On the Geometry and Linear Convergence of Primal-Dual Dynamics

P. Bansode¹, V. Chinde², S. R. Wagh³, R. Pasumarthy⁴, and N. M. Singh⁵

Abstract—The paper proposes a variational-inequality based primal-dual dynamic that has a globally exponentially stable saddle-point solution when applied to solve linear inequality constrained optimization problems. A Riemannian geometric framework is proposed wherein we begin by framing the proposed dynamics in a fiber-bundle setting endowed with a Riemannian metric that captures the geometry of the gradient (of the Lagrangian function). A strongly monotone gradient vector field is obtained by using the natural gradient adaptation on the Riemannian manifold. The Lyapunov stability analysis proves that this adaption leads to a globally exponentially stable saddle-point solution. Further, with numeric simulations we show that the scaling a key parameter in the Riemannian metric results in an accelerated convergence to the saddle-point solution.

I. INTRODUCTION

Saddle-node dynamics has remained a subject of substantial research for many years [1]–[4]. It iteratively seeks a solution to saddle-point problems that arise in a number of disciplines including equilibrium theory, game theory and optimization. However its application to constrained optimization problems has gained wide interest over the recent years, especially in the areas of the power networks [5]–[8] and wireless networks [9]–[11], and building automation systems [12], etc. Within this domain, it is popularly regarded as primal-dual dynamics.

Over the last decade, various formulations of the primal-dual dynamics have been explored in connection with the constrained optimization problems. These algorithms were studied with primary focus on convergence of iterates to the saddle-point solution and its stability. These algorithms either use a framework of hybrid dynamical systems [9], [13] or an augmented Lagrangian technique that involves projections in the Lagrangian function [14]–[16]. The hybrid dynamical systems approach involves switching in the dual dynamics to handle constraint violations. But the inherent discontinuities present in the dual dynamics make it difficult to prove exponential stability of the saddle-point solution. Thus so far, only globally asymptotic stability of the saddle-point solution has been proven. The augmented Lagrangian approach is based on augmentation of the penalty terms in the Lagrangian function [17]. The penalty term is designed in such a way that the resulting dual dynamics is discontinuity-free, meaning that the rate of change of the dual variable does not involve switching terms. The basis of these penalty functions is either a projection or a proximal operator. So far, the latter approach has been successful in proving the global exponential stability of the algorithm when applied to solve linearly constrained optimization problems. In [18], it is shown that the algorithm attains a semi-globally exponential convergence for a more general convex inequality constrained optimization problem.

The algorithm that we propose in this paper uses a variational inequality based projected dynamical systems framework and produces a globally exponentially stable saddle-point solution for a linear inequality constrained optimization problem. The projected dynamical systems have been widely used to solve variational inequalities [19]–[21]. We model the saddle-point problem as a variational inequality and then use the projected dynamical system that singularly handles the constraints without involving discontinuities in the dual dynamics (the detailed approach of the proposed dynamics is documented in our online report [22], which is omitted from this paper due to avoid repetition).

In contrast to the existing research on primal-dual dynamics, our algorithm does not depend on the framework of hybrid dynamical systems or the augmented Lagrangian techniques. We consider differential equations for solution trajectories of variational inequality proposed in [23, Section 5.7.1] as a basis for our algorithm, and equate the time derivative of primal-dual variables to the terms for which the corresponding fixed point problem of the variational inequality fails to be satisfied (proved in our online report [22, Section 2.2]). Instead of tuning free-parameters in the augmented Lagrangian function we start by studying the geometry of gradient vector field defining the primal-dual dynamics. This enables us to carry out the required alterations to the dynamics in the Euclidean space so as to ensure linear convergence rates under suitable assumptions. We first analyze the vector field of the proposed algorithm in the Euclidean space and prove that the Euclidean geometry is not suitable for achieving the same. Then we lay down a procedure to construct a suitable differential geometry for the cause. The geometrization of the problem is based in a fiber bundle with a semi-Riemannian metric imposed by the structure of the primal-dual dynamics. A suitable alteration of the Geometry is achieved by a scaling of the connection leading to a natural gradient in the Riemannian space. As the natural gradient behaves like the Euclidean gradient in the sense that it achieves the steepest descent direction for the gradient descent algorithm [24], [25], the
obtained natural gradient dynamics is substituted for the original Primal Dual dynamics in the Euclidean space to ensure a faster convergence rate (in our case, the contraction region is obtained in the Euclidean domain because we are using the natural gradient as opposed to the contraction region being defined by a Riemann metric). Adapting this into the proposed algorithm results in a strongly monotone gradient vector field with steepest descent and ascent directions along the primal and dual variables, respectively, then we construct a Lyapunov function which shows that for a strongly monotone gradient the proposed algorithm has an exponentially stable saddle-point solution. Further with the help of numeric simulations, it is shown that by appropriate scaling of a key parameter in the natural gradient which in turn leads to the scaling of the connection as alluded to above, the convergence rate of the proposed algorithm can be accelerated.

Notations

The set \( \mathbb{R} \) (respectively \( \mathbb{R}_{\geq 0} \) or \( \mathbb{R}_{>0} \)) is the set of real (respectively non-negative or positive) numbers. If \( f : \mathbb{R}^n \to \mathbb{R} \) is continuously differentiable in \( x \in \mathbb{R}^n \), then \( \nabla_x f : \mathbb{R}^n \to \mathbb{R}^n \) is the gradient of \( f \) with respect to \( x \). \( \| x \| \) denotes the Euclidean norm.

II. PROBLEM FORMULATION

Consider the following constrained optimization problem

\[
\text{minimize } f(x) \\
\text{subject to } x \in X
\]

where

\[
X = \{ x \in \mathbb{R}^n \mid g_i(x) \leq 0, \forall i = 1, \ldots, m \}, \tag{2}
\]

is the domain of the problem \((1)\). The functions \( f : \mathbb{R}^n \to \mathbb{R} \), \( g : \mathbb{R}^n \to \mathbb{R}^m \) are assumed to be continuously differentiable \( (C^2) \) with respect to \( x \), with the following assumptions:

Assumption 1: \( \nabla f : \mathbb{R}^n \to \mathbb{R}^n \) is strongly monotone on \( X \), with \( \mu > 0 \) such that the following holds:

\[
(x_1 - x_2)^T (\nabla f(x_1) - \nabla f(x_2)) \geq \mu \| x_1 - x_2 \|^2.
\]

As a consequence of Assumptions 1, it is derived that the objective function \( f \) is strongly convex in \( x \) with the modulus of convexity given by \( \frac{\mu}{2} \).

Assumption 2: The constraint function \( g(x) = Ax - b \) is linear in \( x \).

Assumption 3: There exists an \( x \in \text{relint } X \) such that \( g_i(x) < 0, \forall i = 1, \ldots, m \).

Assumption 4: Let matrix \( \frac{\partial g}{\partial x} \) have full row rank \( m \leq n \) and \( q_1 I \leq \frac{\partial^2 g}{\partial x^2} \leq q_2 I \), where \( I \) is an identity matrix and \( q_1, q_2 \) are positive constants.

Assumptions 1-3 ensure that \( x \) is strictly feasible and strong duality holds for the optimization problem \((1)\).

Let \( L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \) define the Lagrangian function of the optimization problem \((1)\) as given below

\[
L(x, \lambda) = f(x) + \lambda^T g(x).
\]

Let \( \lambda_i \) be the Lagrange multipliers associated with \( g_i(x) \), then \( \lambda \in \Lambda \subseteq \mathbb{R}^m_+ = \{ \lambda \in \mathbb{R}^m, \lambda_i \geq 0, \forall i = 1, \ldots, m \} \) defines the corresponding vectors of Lagrange multipliers.

The Lagrangian function \( L \) defined in \((3)\) is \( C^2 \)-differentiable convex-concave in \( x \) and \( \lambda \) respectively, i.e., \( L(\cdot, \lambda) \) is convex for all \( \lambda \in \Lambda \) and \( L(x, \cdot) \) is concave for all \( x \in X \). We say that \((x^*, \lambda^*)\) is a saddle-point if the following holds:

\[
L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*)
\]

for all \( x \in X \) and \( \lambda \in \Lambda \).

If \( x^* \) is the unique minimizer of \( L \), then it must satisfy the Karush-Kuhn-Tucker (KKT) conditions stated as follows.

\[
g_i(x^*) \leq 0, \forall i = 1, \ldots, m \tag{5}
\]

\[
\lambda_i^* \geq 0, \forall i = 1, \ldots, m \tag{6}
\]

\[
\lambda_i^* g_i(x^*) = 0, \forall i = 1, \ldots, m \tag{7}
\]

\[
\nabla f(x^*) + \lambda^T \nabla g(x^*) = 0. \tag{8}
\]

Let us define \( z = (x, \lambda) \in \Omega = X \times \Lambda \), where \( \Omega \) by definition is a nonempty and closed convex subset of \( \mathbb{R}^n \) and \( \mathbb{R}^m_0 \). Then \( z^* = (x^*, \lambda^*) \) is the saddle point solution of \((3)\).

Remark 1: Since strong duality holds, the KKT conditions \((5)-(8)\) are necessary and sufficient to guarantee optimality of the problem \((1)\), with \( x^* \) as the unique minimizer of \((1)\) and \( z^* \) as the unique saddle-point of \((3)\).

Let \( G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{n+m} \) define the gradient map of \((3)\) as given below:

\[
G(z) = \nabla_x L = \begin{bmatrix} \nabla_x L(x, \lambda) \\ \nabla_x L(x, \lambda) \end{bmatrix}. \tag{9}
\]

Taking gradient descent and gradient ascent along the direction of \( x \) and \( \lambda \) variables, respectively, we propose an algorithm which in principal uses the framework of projected dynamical systems (for details, please refer to our online report [22, Section 2] or Definition 4.1, Proposition 4.1, and equation (57) in the Appendix section). We designate the algorithm as projected primal-dual dynamics, it is as shown below:

\[
z = \beta \{ P_{\Omega}[z - \alpha G(z)] - z \}, \tag{10}
\]

where \( \alpha, \beta > 0 \) are parameters adjusted to control stability and assure convergence, we set \( \alpha = \beta = 1 \) to avoid confusion. \( P_\Omega \) is a minimum norm projection operator of the form \( P_\Omega = \arg \min_{z \in \Omega} \| z - v \| \). Note that, there is no projection taken w.r.t. the primal variable \( x \) as it belongs to \( \mathbb{R}^n \) but only w.r.t. the dual variable \( \lambda \) which is restricted to \( \mathbb{R}^m_0 \). The following property always holds for projection on \( \Omega \),

\[
[u - P_\Omega(u)]^T [P_\Omega(z) - z] \geq 0, \forall u \in \mathbb{R}^n \times \mathbb{R}^m_0, \forall z \in \Omega. \tag{11}
\]

In what follows, we assess the stability of the proposed algorithm.
A. Stability analysis

Before proceeding to the stability analysis of the proposed dynamics (10), it is worth noting the Definition 4.2 and Proposition 4.3 on the monotonicity property, stated in the Appendix of this paper.

**Lemma 2.1:** If Assumptions (1)-(3) hold, then \( G(z) \) is monotone such that \( [G(z_1) - G(z_2)]^T(z_1 - z_2) \geq 0 \) for every pair of \( z_1, z_2 \in \Omega \).

**Proof:** First we derive the Jacobian matrix of \( G \) as
\[
\nabla G = \begin{bmatrix}
\nabla^2 f(x) + x^T \nabla^2 g(x) & \nabla g(x)^T \\
\nabs{g(x)} & 0
\end{bmatrix}
\] (12)

It is known that \( G \) is monotone if and only if the \( \nabla G \) is positive semidefinite. Following, (26), which implies that \( \nabla G \) must be positive semidefinite \( \frac{1}{2} \nabla G + \frac{1}{2} \nabla G^T \geq 0, \forall z \in \Omega, \forall t \). We verify this property by evaluating the symmetric part of \( \nabla G \):
\[
\nabla G + \nabla G^T = \begin{bmatrix}
\nabla^2 f(x) + x^T \nabla^2 g(x) & \nabla g(x)^T \\
\n0 & 0
\end{bmatrix}
\] (13)

\[
\geq 0, \forall z \in \Omega, \forall t
\] (14)

This proves that \( \nabla G(z) \) is a positive semi-definite matrix. Thus \( G(z) \) is a monotone map.

**Lemma 2.2:** Let \( G(z) \) be continuously differentiable on an open convex subset of \( \mathbb{R}^{n+m} \). If Assumptions (15) hold and \( G \) is monotone for all \( z \in \Omega \), then with \( \alpha > 0 \), the projected PD dynamics (10) is Lyapunov stable.

**Proof:** For the Lagrangian function (3), the following inequalities always hold: \( L(x^*, \lambda^*) - L(x^*, \lambda) \geq 0 \) and \( L(x, \lambda^*) - L(x^*, \lambda^*) \geq 0 \). Let us define the Lyapunov function as follows:
\[
V(z) = (L(x^*, \lambda^*) - L(x^*, \lambda)) + (L(x, \lambda^*) - L(x, \lambda^*)) + \frac{1}{2} \|z - z^*\|^2.
\] (15)

The last term in (15) ensures that \( V(z) \geq \frac{1}{2} \|z - z^*\|^2, \forall z \in \Omega \), thus also ensures the boundedness of the level sets of \( V(z) \).

Differentiating \( V(z) \) along the trajectories of (10) yields,
\[
\dot{V}(z) = \nabla V(z) \dot{z}
\]

\[
= -[\nabla L(x, \lambda^*) - \nabla L(x^*, \lambda)] + z - z^*]^T(z - \dot{z})
\]

\[
= -[G(z) + z - z^*]^T(z - \dot{z})
\] (16)

Using (11) with \( u = z - \alpha G(z) \) and \( z = z^* \), we get the following, (27).
\[
[z - z^* + \alpha G(z)]^T(z - \dot{z}) \geq \|z - \dot{z}\|^2 + \alpha \|z - z^*\|^2 G(z).
\] (17)

Using (17) in (16) yields,
\[
\dot{V}(z) \leq -[z - z^*]^T G(z) - \|z - \dot{z}\|^2
\]

\[
\leq -(z - z^*)^T(G(z) - G(z^*)) - \|z - \dot{z}\|^2
\] (18)

Using Lemma 2.1 in (18), we get \( \dot{V}(z) \leq 0 \). This proves that the projected dynamical system (10) is stable in the sense of Lyapunov.

We see that the results of Lemma 2.1 and 2.2 are not sufficient to prove exponential stability of the proposed dynamics (10). This however is overcome by adapting the framework of natural gradient [24], which allows us to prove that the gradient \( G(z) \) is strongly monotone. This property is later exploited to prove that the proposed dynamics is exponentially stable as discussed in the subsequent section.

B. Geometry of the Primal-Dual Dynamics

First we employ the framework of fiber bundles to understand the geometry of the gradient map (9). Assume that the proposed dynamics is embedded into a tuple \((X, \mathcal{M})\) with \( X \) a manifold and \( \mathcal{M} \) a fiber manifold above \( X \), with projection \( \Pi : \mathcal{M} \rightarrow X \). \( X \) is considered as a state-space of primal variables while the fibers of \( \mathcal{M} \) are along the space of the dual variables, \( \lambda \). If \( x \in X \subset \mathbb{R}^n, \lambda \in \Lambda \subset \mathbb{R}^m, \) then \( \mathcal{M} = (X, \lambda) \) with coordinates \((x, \lambda)\). The tangent space of \( \mathcal{M} \), denoted by \( T_{\mathcal{M}} \) has coordinates \( (\dot{x}, \dot{\lambda}) \).

Consider the function
\[
Q(x, \lambda) = \lambda - g(x).
\] (19)

and the implicit surface \( \mathcal{M} = \{x | \lambda - g(x) = 0\} \). With (19), if \( x \in X \) the primal dynamics in (10) on the manifold \( \mathcal{M} \) reduces to:
\[
\dot{x} = -\nabla f(x) - \left(\frac{\partial g}{\partial x}\right)^T g(x).
\] (20)

It follows that (20) is exponentially stable.

**Lemma 2.3:** The gradient dynamics (20) is exponentially stable.

**Proof:** Assume \( g(x) = Ax - b \) and consider the following Krasovskii-type Lyapunov candidate function \( V(x) = \frac{1}{2} x^T \dot{x} \).

Differentiating it along the trajectories of (20), we get
\[
\dot{V} = \dot{x}^T \left( -\nabla^2 f(x) \dot{x} - \left[ \frac{\partial^2 g}{\partial x^2} \right]^T g(x) \dot{x} - \left[ \frac{\partial g}{\partial x} \right]^T \left[ \frac{\partial g}{\partial x} \right] \dot{x} \right)
\]

\[
\leq \dot{x}^T [ -\nabla^2 f(x) - k \dot{x}]
\]

\[
\leq -\gamma V(x)
\] (21)

where \( \gamma > 0 \) and \( k \) is chosen as
\[
k > \sqrt{g}.
\] (24)

Global attractivity of the manifold \( \mathcal{M} \) would ensure that the proposed dynamics in (10) is globally exponentially stable but since it is known that this is not the case we study the geometry of the problem and identify conditions to improve the convergence rates to the fixed point of (10).

1) Constructing a Riemannian metric: Let \( R \) define a semi-Riemannian metric on space \( T_{\mathcal{M}} \) that endows a semi-Riemannian structure to the fiber-bundle \((x, \lambda) \in \mathbb{R}^{n \times m}, \) as shown below:
\[
R = \nabla Q^T(x) \nabla Q(x)
\] (25)

\[
= \left[ \frac{\partial g}{\partial x} \right]^T \left[ \frac{\partial g}{\partial x} \right] - \left[ \frac{\partial g}{\partial x} \right]^T I
\] (26)

\( R \) is a semi metric on the space \( X \times \Lambda \) with connection \( \left( \frac{\partial g}{\partial x} \right) \). The metric \( R \) can be made symmetric positive definite
by introducing in it the parameter $k$ such that $k$ scales the connection term and for $k = 1$ the original connection as defined by the above semi-Riemann metric is obtained, i.e

$$R = \left[ \frac{\partial^2}{\partial x \partial y} - \frac{\partial^2}{\partial y \partial x} \left( - \frac{\partial}{\partial x} \right)^T \right]. \quad (27)$$

We are now in a position to develop an understanding of the geometry of the proposed dynamics on a Riemannian manifold $(M, R)$. The following definition will be useful in understanding the concept of a linear connection with respect to the manifold $(M, R)$, [28].

**Definition 2.1:** Let $\Pi : M \to X$ be a smooth fiber bundle, a tangent vector $\nu \in T_p M$, $p \in M$, is said to be vertical if $\Pi_\nu(v) = 0$. $V(p)$ denotes the set of all vertical tangent vectors in $P$. A distribution $H$ on $M$ is said to be horizontal if $T_p M = V_p \oplus H(p)$ for all $p \in M$.

**Remark 2:** If $H$ is horizontal, it implies that for all $p \in M$, $H(p)$ is a linear subspace of $T_p M$ with the following properties:

$$\dim H(p) = \dim X \quad (28)$$
$$H(p) \cap M(p) = \Phi \quad (29)$$

$\Pi_{X_p}$ maps $H(p)$ isometrically onto $T_{\Pi(p)} X$.

A connection in the bundle $M$ is due to a unique splitting scheme of the tangent space $T_p M$ into a horizontal and vertical space as shown in Remark 3 in the Appendix section. The preferred direction of the vertical vector is along the fibers of $M$.

2) **Strongly monotone gradient of the Lagrangian:** The natural gradient of $L$ at $z \in M$ is a unique tangent vector $\nabla L$ given as

$$\langle \nabla L, v \rangle = D_z L(v), \forall v \in T_z M. \quad (30)$$

In the matrix notation, (30) implies the following

$$\nabla L = R^{-1} \nabla L^T \quad (31)$$

where $\nabla L = G(z)$ is the gradient vector of $L$ on Euclidean space $\mathbb{R}^{m+n}$.

Denote $G_r(z) = \nabla L$, the linear map $\mathbb{R}^{G_r(z)} : T_z M \to M$ assigned to each point $z \in M$ is defined by the Hessian of $L$, denoted by $\mathbb{H}_{G_r(z)} v = \nabla^2 G_r(z) = R^{-1} \nabla^2 G(z), \forall v \in T_z M$.

The projection operator $P^r_M : \mathbb{R}^{m+n} \to M$ is defined as

$$P^r_M(z) = \arg \min_{v \in T_z M} \| z - v \|^2. \quad (32)$$

Correspondingly, the projected PD dynamics on $M$ is defined as follows:

$$\dot{z} = \beta \{ P^r_M(z - \alpha G_r(z)) \} - z. \quad (33)$$

Replacing $g(x)$ by $Ax - b$ as defined in Assumption 2, we define the gradient vector $G_r(z) \in T_z M$ as follows:

$$G_r(z) = R^{-1} G(z), \quad (34)$$

where $G_r(z) = \nabla L$, and $A^T A = H$ is a symmetric matrix.

In the following section, it is proved that the gradient map $G_r(z)$ is strongly monotone.

**Proposition 2.4:** Consider the optimization problem (1) and let $(M, r)$ be a $n + m$-dimensional smooth manifold. If Assumption 1 and 2 hold for the problem (1), then with the linear map $R^{-1} : T_z M \to T_z M$, the gradient vector $G_r(z)$ is strongly monotone.

**Proof:** For $G_r(z)$ to be strongly monotone, $\nabla z G_r$ must be positive definite [26], i.e., for the symmetric part of $\nabla z G_r$, i.e., $\frac{1}{2} \nabla G_r + \frac{1}{2} \nabla G_r^T$, the following must hold:

$$\nabla G_r + \nabla G_r^T = R^{-1} \nabla G + \nabla G_r^T R^{-1}, \quad (35)$$

where $\nu > 0$ is a constant, $I$ is an identity matrix of appropriate dimensions.

The Jacobian of $G_r(z)$, denoted by $\nabla G_r(z)$ is given below:

$$\nabla G_r(z) = \begin{bmatrix} k\nabla^2 f(x) - A^T A & kA^T \\ A\nabla^2 f(x) - kA & AA^T \end{bmatrix} \quad (36)$$

The symmetric part of $\nabla G_r(z)$ is obtained as:

$$\nabla G_r(z) + \nabla G_r^T(z) = \begin{bmatrix} k\nabla^2 f(x) - A^T A & \frac{1}{2}(A\nabla^2 f(x))^T \\ \frac{1}{2}A\nabla^2 f(x) & AA^T \end{bmatrix} \quad (37)$$

Let $M = \nabla G_r(z) + \nabla G_r^T(z) - q_1 I > 0$. Then

$$M = \begin{bmatrix} 2k\nabla^2 f(x) - 2A^T A - q_1 I & (A\nabla^2 f(x))^T \\ A\nabla^2 f(x) & 2AA^T - q_1 I \end{bmatrix} \quad (38)$$

Further let $S = AA^T$, then the Schur compliment of the block $S$ of the matrix $M$, denoted by $S_{\text{Schur}}$ is derived as

$$S_{\text{Schur}} = 2k\nabla^2 f(x) - 2A^T A - q_1 I - (A\nabla^2 f(x))^T (AA^T)^{-1} A\nabla^2 f(x). \quad (39)$$

Let $H = \nabla^2 f(x)$ for the notational simplicity. Note that in $\frac{k}{2} k \phi > 0, \forall k > 0, 2A^T A \geq 0, q_1 I > 0, 0 \leq HA^T (AA^T)^{-1} AH \leq H^2$. The last terms is a consequence of $A^T (AA^T)^{-1} A \leq I$. Rearranging (37) as given below

$$2kH > 2A^T A + q_1 I + HA^T (AA^T)^{-1} AH \quad (40)$$

allows to choose $k$ such that $S_{\text{Schur}} > 0$. Post multiplying $\frac{k}{2} k I > 2A^T A + q_1 I + HA^T (AA^T)^{-1} AH \quad (41)$

Applying Courant-Fischer theorem [29] to (39) yields the following:

$$\lambda_{\max} (kI) > \lambda_{\max} (AA^T)^{-1} + \frac{1}{2}q_1 H^{-1} + \frac{1}{2}H. \quad (42)$$
Since $\lambda_{\max}(kI) = k$, (40) has the following form:

$$k > \lambda_{\max}(A^T A H^{-1} + \frac{1}{2} q_1 H^{-1} + \frac{1}{2} H). \quad (41)$$

By choosing $k$ as given in (41) ensures that $S_{\text{Schur}} > 0$. But $k$ must also satisfy (24), thus $k$ must be chosen such that the following holds:

$$k > \max\{\sqrt{q_2}, \lambda_{\max}(A^T A H^{-1} + \frac{1}{2} q_1 H^{-1} + \frac{1}{2} H)\} \quad (42)$$

ensures that both (24) and (40) are met. If $k$ is chosen according to (42), then $S_{\text{Schur}} > 0$ holds such that there exists a $\nu \geq \frac{\sqrt{q_2}}{2}$ which implies that

$$\langle \mathbb{E} G_r(z)v, v \rangle_r \geq \nu \|v\|_r^2, \forall v \in T_z \mathcal{M}. \quad (43)$$

Hence it follows that

$$(G_r(z_1) - G_r(z_2), z_1 - z_2)_r \geq \nu \|z_1 - z_2\|_r^2. \quad (44)$$

Hence it is proved that $G_r(z)$ is strongly monotone.

3) Exponential stability: Without loss of generality, let us define $G_r(z)$ similar to (39) as follows:

$$G_r(z) = \begin{bmatrix} \nabla L_r^T L(x, \lambda) & -\nabla L_r^T L(x, \lambda) \end{bmatrix}, \quad (45)$$

where $L(x, \lambda)$ would represent the modified Lagrangian function whose gradient vector field is given by $G_r(z)$. Since, $G_r(z)$ is strongly monotone on $\mathcal{M}$, (32) will converge to a unique saddle-point solution $z^*$.

Theorem 2.5: Let $G_r(z)$ be Lipschitz continuous on an open set including $\mathcal{M}$, then inequality (44) and $\alpha > 0$, imply that the system (32) with $z(0) \in \mathcal{M}$ is globally exponentially stable at the unique solution $z^*$ of (31).

Proof: For each $z(0) \in \mathcal{M}$, there exists a unique solution $z(t)$ of (30), that started from $z(0)$. If $[0, t_f)$ is the maximal interval of $z(t)$, then from Lemma 4.1, $z(t) \in \mathcal{M}$ for all $t \in [0, t_f)$. Since $G_r(z)$ is strongly monotone, the following holds: $L(x^*, \lambda^*) - L(x^*, \lambda) > 0, L(x, \lambda^*) - L(x^*, \lambda^*) > 0$. Let us define the Lyapunov function for the dynamics (32) as follows:

$$V_1(z) = \frac{1}{2} \|z - z^*\|_r^2 + (L(x, \lambda^*) - L(x^*, \lambda^*) + \frac{1}{2} \|z - z^*\|_r^2. \quad (46)$$

It is to be noted that $V_1(z)$ possesses a similar structure as that of $V(z)$ defined in (15), it is also differentiable convex on $\mathcal{M}$, with $V_1(z) \geq \frac{1}{2} \|z - z^*\|_r^2, \forall z \in \mathcal{M}$, thus bounding all level sets of $V_1(z)$.

Differentiating $V_1(z)$ along the trajectories of (32) yields:

$$\dot{V}_1(z) = \nabla V_1(z) \dot{z} = -\nabla L_r^T L(x, \lambda^*) - \nabla L_r^T L(x^*, \lambda) + z - z^*, z - \dot{z}, r = -\langle G_r(z) + z - z^*, z - \dot{z} \rangle_r \quad (47)$$

Substituting $u = z - \alpha G_r(z)$ and $z = z^*$ in (11), yields

$$\langle z - z^* + \alpha G_r(z), z - \dot{z} \rangle_r \geq \|z - \dot{z}\|_r^2 + \langle \alpha(z - z^*), G_r(z) \rangle_r. \quad (48)$$

Using (48) in (47) yields:

$$\dot{V}_1(z) \leq -\langle \alpha(z - z^*), G_r(z) \rangle_r. \quad (49)$$

If $k$ is chosen such that the condition (40) is satisfied then $G_r(z)$ is strongly monotone. The strong monotonicity of $G_r(z)$ leads to the following property of the Lagrangian function $L(z), \forall z \in \mathcal{M}$,

$$\langle z - z^*, G_r(z) \rangle_r \geq L(x, \lambda^*) - L(x^*, \lambda) + \frac{1}{2} \|z - z^*\|_r^2, z \in \mathcal{M}. \quad (50)$$

Using (50), (49) modifies to the following

$$\dot{V}_1(z) \leq -\langle \alpha(z - z^*), G_r(z) \rangle_r, \quad \leq -\alpha \beta [L(x, \lambda^*) - L(x^*, \lambda) + \frac{1}{2} \|z - z^*\|_r^2, z \in \mathcal{M}]. \quad (51)$$

With $\alpha, \beta > 0$, it can be shown that,

$$\dot{V}_1(z) \leq -\beta \min\{1, \alpha \nu\} V(z). \quad (52)$$

Thus, it is proved that the system (10) is exponentially stable at the unique solution $z^*$ of (4). Therefor,

$$\|z(t) - z^*\|_r \leq c e^{-\beta \min\{1, \alpha \nu\} t}, \quad (53)$$

where $c = \sqrt{2V_1(z(0))}$.

Further, if $G_r(z)$ is Lipschitz continuous on $\mathcal{M}$, i.e., $\|G_r(z_1) - G_r(z_2)\| \leq \|L\| \|z_1 - z_2\|_r, \forall z_1, z_2 \in \mathcal{M}$, where $L$ is a Lipschitz constant then by using [27, Theorem 4] the global exponential stability of the projected PD dynamics can be established as follows:

$$\|z(t) - z^*\|_r \leq \|z(0) - z^*\|_r e^{-\beta \min\{1, \alpha \nu\} t}, \forall t \geq 0. \quad (54)$$

If $\alpha < \frac{1}{\|L\|}$, it follows that the projected PD dynamics (32) is globally exponentially stable.

III. SIMULATION RESULTS

This section presents simulation studies of the projected PD dynamics (32). It is known that the Euler discretization of the exponentially stable dynamical system owns geometric rate of convergence [30] for sufficiently small step-sizes. The projected PD dynamics (32) is Euler discretized with a step size $s > 0$ and the following discrete-time projected PD dynamics [20] is obtained.

$$z(\tau + 1) = \beta P_M^T (z(\tau) - \alpha G_r(z)). \quad (53)$$

First example (Example 1) considers an optimization problem of the form (1) with $m = 5$ and $n = 10$. The Hessian matrix is assumed to be $H = 20I$ with $A$ and $b$ taken as Gaussian random matrix and vector respectively. The distance to the primal optimizer $x^*$ for different values of parameter $k$ is shown in Fig. 1, where $\delta = \max\{\sqrt{q_2}, \lambda_{\max}(A^T A H^{-1} + \frac{1}{2} q_1 H^{-1} + \frac{1}{2} H)\}$. It can be seen from the plot that the rate of convergence to the equilibrium point accelerates as the value of $k$ is increased. It implies that increasing the value of $k$ allows increasing
We began by framing the proposed dynamics in a fiber
work with natural gradient adaptation to achieve exponen-
ponentially stable saddle-point solution. We further show ed
that the increasing values of $k$ result in much steeper gradient
that leads to an accelerated convergence to the saddle-point
solution.

Fig. 1. Distance to the primal optimizer $x^*$ for different values of $k$
(Example 1).

Fig. 2. Distance to the primal optimizer $x^*$ ($L_2$ regularized least squares
problem).

the value of $\nu$, which further increases the coefficient of the
negative exponential term in $\mathbf{62}$.

In the second example, an $L_2$ regularized least squares
problem is considered with $m = 30$ and $n = 50$. The
objective function is $f(x) = \|Cx - d\|^2 + \frac{\theta}{2} \|x\|^2$ with
$\theta > 0$, constrained to $Ax \leq b$. Matrices $(C, A) \in \mathbb{R}^{m \times n}$,
and vectors $(d, b) \in \mathbb{R}^{m+1}$ are Gaussian random matrices and vectors, respectively. Parameters $\alpha, \beta$ are chosen as
unity and the proposed dynamics $\mathbf{52}$ is simulated for $k = 1000 \max(\theta)$. A sketch of the error norm as a function of
time is shown in Fig. 2. It can be seen that the error norm $\|x - x^*\|^2$ has geometric rate of convergence.

IV. CONCLUSIONS AND DISCUSSION

In this paper we proposed a Riemannian geometric frame-
work with natural gradient adaptation to achieve exponen-
tially convergent projected primal-dual dynamics when ap-
plied to linear inequality constrained optimization problems.
We began by framing the proposed dynamics in a fiber-
bundle setting endowed with a Riemannian metric $R$ that
captures the geometry of the gradient vector. The metric $R$
induced a unique decomposition of the target space into
a horizontal and vertical distributions. The natural gradient
proved to be strongly monotone on $\mathcal{M}$ leading to an
exponentially stable saddle-point solution. We further showed that the increasing values of $k$ result in much steeper gradient
that leads to an accelerated convergence to the saddle-point
solution.

REFERENCES

[1] Kenneth J Arrow, Leonid Hurwicz, and Hirofumi Uzawa. Studies in
linear and non-linear programming. 1958.
[2] Franco Brezzi. On the existence, uniqueness and approximation of
saddle-point problems arising from lagrangian multipliers. Publica-
tions mathématiques et informatique de Rennes, (S4):1–26, 1974.
[3] Michele Benzi, Gene H Golub, and Jörg Liesen. Numerical solution
of saddle point problems. Acta numerica, 14:1, 2005.
[4] Angelia Nedić and Asuman Ozdaglar. Subgradient methods for saddle-
point problems. Journal of optimization theory and applications,
142(1):205–228, 2009.
[5] Changhong Zhao, Ufuk Topcu, Na Li, and Steven Low. Design and
stability of load-side primary frequency control in power systems.
IEEE Transactions on Automatic Control, 59(5):1177–1189, 2014.
[6] Enrique Mallada, Changhong Zhao, and Steven Low. Optimal load-
side control for frequency regulation in smart grids. IEEE Transactions
on Automatic Control, 62(12):6294–6309, 2017.
[7] Peng Yi, Yiguang Hong, and Feng Liu. Distributed gradient algorithm
for constrained optimization with application to load sharing in power
systems. Systems & Control Letters, 83:45–52, 2015.
[8] Hung D Nguyen, Thanh Long Vu, Konstantin Turitsyn, and Jean-
Jacques Slotine. Contraction and robustness of continuous time primal-
dual dynamics. IEEE Control Systems Letters, 2(4):755–760, 2018.
[9] Diego Feijer and Fernando Paganini. Stability of primal–dual gradient
dynamics and applications to network optimization. Automatica,
46(12):1974–1981, 2010.
[10] Junting Chen and Vincent KN Lau. Convergence analysis of saddle
point problems in time varying wireless systems—control theoretical
approach. IEEE Transactions on Signal Processing, 60(1):443–452,
2012.
[11] Andrés Ferragut and Fernando Paganini. Network resource allocation
for users with multiple connections: fairness and stability. IEEE/ACM
Transactions on Networking (TON), 22(2):349–362, 2014.
[12] Krishna Chaitanya Kosaraju, Venkatesh Chinde, Ramkrishna Pa-
sumarthy, Atul Kelkar, and Navdeep M Singh. Stability analysis of
constrained optimization dynamics via passivity techniques. IEEE
Control Systems Letters, 2(1):91–96, 2018.
[13] Ashish Cherukuri, Enrique Mallada, and Jorge Cortés. Asymptotic
convergence of constrained primal–dual dynamics. Systems & Control
Letters, 87:10–15, 2016.
[14] Neil K Dhingra, Sei Zhen Khong, and Mihailo R Jovanovic. The
proximal augmented lagrangian method for nonsmooth composite
optimization. IEEE Transactions on Automatic Control, 2018.
[15] Guannan Qu and Na Li. On the exponential stability of primal-dual
gradient dynamics. IEEE Control Systems Letters, 3(1):43–48, 2019.
[16] Dongsheng Ding and Mihailo R Jovanovic. Global exponential
stability of primal-dual gradient flow dynamics based on the proximal
augmented lagrangian. In 2019 American Control Conference (ACC),
pages 3414–3419. IEEE, 2019.
[17] Dimitri P Bertsekas. Constrained optimization and Lagrange multi-
plier methods. Academic press, 2014.
[18] Yujie Tang, Guannan Qu, and Na Li. Semi-global exponential stability
of primal-dual gradient dynamics for constrained convex optimization.
arXiv preprint arXiv:1903.09580, 2019.
[19] Terry L Friesz, David Bernstein, Nihal J Mehta, Roger L Tobin,
and Saidi Ganjalizadeh. Day-to-day dynamic network disequilibria
and idealized traveler information systems. Operations Research,
42(6):1120–1136, 1994.
[20] Anna Nagurney and Ding Zhang. Projected dynamical systems and
variational inequalities with applications, volume 2. Springer Science
& Business Media, 2012.
[21] R Tyrrell Rockafellar and Roger J-B Wets. Variational analysis,
volume 317. Springer Science & Business Media, 2009.
[22] PA Bansode, V Chinde, SR Wagh, P Pasumarthy, and NM Singh.
On the exponential stability of projected primal-dual dynamics on a
riemannian manifold. arXiv preprint arXiv:1905.04521, 2019.
[23] Terry L Friesz. Dynamic optimization and differential games, volume 135.
Springer Science & Business Media, 2010.
[24] Shun-Ichi Amari. Natural gradient works efficiently in learning.
Neural computation, 10(2):251–276, 1998.
[25] Shun-Ichi Amari and Scott C Douglas. Why natural gradient? In
Proceedings of the 1998 IEEE International Conference on Acoustics,
Speech and Signal Processing, ICASSP’98 (Cat. No. 98CH36181),
volume 2, pages 1213–1216. IEEE, 1998.
A variational inequality problem (54) is equivalent to a fixed mapping $F$, and there exists precisely one solution $x^* \in X$ such that
\[(x - x^*)^T F(x^*) \geq 0, \forall x \in X.\] (54)

A variational inequality problem is equivalent to a fixed point problem given below:

**Proposition 4.1:** (A Fixed Point Problem, [20])

$x^*$ is a solution to VI$(F, X)$ if and only if for any $\alpha > 0$, $x^*$ is a fixed point of the projection map:
\[x^* = P_X(x^* - \alpha F(x^*))\] (55)

where
\[P_X = \arg \min_{v \in X} \|x - v\|.\] (56)

**Theorem 4.2:** (Uniqueness of the Solution to Variational Inequality, [20])

Suppose that $F(x)$ is strongly monotone on $X$. Then there exists precisely one solution $x^*$ to VI$(F, X)$.

Consider the following globally projected dynamical system proposed in [19]:
\[\dot{x} = \beta \{P_X[x - \alpha F(x)] - x\}\] (57)

where $\beta, \alpha$ are positive constants and $P_X : \mathbb{R}^n \to X$ is a projection operator as defined in (56).

**Remark 3:** (27)

$x^*$ is an equilibrium point of (57) if and only if $x^*$ is a solution of the variational inequality problem (54).

From Remark 3
\[\dot{x} = 0 \implies x^* = P_X(x^* - \alpha F(x^*)).\]

**Definition 4.2:** (Monotone Map, [26])

A mapping $F$ is monotone on $X \subseteq \mathbb{R}^n$, if for every pair of distinct points $x, y \in X$, we have
\[(y - x)^T (F(y) - F(x)) \geq 0.\]

**Definition 4.3:** (Strongly Monotone Map, [26])

A mapping $F$ is strongly monotone on $X \subseteq \mathbb{R}^n$, if there exists $\mu > 0$ such that, for every pair of distinct points $x, y \in X$, we have
\[(y - x)^T (F(y) - F(x)) \geq \mu \|x - y\|^2.\]

The relation between monotonicity of $F$ and positive definiteness of its Jacobian matrix
\[\nabla F(x) = \left( \frac{\partial F_i(x)}{\partial x_j} \right)_{i,j=1,2,...,n},\]
as given below.

**Proposition 4.3:** ((Strongly) Positive Definite Jacobian of $F(x)$ implies (Strongly) Monotone $F(x)$, [20])

Suppose that $F$ is continuously differentiable on $X$.

1) If the Jacobian matrix $\nabla F(x)$ is positive semidefinite, i.e.,
\[y^T \nabla F(x)y \geq 0, \forall y \in \mathbb{R}^n, x \in X,\]
then $F$ is monotone on $X$.

2) If the Jacobian matrix $\nabla F(x)$ is positive definite, i.e.,
\[y^T \nabla F(x)y > 0, \forall y \in \mathbb{R}^n, x \in X,\]
then $F$ is strictly monotone on $X$.

3) If $\nabla F(x)$ is strongly positive definite, i.e.,
\[y^T \nabla F(x)y \geq \mu \|y\|^2, \forall y \in \mathbb{R}^n, x \in X,\]
then $F(x)$ is strongly monotone on $X$.

1) Splitting of the tangent vector: Consider the tangent vector $(\dot{x}, \lambda) \in \mathbb{R}^{n+m}$, with the Euclidean metric $I$, then $v_H = (\dot{x}, 0)$ and $v_V = (0, \lambda)$ such that
\[\langle v_H, v_V \rangle_I = \begin{bmatrix} \dot{x} & 0 \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} 0 \\ \lambda \end{bmatrix} = v_H^T v_V = 0.\]

**Remark 4:** Given a tangent vector $v = (\dot{x}, \lambda) \in T_M$, orthogonality of $v_V$ and $v_H$ is preserved under the new metric $R$.

**Proof:** From (26), we observe that the splitting of the vectors is only dependent on $m_{12}$ or $m_{21}$, and $m_{22}$ in any case. Let us rewrite $R$ as $R = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}$. Now, the relevant splitting of the tangent vectors $(\dot{x}, \lambda)$ confirming the structure of $R$ defined in (27) is as follows:
\[(\dot{x}, \lambda) = v_H \perp v_V\] (59)
\[= \left( \dot{x}, -m_{22}^{-1} m_{21} \dot{x} \right) \perp \left( 0, \lambda + m_{22}^{-1} m_{21} \dot{x} \right)\] (60)
\[= \begin{bmatrix} \dot{x} & -m_{22}^{-1} m_{21} \dot{x} \end{bmatrix} \begin{bmatrix} m_{21} \lambda + m_{22}^{-1} m_{21} \dot{x} \\ m_{22} \lambda + m_{22}^{-1} m_{21} \dot{x} \end{bmatrix}\] (61)
\[= m_{21} \dot{x} \left( \lambda + m_{22}^{-1} m_{21} \dot{x} \right) - m_{21} \dot{x} \left( \lambda + m_{22}^{-1} m_{21} \dot{x} \right)\] (62)
\[= 0\] (63)

**Remark 5:** As the value of parameter $k$ increases the vertical component in $v_H$, i.e., $-m_{22}^{-1} m_{21} \dot{x}$ decreases in value, which causes the trajectories off-manifold $M$ to approach $M$ faster.

**Lemma 4.4:** [31] Assume that $F$ is locally Lipschitz continuous in a domain $D$ that contains $X$. Then the solution...
$x(t)$ of (57) will approach exponentially the feasible set $X$ when the initial point $x^0 \notin X$. Moreover, if $x^0 \in X$, then $x(t) \in X$. 