A First-Principles Analysis of Saddle Flow Dynamics: Stability and Algorithm Design

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

***Abstract*—This work studies the conditions that underlie the asymptotic and exponential convergence of saddle flow dynamics of convex-concave functions to a saddle point. First, we propose an observability-based certificate for asymptotic convergence of saddle flows, directly bridging the gap between the invariant set and the equilibrium set in a LaSalle argument. This certificate generalizes conventional conditions for convergence, e.g., strict convexity-concavity, and leads to a novel, seperable regularization method that requires minimal convexity-concavity for asymptotic convergence. Second, we show that global exponential stability in saddle flows is a direct consequence of strong convexity-strong concavity, which provides a lower-bound estimate of the convergence rate. This insight explains the convergence behavior of proximal gradient algorithms for strongly convex-concave objective functions. Our results generalize to saddle flow dynamics with projections on the vector field and have immediate applications in constrained convex optimization as primal-dual dynamics. In particular, our regularized algorithms can be used to solve linear programs distributedly. Besides, the insight behind strong convexity-strong concavity is further exploited to design a novel alternative conditioned algorithm for inequality-constrained convex problems. Our theoretical results are verified by a network flow problem and a Lasso Regression problem.**

***Index Terms*—Saddle flow dynamics, saddle point, asymptotic convergence, exponential convergence, regularization**

I. INTRODUCTION

Studying optimization algorithms from a dynamical systems view point has become one of the frontier research topics, providing 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, later on, 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 counterpart [16].

One prominent area within this field is the study of saddle flow dynamics, i.e., dynamics in the gradient descent direction on a sub-set of variables and the gradient ascent direction on the complement. Designed for locating min-max saddle points, saddle flow dynamics are particularly suited for solving constrained optimization problems via primal-dual methods [17], and finding Nash equilibria of zero-sum games [18], which lead to a broad application spectrum, including power systems [19], [20], communication networks [21], [22], and cloud computing [23].

A major branch of studies is the asymptotic behavior of saddle flows, with the purpose to identify and potentially relax convergence conditions. The seminal work [24] initially explores the asymptotic convergence within primal-dual algorithms. Since then, advanced analytical tools are used to re-validated conventional conditions for asymptotic convergence. 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. Furthermore, recent studies have demonstrated that asymptotic convergence can be achieved under weaker conditions. One line of conditions focuses on conditions related to convexity, such as local strong convexity-concavity [9], convexity-linearity, or strong quasiconvexity-quasiconcavity [25]. Regularization methods serve as alternative ways to circumvent the above conditions, effectively handling the Lagrangian of constrained convex/linear optimization through various penalty terms on equality constraints or even projected inequality constraints [8], [26], as well as the proximal method [23], [27]. Despite the merit of regularization that relaxes conditions for convergence, the extra penalty terms commonly introduce couplings that may require additional computation and communication overheads when realizing distributed implementation.

Recently, the focus has started to shift toward exponential stability, which is a desirable property of dynamical systems both theoretically and in practice. A series of algorithms have explored the exponential convergence through different techniques, such as the augmented Lagrangian [28], [29], proximal operators [30]–[32], and so on. Special projections are designed as the augmentation of the Lagrangian to avoid discontinuity in recent research [33], [34]. Later on, [28] proposed the projection-free dynamics for linear inequality-constrained cases. A continuing work [29] considers convex inequality constraints with the semi-global exponential convergence depending on the initial point. Another kind of penalty component in the augmented Lagrangian relies on proximal operators. For instance, the framework of integral quadratic constraints has been applied to validate the exponential stability of the proximal gradient flow, as presented in [30]–[32]. However, the current analysis of convergence rates

P. You and Y. Liu are with the College of Engineering, Peking University, Beijing, China (email: {pcyou,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].

mainly centered on primal-dual dynamics of the Lagrangian and remains largely case-specific, highly dependent on the techniques applied. Furthermore, the convergence behavior for general saddle flow dynamics seems not fully understood.

Another subject of analysis concerns how to figure out the less conservative estimation of the convergence rate. Most existing results seem to depend on several parameters. For equality-constrained problems, [35], [36] establish the strict contraction under specific metrics with convergence rates that are related to properties like singular values of the constraint matrix. [37] uses a two-timescale approach which leads to two different convergence regimes depending on the scale ratio. For the extension to linear inequality-constrained problems, a Riemannian geometric framework is proposed in [38] to prove the global exponential stability associated with adjusting parameters by showing the strong monotonicity of gradients endowed with a Riemannian metric. By utilizing the coordinate transformation, our paper enjoys a better lower bound on the convergence rate of the projected primal-dual dynamics for optimization problems with inequality constraints.

In this paper, we propose a systematic methodology for estimating the convergence rate, based purely on the properties of saddle functions. First, we provide a sufficient certificate for the asymptotic convergence of saddle flow dynamics for convex-concave functions. The certificate with observable properties directly connects the invariant set and the equilibrium set through LaSalle's invariance principle, inspiring a novel separable dissipative method that only requires minimal convexity-concavity to establish convergence. Second, we establish the strong convexity-strong concavity condition of objective functions for exponential convergence. Furthermore, it provides a new explanation for the proximal method, showing how proximal operators enhance the concavity. Finally, our results can be generalized to projected saddle flow dynamics on the vector field, and yield novel regularized and conditioned algorithms that have immediate applications in inequality-constrained convex optimization. Compared to most existing ones, our findings exhibit generality in the context of saddle functions, rather than for primal-dual dynamics, directly connecting the properties of the saddle functions to their convergence properties.

Contributions: In particular, our contributions are summarized as follows:

(i) We propose a sufficient observable certificate for asymptotic convergence of saddle flow dynamics of a convex-concave function, which is weaker than existing conditions, e.g., strict convexity-concavity and proximal regularization.

(ii) Furthermore, minimal convexity-concavity conditions of asymptotical convergence are derived through a novel separable dissipative method in a distributed manner.

(iii) We similarly establish a strong convex-strong concave condition, which directly implies the exponential convergence of saddle flow dynamics. Notice that, the convergence purely depends on the structure of the saddle function

(iv) The insight behind the strong convex-concave condition provides a new perspective to explain existing algorithms and inspire new algorithms, such as proximal gradient algorithms for nonsmooth functions and the change of variable for bilinear functions.

(v) Two sufficient conditions can be generalized to accommodate projections on the vector field to solve inequality-constrained optimization problems.

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

Notation: Let $\mathbb{R}$ and $\mathbb{R}_{\geq 0}$ be the sets of real and nonnegative 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^{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 $S : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, we use $\frac{\partial}{\partial x}S(x,y) \in \mathbb{R}^{1\times n}$ and $\frac{\partial}{\partial y}S(x,y) \in \mathbb{R}^{1\times m}$ to denote the partial derivatives with respect to x and y, respectively. We further define $\nabla_x S(x,y) = \left[\frac{\partial}{\partial x}S(x,y)\right]^T$ while $\frac{\partial^2}{\partial x \partial y}S(x,y)$ represents taking the partial derivative of $S(x,y)$ with respect to y first, then with respect to x. Particularly, we further define $\frac{\partial^2}{\partial x \partial y}S(x,y) = \left[\frac{\partial}{\partial y}\nabla_x S(x,y)\right]^T$. Besides, $\mu(A)$ denotes the matrix measure of A corresponding to the norm we use. In particular, for the 2-norm, one has $\mu_2(A) = \lambda_{\max}(A + A^T)/2$ where $\lambda_{\max}$ is the maximal eigenvalue.

II. Problem Formulation

We consider a function $S(x,y)$ with $S : \mathcal{D} \to \mathbb{R}$ where $\mathcal{D} = \mathcal{X} \times \mathcal{Y}$ and 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 $(x_\star, y_\star)$ 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 the purpose of this paper, we focus our attention on functions $S(x,y)$ that are *convex-concave*.

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

We further derive a general definition of the saddle point of a convex-concave function $S(x,y)$ as follows:

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

$$S(x_\star, y) \leq S(x_\star, y_\star) \leq S(x, y_\star) \tag{1}$$

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.*

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 a continuously differentiable surrogate of a continuous convex/concave function by means of the Moreau Envelope [39].

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

$$\dot{x} = -\nabla_x S(x,y)\,, \tag{2a}$$
$$\dot{y} = +\nabla_y S(x,y)\,, \tag{2b}$$

as the saddle flow dynamics of $S(x,y)$. Next, we suppose that such a point $(x_\star, y_\star)$ does exist in the feasible set.

Assumption 2. *$S(x,y)$ is convex-concave, continuously differentiable, and there exists at least an equilibrium point of* (2) *in the domain $\mathcal{D}$.*

Under Assumption 2, we further define stationary points.

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

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

holds.

Remark 1. *If stationary points exist, any stationary point must be a saddle point defined by* (2)*.*

$$\nabla_x S(x_\star, y_\star) = 0 \Rightarrow S(x_\star, y_\star) \le S(x, y_\star), \forall x \in \mathcal{X}$$
$$\nabla_y S(x_\star, y_\star) = 0 \Rightarrow S(x_\star, y) \le S(x_\star, y_\star), \forall y \in \mathcal{Y}$$

However, the converse is not true without Assumption 2.

Due to convexity-concavity, the dynamic law drives the system towards such stationary points in gradient descent and ascent directions, respectively, for x and y. We will mainly work with this standard form of saddle flow dynamics to locate a saddle point of $S(x,y)$. In the following Section III and Section IV, we first consider the case that the feasible domain is full space, i.e., $\mathcal{D} = \mathbb{R}^n \times \mathbb{R}^m$. Then we move to the extension of projected version over the feasible set.

III. Asymptotic Convergence

This section presents an observable certificate that ensures asymptotic convergence of the saddle flow dynamics (2) towards a saddle point of $S(x,y)$. We show that two conventional conditions of strict convexity-concavity and proximal regularization satisfy this certificate as special cases. We further build on this certificate to develop a separable dissipative 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 now describe the proposed observable certificate for the saddle flow dynamics (2) to asymptotically converge to a saddle point of $S(x,y)$.

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

$$\begin{bmatrix} S(x_\star, y_\star) - S(x_\star, y) \\ S(x, y_\star) - S(x_\star, y_\star) \end{bmatrix} \ge h(x,y) \ge 0 \tag{4}$$

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

Remark 2. *We call $h(x,y)$ an observable certificate, due to the second property of Definition 4, which is akin to* (2) *having $h(x,y)$ as an observable output. It is exactly this observability property that will allow us to connect invariant sets with saddle-points.*

Assumption 3. *$S(x,y)$ has an observable certificate $h(x,y)$ as given by Definition 4.*

Checking whether Assumption 3 holds basically requires hunting for a qualified observable certificate $h(x,y)$ of $S(x,y)$. Under this assumption, asymptotic convergence of the saddle flow dynamics (2) is formally stated below.

Theorem 1 (Sufficiency of Observable Certificates)**.** *Let Assumptions 1 and 3 hold. Then the saddle flow dynamics* (2) *asymptotically converge to some saddle point $(x_\star, y_\star)$ of $S(x,y)$.*

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

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

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

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

where the second equality plugs in (2), the first inequality applies the convexity-concavity of $S(x,y)$, and the last inequality follows from the saddle property (1) of $(x_\star, y_\star)$.

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

$$D_0(x_0, y_0) := \{(x,y) \mid V(x,y) \le V(x_0, y_0)\}\,, \tag{6}$$

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

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

Given Assumption 3, (4) implies that $\mathbb{S}$ is indeed a subset of

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

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

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

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

It follows that 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., [40, Corollary 5.2]. Alternatively, notice that $\mathbb{S}$ is compact, and as a result any trajectory within the Ω 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 $(x_\star, y_\star)$ specifically to $(\bar{x}, \bar{y})$ in the definition of $V(x,y)$, it follows that $0 \leq V(x(t),y(t)) \to 0$ holds, which implies $(x(t),y(t)) \to (\bar{x}, \bar{y})$. ■

The existence and characterization of such observable certificates $h(x,y)$ may still be vague from only 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 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 4. *$S(x,y)$ is strictly convex-concave.*

Proposition 2 (Strict Convexity-Concavity)**.** *Let Assumptions 1 and 4 hold. Then the function*

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

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

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

Corollary 3. *Let Assumptions 1 and 4 hold. Then the saddle flow dynamics* (2) *asymptotically converge to some saddle point $(x_\star, y_\star)$ of $S(x,y)$.*

2) Proximal Regularization: In the particular form of saddle flow dynamics known as primal-dual dynamics [17], a proximal regularization method is proposed in [23], [27] to guarantee asymptotic convergence of the regularized saddle flow dynamics, even in the absence of strict convexity-concavity. Specifically, a surrogate differentiable convex-concave function

$$\bar{S}(z,y) := \min_x \left\{ S(x,y) + \frac{1}{2}\|x - z\|^2 \right\}$$

is defined from $S(x,y)$ that maintains the same saddle points [27]. Then the following regularized saddle flow dynamics

$$\dot{z} = -\nabla_z \bar{S}(z,y) , \tag{11a}$$
$$\dot{y} = +\nabla_y \bar{S}(z,y) , \tag{11b}$$

suffice to locate a saddle point. We formalize the connection of this method with our observable certificate as follows.

Proposition 4 (Proximal Regularization)**.** *Let $S(x,y)$ be a Lagrangian function for some constrained convex program and Assumption 1 hold. Then the function*

$$h(z,y) := \begin{bmatrix} \bar{S}(z_\star, y_\star) - \bar{S}(z_\star, y) \\ \frac{1}{2}\|\bar{x}(z, y_\star) - z\|^2 \end{bmatrix} , \tag{12}$$

with $\bar{x}(z, y_\star) := \arg\min_x \left\{S(x, y_\star) + \frac{1}{2}\|x - z\|^2\right\}$ and $(z_\star, y_\star)$ being an arbitrary saddle point of $\bar{S}(z,y)$, is an observable certificate of $\bar{S}(z,y)$.

Details of the proof are omitted here and readers are referred to [27] for more insights. We remark that the identification of this observable certificate (12) does not significantly alleviate the analysis overheads since the complementary equilibrium properties of proximal regularization on the Lagrangian $S(x,y)$ are still crucial to validating the observable certificate (12) and establishing convergence.

Anyhow, the existence of an observable certificate satisfies Assumption 3 for $\bar{S}(z,y)$ and thus asymptotic convergence of the saddle flow dynamics (11) follows immediately from Theorem 1.

Corollary 5. *Let $S(x,y)$ be a Lagrangian function for some constrained convex program and Assumption 1 hold. Then the regularized saddle flow dynamics* (11) *asymptotically converge to some saddle point $(z_\star, y_\star)$ of $\bar{S}(z,y)$, with $(x_\star = z_\star, y_\star)$ being a saddle point of $S(x,y)$.*

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

B. Dissipative Saddle Flow Dynamics

We further design a novel separable regularization method that exploits our observable certificate and only requires Assumption 1 for a regularized version of 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 without altering the positions of the original saddle points. In particular, we propose a regularized surrogate for $S(x,y)$ via the following augmentation

$$S(x,z,y,w) := \frac{\rho}{2}\|x-z\|^2 + S(x,y) - \frac{\rho}{2}\|y-w\|^2, \quad (13)$$

where $z \in \mathbb{R}^n$ and $w \in \mathbb{R}^m$ serve as two new sets of virtual variables and $\rho > 0$ is a constant regularization coefficient. It is straightforward to verify the fixed positions of saddle points between $S(x,y)$ and $S(x,z,y,w)$ with virtual variables aligned with original variables.

Lemma 6 (Saddle Point Invariance)**.** *Let Assumption 1 hold. Then a point $(x_\star, y_\star)$ is a saddle point of $S(x,y)$ if and only if $(x_\star, z_\star, y_\star, w_\star)$ is a saddle point of $S(x,z,y,w)$, with*

$$x_\star = z_\star \quad \text{and} \quad y_\star = w_\star\,. \quad (14)$$

Proof. Recall the saddle property (1) of a saddle point, this theorem follows immediately from

$$\begin{aligned}
&S(x_\star,z_\star,y,w) \le S(x_\star,x_\star,y_\star,y_\star) \le S(x,z,y_\star,w_\star)\\
\iff& S(x_\star,z_\star,y,w) \le S(x_\star,y_\star) \le S(x,z,y_\star,w_\star)\\
\iff& S(x_\star,y) - \frac{\rho}{2}\|y-w\|^2 \le S(x_\star,y_\star) \le S(x,y_\star) + \frac{\rho}{2}\|x-z\|^2\\
\iff& S(x_\star,y) \le S(x_\star,y_\star) \le S(x,y_\star)\,,
\end{aligned}$$

where the first and second steps build upon the definition (13) of $S(x,z,y,w)$, and the third step uses norm non-negativity. ■

The regularized function $S(x,z,y,w)$ is convex in (x,z), concave in (y,w), and continuously differentiable with at least one saddle point, by its definition in (13) and Lemma 6. Therefore, Assumption 1 also holds for $S(x,z,y,w)$. Lemma 6 ensures that whenever we locate a saddle point of $S(x,z,y,w)$, a saddle point of $S(x,y)$ satisfying (1) is attained simultaneously. This motivates us to instead look at the saddle flow dynamics of $S(x,z,y,w)$.

Following (2), this regularized version of saddle flow dynamics are given by

$$\begin{aligned}
\dot{x} &= -\nabla_x S(x,y) - \rho(x-z)\,, && (15a)\\
\dot{z} &= \rho(x-z)\,, && (15b)\\
\dot{y} &= +\nabla_y S(x,y) - \rho(y-w)\,, && (15c)\\
\dot{w} &= \rho(y-w)\,. && (15d)
\end{aligned}$$

Although this dynamic law has twice as many state variables as its prototype (2), it is important to notice that, unlike the proximal gradient algorithm [23], [27], [39] and the equality constrained regularization [8], [26], (15) still preserves the same distributed structure that (2) may have. As a result, it can be implemented in a fully distributed fashion. The distributed property plays a significant role in solving network optimization problems with applications to power systems, wireless systems and bargaining problems.

We are now ready to provide the key result that the regularized saddle flow dynamics (15) asymptotically reach a saddle point as long as the minimal convexity-concavity holds for $S(x,y)$.

Proposition 7 (Separable Regularization)**.** *Let Assumption 1 hold. Then the function*

$$h(x,z,y,w) := \begin{bmatrix} \frac{\rho}{2}\|y-w\|^2 \\ \frac{\rho}{2}\|x-z\|^2 \end{bmatrix} \quad (16)$$

is an observable certificate of $S(x,z,y,w)$.

Proof. The above observable certificate $h(x,z,y,w)$ satisfies (4) in light of the following calculation:

$$\begin{aligned}
&\begin{bmatrix} S(x_\star,z_\star,y_\star,w_\star) - S(x_\star,z_\star,y,w) \\ S(x,z,y_\star,w_\star) - S(x_\star,z_\star,y_\star,w_\star) \end{bmatrix}\\
\ge& \begin{bmatrix} \underbrace{S(x_\star,y_\star) - S(x_\star,y)}_{\ge 0} + \frac{\rho}{2}\|y-w\|^2 \\ \underbrace{S(x,y_\star) - S(x_\star,y_\star)}_{\ge 0} + \frac{\rho}{2}\|x-z\|^2 \end{bmatrix}\\
\ge& \begin{bmatrix} \frac{\rho}{2}\|y-w\|^2 \\ \frac{\rho}{2}\|x-z\|^2 \end{bmatrix}\\
\ge& \; 0\,.
\end{aligned}$$

The fact that $h(x,z,y,w) \equiv 0$ implies $x(t) \equiv z(t)$ and $y(t) \equiv w(t)$ enforces $\dot{z}, \dot{w} \equiv 0$ according to (15b), (15d), and then $\dot{x}, \dot{y} \equiv 0$ is simultaneously guaranteed. ■

Assumption 3 holds for the regularized function $S(x,z,y,w)$ and asymptotic convergence of the regularized saddle flow dynamics (15) follows immediately from Theorem 1.

Corollary 8. *Let Assumption 1 hold. Then the regularized saddle flow dynamics* (15) *asymptotically converge to some saddle point $(x_\star, z_\star, y_\star, w_\star)$ of $S(x,z,y,w)$, with $(x_\star, y_\star)$ being a saddle point of $S(x,y)$.*

Remark 3. *Proposition 7 indicates that only the convexity-concavity of $S(x,y)$ is required to asymptotically arrive at a saddle point through the regularized saddle flow dynamics* (15)*. This condition is significantly milder than most existing ones in the literature, and is in some sense* minimal*, as it includes bi-linear saddle functions as a special case. Unlike the aforementioned proximal regularization method in Section III-A, our separable regularization method applies to saddle flow dynamics of general convex-concave functions.*

IV. Exponential Convergence

We now study the conditions that render the saddle flow dynamics (2) globally exponentially stable. We first show how exponential stability appears as a direct consequence of strong convexity-strong concavity of the saddle function. We then use this insight to explain the exponential convergence of the proximal saddle flow, the extension of the proximal gradient algorithm [32], [41].

A. Strong Convex-Concave Saddle Flow Dynamics

Before moving towards the question of exponential stability, we rewrite the saddle flow dynamics (2) in a more compact form

$$\dot{z} = F(z) \quad (17)$$

with $z = (x, y)$ and

$$F(z) = \begin{bmatrix} -\nabla_x S(x,y) \\ \nabla_y S(x,y) \end{bmatrix} \tag{18}$$

We further introduce an assumption of absolute continuity on $F(z)$ as the minimal requirement for the follow-up proof.

Definition 5 (Absolute Continuity)**.** *Let I be an interval on $\mathbb{R}$. A function $f : I \to \mathbb{R}$ is absolute continues if for every $\varepsilon > 0$, there exists some δ such that whenever a finite sequence of pairwise* disjoint *sub-intervals $[x_k, y_k]$ of I satisfies $\sum_k |y_k - x_k| < \delta$ then $\sum_k |(f(y_k) - f(x_k)| < \epsilon$.*

Assumption 5. *The gradient of $S(x, y)$, i.e., $\nabla S(x, y) := (\nabla_x S(x, y), \nabla_y S(x, y))$, is absolutely continuous.*

Remark 4. *Assumption 5 is slightly weaker than Lipschitz continuity, which is the commonly used assumption in the study of global exponential stability [28], [31], [42].*

Assumption 5 basically enable

$$\frac{\partial}{\partial z}F(z) = \begin{bmatrix} -\frac{\partial^2}{\partial x^2}S(x,y) & -\frac{\partial^2}{\partial x \partial y}S(x,y) \\ \frac{\partial^2}{\partial y \partial x}S(x,y) & \frac{\partial^2}{\partial y^2}S(x,y) \end{bmatrix} \tag{19}$$

and

$$\frac{1}{2}\left(\frac{\partial}{\partial z}F(z) + \frac{\partial}{\partial z}F(z)^T\right) = \begin{bmatrix} -\frac{\partial^2}{\partial x^2}S(x,y) & 0 \\ 0 & \frac{\partial^2}{\partial y^2}S(x,y) \end{bmatrix} \tag{20}$$

We now show strong convexity-strong concavity of the saddle function are conducive to exponential convergence.

Assumption 6 (Strong Saddle)**.** *$S(x, y)$ is μ-strongly convex in x and q-strongly concave in y.*

Remark 5. *One consequence of the strong saddle assumption is that there is a unique saddle point $(x^\star, y^\star)$, and*

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

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

A straightforward consequence of Assumptions 5 and 6 gives the following exponential convergence of the saddle flow dynamics (2).

Theorem 9 (Exponential Convergence)**.** *Let Assumptions 5 and 6 hold. Then the saddle flow dynamics* (2) *are globally exponentially stable. More precisely,*

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

holds with rate

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

Proof. The proof of Theorem 9 features a reformulation of the Lie derivative of (5) based on the fundamental theorem of calculus for absolute continuous functions [43].

We consider again the Lyapunov function

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

Now taking the Lie derivative with respect to (17) gives

$$\begin{aligned} \dot{V}(z) &= (z - z_\star)^T F(z) \\ &= \frac{1}{2}\left((z - z_\star)^T F(z) + F(z)^T(z - z_\star)\right) \end{aligned} \tag{21}$$

Assumption 5 allows us to write $F(z)$ as

$$F(z) = \int_0^1 \frac{\partial}{\partial z}F(z(s))(z - z_\star)ds + \underbrace{F(z_\star)}_{=0} \tag{22}$$

with $z(s) = (z - z_\star)s + z_\star$, where we have used the fact $dz(s) = (z - z_\star)ds$.

Now substituting (22) into (21) gives

$$\begin{aligned} \dot{V}(z) &= (z - z_\star)^T \int_0^1 \frac{1}{2}\left(\frac{\partial}{\partial z}F(z) + \frac{\partial}{\partial z}F(z)^T\right) ds\ (z - z_\star) \\ &= (z - z_\star)^T \int_0^1 \begin{bmatrix} -\frac{\partial^2}{\partial x^2}S(z(s)) & 0 \\ 0 & \frac{\partial^2}{\partial y^2}S(z(s)) \end{bmatrix} ds\ (z - z_\star) \end{aligned} \tag{23}$$

We remark that up to this point all the steps of the calculation follow with an equal sign. That is, $\dot{V}(z)$ is exactly given by (23).

The rest of the proof follows from applying Assumption 6 to (23), i.e.,

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

Therefore, it follows the Comparison Lemma [10] that

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

■

Remark 6. *Contraction theory provides an insight of the exponential convergence. In terms of the induced matrix logarithmic norm $\mu(A) = \lim_{h \to 0^+} \frac{\|I + hA\| - 1}{h}$ and* (18)*, we show the contractivity of saddle flow dynamics:*

$$\begin{aligned} \mu_2(\mathrm{D}_z F(z)) &= \lambda_{\max}\left(\frac{\mathrm{D}_z F(z) + \mathrm{D}_z F(z)^T}{2}\right) \\ &= \lambda_{\max}\left(\begin{bmatrix} -\frac{\partial^2}{\partial x^2}S(x,y) & 0 \\ 0 & \frac{\partial^2}{\partial y^2}S(x,y) \end{bmatrix}\right) \\ &= -\min\{\mu, q\} < 0. \end{aligned} \tag{24}$$

$F(z)$ is infinitesimally contracting with rate $\min\{\mu, q\}$ which implies the exponential convergence to the saddle point.

B. Proximal Regularization for Saddle Flow Dynamics

In this subsection, we gonna to show how proximal methods enhance the concavity when having only one-sided strong convexity. Before moving on, the related assumption is given as follows.

Assumption 7. *The function $S(x, y)$ is μ-strongly convex and l-smooth over x, i.e., $lI \succeq \frac{\partial^2}{\partial x^2}S(x,y) \succeq \mu I$.*

Given a function $S(x,y)$, we introduce an extra proximal variable $u \in \mathbb{R}^n$ and impose a quadratic regularization on the function:

$$\hat{S}(x,u,y) = S(x,y) + \frac{\rho}{2}\|x-u\|^2 \tag{25}$$

where $\rho > 0$ is a constant. Notably, $(x_\star, y_\star)$ is a saddle point if and only if $(x_\star, u_\star = x_\star, y_\star)$ is an optimal solution to (25). Instead of directly applying the standard saddle flow dynamics to $\hat{S}(x,u,y)$, we reduce the modified saddle function by minimizing it over the variable x and attain a proximal saddle function

$$\tilde{S}(u,y) := \min_x \hat{S}(x,u,y). \tag{26}$$

Due to the strong convexity of $\hat{S}(x,u,y)$ in x, (26) can be elaborated as

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

where $x_\star(u,y)$ is the unique minimizer given (u,y) such that

$$\nabla_x S(x_\star(u,y),y) + \rho(x_\star(u,y) - u) = 0 \tag{28}$$

holds.

Although little attention has been paid to the proximal saddle flow dynamics, the proximal method is by far a commonly used method for primal-dual dynamics, especially handling composite optimization problems. In [30], the author has derived the proximal augmented Lagrangian that leads to exponentially convergent primal-dual dynamics. [31] has extended [30] to a discretized version by the explicit forward Euler discretization which remains the exponential convergence. For a set of problems satisfying a structural property, [44] has established the exponential convergence of primal-dual dynamics based on the proximal augmented Lagrangian.

Lemma 10. *$\tilde{S}(u,y)$ is convex in u, concave in y and continuously differentiable on $\mathbb{R}^n \times \mathbb{R}^m$ with gradients:*

$$\nabla_u \tilde{S}(u,y) = \rho u - \rho x_\star(u,y) \tag{29a}$$
$$\nabla_y \tilde{S}(u,y) = \nabla_y S(x_\star(u,y),y) \tag{29b}$$

Refer to [27][Theorem 2] for the proof of Lemma 10.

Theorem 11 (Saddle point characterization of $S(z,y)$)**.** *A point $(x_\star, y_\star)$ is a saddle point of $L(x,y)$ if and only if $(u_\star, y_\star)$ is a saddle point of $S(u,y)$ with $u_\star = x_\star$.*

Theorem 11 follows immediately from Lemma 10 as well as

$$\nabla_u \tilde{S}(u_\star, y_\star) = 0 \iff \nabla_x S(x_\star, y_\star) = 0,$$
$$\nabla_y \tilde{S}(u_\star, y_\star) = 0 \iff \nabla_y S(x_\star, y_\star) = 0,$$

where $x_\star(u_\star, y_\star) = x_\star$ and (28) have been applied. It basically allows us to focus on the proximal saddle flow dynamics of $S(u,y)$, i.e.,

$$\dot{u} = -\nabla_u \tilde{S}(u,y) = -(\rho u - \rho x_\star(u,y)), \tag{30a}$$
$$\dot{y} = +\nabla_y \tilde{S}(u,y) = \nabla_y S(x_\star(u,y),y), \tag{30b}$$

and we next propose a sufficient condition that guarantees its exponential convergence to a saddle point.

Assumption 8. *The Jacobian matrix $\frac{\partial^2}{\partial y \partial x} S(x_\star, y)$ is full row rank with $\sigma I \succeq \left[\frac{\partial^2}{\partial y \partial x} S(x_\star, y)\right]\left[\frac{\partial^2}{\partial x \partial y} S(x_\star, y)\right] \succeq \kappa I$, and locally Lipschitz row-wise.*

Theorem 12. *Let Assumptions 1, 7 and 8 hold. Given any $\rho > 0$, the proximal saddle flow dynamics* (30) *are globally exponentially stable. More precisely, given $z := (u,y)$,*

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

holds with rate

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

Verify assumptions.

Proof. We derive the second-order partial derivatives from (30) as

$$\frac{\partial^2}{\partial u^2}\tilde{S}(u,y) = \rho I - \rho \mathbf{J}^u_{x_\star} \tag{31a}$$

$$\frac{\partial^2}{\partial y^2}\tilde{S}(u,y) = \left[\frac{\partial^2}{\partial x \partial y} S(x_\star, y)\right]^T \mathbf{J}^y_{x_\star} + \frac{\partial^2}{\partial y^2} S(x_\star, y) \tag{31b}$$

where $\mathbf{J}^u_{x_\star}$ and $\mathbf{J}^y_{x_\star}$ are Jacobin matrices of $x_\star$ with respect to u and y. It follows from (28) and Assumption 8 that

$$\frac{\partial^2}{\partial x^2} S(x_\star, y)\mathbf{J}^u_{x_\star} + \rho \mathbf{J}^u_{x_\star} - \rho I = 0 \tag{32a}$$

$$\frac{\partial^2}{\partial x^2} S(x_\star, y)\mathbf{J}^y_{x_\star} + \left[\frac{\partial^2}{\partial y \partial x} S(x_\star, y)\right]^T + \rho \mathbf{J}^y_{x_\star} = 0. \tag{32b}$$

According to (32), we can easily derive that

$$\mathbf{J}^u_{x_\star} = \rho\left(\frac{\partial^2}{\partial x^2} S(x_\star, y) + \rho I\right)^{-1} \tag{33a}$$

$$\mathbf{J}^y_{x_\star} = -\left(\frac{\partial^2}{\partial x^2} S(x_\star, y) + \rho I\right)^{-1}\left[\frac{\partial^2}{\partial y \partial x} S(x_\star, y)\right]^T. \tag{33b}$$

Combining (31) and (33), the second-order partial derivatives can be illustrated as below:

$$\frac{\partial^2}{\partial u^2}\tilde{S}(u,y) = \rho[I - \rho(\frac{\partial^2}{\partial x^2} S(x_\star, y) + \rho I)^{-1}] \tag{34a}$$

$$\begin{aligned}\frac{\partial^2}{\partial y^2}\tilde{S}(u,y) = \frac{\partial^2}{\partial y^2} S(x_\star, y) &- \left[\frac{\partial^2}{\partial x \partial y} S(x_\star, y)\right]^T \\ &\cdot (\frac{\partial^2}{\partial x^2} S(x_\star, y) + \rho I)^{-1}\left[\frac{\partial^2}{\partial y \partial x} S(x_\star, y)\right]^T\end{aligned} \tag{34b}$$

Since $S(x,y)$ is convex-concave, we have $\frac{\partial^2}{\partial y^2}\tilde{S}(x_\star, y) \preceq 0$. Then, Assumption 8 further implies Assumption 6 by

$$\frac{\partial^2}{\partial u^2}\tilde{S} = \rho I - \rho(\frac{\partial^2}{\partial x^2} S(x_\star, y) + \rho I)^{-1} \succeq \frac{\mu\rho}{\mu+\rho} I \succ 0, \tag{35a}$$

$$
\begin{aligned}
\frac{\partial^2}{\partial y^2}\tilde{S} =&\frac{\partial^2}{\partial y^2}S - \left[\frac{\partial^2}{\partial x \partial y}S(x_\star, y)\right]^T \\
&\cdot (\frac{\partial^2}{\partial x^2}S(x_\star, y) + \rho I)^{-1}\left[\frac{\partial^2}{\partial y \partial x}S(x_\star, y)\right]^T \\
\preceq& -\left[\frac{\partial^2}{\partial x \partial y}S(x_\star, y)\right]^T(\frac{\partial^2}{\partial x^2}S(x_\star, y) + \rho I)^{-1} \\
&\cdot \left[\frac{\partial^2}{\partial y \partial x}S(x_\star, y)\right]^T \\
\preceq& -\frac{1}{l+\rho}\left[\frac{\partial^2}{\partial x \partial y}S(x_\star, y)\right]^T\left[\frac{\partial^2}{\partial y \partial x}S(x_\star, y)\right]^T \\
\preceq& -\frac{\kappa}{l+\rho}I \prec 0,
\end{aligned} \tag{35b}
$$

given $\rho > 0$. Therefore, Theorem 12 follows immediately from Theorem 9. ∎

Remark 7. *The second term in* (34b)*, an estimation of the interaction level between* x *and* y*, is positive definite. Therefore, we can adjust the concavity over* y *by choosing proper interaction. If the interaction term dominates the original property, then the concavity can be maintained and even enhanced by the dominating interaction term.*

Remark 8. *In terms of the convergence rate, there is a trade-off between the convexity and the concavity. The proximal method sacrifices some convexity over primal variables to attain strong concavity over dual variables.*

Corollary 13. *The fastest convergence rate can be attained at*

$$c_\star = \frac{2\mu\kappa}{\sqrt{(\mu l - \kappa)^2 + 4\mu^2\kappa} + \mu l + \kappa} < \mu \tag{36}$$

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

$$\frac{\mu\rho_\star}{\mu + \rho^*} = \frac{\kappa}{l + \rho^*}, \tag{37a}$$

The uniqueness of $\rho_\star > 0$ is an immediate result of the observation that for $\rho > 0$, the former bound starts from 0 and keeps increasing to approach m while the latter bound starts from a finite positive value and diminishes to 0, as ρ grows to infinity.

V. Projected Saddle Flow Dynamics

In this section we generalize the results in Section III to account for projections on the vector field of the saddle flow dynamics (2) that are commonly introduced in the case of solving inequality constrained optimization problems.

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

$$\dot{z} = \Pi_\mathcal{D}[z, F(z)] = \begin{bmatrix} \Pi_\mathcal{X}[x, -\nabla_x S(x, y)] \\ \Pi_\mathcal{Y}[y, +\nabla_y S(x, y)] \end{bmatrix}. \tag{38a}$$

Given $x \in \mathcal{X} \subseteq \mathbb{R}^n$ and $v \in \mathbb{R}^n$, the vector projection $\Pi_\mathcal{X}[x, v]$ of v at x with respect to $\mathcal{X}$ is defined as

$$\Pi_\mathcal{X}[x, v] = \lim_{\delta \to 0^+} \frac{\Psi_\mathcal{X}[x + \delta v]}{\delta}, \tag{39}$$

where $\Psi_\mathcal{X}[y] = \arg\min_{z \in \mathcal{X}} \|z - y\|$ denotes the point-wise projection in $\mathcal{X}$ to y. With this projection, $(x(t), y(t))$ is constrained to be in the feasible set $\mathcal{D}$ as long as it starts with a feasible initial point. Accordingly, we slightly modify Assumption 1 to guarantee the existence of such saddle points.

Assumption 9. $S(x, y)$ *is convex-concave, continuously differentiable, and there exists at least one saddle point* $(x_\star, y_\star \geq 0)$ *satisfying* (1).

In this context, saddle points are restrained to ones in the feasible set of (x, y). Therefore, any observable certificate of $S(x, y)$ will be defined on a saddle point $(x_\star, y_\star) \in \mathcal{D}$ in Definition 4. Next, we formally generalize the sufficiency of observable certificates developed in Section III-A.

A. Observable Certificates for Projected Flows

The generalization of Theorem 1 for asymptotic convergence of the projected saddle flow dynamics (41) to a saddle point of $S(x, y)$ is summarized as follows.

Theorem 14 (Sufficiency of Observable Certificates for Projected Flows)**.** *Let Assumptions 3 and 9 hold. Then the projected saddle flow dynamics* (38) *asymptotically converge to some saddle point* $(x_\star, y_\star) \in \mathcal{D}$ *of* $S(x, y)$*.*

The proof requires a lemma regarding the projection $\Pi_\mathcal{X}[x, v]$.

Lemma 15 (Lemma 7 [45])**.** *Given any closed convex set* $\mathcal{K} \subset \mathbb{R}^n$ *and* $a, b \in \mathcal{K}$*,* $v \in \mathbb{R}^n$*, the inner product*

$$\langle b - a, v - \Pi_\mathcal{K}[a, v]\rangle \leq 0$$

holds.

Using this lemma, the proof of Theorem 14 essentially follows from that of Theorem 1 as follows.

Proof. Consider the same quadratic Lyapunov function (5). Taking its Lie derivative along the trajectory $(x(t), y(t))$ of (41) yields

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

where the key step is to use Lemma 15 in the first inequality. The rest of the proof remains almost the same except that the largest invariant set is defined between the on-off switches of the projection. From above, $\dot{V}(x,y) \equiv 0$ additionally implies

$$x(t) \equiv x_\star \text{ or } \Pi_{\mathcal{X}}\left[x, -\nabla_x S(x,y)\right] \equiv -\nabla_x S(x,y),$$

and

$$y(t) \equiv y_\star \text{ or } \Pi_{\mathcal{Y}}\left[y, +\nabla_y S(x,y)\right] \equiv +\nabla_y S(x,y),$$

and an invariance principle for Caratheodory systems [46] can be applied to account for the discontinuities in the vector field due to the projection. ■

B. Exponential Convergence

We next further show how Theorem 9 generalizes to account for the exponential convergence of (38).

Theorem 16. *Let Assumptions 5, 6 and 9 hold. Then the projected saddle flow dynamics* (38) *are globally exponentially stable. More precisely, given $z := (x,y)$,*

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

holds with rate

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

Proof. Lemma 15 allows the proof of Theorem 9 to still apply here. In particular, consider again the quadratic Lyapunov function

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

Taking the Lie derivative with respect to time gives

$$\begin{aligned}
\dot{V}(z) &= (x - x_\star)^T \dot{x} + (y - y_\star)^T \dot{y} \\
&= (x - x_\star)^T \Pi_{\mathcal{X}}\left[x, -\nabla_x S(x,y)\right] \\
&\quad + (y - y_\star)^T \Pi_{\mathcal{Y}}\left[y, +\nabla_y S(x,y)\right] \\
&= (x_\star - x)^T \nabla_x S(x,y) - (y_\star - y)^T \nabla_y S(x,y) \\
&\quad + \underbrace{(x_\star - x)^T \left(-\nabla_x S(x,y) - \Pi_{\mathcal{X}}\left[x, -\nabla_x S(x,y)\right]\right)}_{\leq 0} \\
&\quad + \underbrace{(y_\star - y)^T \left(\nabla_y S(x,y) - \Pi_{\mathcal{Y}}\left[y, \nabla_y S(x,y)\right]\right)}_{\leq 0} \\
&\leq (z - z_\star)^T F(z) \\
&= \frac{1}{2}\left((z - z_\star)^T F(z) + F(z)^T (z - z_\star)\right)
\end{aligned}$$

where

$$F(z) := \begin{bmatrix} -\nabla_x S(x,y) \\ \nabla_y S(x,y) \end{bmatrix} \tag{40}$$

is defined following Theorem 9. The rest of proof is exactly the same. ■

Theorem 16 allows the generalization of the algorithms proposed for equality-constrained optimization problems to convex programs with inequality constraints. Next we explicitly demonstrate their applications.

VI. Applications to Constrained Convex Optimization

In this section, we apply proposed methods to constrained convex optimization problems. Inspired by the Lagrangian of inequality constrained convex problems, the function is convex over $x \in \mathbb{R}^n$ but meanwhile it is just linear with respect to $y \in \mathbb{R}^m_+$. Moreover, we look at a more specific projected version of primal-dual dynamics:

$$\dot{x} = -\nabla_x S(x,y), \tag{41a}$$

$$\dot{y} = \left[+\nabla_y S(x,y)\right]_y^+, \tag{41b}$$

where, without loss of generality, we define the element-wise projection $[\cdot]_y^+$ only on part of the vector field regarding y as

$$\left[\nabla_{y_i} S(x,y)\right]_{y_i}^+ := \begin{cases} \nabla_{y_i} S(x,y), & \text{if } y_i > 0, \\ \max\left\{\nabla_{y_i} S(x,y), 0\right\}, & \text{otherwise.} \end{cases} \tag{42}$$

With this projection, $y(t)$ is constrained to be non-negative as long as it starts with a non-negative initial point. In this context, saddle points are restrained to ones in the non-negative orthant of y. Therefore, any observable certificate of $S(x,y)$ will be defined on a saddle point $(x_\star, y_\star \geq 0)$ in Definition 4. Next we formally generalize the sufficiency of observable certificates developed in Section III-A.

Similarly, the generalization of Theorem 1 for asymptotic convergence of the projected primal-dual dynamics (41) to a saddle point of $S(x,y)$ is summarized as follows.

Theorem 17. *Let Assumptions 3 and 9 hold. Then the projected saddle flow dynamics* (41) *asymptotically converge to some saddle point $(x_\star, y_\star \geq 0)$ of $S(x,y)$.*

The proof requires a lemma regarding the projection $[\cdot]_y^+$.

Lemma 18. *Given any arbitrary $y, y_\star \in \mathbb{R}^m_{\geq 0}$ and $\nu \in \mathbb{R}^m$,*

$$(y - y_\star)^T \left([\nu]_y^+ - \nu\right) \leq 0$$

holds.

We next further show how Theorem 9 generalizes to account for the exponential convergence of (41).

Theorem 19. *Let Assumptions 9, 5 and 6 hold. Then the projected saddle flow dynamics* (41) *are globally exponentially stable. More precisely, given $z := (x,y)$,*

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

holds with rate

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

A. Distributed Solution to Linear Program

Theorem 17 enables the separable regularization method in Section III-B to apply to projected saddle flow dynamics as well since we can still identify the same observable certificate

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

to satisfy Assumption 3. One of its straightforward applications involves solving inequality constrained linear programs in a distributed fashion with guaranteed asymptotic convergence to an optimal solution.

Consider the following problem:

$$\min_{x\in\mathbb{R}^n} \quad c^T x \tag{43a}$$
$$\text{s.t.} \quad Ax-b\leq 0 \ : \ y\in\mathbb{R}^m_{\geq 0} \tag{43b}$$

which corresponds to a bi-linear Lagrangian

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

We introduce virtual variables $z\in\mathbb{R}^n$, $w\in\mathbb{R}^m$ and a constant $\rho>0$ to define

$$S(x,z,y,w):=\frac{\rho}{2}\|x-z\|^2+c^Tx+y^T(Ax-b)-\frac{\rho}{2}\|y-w\|^2$$

to be its augmented Lagrangian. Lemma 6 implies that $(x_\star, y_\star\geq 0)$ is a saddle point of $S(x,y)$, i.e., one optimal solution to (43), if and only if $(x_\star, z_\star=x_\star, y_\star\geq 0, w_\star=y_\star)$ is a saddle point of $S(x,z,y,w)$.

Then an algorithm to optimally solve a linear program of the form (43) follows immediately from asymptotic convergence of the following projected and regularized saddle flow dynamics:

$$\dot{x}=-c-A^Ty-\rho(x-z)\,, \tag{44a}$$
$$\dot{z}=\rho(x-z)\,, \tag{44b}$$
$$\dot{y}=[Ax-b-\rho(y-w)]^+_y\ , \tag{44c}$$
$$\dot{w}=\rho(y-w)\,, \tag{44d}$$

which maintains the distributed structure where each agent $i=1,2,\ldots,n$ may locally manage

$$\dot{x}_i=-c_i-A_i^Ty-\rho(x_i-z_i)\,, \tag{45a}$$
$$\dot{z}_i=\rho(x_i-z_i)\,, \tag{45b}$$

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

$$\dot{y}_j=[A_jx-b_j-\rho(y_j-w_j)]^+_{y_j}\ , \tag{45c}$$
$$\dot{w}_j=\rho(y_j-w_j)\,, \tag{45d}$$

with A_i and A_j being the i^{th} column and the j^{th} row of A, respectively.

B. Proximal Primal-Dual Dynamics of Inequality Constrained Convex Programming

The projected version of proximal saddle flow dynamics in Section IV-B can be applied to handle a convex program with convex inequality constraints:

$$\min_{x\in\mathbb{R}^n} \quad f(x) \tag{46a}$$
$$\text{s.t.} \quad g(x)\leq 0: y\in\mathbb{R}^m_{\geq 0} \tag{46b}$$

where $f(x):\mathbb{R}^n\mapsto\mathbb{R}$ is continuously differentiable and convex while $g(x):\mathbb{R}^n\mapsto\mathbb{R}^m$ consists of locally Lipschitz and convex functions.

Similarly an auxiliary proximal variable $u\in\mathbb{R}^n$ is introduced to formulate the following regularized problem:

$$\min_{x,u\in\mathbb{R}^n} \quad f(x)+\frac{\rho}{2}\|x-u\|^2 \tag{47a}$$
$$\text{s.t.} \quad g(x)\geq 0: y\in\mathbb{R}^m_{\geq 0} \tag{47b}$$

with ρ being a constant and its Lagrangian as

$$\hat{L}(x,u,y):=f(x)+\frac{\rho}{2}\|x-u\|^2+y^Tg(x). \tag{48}$$

By further minimizing $\hat{L}(x,u,y)$ over the original primal variable x, we arrive at a proximal saddle function

$$\begin{aligned}S(u,y):=&\min_x \hat{L}(x,u,y)\\ =&f(x_\star(u,y))+\frac{\rho}{2}\|x_\star(u,y)-u\|^2+y^Tg(x_\star(u,y)),\end{aligned} \tag{49}$$

where $x_\star(u,y)$ is the unique minimizer, or the unique solution to the following equations

$$F(x,u,y)=0, \tag{50}$$

with $F(x,u,y)$ being an implicit function defined as

$$F(x,u,y):=\nabla f(x)+\rho(x-u)+\mathbf{J}_g^T(x)y \tag{51}$$

Denote F_x, F_u, F_y as the partial derivatives of $F(x,u,y)$ with respect to x,u,y, respectively.

Following [27][Theorem 2], the convexity-concavity and continuous differentiability of $S(u,y)$ still apply. Moreover, if the following assumption holds, the strong convexity-concavity is guaranteed for $S(u,y)$.

Assumption 10. *The function $f(x)$ is μ-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 Jacobian matrix $\mathbf{J}_g(x)$ is full row rank with $\sigma I\succeq\mathbf{J}_g(x)\mathbf{J}_g(x)^T\succeq\kappa I$, and locally Lipschitz row-wise.*

Remark 9. *$\mathbf{J}_g(x)$ is full row rank implies that LICQ is satisfied. Therefore, the uniqueness of the saddle point is guaranteed.*

Given Assumption 10, we can derive the partial derivatives of $x_\star(u,y)$ as

$$\frac{\partial x}{\partial u}=-F_x^{-1}F_u=\rho\left(\nabla^2 f(x)+\rho+\sum_{j=1}^m y_j\nabla^2 g_j(x)\right)^{-1}, \tag{52a}$$

$$\frac{\partial x}{\partial y}=-F_x^{-1}F_y=-\left(\nabla^2 f(x)+\rho+\sum_{j=1}^m y_j\nabla^2 g_j(x)\right)^{-1}\mathbf{J}_g^T, \tag{52b}$$

which immediately imply the second-order partial derivatives of $S(u,y)$

$$\frac{\partial^2}{\partial u^2}S=\rho I-\rho\frac{\partial x}{\partial u}\succeq\frac{\mu\rho}{\mu+\rho}I\succ 0, \tag{53a}$$

$$\frac{\partial^2}{\partial y^2}S=\mathbf{J}_g\frac{\partial x}{\partial y}\preceq-\frac{1}{l+\rho+\sum_{j=1}^m y_j\nabla^2 g_j(x)}\mathbf{J}_g\mathbf{J}_g^T\preceq 0, \tag{53b}$$

where we have used the fact of $y_j\geq 0$ and $\nabla^2 g_j(x)\succeq 0$, $\forall j=1,2,\ldots,m$. Recall (23), the above second-order partial derivatives of $S(u,y)$ suffice to guarantee that the standard quadratic Lyapunov function

$$V(u,y):=\frac{1}{2}\|u-u_\star\|^2+\frac{1}{2}\|y-y_\star\|^2 \tag{54}$$

is non-increasing along any trajectories governed by the projected proximal saddle flow dynamics of $S(u, y)$

$$\dot{u} = -\nabla_u S(u, y)\,, \tag{55}$$
$$\dot{y} = [+\nabla_y S(u, y)]_y^+\ . \tag{56}$$

As a result, given an arbitrary initial point (u_0, y_0), the trajectories of (55) are bounded and contained in an invariant domain

$$D_0(u_0, y_0) := \{(u, y) | V(u, y) \leq V(u_0, y_0)\}\ , \tag{57}$$

which is compact and convex since $V(u, y)$ is radially unbounded. Therefore, for any arbitrary point $(u, y) \in D_0(u_0, y_0)$, suppose $\|y\| \leq \bar{y}$ and $\nabla^2 g_j(x) \preceq \gamma I$, $\forall j = 1, 2, \ldots, m$, hold. It implies that along the trajectories starting from (u_0, y_0), which are contained in $D_0(u_0, y_0)$,

$$\frac{\partial^2}{\partial y^2} S \preceq -\frac{\kappa}{l + \rho + m\bar{y}\gamma} I \prec 0 \tag{58}$$

always holds.

According to Theorem 12, the projected proximal saddle flow dynamics of $S(u, y)$, explicitly characterized by

$$\dot{u} \quad = \rho u - \rho x_\star(u, y)\,, \tag{59a}$$
$$\dot{y} \quad = [g(x_\star(u, y))]_y^+\ , \tag{59b}$$

are semi-globally exponentially stable. More precisely, given $z := (u, y)$ and an arbitrary initial point $z_0 := z(0)$,

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

holds with rate

$$c := \min\left\{\frac{\mu\rho}{\mu + \rho}, \frac{\kappa}{l + \rho + m\bar{y}(z_0)\gamma(z_0)}\right\} > 0. \tag{60}$$

Moreover, a point $(u_\star, y_\star)$ is a saddle point of $S(u, y)$ if and only if $(x_\star, y_\star)$ is an optimal primal-dual solution to the original convex program (46) with $x_\star = u_\star$.

Remark 10. *The bound provided by* (60) *on exponential convergence rate is determined by the initial point z_0. Therefore, we can only show semi-global exponential stability of dynamics* (59) *and there is no guarantee for a universal exponential convergence rate.*

C. Regularized Primal-Dual Dynamics

Explain the purpose of ((63)) We have introduced the strong convex-concave condition that directly results in exponential stability in Section IV. Nevertheless, numerous scenarios exist where strongly convex-strongly concave functions are not present. In the context of constrained convex optimization problems, the Lagrangian exhibits bilinearity, and strong concavity cannot be assured for dual variables. To address this challenge, we proceed to employ a change of variables strategy to harness the aforementioned concavity. As a consequence, the resulting novel algorithm only requires milder conditions of objective functions to achieve exponential convergence.

Consider the convex program with affine inequality constraints.

$$\min_{x \in \mathbb{R}^n} \qquad f(x) \tag{61a}$$
$$\text{s.t.} \qquad Ax - b \leq 0 : y \in \mathbb{R}^m_{\geq 0} \tag{61b}$$

Define its (adjusted) Lagrangian as

$$L(x, y) := f(x) + \eta y^T (Ax - b) \tag{62}$$

where $\eta > 0$ is a constant. The following change of variables

$$u := x + \alpha A^T y \tag{63}$$

transforms $L(x, y)$ equivalently into

$$S(u, y) := f(u - \alpha A^T y) + \eta y^T (Au - b) - \eta\alpha \|A^T y\|^2, \tag{64}$$

where $\alpha > 0$ is also a constant.

Assumption 11. *The function $f(x)$ is μ-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 constraint matrix A is full row rank with $\sigma I \succeq AA^T \succeq \kappa I$, and locally Lipschitz row-wise.*

Given the structure of $S(u, y)$, the following theorem characterizes a sufficient condition that guarantees it to be a strong saddle function.

Theorem 20. *Let Assumption 11 hold. Given any $\eta, \alpha > 0$ that satisfy $2\eta > l\alpha$, $S(u, y)$ is μ-strongly convex in u and $(2\eta\alpha - l\alpha^2)\kappa$-strongly concave in y.*

Proof. Assumption 11 allows a straightforward calculation of the second-order partial derivatives

$$\frac{\partial^2}{\partial u^2} S = \nabla^2 f(u - \alpha A^T y) \succeq \mu I \succ 0\,, \tag{65a}$$

$$\begin{aligned}\frac{\partial^2}{\partial y^2} S &= \alpha^2 A \nabla^2 f(u - \alpha A^T y) A^T - 2\eta\alpha AA^T \\ &= A\left(\alpha^2 \nabla^2 f(u - \alpha A^T y) - 2\eta\alpha I\right) A^T \\ &\preceq A(l\alpha^2 - 2\eta\alpha) I A^T \\ &\preceq -\left(2\eta\alpha - l\alpha^2\right)\kappa I \\ &\prec 0\,,\end{aligned} \tag{65b}$$

which justify the strong convexity-strong concavity of $S(u, y)$. ∎

Corollary 21. *Suppose $S(u, y)$ has at least one saddle point $(u_\star, y_\star)$, then its projected saddle flow dynamics*

$$\begin{aligned}\dot{u} &= -\nabla_u S(u, y) \\ &= -\nabla f(u - \alpha A^T y) - \eta A^T y\end{aligned} \tag{66a}$$

$$\begin{aligned}\dot{y} &= [+\nabla_y S(u, y)]_y^+ \\ &= [-\alpha A \nabla f(u - \alpha A^T y) + \eta(Au - b) - 2\eta\alpha AA^T y]_y^+\end{aligned} \tag{66b}$$

are globally exponentially stable. More precisely, given $v := (u, y)$,

$$\|v(t) - v_\star\| \leq \|v(0) - v_\star\| e^{-ct}$$

holds with rate

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

Given the fact that the gradient of $S(u, y)$, predicated on $\nabla f(\cdot)$, is Lipschitz continuous and satisfies Assumption 5, the corollary follows from Theorem 19.

Moreover, the saddle points of $S(u, y)$ are related to those of the original Lagrangian $L(x, y)$ in (62) by

Theorem 22 (Saddle point characterization of $S(u, y)$)**.** *$(u_\star, y_\star)$ is a saddle point of $S(u, y)$ if and only if $(x_\star, y_\star)$ is a saddle point of $S(x, y)$ with $u_\star = x_\star + \alpha A^T y_\star$.*

Theorem 22 is an immediate result of

$$\nabla_x L(x_\star, y_\star) = 0 \iff \nabla_u S(u_\star, y_\star) = 0$$
$$\nabla_y L(x_\star, y_\star) = 0 \iff \nabla_y S(u_\star, y_\star) = 0$$

along with the convexity of $L(x, y)$ and $S(u, y)$ in x and u, respectively, and their concavity in y.

Theorem 22 inspires a novel algorithm of the regularized primal-dual dynamics that builds on the projected saddle flow dynamics of the strong saddle function $S(u, y)$ to solve for an optimal solution to the original convex program (61), i.e., converge to one saddle point of $L(x, y)$. In particular, we propose

$$\begin{aligned}\dot{x} =& -\left(\nabla f(x) + \eta A^T y\right) \\ &- \alpha A^T [-\alpha A \left(\nabla f(x) + \eta A^T y\right) + \eta(Ax - b)]_y^+\end{aligned} \tag{67a}$$

$$\dot{y} = [-\alpha A \left(\nabla f(x) + \eta A^T y\right) + \eta(Ax - b)]_y^+ \tag{67b}$$

as the regularized primal dual dynamics by defining

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

which essentially enforces

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

It implies that the trajectory of $x(t)$ accompanies those of $u(t)$ and $y(t)$, and is eventually driven to $x_\star$.

We next formally state the exponential convergence of the above regularized primal-dual dynamics (67) to one saddle point of the Lagrangian $L(x, y)$.

Theorem 23. *Let Assumptions 1 (in terms of $L(x, y)$) and 11 hold. Given any $\eta, \alpha > 0$ that satisfy $2\eta > l\alpha$, the regularized primal-dual dynamics (67) are globally exponentially stable. More precisely, given $z := (x, y)$,*

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

where $\phi := \max\{2, 2\sigma\alpha^2 + 1\}$, holds with rate

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

Proof. The proof follows Corollary 21 with the following inequalities

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

and

$$\begin{aligned}&\|u(t) - u_\star\|^2 \\ &= \|(x(t) - x_\star) + \alpha A^T (y(t) - y_\star)\|^2 \\ &\le 2\|x(t) - x_\star\|^2 + 2\alpha^2 \|A^T (y(t) - y_\star)\|^2 \\ &\le 2\|x(t) - x_\star\|^2 + 2\sigma\alpha^2 \|y(t) - y_\star\|^2\,,\end{aligned}$$

which imply

$$\begin{aligned}&\|z(t) - z_\star\| \\ &\le \sqrt{2\|u(t) - u_\star\|^2 + (2\sigma\alpha^2 + 1)\|y(t) - y_\star\|^2} \\ &\le \sqrt{\phi}\|v(t) - v_\star\| \\ &\le \sqrt{\phi}\|v(0) - v_\star\| e^{-ct} \\ &\le \sqrt{\phi}\sqrt{2\|x(0) - x_\star\|^2 + (2\sigma\alpha^2 + 1)\|y(0) - y_\star\|^2} e^{-ct} \\ &\le \phi \|z(0) - z_\star\| e^{-ct}\,.\end{aligned}$$

■

Remark 11. *As a matter of fact, we can always pick $\eta, \alpha > 0$ that satisfy $2\eta > l\alpha + \frac{\mu}{\kappa\alpha}$ to guarantee $\left(2\eta\alpha - l\alpha^2\right)\kappa > \mu$ such that*

$$c = \min\left\{\mu, (2\eta\alpha - l\alpha^2)\kappa\right\} = \mu \tag{68}$$

holds.

Remark 12. *In this context, the original Lagrangian exhibits m-strong convexity over x, yet it only demonstrates linearity concerning y. By implementing a coordinate transformation $u = \Phi z$, as described in (63) within this paper, we establish exponential convergence (strict contraction) of the primal-dual dynamics by strong convexity-strong concavity in the new space, primarily utilizing the widely used 2-norm metric. Upon returning to the original space, the resulting novel algorithm undergoes strict contraction in a specific metric $\mu(\Phi^{-T} Q \Phi^{-1})$, which is induced by the coordinate transformation. Here, Q is constructed from the Hessian matrix, the constraint matrix and several parameters.*

D. Reduced Primal-Dual Dynamics

We discuss an alternative to the above proximal regularization for a specific structure of the standard equality-constrained convex program where $f(x)$ is additively separable in terms of $x_s \in \mathbb{R}^{n_s}$ and $x_c \in \mathbb{R}^{n_c}$ with $x =: [x_s^T, x_c^T]^T \in \mathbb{R}^n$ and $n = n_s + n_c$. Under this circumstance, we rewrite the problem as

$$\min_x \quad f(x) := f_s(x_s) + f_c(x_c) \tag{69a}$$
$$\text{s.t.} \quad A_s x_s + A_c x_c - b = 0 : y \tag{69b}$$

with $y \in \mathbb{R}^m$ being the dual variable. Both $f_s(x) : \mathbb{R}^{n_s} \mapsto \mathbb{R}$ and $f_c(x) : \mathbb{R}^{n_c} \mapsto \mathbb{R}$ are assumed to be second-order continuously differentiable and strictly convex. A_s and A_c are submatrices of A that consist of the columns corresponding to x_s and x_c, respectively. Instead of introducing regularization, we use the standard Lagrangian

$$L(x, y) := f_s(x_s) + f_c(x_c) + y^T (A_s x_s + A_c x_c - b), \tag{70}$$

but minimize it over x_s to attain a reduced Lagrangian

$$\begin{aligned}S(x_c, y) &:= \min_{x_s} L(x, y) \\ &= f_s(x_{s*}(y)) + f_c(x_c) + y^T (A_s x_{s*}(y) + A_c x_c - b)\end{aligned} \tag{71}$$

with $x_{s*}(y)$ being the unique minimizer given y such that

$$\nabla_{x_s} f_s(x_{s*}(y)) + A_s^T y = 0 \tag{72}$$

holds.

Lemma 24. *$S(x_c, y)$ is convex in x_c, concave in y and second-order continuously differentiable on $\mathbb{R}^{n_c} \times \mathbb{R}^m$ with gradients:*

$$\nabla_{x_c} S(x_c, y) = \nabla_{x_c} f_c(x_c) + A_c^T y \tag{73a}$$
$$\nabla_y S(x_c, y) = A_s x_{s*}(y) + A_c x_c - b \tag{73b}$$

Proof. It follows from (72) and the strict convexity of $f_s(\cdot)$ that

$$x_{s*}(y) = (\nabla_{x_s} f_s(\cdot))^{-1}(-A_s^T y) =: g(-A_s^T y) \tag{74}$$

is continuously differentiable. Therefore, the gradient of $S(x_c, y)$ can be obtained using the chain rule:

$$\nabla_{x_c} S(x_c, y) = \nabla_{x_c} f_c(x_c) + A_c^T y \tag{75a}$$
$$\begin{aligned}\nabla_y S(x_c, y) &= -A_s \mathbf{J}_g^T \nabla_{x_s} f_s(x_{s*}(y)) - A_s \mathbf{J}_g^T A_s^T y \\ &\quad + A_s x_{s*}(y) + A_c x_c - b \\ &= A_s x_{s*}(y) + A_c x_c - b\end{aligned} \tag{75b}$$

which allows us to further take the second-order partial derivative as

$$\frac{\partial^2}{\partial {x_c}^2} S(x_c, y) = \nabla_{x_c}^2 f_c(x_c) \succ 0 \tag{76a}$$
$$\frac{\partial^2}{\partial y^2} S(x_c, y) = -A_s \mathbf{J}_g^T A_s^T \preceq 0 \tag{76b}$$

where $\nabla_{x_s}^2 f_s(x_s) \succ 0$ and $\nabla_{x_c}^2 f_c(x_c) \succ 0$ follow from their strict convexity and $\mathbf{J}_g^T \succ 0$ follows from $\nabla_{x_s}^2 f_s(x_s) \cdot \mathbf{J}_g^T \equiv I$. ■

Theorem 25 (Saddle point characterization of $S(x_c, y)$)**.** *A point $(x_\star, y_\star)$ is a saddle point of $L(x, y)$ if and only if $(x_{c*}, y_\star)$ is a saddle point of $S(x_c, y)$ with $[x_{s*}^T(y_\star), x_{c*}^T]^T = x_\star$.*

Similarly, Theorem 25 basically follows from Lemma 24 as well as

$$\nabla_{x_c} S(x_{c*}, y_\star) = 0 \iff \nabla_x L(x_\star, y_\star) = 0,$$
$$\nabla_y S(x_{c*}, y_\star) = 0 \iff \nabla_y L(x_\star, y_\star) = 0.$$

Therefore, we primarily concentrate on the reduced primal-dual dynamics of $S(x_c, y)$, i.e.,

$$\dot{x}_c = -\nabla_{x_c} S(x_c, y), \tag{77a}$$
$$\dot{y} = +\nabla_y S(x_c, y). \tag{77b}$$

Assumption 12. *The function $f_s(x)$ is m_s-strongly convex with l_s-Lipschitz gradient, i.e., $l_s I \succeq \nabla^2 f_s(x) \succeq m_s I$ whenever $\nabla^2 f_s(x)$ is defined. Similarly, $l_c I \succeq \nabla^2 f_c(x) \succeq m_c I$ is assumed. The matrix A_s is full row rank with $\sigma_s I \succeq A_s A_s^T \succeq \kappa_s I$, where $\sigma_s := \lambda_{\max}(A_s A_s^T)$ and $\kappa_s := \lambda_{\min}(A_s A_s^T)$ are the largest and smallest eigenvalues of $A_s A_s^T$, respectively.*

Note that Assumption 12 implicitly assumes $n_s \geq m$.

Theorem 26. *Let Assumptions 1 and 12 hold. The reduced primal-dual dynamics* (77) *are globally exponentially stable. More precisely, given $z := (x_c, y)$,*

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

holds with rate

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

Remark 13. *The key enabler for the reduced primal-dual dynamics is the separation of x into x_s and x_c such that all the assumptions hold. It is possible that multiple ways of separation exist, then optimization over separation yields the fastest convergence rate. In this case, contrasting the algorithm with the more general proximal saddle flow dynamics might also suggest a better option between the two since neither of them fully dominates the other in terms of the convergence rate, considering $\kappa_s \leq \kappa$ and $l_s \leq l$.*

VII. Simulation Results

A. Network Flow Optimization

We first illustrate the asymptotic convergence of the distributed algorithm with projections for linear programs, as guaranteed by our observable certificate, through a network flow problem. The goal is to minimize the material transportation cost from suppliers to consumers, while simultaneously adhering to the capacity constraints imposed on each link in the network.

The network $\mathcal{G}(V, \mathcal{E})$ is shown in Figure (1a) with one supplier in red and two consumers in blue. Within this context, V represents the set of nodes while $\mathcal{E}$ is the set of edges. The variable x_{ij} is employed to represent the quantity of transportation from node i to node j. Given the consideration of all relevant node constraints and capacity limitations, the minimum flow cost problem is formulated as follows.

$$\begin{aligned}\min \quad & 4x_{12} + 4x_{13} + 2x_{23} + 2x_{24} + 6x_{25} + x_{34} \\ & +3x_{35} + 3x_{45} + x_{53} \\ \text{s.t.} \quad & x_{12} + x_{13} = 20 \\ & -x_{12} + x_{23} + x_{24} + x_{25} = 0 \\ & -x_{13} - x_{23} + x_{34} + x_{35} - x_{53} = 0 \\ & -x_{24} - x_{34} + x_{45} = -5 \\ & -x_{25} - x_{35} - x_{45} + x_{53} = -15 \\ & x_{12} \leq 15, x_{13} \leq 8, x_{24} \leq 4 \\ & x_{25} \leq 10, x_{34} \leq 15, x_{35} \leq 5, x_{53} \leq 4 \\ & x_{ij} \geq 0, \text{for all } (i,j) \in \mathcal{E}\end{aligned} \tag{78}$$

The equivalent compact form of this problem is

$$\begin{aligned}\min \quad & c^T x \\ \text{s.t.} \quad & Bx = d: \quad y \in \mathbb{R}^m \\ & Ax \leq b: \quad z \in \mathbb{R}_+^p\end{aligned} \tag{79}$$

where $m = 5$ and $p = 16$. We utilize a variant of the algorithm (44) as below

$$\begin{aligned}\dot{x} &= -c - B^T y - A^T z - \rho(x - \bar{x}), & \dot{\bar{x}} &= \rho(x - \bar{x}), \\ \dot{y} &= Bx - d - \rho(y - \bar{y}), & \dot{\bar{y}} &= \rho(y - \bar{y}), \\ \dot{z} &= [Ax - b - \rho(z - \bar{z})]_z^+, & \dot{\bar{z}} &= \rho(z - \bar{z}).\end{aligned} \tag{80}$$

By setting the regularization coefficient $\rho = 0.05$, results in Figure (1b) have shown that all primary variables asymptotically converge to the optimal solution of the minimum flow cost problem.

B. Lasso Regression Problem

We use a classical lasso regression problem to test the performance of algorithms and verify theoretical results. Consider the following problem

$$\min_{x\in\mathbb{R}^n} f(x) + \lambda\|x\|_1, \tag{81}$$

where $f(x)$ is convex. We first handle the nonsmooth term $\|x\|_1$, denoting

$$x_i = z_i^+ - z_i^-, \tag{82a}$$
$$|x_i| = z_i^+ + z_i^-, \tag{82b}$$

where $z_i^+ = \max\{x_i, 0\}$ and $z_i^- = \max\{-x_i, 0\}$. The standard lasso regression problem (81) can be transformed equivalently into

$$\begin{aligned} \min_{x,z} \quad & f(x) + \lambda\mathbf{1}^T z & (83a)\\ \text{s.t.} \quad & Ax + Bz = 0 : y & (83b)\\ & z \geq 0 : w \geq 0 & (83c) \end{aligned}$$

where $z = [z^{+T}, z^{-T}]^T$, $A = I_n$ and $B = [-I_n, I_n]$. Define its Lagrangian as

$$\begin{aligned} L(x,z,y,w) :&= f(x) + \lambda\mathbf{1}^T z + y^T(Ax + Bz) - w^T z \\ &= f(x) + \lambda\mathbf{1}^T z + \begin{bmatrix} y \\ w \end{bmatrix}^T \underbrace{\begin{bmatrix} A & B \\ 0 & -I_{2n} \end{bmatrix}}_{C} \begin{bmatrix} x \\ z \end{bmatrix}, \\ L(u,v) &= g(u) + v^T C u, \end{aligned} \tag{84}$$

with $u = (x, z)$ representing primary variables and $v = (y, w)$ being dual variables.

a) Conditioned Primal-Dual Dynamics: The Lagrangian (84) is affine in v. In order to increase the concavity, we apply the following change of variables

$$\bar{u} := u + \alpha C^T v.$$

Then we obtain an equivalent Lagrangian

$$\bar{L}(\bar{u}, v) = g(\bar{u} - \alpha C^T v) + v^T C\bar{u} - \alpha\|C^T v\|^2. \tag{85}$$

Combining $0 \preceq \nabla^2 g(u) \preceq lI$ and the given structure of $\bar{L}(\bar{u}, v)$, we examine the convexity and concavity of the regularized Lagrangian over the primal-dual variables $(\bar{u}, v)$.

Assumption 13. *The function $f(x)$ is convex with l-Lipschitz gradient, i.e., $lI \succeq \nabla^2 f(x) \succeq 0$ whenever $\nabla^2 f(x)$ is defined. The constraint matrix C is full row rank with $\sigma I \succeq CC^T \succeq \kappa I$, and locally Lipschitz row-wise.*

Lemma 27. *Given any $\alpha > 0$ that satisfy $\alpha < \frac{2}{l}$, $\bar{L}(\bar{u}, v)$ is $(2\alpha - l\alpha^2)\kappa$-strongly concave in v.*

b) Proximal Primal-Dual Dynamics: According to Lemma 27, we cannot illustrate the strongly convexity in $\bar{u}$ just by the change of variables. Next, the proximal regularization is used to increase the convexity. The modified Lagrangian is defined as

$$\tilde{L}(\bar{u}, v, \bar{v}) = g(\bar{u} - \alpha C^T v) + v^T C\bar{u} - \alpha\|C^T v\|^2 - \frac{\rho}{2}\|v - \bar{v}\|^2. \tag{86}$$

By minimizing the modified Lagrangian over the variable x, we attain a proximal saddle function as

$$S(\bar{u}, \bar{v}) := \max_{v\in\mathcal{D}_v} \tilde{L}(\bar{u}, v, \bar{v}), \tag{87}$$

where $\mathcal{D}_v = \mathcal{R}^n \times \mathcal{R}_+^{2n}$. Since $\tilde{L}(\bar{u}, v, \bar{v})$ is strongly concave in v, (87) can be rewritten as

$$\begin{aligned} S(\bar{u}, \bar{v}) = g(\bar{u} - \alpha C^T v_\star(\bar{u}, \bar{v}) + v_\star(\bar{u}, \bar{v})^T C\bar{u} \\ - \alpha\|C^T v_\star(\bar{u}, \bar{v})\|^2 - \frac{\rho}{2}\|v_\star(\bar{u}, \bar{v}) - \bar{v}\|^2. \end{aligned} \tag{88}$$

where $v_\star(\bar{u}, \bar{v})$ is the unique minimizer given $(\bar{u}, \bar{v})$ such that

$$v_\star(\bar{u}, \bar{v}) = \arg\max_{v\in\mathcal{D}_v} \tilde{L}(\bar{u}, v, \bar{v}). \tag{89}$$

The proximal saddle flow dynamics of $S(u, y)$ is as follow:

$$\dot{\bar{u}} = -\nabla g\left(\bar{u} - \alpha C^T v_\star(\bar{u}, \bar{v})\right) - C^T v_\star(\bar{u}, \bar{v}), \tag{90a}$$
$$\dot{\bar{y}} = +\rho v_\star(\bar{u}, \bar{v}) - \rho\bar{v}. \tag{90b}$$

Similarly, $(\bar{u}, \bar{y})$ converges to the saddle point $(x_\star, y_\star)$ due to the invariance of the saddle point.

We illustrate the exponential convergence of (90) via a simple Lasso Regression Problem as:

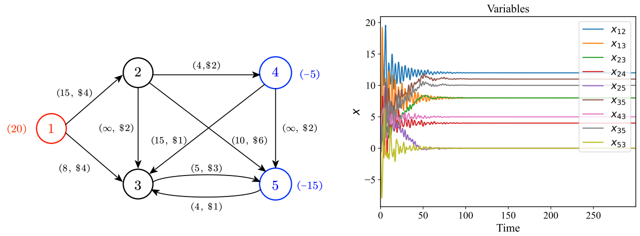
$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_1, \tag{91}$$

where b is observation vector and $A \in \mathbb{R}^{n\times n}$ is known data. The observation is given by $b = Ax + e$ with some noise e and the goal is to estimate x. By applying the algorithm (90) to solve this problem, results are demonstrated in Figure 2. Primal variables and dual variables both converge to the optimal solution with exponential convergence rates.

Remark 14. *The equation (88), representing a constrained saddle function, contrasts with equation (26). This distinction implies that the previously discussed analysis, which hinges on the Jacobian matrix to determine convergence rates, is not directly applicable in this context. Nevertheless, numerical evidence suggests that exponential convergence may still hold. This observation raises the possibility that the analysis could be generalized to encompass a broader range of forms.*

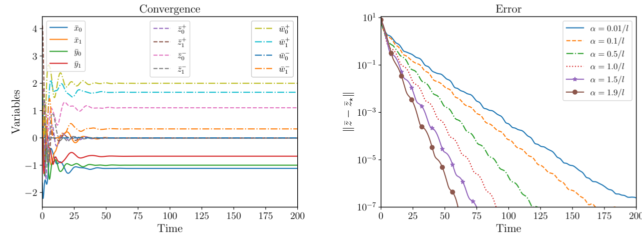
VIII. Conclusion

This paper has advanced the understanding of the convergence properties of saddle flow dynamics, purely depends on the special attributes of saddle functions. Initially, We first propose an observable certificate that directly establishes connection between the invariant set and the equilibrium set for saddle flow dynamics of a convex-concave function such that the asymptotic convergence to a saddle point can be guaranteed. The certificate is rooted in observability, and we identify the existence of such observable certificates in the presence



(a) Network. (b) Variable x ($\rho = 0.05$).

Fig. 1: The network flow problem.



(a) Trajectories ($\alpha = 0.01/l$). (b) Exponential decay of errors.

Fig. 2: The lasso regression problem.

of conventional conditions, e.g., strict convexity-concavity and proximal regularization, as well as the proposed separable regularization method. The novel separable regularization method that builds on our observable certificate requires only minimal convexity-concavity to establish convergence and enjoys a separable structure for potential distributed implementation. Furthermore, our work reveals the direct relationship between global exponential stability in saddle flows and the presence of strong convexity-strong concavity conditions. This insight not only elucidates the convergence properties of established algorithms, such as the proximal gradient method in equality-constrained convex optimization but also underpins our novel conditioned primal-dual gradient algorithms. Practically, our findings further generalize to situations with projections on the vector field of saddle flow dynamics.

Appendix

A. Proof of Lemma 27

Proof. Assumption 13 allows a straightforward calculation of the second-order partial derivatives

$$\frac{\partial^2}{\partial \bar{u}^2}\bar{L}(\bar{u}, v) = \nabla^2 g(\bar{u} - \alpha C^T v) \succeq 0\,, \tag{92a}$$

$$\begin{aligned}\frac{\partial^2}{\partial v^2}\bar{L}(\bar{u}, v) &= \alpha^2 C\nabla^2 g(\bar{u} - \alpha C^T v)C^T - 2\alpha CC^T \\ &= C\left(\alpha^2\nabla^2 g(\bar{u} - \alpha C^T v) - 2\alpha I\right)C^T \\ &\preceq C(l\alpha^2 - 2\alpha)C^T \\ &\preceq -\left(2\alpha - l\alpha^2\right)\kappa I \\ &\prec 0\,,\end{aligned} \tag{92b}$$

which justifies the claim. ■

References

[1] P. You and E. Mallada, "Saddle flow dynamics: Observable certificates and separable regularization," in *American Control Conference (ACC)*, 5 2021, pp. 4817–4823.

[2] T. J. Holding and I. Lestas, "Stability and instability in saddle point dynamics - Part I," *IEEE Transactions on Automatic Control*, pp. 1–1, 2020.

[3] T. J. Holding and I. Lestas, "Stability and instability in saddle point dynamics - Part II: The subgradient method," *IEEE Transactions on Automatic Control*, pp. 1–1, 2020.

[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 accelerated proximal splitting methods," *arXiv preprint arXiv:1908.00865*, 2019.

[6] H. Mohammadi, M. Razaviyayn, and M. R. Jovanovic, "Robustness of accelerated first-order algorithms for strongly convex optimization problems," *IEEE Transactions on Automatic Control*, 2020.

[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. Cortes, "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*. Prentice hall Upper Saddle River, NJ, 2002, vol. 3.

[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*, 2020.

[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 *2020 American Control Conference (ACC)*, 2020, pp. 4677–4682.

[16] H. Lu, "An $$o(sˆr)$$-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. [Online]. Available: https://doi.org/10.1007/s10107-021-01669-4

[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. Cortes, "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, 12 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 (TON)*, vol. 17, no. 5, pp. 1413–1426, 10 2009.

[23] D. Goldsztajn, F. Paganini, and A. Ferragut, "Proximal optimization for resource allocation in distributed computing systems with data locality," in *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 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, vol. 2.

[25] A. Cherukuri, B. Gharesifard, and J. Cortes, "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 *2017 American Control Conference (ACC)*. IEEE, 2017, pp. 3962–3967.
[27] D. Goldsztajn and F. Paganini, "Proximal regularization for the saddle point gradient dynamics," *Working Paper*, 2020.
[28] 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, Jan. 2019.
[29] 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, Oct. 2020.
[30] 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.
[31] D. Ding, B. Hu, N. K. Dhingra, and M. R. Jovanovic, "An exponentially convergent primal-dual algorithm for nonsmooth composite minimization," in *2018 57th IEEE Conference on Decision and Control (CDC)*. Miami Beach, FL: IEEE, Dec. 2018, pp. 4927–4932.
[32] D. Ding and M. R. Jovanovic, "Global exponential stability of primal-dual gradient flow dynamics based on the proximal augmented lagrangian: A lyapunov-based approach," in *2020 59th IEEE Conference on Decision and Control (CDC)*. IEEE, 2020, pp. 4836–4841. [Online]. Available: https://ieeexplore.ieee.org/document/9304313/
[33] S. K. Niederländer, F. Allgöwer, and J. Cortés, "Exponentially fast distributed coordination for nonsmooth convex optimization," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, Dec. 2016, pp. 1036–1041.
[34] J. Cortés and S. K. Niederländer, "Distributed coordination for nonsmooth convex optimization via saddle-point dynamics," *Journal of Nonlinear Science*, vol. 29, no. 4, pp. 1247–1272, Aug. 2019.
[35] 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, Oct. 2018.
[36] 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, Jul. 2022.
[37] I. K. Ozaslan and M. R. Jovanovic, "Tight lower bounds on the convergence rate of primal-dual dynamics for equality constrained convex problems," in *2023 62nd IEEE Conference on Decision and Control (CDC)*. IEEE, Dec. 2023.
[38] 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 preprint arXiv:1905.04521*, 2019.
[39] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in optimization*, vol. 1, no. 3, pp. 127–239, 2014.
[40] 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.
[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. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0005109821001059
[42] X. Chen and N. Li, "Exponential stability of primal-dual gradient dynamics with non-strong convexity," May 2019.
[43] W. Rudin, *Principles of mathematical analysis*, 3rd ed., ser. International series in pure and applied mathematics. New York: McGraw-Hill, 2008.
[44] I. K. Ozaslan and M. R. Jovanovic, "Exponential convergence of primal-dual dynamics for multi-block problems under local error bound condition," in *2022 IEEE 61st Conference on Decision and Control (CDC)*, Dec. 2022, pp. 7579–7584.
[45] T. Zheng, P. You, and E. Mallada. Constrained reinforcement learning via dissipative saddle flow dynamics. [Online]. Available: http://arxiv.org/abs/2212.01505
[46] 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 received her B.S. degree in mechanical engineering from the Department of Mechanics and Engineering Science, College of Engineering, Peking University, Beijing, China, in 2022. She is currently working toward the 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 current research interests include distributed robust and optimal control, saddle flow dynamics and reinforcement learning.

Enrique Mallada (S'09-M'13-SM') 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.