 2016.
- [14] M. Fazlyab, S. Paternain, V. M. Preciado, and A. Ribeiro, "Prediction-correction interior-point method for time-varying convex optimization," *IEEE Transactions on Automatic Control*, vol. 63, no. 7, pp. 1973–1986, 2017.
- [15] T. Zheng, J. Simpson-Porco, and E. Mallada, "Implicit trajectory planning for feedback linearizable systems: A time-varying optimization approach," in *IEEE American Control Conference (ACC)*, 2020, pp. 4677–4682.
- [16] H. Lu, "An  $O(s^r)$ -resolution ODE framework for understanding discrete-time algorithms and applications to the linear convergence of minimax problems," *Mathematical Programming*, vol. 194, no. 1, pp. 1061–1112, 2022.
- [17] A. Cherukuri, E. Mallada, and J. Cortés, "Asymptotic convergence of constrained primal–dual dynamics," *Systems & Control Letters*, vol. 87, pp. 10–15, 2016.
- [18] 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.
- [19] 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.
- [20] E. Mallada, C. Zhao, and S. H. Low, "Optimal load-side control for frequency regulation in smart grids," *IEEE Transactions on Automatic Control*, vol. 62, no. 12, pp. 6294–6309, 2017.
- [21] M. Chiang, S. H. Low, A. R. Calderbank, and J. C. Doyle, "Layering as optimization decomposition: A mathematical theory of network architectures," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 255–312, 2007.
- [22] 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.
- [23] D. Goldsztajn, F. Paganini, and A. Ferragut, "Proximal optimization for resource allocation in distributed computing systems with data locality," in *IEEE Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2019, pp. 773–780.
- [24] K. J. Arrow, H. Azawa, L. Hurwicz, and H. Uzawa, *Studies in Linear and Non-Linear Programming*. Stanford University Press, 1958.



- [25] A. Cherukuri, B. Ghahesifard, and J. Cortés, “Saddle-point dynamics: Conditions for asymptotic stability of saddle points,” *SIAM Journal on Control and Optimization*, vol. 55, no. 1, pp. 486–511, 2017.
- [26] A. Cherukuri, A. D. Domínguez-García, and J. Cortés, “Distributed coordination of power generators for a linearized optimal power flow problem,” in *IEEE American Control Conference (ACC)*, 2017, pp. 3962–3967.
- [27] 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, 2020.
- [28] J. Cortés and S. K. Niederländer, “Distributed coordination for non-smooth convex optimization via saddle-point dynamics,” *Journal of Nonlinear Science*, vol. 29, no. 4, pp. 1247–1272, 2019.
- [29] 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.
- [30] 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.
- [31] N. K. Dhingra, S. Z. Khong, and M. R. Jovanović, “The proximal augmented lagrangian method for nonsmooth composite optimization,” *IEEE Transactions on Automatic Control*, vol. 64, no. 7, pp. 2861–2868, 2018.
- [32] D. Ding, B. Hu, N. K. Dhingra, and M. R. Jovanović, “An exponentially convergent primal-dual algorithm for nonsmooth composite minimization,” in *IEEE Conference on Decision and Control (CDC)*, 2018, pp. 4927–4932.
- [33] D. Ding and M. R. Jovanović, “Global exponential stability of primal-dual gradient flow dynamics based on the proximal augmented lagrangian: A Lyapunov-based approach,” in *IEEE Conference on Decision and Control (CDC)*, 2020, pp. 4836–4841.
- [34] 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.
- [35] P. Cisneros-Velarde, S. Jafarpour, and F. Bullo, “A contraction analysis of primal-dual dynamics in distributed and time-varying implementations,” *IEEE Transactions on Automatic Control*, vol. 67, no. 7, pp. 3560–3566, 2022.
- [36] P. Bansode, V. Chinde, S. Wagh, R. Pasumarthy, and N. Singh, “On the exponential stability of projected primal-dual dynamics on a Riemannian manifold,” *arXiv:1905.04521*, 2019.
- [37] I. K. Ozaslan and M. R. Jovanović, “Tight lower bounds on the convergence rate of primal-dual dynamics for equality constrained convex problems,” in *IEEE Conference on Decision and Control (CDC)*, 2023, pp. 7318–7323.
- [38] N. Parikh and S. Boyd, “Proximal algorithms,” *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.
- [39] S. P. Bhat and D. S. Bernstein, “Nontangency-based Lyapunov tests for convergence and stability in systems having a continuum of equilibria,” *SIAM Journal on Control and Optimization*, vol. 42, no. 5, pp. 1745–1775, 2003.
- [40] T. Zheng, N. Loizou, P. You, and E. Mallada, “Dissipative gradient descent ascent method: A control theory inspired algorithm for min-max optimization,” *IEEE Control Systems Letters*, vol. 8, pp. 2009–2014, 2024.
- [41] 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.
- [42] W. Rudin, *Real and Complex Analysis*, 3rd ed. McGraw-Hill, 1987.
- [43] X. Chen and N. Li, “Exponential stability of primal-dual gradient dynamics with non-strong convexity,” in *IEEE American Control Conference (ACC)*, 2020, pp. 1612–1618.
- [44] J. Hubbard and B. Hubbard, *Vector Calculus, Linear Algebra, and Differential Forms: A Unified Approach*. Prentice Hall, 1999.
- [45] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. McGraw-Hill, 2008.
- [46] F. Bullo, *Contraction Theory for Dynamical Systems*, 1.2 ed. Kindle Direct Publishing, 2024.
- [47] I. K. Ozaslan and M. R. Jovanović, “Exponential convergence of primal-dual dynamics for multi-block problems under local error bound condition,” in *IEEE Conference on Decision and Control (CDC)*, 2022, pp. 7579–7584.
- [48] 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.
- [49] A. Nagurny and D. Zhang, *Projected Dynamical Systems and Variational Inequalities with Applications*. Springer US, 1996.
- [50] J.-P. Aubin and C. Arrigo, *Differential Inclusions: Set-Valued Maps and Viability Theory*. Springer Science & Business Media, 1984, vol. 264.
- [51] A. Bacciotti and F. Ceragioli, “Nonpathological Lyapunov functions and discontinuous Carathéodory systems,” *Automatica*, vol. 42, no. 3, pp. 453–458, 2006.



**Pengcheng You** (S'14-M'18) is an Assistant Professor at the Department of Industrial Engineering and Management, Peking University. He also holds a joint appointment at the National Engineering Laboratory for Big Data Analysis and Applications, Peking University. Prior to joining PKU, he was a Postdoctoral Fellow at the ECE and ME Departments, Johns Hopkins University. He earned his Ph.D. and B.S. degrees both from Zhejiang University, China. During the graduate studies, he was a visiting student at Caltech and a research intern at PNNL. His research interests include control, optimization, reinforcement learning and market mechanism, with main applications in power and energy systems.



**Yingzhu Liu** (S'24) received her B.S. degree in mechanical engineering from Peking University, Beijing, China, in 2022. She is currently working toward the Ph.D. degree in Dynamical Systems and Control with the Department of Mechanics and Engineering Science, College of Engineering, Peking University, Beijing, China.

Her research interests include saddle flow dynamics, minimax optimization, control theory, and reinforcement learning.



**Enrique Mallada** (S'09-M'13-SM'19) is an Associate Professor of Electrical and Computer Engineering at Johns Hopkins University. He was an Assistant Professor in the same department from 2016 to 2022. Prior to joining Hopkins, he was a Post-Doctoral Fellow in the Center for the Mathematics of Information at Caltech from 2014 to 2016. He received his Ingeniero en Telecomunicaciones degree from Universidad ORT, Uruguay, in 2005 and his Ph.D. degree in Electrical and Computer Engineering with a minor in Applied Mathematics from Cornell University in 2014. Dr. Mallada was awarded the NSF CAREER award in 2018, the ECE Director's PhD Thesis Research Award for his dissertation in 2014, the Center for the Mathematics of Information (CMI) Fellowship from Caltech in 2014, and the Cornell University Jacobs Fellowship in 2011. His research interests lie in the areas of control, dynamical systems, optimization, machine learning, with applications to engineering networks such as power systems and the Internet.