Contractivity of Distributed Optimization and Nash Seeking Dynamics

Anand Gokhale, Alexander Davydov, Francesco Bullo

Abstract— In this letter, we study distributed optimization and Nash equilibrium-seeking dynamics from a contraction theoretic perspective. Our first result is a novel bound on the logarithmic norm of saddle matrices. Second, for distributed gradient flows based upon incidence and Laplacian constraints over arbitrary topologies, we establish strong contractivity over an appropriate invariant vector subspace. Third, we give sufficient conditions for strong contractivity in pseudogradient and best response games with complete information, show the equivalence of these conditions, and consider the special case of aggregative games.

I. INTRODUCTION

Problem Description and Motivation: The past decade has witnessed rapid progress in distributed optimization problems, spurred on by the seminal paper [14]. This growing body of literature is surveyed in [21]. A significant majority of the literature shows the convergence of distributed optimization algorithms to a fixed point. Beyond convergence, practical implementations of these algorithms may require robustness to noise, delays, and unmodeled dynamics, and may involve time-varying parameters. No unifying framework currently exists that achieves these favorable properties in distributed optimization. Parallel to these developments, game theory and its applications in multi-agent systems have seen recent significant advancements [17] and there continues to be great interest for a "unified mathematical framework for learning a Nash equilibrium in games". Recently, game theoreticians have utilized passivity and its derivatives to study the convergence of Nash equilibrium (NE) seeking dynamics [8], [17].

In recent years, contraction theory [13] has emerged [3] as a promising framework for a broad range of dynamical systems, including both distributed optimization dynamics and NE-seeking dynamics. Specifically, the strong contractivity property is known to guarantee exponential convergence, robustness in time-varying optimization problems, periodic entrainment, and robustness to noise [3, Section 3.4]. For NE-seeking systems with time-invariant dynamics, strong contractivity implies the existence and uniqueness of a Nash equilibrium, associated with two Lyapunov functions. In time-varying systems, the error between the equilibrium trajectory and the system trajectory is explicitly and uniformly

upper bounded [6]. Finally, contracting systems support the use of efficient numerical integration algorithms [10].

Strong infinitesimal contraction is a strict property. This letter employs a relaxed version of contractivity called semi-contractivity, whereby the system is proved to be strongly contracting in a vector subspace of the system; see the recent theoretical works [9], [7], [3], [20].

Related Work: Here, we study continuous-time dynamics for distributed optimization problems. Such dynamics have previously been discussed in [11], [5]. The distributed optimization problem may be recast as a linearly constrained optimization problem, which in turn may be solved using a primal-dual approach [1]. Previous work in establishing the contraction of linearly constrained primal-dual setups include [18], [6], [15], but is limited to full row rank constraint matrices. The exponential stability of such setups with rank-deficient constraints is established in [16]. Here, we extend these results to contractivity in the case of non-full rank constraints. Our analysis requires tools from semicontraction theory [9], [3], [7], [20].

An early influential reference on the application of contraction analysis to games is [12]. More recently, several works invoke the use of passivity-based tools [17], [8] in multi-agent game theoretic systems for seeking a Nash equilibrium. With a fresh look at these dynamics, we approach it using ideas from interconnections of contracting systems [3].

Contributions: Our main contributions in this paper are threefold. (i) We start with a general result about the lognorms of saddle matrices that appear while studying continuous time primal-dual dynamics with a linear equality constraint. Specifically, we show that these dynamics are semicontracting in the presence of redundant constraints, while also improving upon the best-known rate in [18]. The use of redundant constraints allows the extension of contraction theory to distributed optimization. We also consider a non-symmetric (1,1) block in the saddle matrix as this is useful for developing Nash-seeking dynamics. (ii) Next, we consider the applications of these contracting dynamics in different optimization settings. We show that the Jacobian matrices obtained by linearizing the linearly constrained primal-dual flow and distributed optimization both form a saddle matrix and thus have semicontracting dynamics. We show the relationship of our contraction rate with network synchronizability and study the effect of modeling network constraints using Laplacian and incidence matrix-based constraints. (iii) Finally, we explore Nash-seeking dynamics in game theoretic setups. We identify equivalent sufficient conditions for the contraction of best response and pseudogradient dynamics. We also extend these results to

Anand Gokhale, Alexander Davydov, and Francesco Bullo are with the Center for Control, Dynamical Systems, and Computation, UC Santa Barbara, Santa Barbara, CA 93106 USA. anand_gokhale@ucsb.edu, davydov@ucsb.edu, bullo@ucsb.edu.

aggregative games obtaining explicit conditions for relating the Lipschitz constants and strong convexity parameter of the cost functions to the contractivity of the system.

The paper is organized as follows, some preliminaries are discussed in Section II. Then we study saddle matrices in Section III. Next, we discuss the contractivity of primal-dual dynamics and distributed optimization in Section IV and NE-Seeking dynamics in Section V.

II. PRELIMINARIES

Notation : Given $x \in \mathbb{R}^n$, we define $[x] \in \mathbb{R}^{n \times n}$ to be a diagonal matrix whose diagonal entries equal to x. We denote a vector of ones by $\mathbf{1}_n \in \mathbb{R}^n$. For a matrix $A \in \mathbb{R}^{n \times n}$, let $\alpha(A)$ denote its spectral abscissa. We denote the inner product $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$.

A. Semicontraction theory

Definition 1 (Seminorm): A function $\|\|\cdot\|\| : \mathbb{R}^n \to \mathbb{R}$ is a seminorm on $\mathbb{R}^n$ if , for all $v, w \in \mathbb{R}^n$ and $a \in \mathbb{R}$:

(i) $\|\|av\|\| = |a|\|\|v\|\|$.
(ii) $\|\|v + w\|\| \leq \|\|v\|\| + \|\|w\|\|$.

The kernel of a seminorm $\|\|\cdot\|\|$ is given by $\mathcal{K} = \{v \in \mathbb{R}^n : \|\|v\|\| = 0\}$.

Definition 2 (Matrix induced Seminorm): Let $\|\|\cdot\|\|$ be a seminorm on $\mathbb{R}^n$. Then, the induced matrix seminorm associated with $\|\|\cdot\|\|$ is given by

$$\|\|A\|\| = \max\{\|\|Av\|\| \mid \|\|v\|\| = 1 \text{ and } v \perp \mathcal{K}\}.$$

Definition 3 (Matrix log seminorm): Let $\|\|\cdot\|\|$ be a seminorm on $\mathbb{R}^n$ and its corresponding induced seminorm on $\mathbb{R}^{n \times n}$. The *log seminorm* associated with $\|\|\cdot\|\|$ is defined by

$$\mu_{\|\|\cdot\|\|}(A) = \lim_{h \to 0^+} \frac{\|\|I + hA\|\| - 1}{h}.$$

Definition 4 (Semicontracting and contracting systems): A system $\dot{x} = F(x)$ is said to be

(i) strongly infinitesimally semicontracting with respect to a seminorm $\|\|\cdot\|\|$ with rate $c \geq 0$ if

$$\sup_x \mu_{\|\|\cdot\|\|}(DF(x)) \leq -c. \tag{1}$$

(ii) strongly infinitesimally contracting with respect to a norm $\| \cdot \|$ with rate $c \geq 0$ if (1) holds and $\|\|\cdot\|\| = \| \cdot \|$.

In this paper, we will consider weighted Euclidean seminorm, i.e., seminorms of the form $\|\|x\|\| = \|Rx\|_2$, where R is some matrix. We present an equivalent condition for the semicontraction of a system in such norms presented in [9].

Lemma 5 (Demidovich Condition): Let $\|\|\cdot\|\|$ be a seminorm associated with a log seminorm μ, $R \in \mathbb{R}^{k \times n}, k \leq n$ be a full rank matrix, and $P = R^\top R \in \mathbb{R}^{n \times n}$. For each $A \in \mathbb{R}^{n \times n}$, and $c \in \mathbb{R}$ and if $\ker(R)$ is invariant under A, then,

$$\mu_{2,R}(A) \leq c \iff PA + A^\top P \preceq 2cP$$

We also consider the following result relating semicontracting systems with an invariance property and contracting systems from [7].

Lemma 6 (Semicontraction with invariance property): Consider a system $\dot{x} = f(x)$. Let $\mathcal{K} \subset \mathbb{R}^n$ be such that $\mathcal{K}^\perp$ is an $f-$invariant subspace. Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be strongly infinitesimally semicontracting with rate $c > 0$ with respect to a seminorm $\|\|\cdot\|\|$ on $\mathbb{R}^n$ with kernel $\mathcal{K}$. Then, the system admits the cascade decomposition,

$$\dot{x}_\| = f_\|(x_\| + x_\perp) \tag{2a}$$
$$\dot{x}_\perp = f_\perp(x_\perp), \tag{2b}$$

and the perpendicular dynamics (2b) are strongly infinitesimally contracting on $\mathcal{K}_\perp$ with rate c, with respect to the norm given by the domain restriction of $\|\|\cdot\|\|$ to $\mathcal{K}$.

B. Network Interconnection Theorem

Theorem 7 (Network Contraction Theorem [3, Theorem 3.22]): Consider an interconnection of r dynamical systems,

$$\dot{x}_i = f_i(t, x_i, x_{-i}) \quad \text{for } i \in \{1, 2, \ldots, r\}$$

where $x_i \in \mathbb{R}^{n_i}$ and $x_{-i} \in \mathbb{R}^{n-n_i}$. We let x_{-i} denote the vector with all components x_j of x, except x_i. We assume

(i) at fixed x_{-i} and t, each map $x_i \to f_i(t, x_i, x_{-i})$ is strongly infinitesimally contracting with rate c_i with respect to $\| \cdot \|_i$.
(ii) at fixed x_i and t, each map $x_{-i} \to f_i(t, x_i, x_{-i})$ is Lipschitz with constant γ_{ij}.
(iii) The gain matrix

$$\Gamma = \begin{bmatrix} -c_1 & \cdots & \gamma_{1r} \\ \vdots & \ddots & \vdots \\ \gamma_{r1} & \cdots & -c_r \end{bmatrix}$$

is Hurwitz.

Then, for every $\epsilon \in]0, \alpha(\Gamma)[$, there exists a norm such that the interconnected system is strongly infinitesimally contracting with a rate $|\alpha(\Gamma) + \epsilon|$.

III. LOGARITHMIC NORM OF SADDLE MATRICES WITH REDUNDANT CONSTRAINTS, AND ASYMMETRY

We begin with a general result about Hurwitz saddle matrices. We improve upon the bound for the contraction rate presented in [6] using a similar proof technique, with extensions to allow for redundant constraints by modifying the weight matrix.

Theorem 8 (Semicontractivity of Saddle Matrices): Given $Q \in \mathbb{R}^{n \times n}$, $A \in \mathbb{R}^{m \times n}$, and a parameter $\tau > 0$, consider the *saddle matrix*

$$S = \begin{bmatrix} -Q & -A^\top \\ \tau^{-1}A & 0 \end{bmatrix} \in \mathbb{R}^{(m+n) \times (m+n)}.$$

(i) Let $\lambda_{\min}(Q + Q^\top)/2 = q_{\min} > 0$ and $Q^\top Q \preceq q_{\max}(Q + Q^\top)/2$, that is, $q_{\max} = \sigma_{\max}^2(Q)/q_{\min}$.
(ii) Let $AA^\top \succeq a_{\min}\Pi_A$, with $a_{\min} > 0$, and $AA^\top \preceq a_{\max}I_m$, where the matrix $\Pi_A \in \mathbb{R}^{m \times m}$ is the orthogonal projection onto the image of A, that is, in the linear map interpretation, $\Pi_A \colon \mathbb{R}^m \to \text{img}(A)$. Note $a_{\min} \leq a_{\max}$.

The following *semicontractivity LMI* holds:

$$\mathcal{S}^\top P + P\mathcal{S} \preceq -2cP, \tag{3}$$

where $P \in \mathbb{R}^{(n+m)\times(n+m)}$ is defined by

$$P = \begin{bmatrix} I_n & \alpha A^\top \\ \alpha A & \tau \Pi_A \end{bmatrix} \succeq 0 \tag{4}$$

such that $\mathcal{S}\ker(P) \subseteq \ker(P)$, with

$$\alpha = \frac{1}{2}\min\left\{\frac{1}{q_{\max}}, \tau\frac{q_{\min}}{a_{\max}}\right\}, \quad \text{and} \quad c = \frac{1}{2}\tau^{-1}\alpha a_{\min}.$$

Proof: The proof is deferred to the Appendix A. ∎

Next, we discuss methods to sharpen the rate obtained. First, we identify an α that achieves a better rate, in exchange for a more complicated expression. Notably, this rate improves upon the previous rate by a factor of $4/3$ when τ is arbitrarily small, i.e. the dual dynamics are much faster than the primal dynamics.

Theorem 9 (Semicontractivity of Saddle Matrices for small τ): Given $Q \in \mathbb{R}^{n\times n}$, $A \in \mathbb{R}^{m\times n}$, and a parameter $\tau > 0$, consider the *saddle matrix*

$$\mathcal{S} = \begin{bmatrix} -Q & -A^\top \\ \tau^{-1}A & 0 \end{bmatrix} \in \mathbb{R}^{(m+n)\times(m+n)}.$$

(i) Let $\lambda_{\min}(Q+Q^\top)/2 = q_{\min} > 0$ and $Q^\top Q \preceq q_{\max}(Q+Q^\top)/2$, that is, $q_{\max} = \sigma^2_{\max}(Q)/q_{\min}$. Note $q_{\min} \le q_{\max}$.

(ii) Let $AA^\top \succeq a_{\min}\Pi_A$, with $a_{\min} > 0$, and $AA^\top \preceq a_{\max}I_m$, where the matrix $\Pi_A \in \mathbb{R}^{m\times m}$ is the orthogonal projection onto the image of A, that is, in the linear map interpretation, $\Pi_A \colon \mathbb{R}^m \to \mathrm{img}(A)$. Note $a_{\min} \le a_{\max}$.

Then, the following *semicontractivity LMI* holds:

$$\mathcal{S}^\top P + P\mathcal{S} \preceq -2cP,$$

where $P \in \mathbb{R}^{(n+m)\times(n+m)}$ is defined by as per equation (4). The rate $c = \frac{1}{2}\tau^{-1}\alpha a_{\min}$, and

$$\alpha = \min\left\{\frac{\epsilon}{q_{\max}}, \frac{2-\epsilon}{3}\tau\frac{q_{\min}}{a_{\max}}\right\}$$

for $0 < \epsilon < 2$.

Proof: The proof follows similar to that of Theorem 8, and the differences in proofs are discussed in Appendix B. ∎

Remark 10: In the case where $\tau \to 0$, we get an improved rate of convergence c that approaches $\frac{1}{3}\frac{a_{\min}}{a_{\max}}q_{\min}$, for an appropriately chosen ϵ.

Finally, we identify a sharper bound that depends on a quadratic constraint, complicating the expression of the bound even further.

Theorem 11 (Semicontractivity of Saddle Matrices with a sharper rate): Given $Q \in \mathbb{R}^{n\times n}$, $A \in \mathbb{R}^{m\times n}$, and a parameter $\tau > 0$, consider the *saddle matrix*

$$\mathcal{S} = \begin{bmatrix} -Q & -A^\top \\ \tau^{-1}A & 0 \end{bmatrix} \in \mathbb{R}^{(m+n)\times(m+n)}.$$

(i) Let $\lambda_{\min}(Q+Q^\top)/2 = q_{\min} > 0$ and $Q^\top Q \preceq q_{\max}(Q+Q^\top)/2$, that is, $q_{\max} = \sigma^2_{\max}(Q)/q_{\min}$. Note $q_{\min} \le q_{\max}$.

(ii) Let $AA^\top \succeq a_{\min}\Pi_A$, with $a_{\min} > 0$, and $AA^\top \preceq a_{\max}I_m$, where the matrix $\Pi_A \in \mathbb{R}^{m\times m}$ is the orthogonal projection onto the image of A, that is, in the linear map interpretation, $\Pi_A \colon \mathbb{R}^m \to \mathrm{img}(A)$. Note $a_{\min} \le a_{\max}$.

Then, the following *semicontractivity LMI* holds:

$$\mathcal{S}^\top P + P\mathcal{S} \preceq -2cP,$$

where $P \in \mathbb{R}^{(n+m)\times(n+m)}$ is defined by as per equation (4). The rate $c = \frac{1}{2}\tau^{-1}\alpha a_{\min}$, and α depends on the roots of the following quadratic equation,

$$2 - \alpha(q_{\max} + 3\tau^{-1}\frac{a_{\max}}{q_{\min}}) + \tau^{-1}\alpha^2 a_{\min} = 0. \tag{5}$$

This quadratic has two roots, $\beta_1 < \beta_2$, and $\alpha = \min\left\{\beta_1, \frac{\tau q_{\min}}{a_{\min}}\right\}$.

Proof: The proof follows similar to that of Theorem 8, and the differences in proofs are discussed in Appendix C. ∎

Our rate improves upon the rate presented in [18]. In order to work with a simple interpretable expression, our analysis in the rest of the paper uses the result in Theorem 8. Since our norm allows the use of redundant constraints, we may consider problems with a consensus constraint over a graph, such as distributed optimization problems. Further, nonsymmetry of the $(1,1)$ block in the saddle matrix finds applications in NE-seeking dynamics in games.

IV. CONTRACTIVITY IN OPTIMIZATION DYNAMICS

In this section, we study the implications of Theorem 8 on constrained optimization problems. Specifically, we establish contractivity for the linearly constrained primal-dual flow and the distributed optimization problem. We prove contractivity on a subspace which includes the primal space by proving semicontractivity on the full space. Apart from guaranteeing convergence, showing contractivity leads to useful properties, such as robustness to noise, delays, and strong guarantees over equilibrium tracking and periodic entrainment to time-varying parameters [3, Section 3.4].

A. Standard Primal-Dual Dynamics

Consider the primal-dual flow for a linearly constrained optimization problem [1],

$$\dot{x} = -\nabla f(x) - A^\top \lambda \tag{6a}$$
$$\tau\dot{\lambda} = Ax - b \tag{6b}$$

This system, when linearized results in a saddle matrix, therefore, we may use Theorem 8 in order to show that these dynamics are semicontracting.

Lemma 12 (Semicontraction of primal-dual dynamics): If the function $f(x)$ is strongly convex with parameter μ and Lipschitz with constant ℓ, and $a_{\min}, a_{\max}$ respectively are the smallest and largest nonzero eigenvalues of $AA^\top$, then the dynamics (6) that solve the primal-dual optimization problem in continuous time are semicontracting in the Euclidean norm weighted by P, where P is defined in (4), with a rate

$$c = \frac{1}{4}\min\left\{\frac{a_{\min}}{\tau\ell}, \frac{a_{\min}}{a_{\max}}\mu\right\}.$$

Proof: The Jacobian of the system is given by

$$\mathcal{S} = \begin{bmatrix} -\nabla^2 f(x) & -A^\top \\ \tau^{-1}A & 0 \end{bmatrix}.$$

The Jacobian is in the form of a saddle matrix, and $\mu I \preceq \nabla^2 f(x) \preceq \ell I$. Using Theorem 8, we get the required result. ∎

B. Distributed Optimization Dynamics

In our distributed optimization setup, a team of agents seeks to minimize a decomposable function of the form $f(x) = \sum_{i=1}^{N} f_i(x)$, where $f_i : \mathbb{R}^n \to \mathbb{R}$, and the sum of all cost functions is *strongly convex*. In this setup, each agent has a local estimate $x_{[i]} \in \mathbb{R}^n$ of the solution to the problem, and agents exchange information over a network with a *strongly connected graph* with a Laplacian L. This setup can be represented as the following optimization problem.

$$\min_{x \in \mathbb{R}^{nN}} \quad \sum_{i=1}^{N} f_i(x_{[i]})$$
$$\text{subject to} \quad (L \otimes I_n)x = 0$$

We observe that this is similar to a linearly constrained convex optimization problem. Therefore, we may use a primal-dual flow to solve this system. The corresponding dynamics to solve this problem are of the form,

$$\dot{x} = -\nabla f(x) - (L \otimes I_n)^\top \lambda \tag{8a}$$
$$\tau\dot{\lambda} = (L \otimes I_n)x \tag{8b}$$

Note that these dynamics allow the graph to be directed. However, to obtain an accurate minimizer for the cost function, the graph must be strongly connected. We also make the following assumption about the cost function.

Assumption 13: $\sum_{i=1}^{N} f_i(x_{[i]})$ is twice differentiable, strongly convex with parameter μ and is Lipschitz in x with parameter ℓ.

We may now study dynamics (8) using Theorem 8.

Theorem 14 (Distributed Optimization with Laplacian Constraints): The continuous time dynamics (8) are semi-contracting in a weighted L_2 norm with weight

$$P = \begin{bmatrix} I & \alpha(L \otimes I_n)^\top \\ \alpha(L \otimes I_n) & \tau\Pi_{L \otimes I_n} \end{bmatrix}$$

with a rate

$$c = \frac{1}{4}\min\left\{\frac{\lambda_2^2}{\tau\ell}, \frac{\lambda_2^2}{\lambda_n^2}q_{\min}\right\}.$$

where λ_2 and λ_n are the smallest and largest non-zero eigenvalues of L respectively.

Proof: The linearized version of this system is

$$\begin{bmatrix} \dot{\delta x} \\ \dot{\delta \lambda} \end{bmatrix} = \begin{bmatrix} -\nabla^2 f(x) & -(L \otimes I_n)^\top \\ \tau^{-1}(L \otimes I_n) & 0 \end{bmatrix} \begin{bmatrix} \delta x \\ \delta \lambda \end{bmatrix}$$

We know that $\nabla^2 f(x)$ is strongly convex with parameter μ and Lipschitz with parameter ℓ. Therefore, $q_{\min} = \mu, q_{\max} = \ell$. Further, from the properties of the Kronecker product,

$$(L \otimes I)(L \otimes I_n)^\top = (L \otimes I_n)(L^\top \otimes I_n) = LL^\top \otimes I_n$$

and $LL^\top \otimes I_n$ has the same eigenvalues as the $LL^\top$, but each with multiplicity n times the original multiplicity. The result follows from an application of Theorem 8. ∎

Alternatively, we may model the graph using the incidence matrix $B \in \mathbb{R}^{N \times M}$. Each column of the incidence matrix $B \in \mathbb{R}^{N \times M}$ is associated with an edge in the original graph, and each row is associated with a row in the original matrix. The problem is now represented as,

$$\min_{x \in \mathbb{R}^{nN}} \quad \sum_{i=1}^{N} f_i(x_{[i]})$$
$$\text{subject to} \quad (B^\top \otimes I_n)x = 0$$

The corresponding flow dynamics for this problem are,

$$\dot{x} = -\nabla f(x) - (B^\top \otimes I_n)^\top \lambda \tag{9a}$$
$$\tau\dot{\lambda} = (B^\top \otimes I_n)x \tag{9b}$$

Theorem 15 (Distributed Optimization with incidence matrix constraints): The continuous time dynamics (9) are semicontracting in a weighted L_2 norm with weight

$$P = \begin{bmatrix} I & \alpha(B^\top \otimes I_n)^\top \\ \alpha(B^\top \otimes I_n) & \tau\Pi_{L \otimes I_n} \end{bmatrix}$$

with a rate

$$c = \frac{1}{4}\min\left\{\frac{\lambda_2}{\tau\ell}, \frac{\lambda_2}{\lambda_n}q_{\min}\right\}.$$

where λ_2 and λ_n are the smallest and largest non-zero eigenvalues of L respectively.

Proof: The linearized version of this system is

$$\begin{bmatrix} \dot{\delta x} \\ \dot{\delta \lambda} \end{bmatrix} = \begin{bmatrix} -\nabla^2 f(x) & -(B \otimes I_n) \\ \tau^{-1}(B^\top \otimes I) & 0 \end{bmatrix} \begin{bmatrix} \delta x \\ \delta \lambda \end{bmatrix}$$

We know that $\nabla^2 f(x)$ is strongly convex with parameter μ and Lipschitz with parameter ℓ. Therefore, $q_{\min} = \mu, q_{\max} = \ell$. Further, from the properties of the Kronecker product,

$$(B^\top \otimes I_n)(B \otimes I_n) = B^\top B \otimes I_n$$

Now, we know that $B^\top B$ and $BB^\top$ have the same nonzero eigenvalues. Further, $L = BB^\top$. Therefore, $B^\top B \otimes I_n$ has the same nonzero eigenvalues as the L, but each with multiplicity n times the original multiplicity. The result follows from an application of Theorem 8. ∎

Remark 16: In both Theorem 14 and 15, the contraction rate depends on λ_2/λ_n. Notably, this parameter appears also in the study of oscillator networks and their optimal synchronizability [4] via the master stability function approach [2]. A higher value for λ_2/λ_n implies both better synchronizability in oscillator networks as well as an improved estimate for the contraction rate in distributed optimization.

C. Numerical Simulations

The lower bound we achieve for the convergence rate using the incidence matrix is faster than the Laplacian constraints, as $\lambda_2/\lambda_n < 1$. To study this comparison further, we considered a simple distributed optimization problem, where each agent attempts to minimize a local function $q_i(x_i - v_i)^2$, for $x_i, v_i, q_i \in \mathbb{R}, q_i > 0$, and v_i and q_i are
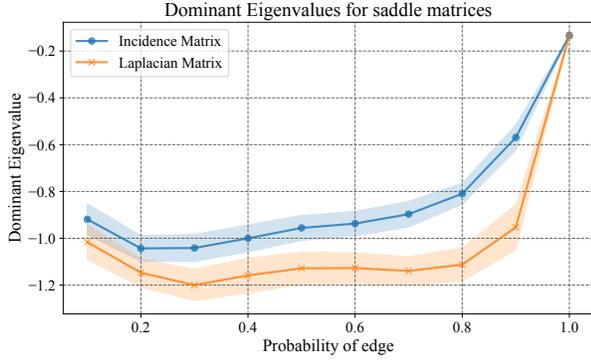
Fig. 1. Plot comparing mean dominant eigenvalues of saddle matrices from Laplacian and incidence matrix constraints in a distributed optimization setup on Erdős-Rényi graphs with varying edge probabilities.

sampled from uniform distributions in $[0, 10]$, and τ is chosen sufficiently small. The underlying network is a symmetric connected Erdős-Rényi graph with $N = 40$ nodes representing agents, with varying edge probability parameters. We construct the saddle matrix of this graph using both the incidence and Laplacian constraints and compute their dominant eigenvalue. On considering 50 graphs for each probability value, the average dominant eigenvalue and its confidence bound are presented in Figure 1. The Laplacian-based dynamics are faster than the incidence matrix-based dynamics, which implies that the bounds we obtain for the semi-lognorms are weak for this example system. There appears to be an inherent tradeoff between the generality of a bound and its tightness, as observed here, and it remains an open problem to identify better bounds for specific cases.

V. CONTRACTIVITY IN NASH SEEKING DYNAMICS

In this section, we discuss the contractivity of NE-seeking dynamics that are known to converge to a Nash equilibrium. Contractivity in such systems can be used to prove the robustness of such equilibrium points to delays and noise.

We consider a multiplayer continuous-time game among N players with perfect information exchange. At each instant of time, players pick an action $x_i \in \mathbb{R}^n$ in order to minimize a cost $J_i : \mathbb{R}^n \times \mathbb{R}^{n(N-1)} \to \mathbb{R}$, which is a function of both the player's actions and the actions of other players. We consider both pseudogradient and best response plays, and use the network contraction theorem to find conditions where they represent strongly infinitesimally contracting dynamics.

First, we consider the evolution of pseudogradient plays. The pseudogradient dynamics are given by

$$\dot{x}_i = \mathrm{F_{PseudoG}}(x) = -\nabla_{x_i} J_i(x_i, x_{-i}) \qquad (10)$$

where x_{-i} represents the current actions of the other players in the game.

Theorem 17 (Contraction of Pseudogradient Plays): Consider a game with N agents, where each agent aims to minimize a cost $J_i : \mathbb{R}^n \times \mathbb{R}^{n(N-1)} \to \mathbb{R}$, following the dynamics (10). Then, if J_i is

(i) μ_i strongly convex with respect to x_i for each i,

(ii) Lipschitz with respect to x_j for each $j \neq i$ with a constant ℓ_{ij}, and

(iii) the gain matrix for the system,

$$\Gamma_{\mathrm{PseudoG}} = \begin{bmatrix} -\mu_1 & \dots & \ell_{1r} \\ \vdots & \ddots & \vdots \\ \ell_{r1} & \dots & -\mu_r \end{bmatrix} \text{ is Hurwitz.}$$

Then the system is infinitesimally contracting with respect to some norm, with a rate $|\alpha(\Gamma) + \epsilon|$, for $-\alpha(\Gamma) > \epsilon > 0$.

Proof: The proof for this statement is a direct application of the network contraction theorem. ∎

Next, we consider the evolution of unconstrained best response plays. We construct a best response cost function, $\mathrm{BR}_i : x_i \to \arg\min_{x_i} J_i(x_i, x_{-i})$, which represents the best response of player i with respect to the actions of other players x_{-i}. We represent the best response dynamics,

$$\dot{x} = \mathrm{F_{BR}}(x) = \mathrm{BR}(x) - x \iff \dot{x}_i = \mathrm{BR}_i(x_{-i}) - x_i \quad (11)$$

Theorem 18 (Contraction of Best Response Plays): Consider a game with N agents, where each agent aims to minimize a cost $J_i : \mathbb{R}^n \times \mathbb{R}^{n(N-1)} \to \mathbb{R}$, following the dynamics (11). Then, if J_i is

(i) μ_i strongly convex with respect to x_i for each i,

(ii) Lipschitz with respect to x_j for each $j \neq i$ with a constant ℓ_{ij}, and

(iii) the gain matrix for $\mathrm{F_{BR}}$, $\Gamma_{\mathrm{BR}} = \begin{bmatrix} -1 & \dots & \frac{\ell_{1r}}{\mu_1} \\ \vdots & \ddots & \vdots \\ \frac{\ell_{r1}}{\mu_n} & \dots & -1 \end{bmatrix}$ is Hurwitz.

Then the system is infinitesimally contracting with respect to some norm, with a rate $|\alpha(\Gamma) + \epsilon|$, for $-\alpha(\Gamma) > \epsilon > 0$.

Proof: The gain matrix may be computed using the results in [6, Lemma 1], and [3, Equation (3.39)]. On using the network contraction theorem for this system, we get our required results. ∎

Next, we show a result about the equivalences of the sufficient conditions for contraction in pseudogradient and best response plays.

Theorem 19 (Best Response-Pseudogradient Equivalence): If each J_i is

(i) μ_i strongly convex with respect to x_i for each i,

(ii) Lipschitz with respect to x_j for each $j \neq i$ with a constant ℓ_{ij}

The following statements are equivalent

(i) The gain matrix for $\mathrm{F_{PseudoG}}$, $\Gamma_{\mathrm{PseudoG}}$ is Hurwitz.

(ii) The gain matrix for $\mathrm{F_{BR}}$, Γ_{BR} is Hurwitz.

(iii) The discrete-time $\mathrm{F_{BR}}$ gain matrix

$$\Gamma_{\mathrm{BR,d}} = \begin{bmatrix} 0 & \dots & \frac{\ell_{1r}}{\mu_1} \\ \vdots & \ddots & \vdots \\ \frac{\ell_{r1}}{\mu_n} & \dots & 0 \end{bmatrix} \text{ is Schur.}$$

Proof: First, we note that $\Gamma_{\mathrm{PseudoG}} = [\mu]\Gamma_{\mathrm{BR}}$, where $\mu = [\mu_1, \mu_2, \cdots, \mu_r]^\top \in \mathbb{R}^n$. Next, we prove $(i) \implies (ii)$. Since $\Gamma_{\mathrm{PseudoG}}$ is Metzler and Hurwitz, $\Gamma_{\mathrm{PseudoG}}$ is Lyapunov diagonally stable, i.e. there exists a diagonal matrix $[p]$ such that $\Gamma_{\mathrm{PseudoG}}^\top [p] + [p]\Gamma_{\mathrm{PseudoG}} \preceq 0$. Using the relationship between

Γ_{PseudoG} and Γ_{BR}, we note that Γ_{BR} is Lyapunov diagonally stable with weight $[p][\mu]$. Therefore, Γ_{BR} is Hurwitz. We may use a similar argument to show that $(ii) \implies (i)$.

Next, to show that $(ii) \iff (iii)$, we note that $\Gamma_{\text{BR}} = -I + \Gamma_{\text{BR,d}}$, and $\Gamma_{\text{BR,d}}$ is a non-negative matrix. We know that Γ_{BR} is Hurwitz iff $\alpha(\Gamma_{\text{BR}}) < 0$. Further, since $\Gamma_{\text{BR,d}}$ is non-negative, due to the Perron-Frobenius Theorem, we have $\alpha(\Gamma_{\text{BR}}) < 0$ iff $\rho(\Gamma_{\text{BR,d}}) < 1$. ∎

Next, we discuss the contractivity of aggregative games.

Theorem 20 (Aggregative games): For a pseudogradient play, If $J_i(x_i, x_{-i})$ is of the form $f_i(x_i, \frac{1}{n}\sum_{j=1}^{n} x_j)$, and for each agent i, if

(i) f_i is μ_i-strongly convex with respect to its first argument
(ii) f_i is Lipschitz in its second argument with constant ℓ_i,
(iii) $\mu_i > \ell_i$.

Then, the gain matrix of the system is Hurwitz, and the system is contracting.

Proof: First, we claim that the Lipschitz constant for in x_j for $j \neq i$ is ℓ_i/n. This follows from the fact that the second argument of f_i varies linearly with constant $\frac{1}{n}$ in x_j, and our assumption regarding the Lipschitz constant of f_i. Further, since x_i appears both in the first and second argument, we may claim that

$$\langle x_i - y_i, \nabla f(x_i, \frac{x_i}{n} + \frac{1}{n}\sum_{j\neq i} x_j) - \nabla f_i(y_i, y_i + \frac{1}{n}\sum_{j\neq i} x_j)\rangle$$
$$\leq -\mu_i \|x_i - y_i\|^2 + \frac{\ell_i}{n}\|x_i - y_i\|^2.$$

Now, we may construct the gain matrix of an aggregative game, $\Gamma_{\text{Agg}} \in \mathbb{R}^{n \times n}$,

$$(\Gamma_{\text{Agg}})_{ij} = \begin{cases} -\mu_i + \ell_i/n & i = j, \\ \ell_i/n & i \neq j \end{cases}$$

Since $\mu_i > \ell_i$, $\Gamma_{\text{Agg}}\mathbf{1}_n < 0$, in an elementwise manner. Using properties of Metzler matrices [3, Theorem 2.7], Γ_{Agg} is Hurwitz, and the system is contracting by Theorem 17. ∎

VI. Conclusion

In this work, we propose the use of contraction theory as a framework to analyze game theoretic and distributed optimization dynamics. We identify a general result for saddle matrices and use it to show contraction in distributed optimization dynamics. Our numerical results suggest that our estimate for the contraction rate is not sharp. Next, we use the network interconnection theorem 7 to show contractivity in some basic game setups. We believe that this work opens up new avenues to explore in both distributed optimization and game theory. In distributed optimization, we plan to study other distributed optimization flows [21]. In game theory, we wish to study setups with partial information and fictitious play [19]. We also intend to relax our assumption on strong convexity in future work.

Appendix

A. Proof of Theorem 8

We begin by using the Schur Complement to show that $P \succeq 0$. Clearly the $(1,1)$ block is positive definite. Therefore,

$$P \succeq 0 \iff \tau\Pi_A - \alpha^2 AA^\top \succ 0$$
$$\iff \tau - \alpha^2 a_{\max} > 0 \iff \alpha^2 < \tau/a_{\max}.$$

The inequality $\alpha^2 < \tau/a_{\max}$ follows from the stronger inequality $(2\alpha)^2 < \tau/a_{\max}$ with the following argument:

$$(2\alpha)^2 \leq \min\left\{\frac{1}{q_{\max}}, \tau\frac{q_{\min}}{a_{\max}}\right\} \cdot \max\left\{\frac{1}{q_{\max}}, \tau\frac{q_{\min}}{a_{\max}}\right\}$$
$$= \frac{q_{\min}}{q_{\max}} \cdot \frac{\tau}{a_{\max}} \leq \frac{\tau}{a_{\max}}.$$

Finally, we need to show that $\mathcal{S}\ker(P) \subseteq \ker(P)$. Let $[x^\top \ y^\top]^\top \in \ker(P)$. Then $x = -\alpha A^\top y$ and $y \in \ker(-\alpha AA^\top + \tau\Pi_A)$. Note that Π_A has the same kernel as $AA^\top$. Therefore, $y \in \ker(AA^\top) = \ker(A^\top)$. Next, to show the invariance of the norm with respect to the dynamics as per Lemma 6,

$$\mathcal{S}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -A^\top y \\ 0 \end{bmatrix} = 0 \in \ker(P).$$

Next, we aim to show the LMI (3). After some bookkeeping, we compute $\mathcal{Q} = -\mathcal{S}^\top P - P\mathcal{S} - 2cP$, and using the fact that $\Pi_A A = A$,

$$\mathcal{Q} = \begin{bmatrix} Q + Q^\top - 2\tau^{-1}\alpha A^\top A - 2cI_n & \alpha Q^\top A^\top - 2c\alpha A^\top \\ \alpha AQ - 2c\alpha A & 2\alpha AA^\top - 2c\tau\Pi_A \end{bmatrix}.$$

The (2,2) block satisfies the lower bound

$$2\alpha AA^\top - 2c\tau\Pi_A = 2\left(\frac{1}{2}\alpha AA^\top - c\tau\Pi_A\right) + \alpha AA^\top$$
$$\succeq 2\left(\frac{1}{2}\alpha a_{\min} - c\tau\right)\Pi_A + \alpha AA^\top$$
$$= \alpha AA^\top \succ 0.$$

Given this lower bound, we can factorize the resulting matrix as follows, setting $\mathcal{A} = \begin{bmatrix} I_n & 0 \\ 0 & A \end{bmatrix}$,

$$\mathcal{Q} \succeq$$
$$\mathcal{A}\begin{bmatrix} Q + Q^\top - 2(\tau^{-1}\alpha A^\top A + cI_n) & \alpha Q^\top - 2c\alpha I_n \\ \alpha Q - 2c\alpha I_n & \alpha I_n \end{bmatrix}\mathcal{A}^\top.$$

Since $\alpha I_n \succ 0$, it now suffices to show that the Schur complement of the (2,2) block of $n \times n$ matrix is positive semidefinite. A sufficient condition for the positive semidefiniteness of the Schur complement of the $(2,2)$ block is,

$$Q + Q^\top - \alpha Q^\top Q \succeq 2(\tau^{-1}\alpha A^\top A + cI_n) \quad (12)$$
$$\text{and} \quad 2\alpha c(Q + Q^\top) \succeq 4\alpha c^2 I_n. \quad (13)$$

To prove (12), we upper bound the right-hand side as follows:

$$2(\tau^{-1}\alpha A^\top A + cI_n) \preceq 2(\tau^{-1}\alpha a_{\max} + c)I_n$$
$$= \tau^{-1}\alpha(2a_{\max} + a_{\min})I_n$$
$$\preceq \frac{1}{2}\frac{q_{\min}}{a_{\max}}(2a_{\max} + a_{\min})I_n \preceq \frac{3}{2}q_{\min}I_n.$$

where the second equality follows from the definition of c and the last inequality follows as $\alpha \leq \frac{1}{2}\tau q_{\min}/a_{\max}$. Next, since $\alpha \leq \frac{1}{2q_{\max}}$, we know $-\alpha q_{\max} \geq -\frac{1}{2}$. We then lower bound the left hand side of (12) as follows:

$$Q + Q^\top - \alpha Q^\top Q \succeq Q + Q^\top - \alpha q_{\max}(Q + Q^\top)/2$$
$$\succeq (2 - \frac{1}{2})\frac{1}{2}(Q + Q^\top) \succeq \frac{3}{2}q_{\min}I_n.$$

Finally, it remains to prove (13) that is, $2\alpha c(Q + Q^\top) \succeq 4\alpha c^2 I_n$. This inequality is equivalent to $Q + Q^\top \succeq 2cI_n$ and follows from noting $c \leq \frac{1}{2}\frac{a_{\min}}{a_{\max}}q_{\min} < q_{\min}$.

B. Proof of Theorem 9

The proof is similar to the previous Theorem 8, with one key difference. The proof that $\alpha^2 < \tau/a_{\max}$ is modified as follows,

$$\alpha^2 \leq$$
$$\min\left\{\frac{\epsilon}{q_{\max}}, \tau\frac{(2-\epsilon)q_{\min}}{3a_{\max}}\right\} \cdot \max\left\{\frac{\epsilon}{q_{\max}}, \tau\frac{(2-\epsilon)q_{\min}}{3a_{\max}}\right\}$$
$$= \frac{q_{\min}}{q_{\max}} \cdot \frac{\tau}{a_{\max}} \cdot \frac{\epsilon(2-\epsilon)}{3} \leq \frac{\tau}{3a_{\max}},$$

which is true since $\epsilon(2-\epsilon) \leq 1$ for $\epsilon \in (0,2)$. The rest of the proof follows similar to the original proof with some algebraic differences.

C. Proof of Theorem 11

The proof is very similar to the proof of Theorem 8 with the following differences. First, we identify a tighter sufficient condition than the one presented in (12) and (13).

$$(1 + \alpha c)(Q + Q^\top) - \alpha Q^\top Q \succeq 2(\tau^{-1}\alpha A^\top A + cI_n) \tag{14}$$

$$\text{and } \alpha c(Q + Q^\top) \succeq 4\alpha c^2 I_n. \tag{15}$$

The inequality (15), $\alpha c(Q + Q^\top) \succeq 4\alpha c^2 I_n$ is equivalent to $Q + Q^\top \succeq 2cI_n$ and follows from noting that $c \leq \frac{1}{2}q_{\min}$, since we assume $\alpha \leq \tau q_{\min}/a_{\min}$. On simplifying (14) using similar arguments to the previous proof, we obtain the following condition on α,

$$(2 - \alpha q_{\max} + \tau^{-1}\alpha^2 a_{\min})q_{\min} \geq 3\tau^{-1}\alpha a_{\max}. \tag{16}$$

Next, we note that

$$2 - \alpha\left(q_{\max} + 3\tau^{-1}\frac{a_{\max}}{q_{\min}}\right) + \tau^{-1}\alpha^2 a_{\min} < 0$$

for $\alpha = \sqrt{\frac{\tau}{a_{\max}}}$. This is because the inequality simplifies to

$$q_{\max}\sqrt{\frac{\tau}{a_{\max}}} + \sqrt{\frac{a_{\max}}{\tau}}\frac{3}{q_{\min}} > 2 + \frac{a_{\min}}{a_{\max}}$$

Now, through an AM-GM inequality, the LHS is lower bounded by $2\sqrt{3\frac{q_{\max}}{q_{\min}}} \geq 2\sqrt{3}$, and the RHS is upper bounded by 3. Since $2\sqrt{3} > 3$, the quadratic equation is guaranteed to have a root. Since $\alpha \leq \beta_1 < \sqrt{\frac{\tau}{a_{\max}}}$, the quadratic constraint is guaranteed to hold, along with a guarantee of the positive definiteness of P.

REFERENCES

[1] K. J. Arrow, L. Hurwicz, and H. Uzawa, editors. *Studies in Linear and Nonlinear Programming*. Standford University Press, 1958.

[2] M. Barahona and L. M. Pecora. Synchronization in small-world systems. *Physical Review Letters*, 89:054101, 2002. doi:10.1103/PhysRevLett.89.054101.

[3] F. Bullo. *Contraction Theory for Dynamical Systems*. Kindle Direct Publishing, 1.1 edition, 2023, ISBN 979-8836646806. URL: http://motion.me.ucsb.edu/book-ctds.

[4] G. Chen. Searching for best network topologies with optimal synchronizability: A brief review. *IEEE/CAA Journal of Automatica Sinica*, 9(4):573–577, 2022. doi:10.1109/JAS.2022.105443.

[5] J. Cortés and S. K. Niederländer. Distributed coordination for nonsmooth convex optimization via saddle-point dynamics. *Journal of Nonlinear Science*, 29(4):1247–1272, 2019. doi:10.1007/s00332-018-9516-4.

[6] A. Davydov, V. Centorrino, A. Gokhale, G. Russo, and F. Bullo. Contracting dynamics for time-varying convex optimization, 2023. arXiv:2305.15595.

[7] G. De Pasquale, K. D. Smith, F. Bullo, and M. E. Valcher. Dual seminorms, ergodic coefficients, and semicontraction theory. *IEEE Transactions on Automatic Control*, December 2022. doi:10.48550/arXiv.2201.03103.

[8] D. Gadjov and L. Pavel. A passivity-based approach to nash equilibrium seeking over networks. *IEEE Transactions on Automatic Control*, 64(3):1077–1092, 2019. doi:10.1109/TAC.2018.2833140.

[9] S. Jafarpour, P. Cisneros-Velarde, and F. Bullo. Weak and semicontraction for network systems and diffusively-coupled oscillators. *IEEE Transactions on Automatic Control*, 67(3):1285–1300, 2022. doi:10.1109/TAC.2021.3073096.

[10] S. Jafarpour, A. Davydov, A. V. Proskurnikov, and F. Bullo. Robust implicit networks via non-Euclidean contractions. In *Advances in Neural Information Processing Systems*, December 2021. doi:10.48550/arXiv.2106.03194.

[11] S. S. Kia, J. Cortes, and S. Martinez. Distributed convex optimization via continuous-time coordination algorithms with discrete-time communication. *Automatica*, 55:254–264, 2015. doi:10.1016/j.automatica.2015.03.001.

[12] S. Li and T. Başar. Distributed algorithms for the computation of noncooperative equilibria. *Automatica*, 23(4):523–533, 1987. doi:10.1016/0005-1098(87)90081-1.

[13] W. Lohmiller and J.-J. E. Slotine. On contraction analysis for nonlinear systems. *Automatica*, 34(6):683–696, 1998. doi:10.1016/S0005-1098(98)00019-3.

[14] A. Nedić and A. Ozdaglar. Distributed subgradient methods for multi-agent optimization. *IEEE Transactions on Automatic Control*, 54(1):48–61, 2009. doi:10.1109/TAC.2008.2009515.

[15] H. D. Nguyen, T. L. Vu, K. Turitsyn, and J.-J. E. Slotine. Contraction and robustness of continuous time primal-dual dynamics. *IEEE Control Systems Letters*, 2(4):755–760, 2018. doi:10.1109/LCSYS.2018.2847408.

[16] I. K. Ozaslan and M. R. Jovanović. On the global exponential stability of primal-dual dynamics for convex problems with linear equality constraints. In *American Control Conference*, pages 210–215, San Diego, USA, 2023. doi:10.23919/ACC55779.2023.10156504.

[17] L. Pavel. Dissipativity theory in game theory: On the role of dissipativity and passivity in Nash equilibrium seeking. *IEEE Control Systems Magazine*, 42(3):150–164, 2022. doi:10.1109/MCS.2022.3157119.

[18] G. Qu and N. Li. On the exponential stability of primal-dual gradient dynamics. *IEEE Control Systems Letters*, 3(1):43–48, 2019. doi:10.1109/LCSYS.2018.2851375.

[19] J. S. Shamma and G. Arslan. Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria. *IEEE Transactions on Automatic Control*, 50(3):312–327, 2005.

[20] W. Wang and J. J. Slotine. On partial contraction analysis for coupled nonlinear oscillators. *Biological Cybernetics*, 92(1):38–53, 2005. doi:10.1007/s00422-004-0527-x.

[21] T. Yang, X. Yi, J. Wu, Y. Yuan, D. Wu, Z. Meng, Y. Hong, H. Wang, Z. Lin, and K. H. Johansson. A survey of distributed optimization. *Annual Reviews in Control*, 47:278–305, 2019. doi:10.1016/j.arcontrol.2019.05.006.