On the exact convergence to Nash equilibrium in hypomonotone regimes under full and partial-information

Dian Gadjov and Lacra Pavel

Abstract—In this paper, we consider distributed Nash equilibrium seeking in monotone and hypomonotone games. We first assume that each player has knowledge of the opponents’ decisions and propose a passivity-based modification of the standard gradient-play dynamics, that we call “Heavy Anchor”. We prove that Heavy Anchor allows a relaxation of strict monotonicity of the pseudo-gradient, needed for gradient-play dynamics, and can ensure exact asymptotic convergence in merely monotone regimes. We extend these results to the setting where each player has only partial information of the opponents’ decisions. Each player maintains a local decision variable and an auxiliary state estimate, and communicates with their neighbours to learn the opponents’ actions. We modify Heavy Anchor via a distributed Laplacian feedback and show how we can exploit equilibrium-independent passivity properties to achieve convergence to a Nash equilibrium in hypomonotone regimes.

I. INTRODUCTION

Recent years have seen a flurry of research papers on distributed Nash equilibrium seeking, due to the increase of distributed systems. There are a broad range of networked scenarios that involve strategic interacting agents, where centralized approaches are not suitable. Some examples are demand-side management for smart grids, electric vehicles, competitive markets, network congestion control, power control and resource sharing in wireless/wired peer-to-peer networks, control, cognitive radio systems, etc.

Classically, Nash Equilibrium (NE) seeking algorithms assume that each player has knowledge of every other agent’s decision/action, the so called full-decision information setting. In this setting, there are many well known algorithms that find the NE under various assumptions. However, these methods typically require more complex computations, such as the forward-backward-forward algorithm, Tikhonov proximal-point algorithm in [12], inexact proximal best-response in [5], inexact proximal-point/resolvent computation (e.g. Douglas-Rachford splitting), [8]. Even though proximal-point algorithms or Douglas-Rachford splitting can achieve exact convergence to a NE, they are computationally expensive since each step involves solving an optimization problem. These methods are only applicable in games with easily computed prox (resolvent) operators, regularization methods, such as the Tikhonov regularization [12], or continuous-time mirror-descent dynamics, are simpler, but require diminishing step-sizes (very slow convergence), or ensure convergence to only an approximate NE. We emphasize that all these existing methods for monotone games, assume agents have perfect knowledge of the actions of the other agents. Additionally, none of these methods deal with hypomonotone games for either the full or partial information setting.

An extremum seeking method for continuous time monotone games has been proposed. However, this method only converges to an \( \epsilon \) neighborhood of the NE. A payoff based method for discrete time is recently proposed to find the NE in monotone games. Agents random perturbation their action and moves in the direction of improvement. The random nature of the algorithm with the diminishing step sizes result in a slow method to converge to the NE but at the benefit of just using payoff information.

Contributions. Recognizing the lack of results for (hypo)monotone games, in this paper, we consider NE seeking for games with a monotone or hypomonotone pseudo-gradient. We propose an algorithm we call “Heavy Anchor”, constructed by a passivity-based modification of the standard gradient-play dynamics. We demonstrate that in the full-decision information setting, Heavy Anchor ensures exact convergence to a NE for any positive parameter values. Additionally, we show that under a carefully chosen change of coordinates, and conditions on the parameters, Heavy Anchor converges in hypomonotone games. Furthermore, we extend the result to the partial-decision information setting, by using a distributed Laplacian feedback. More specifically, we prove convergence for monotone extended pseudo-gradient, or (hypo)monotone and inverse Lipschitz
pseudo-gradient. To the best of our knowledge these are the first such results in the literature. Lastly, we look at quadratic games, an important subclass of games, and derive tighter conditions for the full information (hypo)monotone setting and the partial information (hypo)monotone setting.

Heavy Anchor can be interpreted as modifying the standard gradient method with a term approximating the derivative of the agent’s own action as predictive term. We use the approximation as a frictional force to improve stability. Heavy Anchor is similar to [19], [20] used in saddle-point problems in the full information setting. However, [19], [20], approximate the derivative of the other agents’ actions. Furthermore, our convergence results are global, unlike the local results in [20].

In the physics literature, similar dynamics were investigated for stabilizing unknown equilibrium in chaotic systems [21], [22]. However, they do not provide a rigorous characterization describing when the equilibrium is stabilized. Finally, Heavy Anchor is also related to second-order dynamics used in the optimization literature, e.g. [23]. If we discretize Heavy Anchor and restrict the parameter values, we can recover the optimistic gradient-descent/ascent (OGDA) [24] or the shadow Douglas Rachford [25]. However, all these optimization methods assume that the map is the gradient of a convex function. This does not hold in a game typically - the game map a pseudo-gradient rather than a full gradient, and thus convergence results are not applicable in a game context. Moreover, all these results are for the full information setting.

In [26] we presented the algorithm and proved convergence for monotone games. In this paper, we extend our results to hypomonotone games and provide additional analysis of inverse Lipschitz operators. Furthermore, we derive tighter conditions for the class of quadratic games, which were not analyzed in [26].

The paper is organized as follows. Section II gives preliminary background. Section III formulates the problem, standing assumptions and introduces the NE seeking algorithm for the full information case. The convergence analysis is presented in Section IV. Section V presents the partial-information version of the algorithm. Section VI investigates a property we are calling “inverse Lipschitz”, critical in our analysis of the partial information setting and hypomonotone games. Section VII proves convergence for the partial information setting. Section VIII derives tighter conditions for the class of quadratic games. Section IX shows simulations of our proposed algorithm and concluding remarks are given in Section X.

Notations. For \( x \in \mathbb{R}^n \), \( x^T \) denotes its transpose and \( |x| = \sqrt{x^T x} \) the norm induced by inner product \( \langle \cdot, \cdot \rangle \). For a matrix \( A \in \mathbb{R}^{n \times n} \), \( \lambda(A) = \{\lambda_1, \ldots, \lambda_n\} \) and \( \sigma(A) = \{\sigma_1, \ldots, \sigma_n\} \) denotes its eigenvalue and singular value set, respectively. Given \( A, B \in \mathbb{R}^{n \times n} \), let \( A \succeq B \) denote that \( A - B \) is positive semidefinite.

For \( \mathcal{N} = \{1, \ldots, N\} \), \( \text{col}(x_i)_{i \in \mathcal{N}} = [x_i^T, \ldots, x_N^T]^T \) denotes the stacked matrix with \( x_i \) along the diagonal. \( I_n, 1_n \) and \( 0_n \) denote the identity matrix, the all-ones and the all-zeros vector of dimension \( n \), and \( \otimes \) denotes the Kronecker product. Lastly, we denote \( i = \sqrt{-1} \).

II. BACKGROUND

A. Monotone Operators

The following are from [8]. Let \( T : \mathcal{H} \to 2^\mathcal{H} \) be an operator, where \( \mathcal{H} \) is a Hilbert space. Its graph is denoted by \( \text{gra} T = \{(x, y) \in \mathcal{H} \times \mathcal{H} \mid y \in Tx\} \). An operator \( T \) is \( \mu \)-strongly monotone and monotone, respectively, if it satisfies, \( \langle Tx - Ty, x - y \rangle \geq \mu \|x - y\|^2 \) \( \forall x, y \in \mathcal{H} \), where \( \mu > 0 \) and \( \mu = 0 \), respectively. Additionally, we say an operator \( T \) is \( \mu \)-hypomonotone if \( \langle Tx - Ty, x - y \rangle \geq -\mu \|x - y\|^2 \) \( \forall x, y \in \mathcal{H} \), where \( \mu \geq 0 \). \( T \) is maximally monotone if \( \langle x, y \rangle \in \mathcal{H} \times \mathcal{H} \Rightarrow \forall (u, v) \in \text{gra} T \), \( \langle x - u, y - v \rangle \geq 0 \). The resolvent of a monotone operator \( T \) is denoted by \( J_{\alpha T} = (I + \alpha T)^{-1}, \alpha > 0 \), where Id is the identity operator. Fixed points of \( J_{\alpha T} \) are identical to zeros of \( T \) (Prop. 23.2, [8]). An operator \( T \) is \( L \)-Lipschitz if, \( ||Tx - Ty|| \leq L \|x - y\| \forall x, y \in \mathcal{H} \). An operator \( T \) is \( C \)-cocoercive (\( C \)-inverse strongly monotone) if \( \langle Tx - Ty, x - y \rangle \geq C \|Tx - Ty\|^2 \forall x, y \in \mathcal{H} \).

B. Equilibrium Independent Passivity

The following are from [27]. Consider a system,

\[
\begin{align*}
\dot{x} &= f(x, u) \\
y &= h(x, u)
\end{align*}
\]

with \( x \in \mathbb{R}^n, u \in \mathbb{R}^q \) and \( y \in \mathbb{R}^r \), \( f \) locally Lipschitz and \( h \) continuous. For a differentiable function \( V : \mathbb{R}^n \to \mathbb{R} \), the time derivative of \( V \) along solutions of (1) is denoted by \( \dot{V}(x) = \nabla^T V(x)f(x, u) \) or just \( \dot{V} \). Let \( \overline{\pi}, \overline{\tau}, \overline{\gamma} \) be an equilibrium condition, such that \( 0 = f(\overline{\pi}, \overline{\tau}) \), \( \overline{\gamma} = h(\overline{\pi}, \overline{\tau}) \). Equilibrium independent passivity (EIP) requires a system to be passive independent of the equilibrium point.

Definition 1: System (1) is Equilibrium Independent Passive (EIP) if it is passive with respect to \( \overline{\pi} \) and \( \overline{\gamma} \); that is for every \( \overline{\pi} \in \mathbb{U} \) there exists a differentiable, positive semi-definite storage function \( V : \mathbb{R}^n \to \mathbb{R} \) such that \( \dot{V}(\overline{\pi}) = 0 \) and \( \forall u \in \mathbb{R}^q, x \in \mathbb{R}^n, \dot{V}(x) \leq \langle y - \overline{\gamma}, u - \overline{\pi} \rangle \). The system is Output-strictly EIP if, \( \dot{V} \leq \langle y - \overline{\gamma}, u - \overline{\pi} \rangle - \rho \|y - \overline{\gamma}\|^2 \) where \( \rho > 0 \).

C. Graph Theory

Let the graph \( G = (\mathcal{N}, \mathcal{E}) \) describe the information exchange among a set \( \mathcal{N} \) of agents, where \( \mathcal{E} \subset \mathcal{N} \times \mathcal{N} \). If agent \( i \) can get information from agent \( j \), then \( (i, j) \in \mathcal{E} \) and agent \( j \) is in agent \( i \)'s neighbour set \( \mathcal{N}_i = \{j \mid (i, j) \in \mathcal{E}\} \). \( G \) is undirected when \( (i, j) \in \mathcal{E} \) if and only if \( (j, i) \in \mathcal{E} \). \( G \) is connected if there is a path between any two nodes. Let \( W = [w_{ij}] \in \mathbb{R}^{N \times N} \) be the weighted adjacency matrix, with \( w_{ij} > 0 \) if \( j \in \mathcal{N}_i \) and \( w_{ij} = 0 \) otherwise. Let \( D_{\mathcal{E}G} = \text{diag}(d_i)_{i \in \mathcal{N}}, \) where \( d_i = \sum_{j=1}^{N} w_{ij} \). Assume that \( W = W^T \) so the weighted Laplacian of \( G \) is \( L = D_{\mathcal{E}G} - W \). When \( G \) is connected and undirected, \( 0 \) is a simple eigenvalue of \( L, L1_N = 0, 1_N^T L = 0^T \), and all other eigenvalues are positive, \( 0 < \lambda_2(L) \leq \cdots \leq \lambda_N(L) \).
III. Problem Setup

Consider a set \( N = \{1, \ldots, N\} \) of \( N \) players (agents) involved in a game. Each player \( i \in N \) controls its action or decision \( x_i \in \Omega_i \subseteq \mathbb{R}^n \). The action set of all players is the Cartesian product \( \Omega = \prod_{i \in N} \Omega_i \subseteq \mathbb{R}^n \), \( n = \sum_{i \in N} n_i \). Let \( x = (x_1, \ldots, x_N) \in \Omega \) denote all agents’ action profile or \( N \)-tuple, where \( x_{-i} \) is the \( (N-1) \)-tuple of all agents’ actions except agent \( i \)’s. Alternatively, \( x \) is represented as a stacked vector \( x = [x_1^T \ldots x_N^T]^T \in \Omega \subseteq \mathbb{R}^n \). Each player (agent) \( i \) aims to minimize its own cost function \( J_i \), satisfies the variational inequality (VI) (Proposition 1.4.2, [7]),

\[
(x - x^*)^T F(x^*) \geq 0 \quad \forall x \in \Omega
\]

where \( F : \Omega \to \mathbb{R}^n \) is the pseudo-gradient (game) map defined by stacking all agents’ partial gradients,

\[
F(x) = [\nabla x_1 J_1^T(x), \ldots, \nabla x_N J_N^T(x)]^T
\]

with \( \nabla x_i J_i(x_i, x_{-i}) = \frac{\partial J_i}{\partial x_i}(x_i, x_{-i}) \in \mathbb{R}^{n_i} \), the partial-gradient of \( J_i(x_i, x_{-i}) \) with respect to its own action \( x_i \).

We use the following basic convexity and smoothness assumption, which ensures the existence of a NE.

**Assumption 1:** For every \( i \in N \), \( \Omega_i = \mathbb{R}^{n_i} \) and the cost function \( J_i : \Omega \to \mathbb{R} \) is \( C^1 \) in its arguments, convex and radially unbounded in \( x_i \), for every \( x_{-i} \in \Omega_{-i} \).

Under Assumption 1 from Corollary 4.2 in [28] it follows that a NE \( x^* \) exists. Furthermore, the VI (2) reduces to \( F(x^*) = 0 \).

A standard method for reaching a Nash Equilibrium (NE) is using gradient-play dynamics [29], i.e.,

\[
\dot{x}_i = -\nabla x_i J_i(x_i, x_{-i}), \quad \forall i \in N, \quad \text{or} \quad \dot{x} = -F(x)
\]

This algorithm converges to the NE if the pseudo-gradient is strictly monotone but may fail if the pseudo-gradient is only monotone. For example, consider a 2-player zero-sum game where the cost functions are \( J_1(x_1, x_2) = x_1 x_2 \), \( J_2(x_1, x_2) = -x_1 x_2 \). The pseudo-gradient is,

\[
F(x) = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} x
\]

which is monotone and the NE is \((0, 0)\). If the initial state \( x(0) \neq (0, 0) \) then \( (4) \) will cycle around the NE and never converge, i.e., Figure 1. In this paper, we are interested in monotone games.

**Assumption 2:** The pseudo-gradient \( F \) is monotone.

Under Assumption 1 and 2 the set of NE is convex, (cf. Theorem 3, [5]), characterized by \( \{x^* \mid F(x^*) = 0\} \).

A. Proposed Algorithm

The dynamics (4) can be viewed as an open-loop system with no feedback. We propose a new algorithm, what we are calling “Heavy Anchor”, by modifying the feedback path with a bank of high-pass filters as depicted in Figure 2 below, with \( u_c = 0 \). We call it Heavy Anchor because we show that it looks like Poljak’s heavy ball method but with the momentum term having the opposite sign.

![Fig. 2: Block diagram of HA_F](image)

Explicitly the dynamics are,

\[
\begin{align*}
\dot{y}_1 &= -F(x) + u_1 \\
y_1 &= x \\
\dot{y}_2 &= -\alpha \dot{r} + \alpha u_2 \\
y_2 &= -\beta \dot{r} + \beta u_2
\end{align*}
\]

with \( \alpha, \beta \in \mathbb{R}_{++} \) and \( r \in \mathbb{R}^n \) are auxiliary variables. The individual agent dynamics are,

\[
\begin{align*}
\dot{r}_1 &= \alpha (x_1 - r_1) \\
\dot{r}_2 &= \alpha (x_2 - r_2) \\
\dot{x}_i &= -\nabla x_i J_i(x_i, x_{-i}) - \beta (x_i - r_i)
\end{align*}
\]

The new dynamics have a gradient-play component with a dynamic estimation of the own action derivative. Figure 3 shows the decision trajectories \( x \) under Heavy Anchor for the two player zero-sum game (5).

B. Connections to Other Dynamics/Algorithms

Our proposed dynamics (HA_F) is related to other continuous-time dynamics or discrete-time algorithms. First, (HA_F) can be written as the second-order dynamics,

\[
\ddot{x} + (\nabla F(x) + \beta + \alpha) \dot{x} + \alpha F(x) = 0.
\]
Under appropriate restrictions on the values of $\alpha$ and $\beta$, this dynamics recovers other dynamics/algorithms. For example, similar dynamics appears in stabilizing unknown equilibrium in chaotic systems or saddle functions [21], [22]. However, these works do not rigorously characterize stability/convergence. As another example, consider

\[ \ddot{x} + \alpha \dot{x} + \beta \nabla^2 f(x) \dot{x} + \nabla f(x) + \nabla \Psi(x) = 0, \]

where $f$ is a convex function, as considered in the optimization literature, [23], [30]. If $F = \nabla f$, (6) can be written as the above (with $\Psi(x) \equiv 0$). However, in a game $F$ is not a true gradient (unless the game is a potential game), but rather a pseudo-gradient, so convergence results are not applicable.

Next, we relate (HA) to some existing discrete-time algorithms. Performing an Euler discretization of (HA) gives,

\[ x_{k+1} = x_k - sF(x_k) - \beta(x_k - r_k) \]
\[ r_{k+1} = r_k + s\alpha(x_k - r_k) \]

where $s > 0$ is the step size, which after some manipulations yields the second-order difference equation,

\[ x_{k+2} = x_{k+1} - \alpha s^2 F(x_{k+1}) + (1 - \alpha s - \beta) (x_{k+1} - x_k) - s(1 - \alpha s) (F(x_{k+1}) - F(x_k)). \]  

Depending on how the parameters $\alpha$ and $\beta$ are selected we can recover some known algorithms. If $\alpha = \beta = \frac{1}{2\alpha}$ and $F = \nabla f$ for some convex function $f$ then (7) becomes,

\[ x_{k+2} = x_{k+1} - \bar{s} (2\nabla f(x_{k+1}) - \nabla f(x_k)) \]

where $\bar{s} = \frac{s}{2}$ gives the optimistic gradient-descent-ascent (OGDA) [24], shadow Douglas Rachford [25], or the forward-reflected backward method [13]. On the other hand, if $\alpha = \frac{1}{2}$, and $F = \nabla f$ then (7) becomes,

\[ x_{k+2} = x_{k+1} - s (\nabla f(x_{k+1}) + \beta (x_{k+1} - x_k)) \]

where $\beta < 0$ gives Polyak’s heavy-ball method, [31].

**IV. CONVERGENCE UNDER PERFECT INFORMATION**

In this section we consider that each agent knows all $x_{-i}$ (actions that his cost depends on), hence the full (perfect) decision information setting. In Theorem 1 we show that the continuous-time dynamics (6) converges for all $\alpha, \beta > 0$, in this full information setting. Our idea is to see (HA) as an (EIP) passivity-based feedback modification of (4). To prove that $x$ in (HA) converges to a Nash Equilibrium in monotone games, we decompose the system into a feedback interconnection between two subsystems (see Fig. 2).

We show that each subsystem is EIP and use their storage functions to construct an appropriate Lyapunov function to prove that the equilibrium point of the interconnected system (which is a NE) is asymptotically stable.

**Lemma 1:** Under Assumption 2 the following system,

\[ \dot{x} = -F(x) + u_1 \]
\[ y_1 = x. \]

is EIP with respect to $u_1$ and $y_1$.

**Proof:** Let $\bar{x}$ be the equilibrium of (8) for input $\bar{u}_1$, and $\bar{y}_1$ the corresponding output. Consider the storage function $V_1(x) = \frac{1}{2} \| x - \bar{x} \|^2$. Then, along the solutions of (8),

\[ \dot{V}_1(x) = \langle x - \bar{x}, -F(x) + u_1 - F(\bar{x}) - \bar{u}_1 \rangle = -\langle x - \bar{x} | F(x) - F(\bar{x}) \rangle + \langle u_1 - \bar{u}_1 | y_1 - \bar{y}_1 \rangle \]

By Assumption 2 the first term is $\leq 0$ and the system is EIP.

**Lemma 2:** For any $\alpha, \beta > 0$ the following system,

\[ \dot{r} = -\alpha r + \alpha u_2 \]
\[ y_2 = -\beta r + \beta u_2. \]

is OSEIP with respect to $u_2$ and $y_2$.

**Proof:** Let $\bar{r}$ be the equilibrium of (10) for the input $\bar{u}_2$ and let $\bar{y}_2$ be the corresponding output. Consider the storage function $V_2(r) = \frac{\beta}{2\alpha} \| r - \bar{r} \|^2$. Then, along solutions of (10),

\[ \dot{V}_2(r) = \frac{\beta}{\alpha} \langle r - \bar{r}, -\alpha r + \alpha u_2 + \alpha \bar{r} - \alpha \bar{u}_2 \rangle = \frac{\beta}{\alpha} \langle u_2 - \frac{1}{\beta} \bar{y}_2 - \bar{u}_2, u_2 - \frac{1}{\beta} \bar{y}_2 \rangle = \langle u_2 - \bar{u}_2, y_2 - \bar{y}_2 \rangle - \frac{1}{\beta} \| y_2 - \bar{y}_2 \|^2 \]

Therefore, the system is OSEIP for any $\beta > 0$.

We now turn to the interconnected system (HA). We show first that any equilibrium of (HA) is a NE. Then, using the two storage functions from Lemma 1 and 2 we show that any equilibrium point (HA) is asymptotically stable.

**Lemma 3:** Any equilibrium of (HA) is $(x^*, x^*)$ where $x^*$ is a Nash equilibrium of the game.

**Proof:** Let the equilibrium point of (HA) be denoted $(\bar{x}, \bar{r})$. Then $0 = \alpha (\bar{x} - \bar{r})$ implies that $\bar{x} = \bar{r}$ and $0 = -F(\bar{x}) - \beta (\bar{x} - \bar{r}) = -F(\bar{x})$. An equilibrium $x^*$ of (4) is such that $F(x^*) = 0$ therefore $\bar{x} = \bar{r} = x^*$ a NE.

**Theorem 1:** Consider a game $G(N, J_i, \Omega_i)$ under Assumption 1 and 2. Let the overall dynamics of the agents
be given by $\mathcal{HA}_F$. Then, for any $\alpha, \beta > 0$, the set of Nash equilibrium points $\{x^* \mid F(x^*) = 0\}$ is globally asymptotically stable.

**Proof:** Note that $\mathcal{HA}_F$ is the system in Fig. [5] with $u_c = 0$. Consider the following candidate Lyapunov function $V(x, r) = V_1(x) + V_2(r)$ where $V_1(x) = \frac{1}{2} \|x - \bar{x}\|^2$ and $V_2(r) = \frac{\alpha}{\beta} \|r - \bar{r}\|^2$, where cf. Lemma 3 $\bar{x} = \bar{r} = x^*$. Along the solutions of $\mathcal{HA}_F$, from Lemma 1 (9), and Lemma 2 (11), using $u_1 = -y_2$, $u_2 = y_1$, $\bar{x} = \bar{r}$ and cancelling terms, we obtain,

$$
\dot{V}(x, r) = -\langle x - \bar{x} \mid F(x) - F(\bar{x}) \rangle - \frac{1}{\beta} \|y_2 - \bar{y}_2\|^2
= -\langle x - \bar{x} \mid F(x) - F(\bar{x}) \rangle - \beta \|x - r\|^2 
$$

(12)

By Assumption 2 it follows that $\dot{V} \leq 0$. We resort to LaSalle’s Invariance Principle [32]. Note that $\dot{V} = 0$ implies $x - r = 0$. On $x = r$ the dynamics $\mathcal{HA}_F$ reduces to,

$$
0 = x - r = -F(x) - \beta(x - r) - \alpha(x - r) = -F(x),
$$

hence the largest invariant set is $\{x \mid F(x) = 0\}$. Since $V$ is radially unbounded, the conclusion follows.

V. PARTIAL INFORMATION

In Section [4] we considered that each agent knows all others’ decisions $x_{-i}$. In this section we propose a version of $\mathcal{HA}_F$, that works in the partial information setting, i.e. when agents do not know all others’ decisions and instead estimate them based on communicating with their neighbors over a communication graph $G_c$.

**Assumption 3:** $G_c = (N, \mathcal{E})$ is undirected and connected.

Assume that each agent $i$ maintains an estimate vector $x^i = col(x^i_j)_{j \in \mathcal{N}^i} \in \mathbb{R}^{n_i}$ where $x^i_j$ is agent $i$’s estimate of player $j$’s action. Note that $x^i_j = x_i$ is player $i$’s actual action. Let $x = col(x^i)_{i \in \mathcal{N}^e} \in \mathbb{R}^{n_e}$ represent all agents’ estimates stacked into a single vector. Similarly, define the auxiliary variable $r^i \in \mathbb{R}^{n_i}$ for each agent $i.$ Let the extended pseudo-gradient be denoted as $F(x) = col(\nabla x_i, J_i(x^i))_{i \in \mathcal{N}^e}$, where each agent uses its estimate of others’ decisions instead of true decisions. Note that at consensus of estimates, $x^i = x$, for all $i \in \mathcal{N}$, and $F(1_N \otimes x) = F(x)$, for any $x \in \mathbb{R}^{n}$. Let the matrix $\mathcal{R} = diag(\mathcal{R}_i)_{i \in \mathcal{N}^e}$, where $\mathcal{R}_i = [0_{n_i \times n_i} \otimes I_{n_i - n_i}]$, and $n < i = \sum_{j < i} n_j$, $n < i = \sum_{j > i} n_j$, $i, j \in \mathcal{N}$. The matrix $\mathcal{R}_i$ is used to get the component of a vector that belongs to agent $i$, i.e., $x_i = \mathcal{R}_i x^i$ and $x = \mathcal{R} x$. The operation $x = \mathcal{R}^T x^i$ sets $x^i_j = x_i$ and $x^i_j = 0$ for all $j \neq i$.

The problem is thus lifted into an augmented space of decisions, estimates and auxiliary variables $(x, r)$, with the original space being its consensus subspace. Consider the partial information version of $\mathcal{HA}_F$, over $G_c$, where the individual agent dynamics is given as,

$$
\dot{r}^i = \alpha(x^i - r^i) \tag{13}
$$

$$
\dot{x}^i = -\mathcal{R}_i^T \nabla x_i J_i(x^i) - \beta(x^i - r^i) - c \sum_{j \in \mathcal{N}} w_{ij} (x^i - x^j)
$$

or, in compact (stacked) form, as

$$
\dot{\bar{x}} = \alpha(\bar{x} - \bar{r}) \tag{13}
$$

$$
\dot{x} = -\mathcal{R}^T F(x) - \beta(x - r) - c L x
$$

(\mathcal{HA}_F)

where $c > 0$ is a scaling factor and $L = L \otimes I_n$. The individual agent dynamics (13) is the augmented version of $\mathcal{HA}_F$ with a Laplacian (consensus) correction for the estimates. Note that the dynamics $\mathcal{HA}_F$ is similar to Fig. [5] but with an augmented state $(r, x)$, and with feedback loop closed with $u_c = -c L x$. At consensus, $x = 1_N \otimes x$, $r = 1_N \otimes r$, and $\mathcal{HA}_F$ recovers $\mathcal{HA}_E$. First we show that any equilibrium point of $\mathcal{HA}_F$ is a NE.

**Lemma 4:** Consider a game $G(N, J_i, \Omega_i)$ under Assumption 1 over a communication graph $G_c = (N, \mathcal{E})$ satisfying Assumption 3. Let each agents’ dynamics be as in (13) or overall as $\mathcal{HA}_F$. Then, any equilibrium $(\bar{x}, \bar{r})$ of $\mathcal{HA}_F$ satisfies $\bar{x}^i = \cdots = \bar{x}^N = \bar{r}^1 = \cdots = \bar{r}^N = x^* \in \mathbb{R}^{n_e}$ where $x^*$ is a NE.

**Proof:** Let $(\bar{x}, \bar{r})$ denote an equilibrium of $\mathcal{HA}_F$. Then at equilibrium we have $\bar{x} = \bar{r}$ and $0_{N_N} = -\mathcal{R}^T F(\bar{x}) - \bar{L} x$. Pre-multiplying both sides by $(1_N^T \otimes I_n)$ yields $0_{n} = F(\bar{x})$ and therefore, $0_{N_N} = -\bar{L} x$. By Assumption 3 $0_{N_N} = -\bar{L} x$ when $\bar{x}^i = \cdots = \bar{x}^N i.e., \bar{x} = 1_N \otimes \bar{x}$ for some $\bar{x} \in \mathbb{R}^{n_e}$. Therefore, $0_{N} = F(\bar{x}) = F(1_N \otimes \bar{x}) = F(\bar{x})$ hence, $\bar{x} = x^*$, where $x^*$ is a Nash Equilibrium.

**Remark 1:** We note that in the full decision information case, monotonicity of $F$ was instrumental (see Theorem 1). In the augmented space monotonicity does not necessarily hold, even if on the consensus subspace it does cf. Assumption 2 see [33]. This is unlike distributed optimization, where due to separability, the extension of monotonicity/convexity properties to the augmented space holds automatically. This is the main technical difficulty in developing NE seeking dynamics in partial-information settings.

Our first result is proved under a monotonicity assumption on the extended pseudo-gradient $F$.

**Assumption 4:** The extended pseudo-gradient is monotone, $(x - x')^T (\mathcal{R}^T (F(x) - F(x'))) \geq 0, \forall x, x'$.

Assumption 4 has been also used in Thm.1, [34], or [35] (as cocoercivity). It represents extension of monotonicity off the consensus subspace. Note that on the consensus subspace $(x = 1_N \otimes x)$, it is automatically satisfied by Assumption 2.

Under Assumption 4 the following result can be immediately obtained by exploiting EIP properties.

**Theorem 2:** Consider a game $G(N, J_i, \Omega_i)$ under Assumption 1 and 4 over a communication graph $G_c = (N, \mathcal{E})$ satisfying Assumption 3. Let the overall dynamics of the agents be given by $\mathcal{HA}_F$ or (13). Then, for any $\alpha, \beta > 0$, $\mathcal{HA}_F$ converges asymptotically to $(1_N \otimes x^*, 1_N \otimes x^*)$ where $x^*$ is a NE.

**Proof:** Note that $\mathcal{HA}_F$ is similar to a dynamics as in Fig. 2 but with an augmented state $x$ (decisions and estimates), and with feedback loop closed with $u_c = -c L x$. We exploit the EIP properties of the two, forward and feedback, subsystems. Namely, consider $V(x, r) = \frac{1}{2} \|x - \bar{x}\|^2 + \frac{c}{2\alpha} \|r - \bar{r}\|^2$, where $\bar{x} = \bar{r} = 1_N \otimes \bar{x}$ (cf. Lemma 3). Then,
\[ \frac{\dot{V}(\mathbf{x}, \mathbf{r})}{V} = -\langle \mathbf{x} - \tilde{\mathbf{x}} \mid \mathcal{R}^TF(\mathbf{x}) - \mathcal{R}^TF(\tilde{\mathbf{x}}) \rangle + \langle \mathbf{x} - \tilde{\mathbf{x}} \mid u_c - \tilde{u}_c - \beta \| \mathbf{x} - \mathbf{r} \|^2 \]  \tag{14}

where \( u_c = -cL\mathbf{x}. \) The first term is nonpositive under Assumption 4. For any \( \alpha, \beta > 0, \) the system is strictly EIP from \( u_c \) to \( y_1 = \mathbf{x}, \) and with \( u_c = -cL\mathbf{x}, \) since \( \mathcal{L} \) is positive semidefinite, it follows that \( \dot{V} \leq 0. \) We use LaSalle’s Invariance Principle and find the largest invariant set [32]. Note that \( \dot{V} = 0 \) implies that \( \mathbf{x} = \mathbf{r} \) and \( \mathcal{L}\mathbf{x} = \mathcal{L}\tilde{\mathbf{x}}. \) Since \( \tilde{x} = 1_N \otimes \tilde{x} \) (cf. Lemma 2), \( \mathcal{L}\mathbf{x} = \mathcal{L}\tilde{\mathbf{x}} = 0, \) hence \( \mathbf{x} = 1_N \otimes \mathbf{x} \) for some \( \mathbf{x} \in \mathbb{R}^n. \) Then, on \( \mathbf{x} = \mathbf{r}, \) the dynamics \( \frac{\dot{\mathbf{x}}}{\dot{\mathbf{x}}} \) reduces to 0 = \( \dot{\mathbf{x}} - \dot{\mathbf{r}} = -\mathcal{R}^TF(1_N \otimes \mathbf{x}) = -\mathcal{R}^TF(\mathbf{x}), \) which implies \( F(\mathbf{x}) = 0, \) hence the largest invariant set is the NE set. Since \( V \) is radially unbounded, the conclusion follows.

On the other hand, Assumption 4 can be quite restrictive. Instead of this assumption on \( \mathcal{F}, \) we will use a weaker additional condition, this time on the pseudo-gradient \( F. \) This is the inverse Lipschitz property. In the next section we discuss this property.

VI. INVERSE LIPSCHITZ

In convex analysis and monotone operator theory there are three properties on an operator \( T \) that are frequently used and are important. These three properties are: \( \mu \)-strongly monotone, \( L \)-Lipschitz, and \( C \)-cocoercive, which describe upper and lower bounds on an operator \( T. \) However, it appears there is a natural definition missing.

Definition 3: An operator \( T : \mathcal{H} \rightarrow \mathbb{H}^2 \) is \( R \)-inverse Lipschitz if,

\[ \| x - y \| \leq R \| Tx - Ty \|, \quad \forall x, y \in \mathcal{H}, \]

related to the condition used by Rockafellar [36].

Remark 2: A \( C \)-cocoercive operator is also called \( C \)-inverse strongly monotone, because if \( x = T^{-1}y \) and \( y = T^{-1}v \),

\[ \langle Tx - Ty \mid x - y \rangle \geq C \| Tx - Ty \|^2 \]
\[ \langle u - v \mid T^{-1}u - T^{-1}v \rangle \geq C \| u - v \|^2 \]

This is the same as the inverse operator \( T^{-1} \) being \( C \)-strongly monotone. In the same spirit, we call \( T \) a \( R \)-inverse Lipschitz operator because it is the same as the inverse \( T^{-1} \) being \( R \)-Lipschitz, i.e.,

\[ \| x - y \| \leq R \| Tx - Ty \| \]
\[ \| T^{-1}u - T^{-1}v \| \leq R \| u - v \| \]

A. Similarities

1) Similarities in monotone operator theory: The property of inverse Lipschitz is closely related to coercive or radially unbounded property.

Definition 4: A function \( f : \mathcal{H} \rightarrow \mathbb{R} \) is coercive (radially unbounded) if,

\[ \lim_{\| x \| \rightarrow \infty} f(x) = +\infty \]

Since coercive functions are real valued and \( T \) is in general vector valued, taking the norm can be thought of as an extension of the definition, i.e., \( \lim_{\| x \| \rightarrow \infty} \| Tx \| = +\infty. \) If \( y = 0 \) and define \( \tilde{T}x = Tx - T0 \) then from the \( R \)-inverse Lipschitz definition of \( T, \) we see that \( \| x \| \leq R \| \tilde{T}x \|. \)

Therefore, \( R \)-inverse Lipschitz is a stronger growth condition relating the input to the output, similar to coercivity and implies that \( \tilde{T} \) is coercive.

2) Similarities to optimization: In optimization there are weaker conditions than strong convexity that can get linear convergence rates [37]. One of these conditions is the Polyak-Lojasiewicz (PL) inequality. A function \( f \) satisfies the (PL) inequality if \( \forall x \in X, \frac{1}{2} \| \nabla f(x) \|^2 \geq \mu (f(x) - f^*), \) where \( f^* \) is the value of \( f \) at the optimal solution. Theorem 2 [37] shows that if \( f \) has a Lipschitz-continuous gradient then (PL) is equivalent to the Error Bound (EB) inequality, \( \forall x \in X, \| \nabla f(x) \| \geq \mu \| x - \text{Proj}_X(x) \|. \)

If \( X = \mathbb{R}^n \) then \( \nabla f(\text{Proj}_X(x)) = 0 \) and the condition can be written as, \( \| \nabla f(x) - \nabla f(y) \| \geq \mu \| x - y \| \) where \( y = \text{Proj}_X(x), \) hence \( \nabla f \) is \( \frac{1}{\mu} \)-inverse Lipschitz. Thus, the \( R \)-inverse Lipschitz condition is the monotone operator equivalent to the Error Bound inequality / Polyak-Lojasiewicz inequality. If we replace \( \nabla f \) with a monotone operator then this condition is the condition used by Rockafellar in [36],

\[ \| x - y \| \leq R \| Tx - Ty \| \]

\( \forall x \in \mathcal{H} \) and \( y = Ty \) (restricted). If we remove the restriction of \( y = Ty \) then we get the definition of \( R \)-inverse Lipschitz.

3) Similarities in control / passivity: Passive systems with an inverse Lipschitz property have been analyzed in Chapter 6, section 11 [38]. However, the analysis is only for the case when the operator is strongly monotone.

B. Relations / Properties

The following diagram shows the relationship between \( R \)-inverse Lipschitz and the other properties.

Fig. 4: \( \text{cex}: \) set of convex functions, \( \mu: \) set of \( \mu \)-strongly monotone operators, \( C: \) set of \( C \)-cocoercive operators, \( M: \) set of monotone operators, \( R: \) set of \( R \)-inverse Lipschitz operators, \( L: \) set of Lipschitz operators

Proposition 1: If an operator \( T : \mathcal{H} \rightarrow \mathbb{H}^2 \) is \( C \)-inverse strongly monotone then it is \( \frac{1}{\mu} \)-Lipschitz

Proof: Note that a \( C \)-inverse strongly monotone satisfies, \( C \| Tx - Ty \|^2 \leq \langle Tx - Ty \mid x - y \rangle \leq \| Tx - Ty \| \| x - y \|, \) therefore \( \| Tx - Ty \| \leq \frac{1}{\mu} \| x - y \|. \)
Proposition 2 (Bailon-Haddad [39]): If an operator $T : \mathcal{H} \rightarrow 2^\mathcal{H}$ is $\frac{1}{C}$-Lipschitz and is the gradient of a convex function then it is $C$-inverse strongly monotone.

Proposition 3: If an operator $T : \mathcal{H} \rightarrow 2^\mathcal{H}$ is $\mu$-strongly monotone then it is $\frac{1}{\mu}$-inverse Lipschitz

Proof: Note that a $\mu$-inverse strongly monotone satisfies, $\mu \|x - y\|^2 \leq \langle Tx - Ty, x - y \rangle \leq \|Tx - Ty\| \|x - y\|$, therefore $\|x - y\| \leq \frac{1}{\mu} \|Tx - Ty\|$. \n
Proposition 4: If an operator $T : \mathcal{H} \rightarrow 2^\mathcal{H}$ is $\frac{1}{\mu}$-inverse Lipschitz and is the gradient of a convex function then it is $\mu$-strongly monotone.

Proof: Let $\partial f = T$. If $\partial f$ is $\frac{1}{\mu}$-inverse Lipschitz then $(\partial f)^{-1} = \partial f^*$ is $\frac{1}{\mu}$-Lipschitz. From Prop. 12.60(a,b) [40], if a function $f$ is convex and $\partial f$ is $\frac{1}{\mu}$-Lipschitz then $f^*$ is $\mu$-strongly monotone. Since $\partial f^*$ is $\frac{1}{\mu}$-Lipschitz and $f^* = f$ we can conclude that $f$ is $\mu$-strongly monotone. \n
Proposition 5: If an operator $T : \mathcal{H} \rightarrow 2^\mathcal{H}$ is $\mu$-strongly monotone and $L$-Lipschitz then it is $\frac{1}{\mu + L}$-inverse strongly monotone, or cocoercive with $C = \frac{1}{\mu + L}$.

Proposition 6: An $\mu$-strongly monotone operator satisfies, $\mu \|x - y\|^2 \leq \langle Tx - Ty, x - y \rangle$ and a $L$-Lipschitz operator satisfies, $\frac{1}{\mu + L} \|Tx - Ty\|^2 \leq \|x - y\|^2$. Combining these together gives $\frac{1}{\mu + L} \|Tx - Ty\|^2 \leq \langle Tx - Ty, x - y \rangle$. \n
Proposition 7: Let $T : \mathcal{H} \rightarrow 2^\mathcal{H}$ be a maximally $\mu$-hypomonotone operator that is $R$-inverse Lipschitz. Then for any $\lambda \geq 0$ such that $\mu R^2 \leq \lambda \leq \frac{1}{\mu}$, the following hold for the resolvent of $T$, $J_{\lambda T} = (\text{Id} + \lambda X T)^{-1}$

(i) $J_{\lambda T}$ is maximaly monotone.

(ii) $J_{\lambda T}$ is $L_\lambda$-Lipschitz, $\|J_{\lambda T} x - J_{\lambda T} y\| \leq L_\lambda \|x - y\|$, where, $L_\lambda \triangleq \sqrt{\frac{R^2}{\sqrt{\mu^2 + \lambda^2} - 2 \mu \lambda R}}$.

(iii) $\langle x - y \mid J_{\lambda T} x - J_{\lambda T} y \rangle \leq \kappa_\lambda \|x - y\|^2$ where, 

$$\kappa_\lambda \triangleq \begin{cases} \frac{R^2 (1 - \mu \lambda)}{\sqrt{\mu^2 + \lambda^2} - 2 \mu \lambda R} & R \geq \lambda \\ \frac{\lambda R^2}{\sqrt{\mu^2 + \lambda^2} - 2 \mu \lambda R} & \lambda \geq R \end{cases}$$

Proof: Found in the Appendix.

Remark 3: The Lipschitz constant from Lemma 7(ii) can upper bound the inner product $\langle x - y \mid J_{\lambda T} x - J_{\lambda T} y \rangle$, but Lemma 7(iii) provides a tighter bound.

When $T$ is differentiable some sufficient conditions for these properties are given next.

Proposition 8: Let $T$ be a differentiable operator and the Jacobian of $T$ be denoted $JT(x)$. Then $T$ is,

1) $\mu$-strongly monotone if: $\frac{1}{2} \langle JT(x) + JT(x)^T, x \rangle \geq \mu I$

2) $C$-cocoercive if: $\langle JT(x) + JT(x)^T, x \rangle \leq \mu I$

3) $L$-Lipschitz if: $\|JT(x) + JT(x)^T\| \leq L^2 I$

4) $R$-inverse Lipschitz if: $\langle JT(x)^T + JT(x), x \rangle \geq \frac{1}{R} I$

Proof: Found in the Appendix.

C. Examples

Example 1: The operator $Tx = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} x$ is monotone but is not strongly monotone nor cocoercive. It is $1$-Lipschitz and $1$-inverse Lipschitz.

Example 2: The operator $T : [-1, 1] \rightarrow [-1, 1]$, $Tx = x^3$ is $\frac{1}{3}$-cocoercive and is not strongly monotone nor inverse Lipschitz.

Example 3: The operator $Tx = \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix} x$ is $2$-strongly monotone, $\sqrt{15 + \sqrt{29}}$-Lipschitz, $\frac{1}{\sqrt{15 + \sqrt{29}}}$-cocoercive and $\sqrt{15 - \sqrt{29}}$-inverse Lipschitz

Example 4: The operator $T : [0, \infty) \rightarrow [1, \infty)$, $Tx = e^x$ is $1$-strongly monotone and $1$-inverse Lipschitz, and is not cocoercive nor Lipschitz.

Example 5: The operator $Tx = \sin(x)$ is $1$-Lipschitz and is not strongly monotone, cocoercive or inverse Lipschitz.

Example 6: The operator $T : (0, 1) \rightarrow (0, \infty)$, $Tx = \frac{1}{x}$ is $1$-inverse Lipschitz and is not strongly monotone, cocoercive or Lipschitz.

VII. Convergence under Partial Information

We will now show that $\{H\Delta X\}$ converges to the NE when the monotonicity of the extended pseudo-gradient, Assumption 3 is replaced by a weaker assumption only on the pseudo-gradient.

Assumption 5: The pseudo-gradient $F$ is $L_F$-Lipschitz, $R$-inverse Lipschitz, and $\mu$-hypomonotone, i.e., $\langle Fx - Fy, x - y \rangle \geq -\mu \|x - y\|^2$.

Remark 4: Note that $F$ may not be monotone. For example, $F(x) = \begin{bmatrix} -1 & 1 \\ -1 & -1 \end{bmatrix}$ is $1$-hypomonotone, $\sqrt{2}$-Lipschitz and $\sqrt{2}$-inverse Lipschitz.

When the extended monotonicity property (Assumption 3) does not hold, we use Assumption 5 and take advantage of properties of the dynamics on the augmented consensus subspace and its orthogonal complement. Our idea is to use a change of coordinates and in these coordinates show that, under Assumptions 2 and 5 the dynamics restricted to the consensus subspace satisfies a property similar to strict EIP for $\alpha$ parameters selected in a certain range (Lemma 5). Then, for the overall dynamics, we exploit this property together with the excess passivity of the Laplacian to balance the coupling terms off the consensus subspace and show that $\{H\Delta X\}$ converges to a Nash Equilibrium (Theorem 3).

We first decompose the system into consensus and orthogonal component dynamics. Let $x$ and $r$ be decomposed into consensus and orthogonal components, i.e.,

$x = x^c + x^o$, $x^c = \Pi_r x$, $x^o = x - x^c$

$r = r^c + r^o$, $r^c = \Pi_r r$, $r^o = r - r^c$

where $\Pi_r = \frac{1}{N} (I_N \otimes I_N \otimes I_r)$ and $\Pi_r = I_{NN} - \Pi_r$, $x^c = I_N \otimes x$, $r^o = I_N \otimes r$, for some $x, r \in \mathbb{R}^N$. The overall
dynamics (HA) can be decomposed into the (augmented) consensus component dynamics,
\[
\dot{\bar{x}}^\parallel = \alpha \left( x^\parallel - r^\parallel \right)
\]
and the orthogonal component dynamics,
\[
\dot{\bar{x}}^\perp = -\Pi_{\perp} R^T \bar{F} \left( x^\parallel + x^\perp \right) - \beta \left( x^\perp - r^\perp \right)
\]
which are coupled one to another via \( x^\perp \) and \( x^\parallel \).

Consider the dynamics (18) restricted to the consensus subspace. Then from (15), it follows that
\[
\dot{\bar{x}}^\parallel = \alpha \left( x^\parallel + h(r^\parallel) - r^\parallel \right)
\]  
(18)

Consider the dynamics (15) restricted to the consensus subspace, i.e., when \( x^\parallel = 0 \), which is given as
\[
\dot{\bar{x}}^\parallel = \alpha \left( x^\parallel + h(r^\parallel) - r^\parallel \right)
\]  
(19)

Note that an equilibrium point for these dynamics is \((z^\parallel, \bar{r}^\parallel, \bar{x}^\parallel, \bar{r}^\parallel) = (0, 1_N \otimes x^*, 0_N, 0_N)\), where \( F(x^*) = 0 \) (\( x^* \) is a NE), cf. Lemma 4.

Consider the dynamics (15) restricted to the consensus subspace, i.e., when \( x^\parallel = 0 \), which is given as
\[
\dot{\bar{x}}^\parallel = \alpha \left( x^\parallel + h(r^\parallel) - r^\parallel \right)
\]  
(18)

\[\begin{align*}
\bar{x}^\parallel &= -\frac{1}{\beta N} \Pi_{\perp} R^T \bar{F} \left( x^\parallel + h(r^\parallel) + x^\perp \right) \\
&= -\left( \beta + \frac{\partial h}{\partial \bar{r}} \right) \left( \bar{x}^\parallel + h(r^\parallel) - r^\parallel \right)
\end{align*}\]

Therefore, the dynamics (HA) can be equivalently represented as (16) and (18). Note that an equilibrium point for these dynamics is \((\bar{z}^\parallel, \bar{r}^\parallel, \bar{x}^\parallel, \bar{r}^\parallel) = (0, 1_N \otimes x^*, 0_N, 0_N)\), where \( F(x^*) = 0 \) (\( x^* \) is a NE), cf. Lemma 4.

Lemma 5: Consider (19), under Assumption 1 and 5. For any \( 0 < d < 1 \), let \( \beta \in \left( \frac{\mu}{N}, \frac{1}{\beta N} \right) \) and
\[
0 < \alpha < \frac{4d}{\beta N} \left( \beta - \frac{\mu}{N} \right) \left( 1 + \frac{d}{N} \right)
\]
where \( \kappa_J \) and \( L_J \) are obtained from Lemma 7 for the pseudo-gradient \( \bar{F} \) and \( \lambda = \frac{\beta}{\beta N} \). Let,
\[
V^\parallel(z^\parallel, r^\parallel) = \frac{1 - d}{2} \left\| r^\parallel - \bar{r}^\parallel \right\|^2 + \frac{d}{2} || z^\parallel ||^2
\]
\[
\Phi = \left[ \frac{(1-d)\alpha(1-\kappa_J)}{\alpha + (L_J + L_J - 1) \frac{d}{2} \left( \beta - \frac{\mu}{N} \right)} \right]
\]
and the matrix \( \Phi \) is positive definite.

Using this Lemma we can show that (HA), in the full information case, converges for hypomonotone games instead of just monotone.

Lemma 6: Consider (HA), under Assumption 1 and 5. For any \( 0 < d < 1 \), let \( \beta \in \left( \frac{\mu}{N}, \frac{1}{\beta N} \right) \) and
\[
0 < \alpha < \frac{4d(1 - d)\left( \beta - \frac{\mu}{N} \right) \left( 1 + \kappa_J \right)}{\left( (1 - d) + d(L_J + L_J - L_J^2) \right)^2}
\]
where \( \kappa_J \) and \( L_J \) are obtained from Lemma 7 for the pseudo-gradient \( \bar{F} \) and \( \lambda = \frac{1}{\beta} \). Then, the dynamics (HA) globally converge to a NE \( x^* \).

Next, we now show that (HA) converges to a NE in the partial information case.

Theorem 3: Consider a game \( G(N, J, \Omega_c) \) over a communication graph \( G_c = (\mathcal{N}, \mathcal{E}) \), under Assumption 1, 3 and 5. Let the overall dynamics of the agents be given by (HA) or, equivalently, (16) and (18). Given any \( 0 < d < 1 \), set \( \alpha, \beta \) to satisfy the conditions in Lemma 5. Set \( c \) such that,
\[
c\lambda_2(L) > \frac{\eta_1 + \eta_2}{4\det(\Phi)} L_F^2 + L_F
\]  
(21)

where \( \Phi \) is defined in (20) and
\[
\eta_1 = \alpha(1 - d) \left( \beta - \frac{\mu}{N} \right) \left( 1 + \frac{d}{N} \right) + \frac{d}{2} \left\| z^\parallel \right\|^2
\]

\[
\eta_2 = \alpha \left( 1 + \frac{L_J^2 + L_J - 1}{2} \right) \left( 1 + \frac{d}{N} \right) L_J
\]

Then, the dynamics (HA) globally converges to a NE \( x^* \).

Proof: Consider the candidate Lyapunov function,
\[
V(z^\parallel, r^\parallel, x^\parallel, r^\parallel) = \frac{1 - d}{2} \left\| r^\parallel - \bar{r}^\parallel \right\|^2 + \frac{d}{2} \left\| z^\parallel \right\|^2 + \frac{1}{2} \left\| x^\parallel \right\|^2 + \frac{\beta}{\alpha} \left\| r^\parallel \right\|^2
\]

where \( \bar{r}^\parallel = 1_N \otimes x^* \) and \( F(r^\parallel) = F(x^*) = 0 \). Along (16) and (18), after re-grouping terms we can write,
\[
\dot{V} = \alpha(1 - d) \left\langle r^\parallel - \bar{r}^\parallel, z^\parallel + h(r^\parallel) - r^\parallel \right\rangle
\]

\[
- \frac{d}{N} \left\langle z^\parallel, 1_N \otimes F(z^\parallel + h(r^\parallel)) \right\rangle
\]

\[
- \frac{d}{N} \left\langle z^\parallel, \beta(z^\parallel + h(r^\parallel) - r^\parallel) + \frac{\partial h}{\partial \bar{r}} \bar{r}^\parallel \right\rangle
\]

\[
- \frac{d}{N} \left\langle z^\parallel, 1_N \otimes F(z^\parallel + h(r^\parallel) + x^\parallel) \right\rangle
\]

\[
+ \frac{d}{N} \left\langle z^\parallel, 1_N \otimes F(z^\parallel + h(r^\parallel)) \right\rangle - \beta \left\| x^\parallel - r^\parallel \right\|^2
\]

\[
- \left\langle x^\parallel, \Pi_{\perp} R^T F(z^\parallel + h(r^\parallel) + x^\parallel) + cL x^\parallel \right\rangle
\]

Note that the first three terms correspond to \( \dot{V}^\parallel \) along (19) in Lemma 5 and \( \beta > 0 \). Therefore, using Lemma 5 yields,
\[
\dot{V} \leq -\omega^T \Phi \omega
\]

\[
- \frac{d}{N} \left\langle z^\parallel, 1_N \otimes F(z^\parallel + h(r^\parallel) + x^\parallel) - 1_N \otimes F(z^\parallel + h(r^\parallel)) \right\rangle
\]

\[
- \left\langle x^\parallel, \Pi_{\perp} R^T F(z^\parallel + h(r^\parallel) + x^\parallel) \right\rangle - c\lambda_2(L) \left\| x^\parallel \right\|^2
\]
where $\omega = (\|x^+\| - \|\tilde{r}\|, \|z^+\|)$. Under Assumption 5 it follows that $F$ is also $L_F$-Lipschitz (cf. Lemma 3, [41] or Lemma 1,[42]). Using this and Cauchy-Schwarz inequality, as well as $\Pi_i R^T F(h(\tilde{r})) = 0_{N_i}$ yields,

\[
\dot{V} \leq -\omega^T \Phi \omega + \frac{d}{N} \sqrt{NL_F} \|z^+\| \|x^+\| - c\lambda_2(L) \|x^+\|^2
\]

+ $\|x^+\| \|\Pi_i R^T\| \|F(z^+ + h(\tilde{r}^+) + x^+) - F(h(\tilde{r}^+))\|

which, with $\|\Pi_i R^T\| \leq 1$ and Lemma 7(ii) for $h$, leads to,

\[
\dot{V} \leq -\omega^T \Phi \omega + \frac{d}{N} \sqrt{NL_F} \|z^+\| \|x^+\| - c\lambda_2(L) \|x^+\|^2
\]

+ $\|x^+\| L_F \left(\|z^+\| + L_\beta \|x^+\| - \|\tilde{r}\| + \|x^+\|\right)

Therefore,

\[
\dot{V} \leq -\omega^T \left[ \Phi - \frac{L_F L_\beta}{2} - \frac{L_F(\sqrt{N_i}+d)}{2} c\lambda_2(L) - L_F \right] \omega
\]

where $\omega = (\|z^+\|, \|x^+\|) = (\|x^+ - \tilde{r}\|, \|z^+\|, \|x^+\|)$. The block matrix is positive definite if its Schur complement is positive definite, i.e., if

\[
c\lambda_2(L) > \frac{\eta_1 + \eta_2}{4 det(\Phi)} L_F^2 + L_F
\]

where $\eta_1, \eta_2$ are as in the statement. Therefore, $\dot{V} \leq 0$ and $V = 0$ only if $x^+ = \tilde{r}^+ = 1_N x^* = 0, x^+ = 0$, i.e., $x^+ = 0 + h(\tilde{r}) = h(1_N x^*) = 1_N \otimes h(x^*) = 1_N \otimes x^*$, where since $F(x^*) = 0$, $x^*$ is a NE. The conclusion follows by a LaSalle argument [32].

The conditions that we obtain for Theorem 3 are conservative. In the following section we restrict our attention to an important subclass of games called quadratic games and derive tighter conditions on the parameters $\alpha, \beta$ to ensure convergence.

### VIII. Quadratic Hypomonotone Games

In this section, we consider a quadratic game $J_i(x_i, x_{-i}) = \frac{1}{2} x_i^T Q_i x_i + l_i^T x_i + c_i$ where $Q_i \in \mathbb{R}^{n_i \times n_i}$, $l_i \in \mathbb{R}^n$, and $c_i \in \mathbb{R}$. The gradient of agents cost function with respect to their own action is, $\nabla x_i J_i(x) = Q_i x_i + l_i$ and the pseudo-gradient is,

\[
F(x) = Ax + b, \quad A \triangleq \begin{bmatrix} Q_1 & l_1 \\ Q_2 & l_2 \\ \vdots \\ Q_N & l_N \end{bmatrix}, \quad b \triangleq \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_N \end{bmatrix}
\]

For the perfect information case, algorithm (HAP), after the change of coordinates, $\hat{x} = x - x^*$ and $\hat{r} = r - r^*$, is written as,

\[
\dot{\hat{w}} = \begin{bmatrix} \hat{x} \\ \hat{r} \end{bmatrix} = \begin{bmatrix} -A - \beta I & \beta I \\ \alpha I & -\alpha I \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{r} \end{bmatrix} \triangleq M \hat{w}
\]

The following lemma relates the eigenvalues of $A$ to the eigenvalues of the overall $M$. [23]

**Lemma 7:** Let $A \in \mathbb{R}^{n \times n}$ be a matrix where the $i^{th}$ eigenvalue of $A$ is denoted $\rho_i$. Then the eigenvalues of $M$, [23], are,

\[
\lambda_i = \frac{-(\alpha + \beta + \rho_i) \pm \sqrt{(\alpha + \beta + \rho_i)^2 - 4\alpha \rho_i}}{2}
\]

for all $i \in \{1, \ldots, n\}$.

**Proof:** Found in the Appendix

The following Lemma gives conditions for the eigenvalues of $M$ to be in the OLHP.

**Lemma 8:** Let $A \in \mathbb{R}^{n \times n}$ be a matrix where the $i^{th}$ eigenvalue of $A$ is denoted $\rho_i = r_i + jk_i$ where $r_i (k_i)$ is the real (imaginary) part of $\rho_i$ and $j = \sqrt{-1}$.

(i) If $\rho_i = 0$ and $\alpha, \beta > 0$, then $\lambda_i$ from [24] are 0 and $-(\alpha + \beta)$.

(ii) If $\rho_i \neq 0$, $r_i \geq 0$ and $\alpha, \beta > 0$, then $\lambda_i$ from [24] are complex conjugate with real part less than 0.

(iii) If $r_i < 0$, $\beta \in \left(-r_i, \frac{k_i^2 + r_i^2}{-r_i}\right)$ and

\[
\alpha \in \left(0, -(\beta + r_i) + \sqrt{\frac{(\beta + r_i)k_i^2}{-r_i}}\right)
\]

then $\lambda_i$ from [24] have real part less than 0.

**Proof:** Found in the Appendix

**Remark 5:** Note that if the eigenvalues of $A$ fall only in case (i) and (ii) then $F$ is monotone. Additionally, the conditions $\alpha, \beta \geq 0$ are the same conditions as for the nonlinear case, Theorem 1. If $A$ has eigenvalues in case (iii) then $F$ is hypomonotone. If the eigenvalues of $A$ are $-r \pm jk$, then $F$ is $r$-hypomonotone and $R = \frac{1}{\sqrt{r^2 + k^2}}$-inverse Lipschitz. From Lemma 8 $\beta \in \left(\mu, \frac{1}{\sqrt{r^2 + k^2}}\right)$ is the same condition on $\beta$ as in Lemma 6 for the nonlinear case.

**Theorem 4:** Consider a quadratic game $G(N, J_i, \Omega_i)$ under Assumption 1. Let the overall dynamics of the agents be given by (HAP). For the matrix $A$ given in [22] with eigenvalues $\rho_i = r_i + jk_i$, let $I = \{i \in \{1, \ldots, n\} \mid r_i < 0\}$. If $I = \emptyset$ then set $\alpha, \beta > 0$ else,

\[
\beta \in \bigcap_{i \in I} \left(-r_i, \frac{k_i^2 + r_i^2}{-r_i}\right)
\]

\[
\alpha \in \bigcap_{i \in I} \left(0, -(\beta + r_i) + \sqrt{\frac{(\beta + r_i)k_i^2}{-r_i}}\right)
\]

Then, the set $\{(x^*, x^*) \mid F(x^*) = 0\}$ is globally asymptotically stable.

**Conjecture 1:** For the class of quadratic games where $F$ is $R$-inverse Lipschitz (for the perfect information setting) the optimal convergence rate is $exp(\frac{\alpha}{3\beta} t)$ when $\alpha_0 = \frac{1}{3\beta}$ and $\beta = \frac{1}{3R}$.

**A. Partial Information**

In the partial information case the dynamics (HAP) are,

\[
\dot{x} = -R^T (Ax + b) - \beta(x - r) - cLx
\]

\[
\dot{r} = \alpha(x - r)
\]
Theorem 5: Consider a game $\mathcal{G}(\mathcal{N}, J_i, \Omega_i)$ under Assumption [1] [3] and [5]. Let the overall dynamics of the agents be given by (HAX). Let $\alpha, \beta$ be selected as in [25] and scaled by $\frac{1}{N}$, and $c$ such that,

$$c \lambda_2(L) \geq L_A + \left( \frac{p}{\sqrt{N}} + \frac{1}{2} \right)^2$$

(27)

where $L_A = \|A\|$, $p = \|P\|$ where $P > 0$ satisfies the Lyapunov equation $PM + M^T P = -I$ and

$$\tilde{M} = \left[ \begin{array}{cc} -\frac{1}{N} A - \beta I & \beta I \\ \frac{1}{N} A + \beta I & -\beta I \end{array} \right]$$

Then, the set $\{(1 \otimes x^*, 1 \otimes x^*) \mid F(x^*) = 0\}$ is globally asymptotically stable.

Proof: Found in the Appendix.

Remark 6: Note that Theorem 5 requires $\|P\| = p$, if we restrict $\beta = \alpha$, and use Corollary 1 [43] and Corollary 2.10 [44], we can obtain the simpler bound

$$p \leq \frac{N}{2L_A + 4\alpha N}$$

(28)

B. Comparing Results For Quadratic vs General Games

For perfect information quadratic games with monotone pseudo-gradient, notice that Theorem 4 requires that $\alpha, \beta > 0$ and the rate of convergence can be determined by Lemma 7. For perfect information general games with monotone pseudo-gradient, Theorem 1 also requires $\alpha, \beta > 0$ but with no rate of convergence.

For the partial information quadratic games with monotone pseudo-gradient, Theorem 5 again requires that $\alpha, \beta > 0$. Additionally, the theorem requires that $c$ is larger than a function of the Lipschitz constant of the pseudo-gradient. For partial information general games with monotone pseudo-gradient, Theorem 1 allows $\beta > 0$ but $\alpha$ is now restricted by a function of $\beta$. Additionally, the $c$ term is larger than the one obtained for quadratic games.

For perfect information quadratic games with hypomonotone pseudo-gradient, Lemma 8 and Theorem 5 provides tight conditions on the range of values of $\alpha$ and $\beta$ for convergence to a NE. Note that for quadratic games, we are able to use the same method of analyzing the eigenvalues for both monotone and hypomonotone games. On the other hand for general games, the EIP analysis cannot be extended to the hypomonotone case and a different method is used to prove convergence. The analysis ends up having restrictions on $\alpha$ that don’t appear for the quadratic case. The quadratic game case suggests that there might be a better Lyapunov function that could remove or relax the condition on $\alpha$ for general games.

IX. Simulations

In this section we first consider three hypomonotone quadratic games between $N = 10$ agents communicating over a ring $G_c$ graph. We index each game by $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3$. In game $\mathcal{G}_j$, the cost function for agent $i$ is $J_i(x) = w^T_i x_T \left[ \begin{array}{c} 5 \\ 1 \end{array} \right] x_{N+1-i}$ where $w^T_i = [w^T_i, \ldots, w^T_N]$ is equal to

$$w^T_i = \left[ \begin{array}{c} 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \end{array} \right]$$

For game $\mathcal{G}_1$, the eigenvalues of $A$ from [22] are $1 \pm j\sqrt{5}$; $\mathcal{G}_2$ the eigenvalues are $\pm 1 \pm j5$, $\pm \frac{5}{2} \pm j\frac{5}{2}$, $\pm \frac{5}{2} \pm j\frac{5}{2}$, and $\pm \frac{3}{2} \pm j\frac{3}{2}$; and for $\mathcal{G}_3$, the eigenvalues are $\pm 2 \pm j0$ and $\pm 1 \pm j5$. For all three games the Nash equilibrium is the origin. The following table contains information about the parameter values as in Theorem 3 and 5. For game $\mathcal{G}_2$, the conditions of Lemma 7 are not satisfied and hence the column is empty. The $\beta$ values are selected as $0.9\beta_{\text{min}} + 0.1\beta_{\text{max}}$ and the $\alpha$ values are selected as $0.5\alpha_{\text{min}} + 0.5\alpha_{\text{max}}$.

| param. | $\mathcal{G}_1$ | $\mathcal{G}_2$ | $\mathcal{G}_3$ |
|--------|----------------|----------------|----------------|
| $\beta_{\text{min}}$ | 0.1 | 0.1 | 0.2 |
| $\beta_{\text{max}}$ | 2.6 | 2.6 | 2.6 |
| $\alpha$ | 0.35 | 0.35 | 0.44 |

Figure 5 shows the action trajectories for game $\mathcal{G}_1$ under (HAX) for the parameters $\alpha, \beta, c$ satisfying Theorem 5 where the initial conditions $x(0)$, $r(0)$ are randomly selected with components between $-10$ to $10$. Notice in $\mathcal{G}_1$ that $\beta$ used is the same for Theorem 3 and Theorem 5. However, the $\alpha$ obtained from Theorem 5 gives a conservative value for $\alpha$ and is an order of magnitude smaller than Theorem 5.

The figures for the other examples are similar and are omitted.

A. Nonquadratic Example

The following example is a non quadratic game where Theorem 5 no longer applies. Consider a hypomonotone
game between \( N = 10 \) agents communicating over a ring \( G_c \) graph. The cost function for agent \( i \) is \( J_i(x) = w_i^1 x_i^1 + w_i^2 x_i^2 \left[ \sin(x_{N+1-i,2}) - \sin(x_{N+1-i,1}) \right] \) where \( x_{i,j} \) is the \( j \)th component of the vector \( x_i \). For this game the pseudo-gradient is 1-hypomonotone, \( \frac{1}{2} \)-inverse Lipschitz, and \( 6 \)-Lipschitz.

Using Theorem 3, \( \beta_{\min} = 0.1, \beta_{\max} = 1.6, \) and we selected \( \beta = 0.9 \beta_{\min} + 0.1 \beta_{\max} = 0.25. \) Using \( d = 0.5 \) we obtain that \( \alpha_{\min} = 0, \alpha_{\max} = 0.095 \) and \( \alpha = 0.5 \alpha_{\min} + 0.5 \alpha_{\max} = 0.0478. \) Lastly, for a ring communication graph we obtain that \( c_{\min} = 3417. \) Figures 6 shows the action trajectories and convergence to the NE.

X. Conclusion

In this paper, we considered monotone games and proposed a continuous-time dynamics constructed via passivity-based modification of a gradient-play scheme. We showed that in the full-information setting it converges to a Nash equilibrium in merely monotone games, for any positive parameter values. Under different assumptions we provided extensions to the partial-decision information case and extensions to hypomonotone games. Among future interesting problems we mention, extensions to directed communication graphs or, with adaptive gains, as well as to generalized Nash equilibrium problems.

References

[1] W. Saad, Z. Han, H. Poor, and T. Başar, “Game theoretic methods for the smart grid,” IEEE Signal Processing Magazine, pp. 86–105, 2012.
[2] F. Parise, M. Colombino, S. Grammatico, and J. Lygeros, “Mean field constrained charging policy for large populations of plug-in electric vehicles,” in IEEE Conf. Decision and Control, 2014, pp.5101–5106.
[3] N. Li, L. Chen, M. Dahleh “Demand response using linear supply function bidding,” IEEE Trans. Smart Grid, vol.6(4), pp.1827-1838,2015.
[4] J. Barrera and A. Garcia, “Dynamic incentives for congestion control,” IEEE Trans. on Automatic Control, vol. 60, no. 2, pp. 299–310, 2015.
[5] G. Scutari, F. Facchinei, J.-S. Pang, and D. P. Palomar, “Real and Complex Monotone Communication Games,” IEEE Trans. on Information Theory, vol. 60(7), pp. 4197–4231, 2014.
[6] S. Li and T. Basar, “Distributed algorithms for the computation of noncooperative equilibria,” Automatica, vol. 23(4), pp. 523–533, 1987.
[7] F. Facchinei and J. Pang, Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer New York, 2007.
[8] H. Bauschke and P. Combettes, Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, 2017.
[9] G. Belgioioso and S. Grammatico, “A Douglas-Rachford splitting for semi-decentralized equilibrium seeking in generalized aggregate games,” in Proc. of the 57th IEEE CDC, Dec 2018.
[10] J. Koshal, A. Nedić, and U. Shanbhag, “Distributed Algorithms for Aggregative Games on Graphs,” Oper. Research vol.64,680-704,2016.
[11] D. Gadjoov and L. Pavel, “Single-timescale distributed gne seeking for aggregative games over networks via forward-backward operator splitting,” IEEE Transactions on Automatic Control, pp. 1–1, 2020.
[12] A. Kannan and U. V. Shanbhag, “Distributed Computation of Equilibrium in Monotone Nash Games via Iterative Regularization Techniques,” SIAM Journal on Optimization, vol. 22, no. 4, pp. 1177–1205, 2012.
[13] Y. Malitsky and M. K. Tam, “A Forward-Backward Splitting Method for Monotone Inclusions Without Cocoercivity,” SIAM Journal on Optimization, vol. 30, no. 2, pp. 1451–1472, 2020.
[14] B. Franci, M. Staudigl, and S. Grammatico, “Distributed forward-backward (half) forward algorithms for generalized Nash equilibrium seeking,” in 2020 European Control Conference (ECC), 2020, pp. 1274–1279.
[15] P. Yi and L. Pavel, “Distributed Generalized Nash Equilibrium Computation of Monotone Games via Double-Layer Preconditioned Proximal-Point Algorithms,” IEEE Trans. on Control of Network Systems, vol. 6(1), pp. 299–311, 2019.
[16] B. Gao and L. Pavel, “Continuous-time Discounted Mirror-Descent Dynamics in Monotone Concave Games,” IEEE Transactions on Automatic Control, pp. 1–1, 2020.
[17] Paul Frihauf and Miroslav Krstic and Tamer Basar, “Nash Equilibrium Seeking for Dynamic Systems with Non-Quadratic Payoffs,” IFAC Proceedings Volumes, vol. 44, no. 1, pp. 3605–3610, 2011, 11th IFAC World Congress.
[18] Tatiana Tatarenko and Maryam Kamgarpour, “Bandit Online Learning of Nash Equilibria in Monotone Games,” 2021.
[19] A. S. Antipin, “Feedback-Controlled Saddle Gradient Processes,” Automation and Remote Control, vol. 55, no. 3, pp. 311–320, 2003.
[20] J. S. Shamma and G. Arslan, “Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria,” IEEE Trans. on Automatic Control, vol. 50(3), pp. 312–327, 2005.
[21] E. Adomaitienė, G. Mykolaïtis, S. Bumeliene, and A. Tamaševičius, “Adaptive nonlinear controller for stabilizing saddle-type steady states of dynamical systems,” Nonlin. Dynamics, vol. 82(4),1743-1753, 2015.
[22] Y. Wu and W. Lin, “Adaptively locating unknown steady states: Formalism and basin of attraction,” Physics Lett. A, vol. 375(37), pp. 3279–3289, 2011.
[23] H. Attouch, J. Peyrouquet, and P. Redont, “A Dynamical Approach to an Inertial Forward-Backward Algorithm for Convex Minimization,” SIAM Journal on Optimization, vol. 24(1), pp. 232–256, 2014.
[24] S. Mohitkar, A. Ozdaglar, and S. Patrallil, “A Unified Analysis of Extra-gradient and Optimistic Gradient Methods for Saddle Point Problems: Proximal Point Approach,” in Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics, ser. Proceedings of Machine Learning Research, S. Chiappa and R. Calandra, Eds., vol. 108. PMLR, 26–28 Aug 2020, pp. 1497–1507.
[25] E. R. Csetnek, Y. Malitsky, and M. K. Tam, “Shadow Douglas–Rachford Splitting for Monotone Inclusions,” Applied Mathematics & Optimization, vol. 80, no. 3, pp. 665–678, Dec 2019.
[26] D. Gadjoov and L. Pavel, “On the exact convergence to Nash equilibrium in monotone regimes under partial-information,” in 2020 59th IEEE Conference on Decision and Control (CDC), 2020, pp. 2297–2302.
[27] G. H. Hines, M. Arcak, and A. K. Packard, “Equilibrium-independent passivity: A new definition and numerical certification,” Automatica, vol. 47(9), pp. 1949–1956, 2011.
[28] T. Başar and G. Olsder, Dynamic Noncooperative Game Theory, 1999.
[29] Flam, S. D., “Equilibrium, Evolutionary Stability and Gradient Dynamics,” Int. Game Theory Review, vol. 4(4), pp. 357–370, 2002.
[30] B. I. Bot and E. R. Csetnek, “A second-order dynamical system with Hessian-driven damping and penalty term associated to variational inequalities,” Optimization, vol. 68(7), pp. 1265–1277, 2019.
[31] B. Polyak, “Some methods of speeding up the convergence of iteration methods,” USSR Comp. Math. & Math. Phys., vol. 4(5), pp. 1–17, 1964.
[32] H. Khalil, Nonlinear Systems. Prentice Hall, 2002.
[33] L. Pavel, “Distributed GNE seeking under partial-decision information over networks via a doubly-augmented operator splitting approach,” IEEE Trans. on Automatic Control, p. DOI:10.1109/TAC.2020.2978037, 2019.
(iii): Assume that $R \geq \lambda$ and let $u = J_{\lambda T} x$, $v = J_{\lambda T} y$, and $c = \frac{R^2(1-\mu\lambda)}{R^2+\lambda^2-2\mu R^2}$. Then,
\[
c \|x - y\|^2 = c \|(I + \lambda T)u - (I + \lambda T)v\|^2 \\
= c \|u - v\|^2 + c\lambda^2 \|T u - T v\|^2 + 2c\lambda (T u - T v \mid u - v) \\
- \langle u - v \mid (1 + \lambda T)u - (1 + \lambda T)v \rangle \\
+ \langle u - v \mid (1 + \lambda T)u - (1 + \lambda T)v \rangle \\
= (c - 1) \|u - v\|^2 + c\lambda^2 \|T u - T v\|^2 \\
+ \lambda(2c - 1)(T u - T v \mid u - v) \\
+ \langle u - v \mid (1 + \lambda T)u - (1 + \lambda T)v \rangle \quad (29)
\]

Note that $2c - 1 \geq 0$ for all $R \geq \lambda$ and $1 > \mu \lambda$, so that
\[
c \|x - y\|^2 \geq c \left( \frac{R^2 + \lambda^2 - 2\mu R^2}{R^2} \right) \|u - v\|^2 \\
- (1 - \mu\lambda) \|u - v\|^2 + (u - v \mid (1 + \lambda T)u - (1 + \lambda T)v) \\
= \langle u - v \mid (1 + \lambda T)u - (1 + \lambda T)v \rangle \\
= (J_{\lambda T} x - J_{\lambda T} y \mid x - y) \\

Now assume that $\lambda \geq R$ and $c = \frac{R^2(1-\lambda)}{R^2+\lambda^2-2\lambda R}$ then $2c - 1 \leq 0$. Continuing from (29) and using the fact that $-\langle a \mid b \rangle \geq -\frac{1}{2} \|a\|^2 - s \|b\|^2$, yields
\[
c \|x - y\|^2 \geq (c - 1) \|u - v\|^2 + c\lambda^2 \|T u - T v\|^2 \\
+ \lambda(2c - 1) \left( \frac{R}{2} \|T u - T v\|^2 + \frac{1}{2R} \|u - v\|^2 \right) \\
+ \langle u - v \mid (1 + \lambda T)u - (1 + \lambda T)v \rangle \\
= c \left( \frac{R^2 + \lambda^2 + 2\lambda R}{R^2} \right) \|u - v\|^2 - \left( \frac{\lambda}{R} \right) \|u - v\|^2 \\
+ \langle u - v \mid (1 + \lambda T)u - (1 + \lambda T)v \rangle \\
= (J_{\lambda T} x - J_{\lambda T} y \mid x - y)
\]

Proof of Proposition 8
(i) From [7] Prop 2.3.2 (c).
(ii) From [7] Prop 2.9.25 (a).
(iii) From [7] Prop 2.3.2 (c).
(iv) Note that,
\[
\|T x - T y\|^2 = \left\| \left( \int_0^1 J T(x + t(y - x)) \partial t \right)(y - x) \| (y - x) \|^2 \\
\leq \max_z \|J T(z)\|^2 \|x - y\|^2 \\
\leq L^2 \|x - y\|^2
\]

(iv) Note that,
\[
\|T x - T y\|^2 = \left\| \left( \int_0^1 J T(x + t(y - x)) \partial t \right)(y - x) \| (y - x) \|^2 \\
\geq \min_z \left\| \left( \int_0^1 J T(z) \partial t \right)(y - x) \| (y - x) \|^2 \\
= \min_z (y - x)^T \left( J T(z) J T(z) \right)(y - x) \\
\geq \frac{1}{R^2} \|y - x\|^2
\]
Proof of Lemma 5

First, note that \( F(\bar{r}) = F(x) = 0 \) for \( \bar{r} = 1_N \otimes x^\ast \). Therefore, since \( h \) is the resolvent of \( 1_N \otimes F \) on the consensus subspace, and zeros of \( 1_N \otimes F \) are fixed points of the resolvent, \( \bar{r} = h(\bar{r}) \). Using this, along \( (19) \), we can write
\[
\left< r^\| - \bar{r}^\|, \bar{r}^\| \right> = \alpha \left< r^\| - \bar{r}^\|, z^\| \right> - \alpha \| r^\| - \bar{r}^\| \|^2
\]
\[
+ \left< r^\| - \bar{r}^\|, h(\bar{r}^\|) - h(\bar{r}) \right>.
\]

To bound the last term we use Lemma 7 as follows. For any \( r \in \mathbb{R}^n \) let \( h(r) := \frac{1}{\lambda} F(r) \) the resolvent of \( F \). Using \( F(h(r)) = F(1_N \otimes h(r)) \), we can write \( \alpha \). Using \( \frac{1}{\lambda} F(h(r)) = r \) and \( (17) \), this is equivalent to \( h(1_N \otimes r) = 1_N \otimes h(r) \). As \( h \) is the resolvent of \( F \), under Assumption 2 and 3, we apply Lemma 7 to \( \bar{r} \) with \( \lambda = \frac{1}{\alpha} \).

Therefore, since for any \( r^\| = 1_N \otimes r \), \( h(r^\|) = 1_N \otimes h(r) \) the bounds from Lemma 7 (ii) and (iii) hold, and it follows that the same bounds hold for \( \bar{r} \). Using Lemma 7 (iii) in the last term of (30) yields,
\[
\left< r^\| - \bar{r}^\|, \bar{r}^\| \right> \leq \alpha \| r^\| - \bar{r}^\| \| z^\| \|
\]
\[
- \alpha (1 - \kappa_\beta) \| r^\| - \bar{r}^\| \|^2
\]

Similarly, using (19), we can write,
\[
\left< z^\|, \bar{z}^\| \right> = -\frac{1}{\lambda} \left< z^\|, 1_N \otimes F(z^\| + h(r^\|)) \right>
\]
\[
- \beta \left< z^\|, z^\| + h(r^\|) \right> - \alpha \left< z^\|, \left( \frac{\partial h}{\partial r} \right)(z^\| + h(r^\|) - r^\|) \right>
\]

Substituting \( r^\| = (1_N \otimes \bar{r}^\|) F(1_N \otimes h(r^\|)) \) (cf. (17)) in the middle term and combining terms yields,
\[
\left< z^\|, \bar{z}^\| \right>
\]
\[
= -\frac{1}{\lambda} \left< z^\|, 1_N \otimes F(z^\| + h(r^\|)) - 1_N \otimes F(h(r^\|)) \right>
\]
\[
- \beta \left< z^\|, z^\| \right> - \alpha \left< z^\|, \left( \frac{\partial h}{\partial r} \right)(z^\| + h(r^\|) - r^\|) \right>
\]

The first term is non-negative since \( z^\|, z^\| + h(r^\|) \) and \( h(r^\|) \) are on the consensus subspace and \( 1_N \otimes F \) evaluates to just \( 1_N \otimes F \), which is \( \mu \)-hypo-monotone by Assumption 5. Adding and subtracting \( \bar{r}^\| = h(\bar{r}^\|) \) in the last term, we can then write
\[
\left< z^\|, \bar{z}^\| \right> \leq -\left( \beta - \frac{\mu}{\lambda} \right) \left< z^\|, z^\| \right> - \alpha \left< z^\|, \left( \frac{\partial h}{\partial r} \right) z^\| \right>
\]
\[
+ \alpha \left< z^\|, \frac{\partial h}{\partial r} \right> \left< h(r^\|) - h(\bar{r}^\|) \right>
\]
\[
+ \alpha \left< z^\|, \frac{\partial h}{\partial r} \right> \left< r^\| - \bar{r}^\| \right>.
\]

The second term is non-positive since \( h \) is monotone by Lemma 7 (i) and \( \frac{\partial h}{\partial r} \) is positive semidefinite (cf. Proposition 2.3.2 [7]). Using
\[
\left< h(r^\|) - h(\bar{r}^\|), \bar{r}^\| \right> \leq L_J \| r^\| - \bar{r}^\| \|
\]
\[
\text{and} \quad \| \frac{\partial h}{\partial r} \| \leq L_J \text{ from Lemma 2 (ii)},
\]

yields
\[
\left< z^\|, \bar{z}^\| \right> \leq - \left( \beta - \frac{\mu}{\lambda} \right) \| z^\| \|^2
\]
\[
+ \alpha \left( L_J + L_J^2 \right) \| r^\| - \bar{r}^\| \|^2
\]

Finally, for \( V^\| \) as in the lemma, using the bounds in (31), (32), along the solution of (19), we can write \( V^\| (z^\|, r^\|) \leq -\omega^T \rho \Phi \), where \( \Phi \) as in (20) and \( \omega = \left( \| r^\| - \bar{r}^\| \|, \| z^\| \| \right) \).

It can be easily seen that for any given \( d \in (0,1) \) and \( \alpha \) as in the lemma, \( \Phi \) is positive definite.

Proof of Lemma 6

Note that if we start with \( (H \otimes T) \) and do the change of coordinates \( z = x - J_{\frac{1}{\lambda} F} r \) we get (19) but with \( r^\| \) replaced with \( r, x^\| \) replaced with \( x, z^\| \) with \( z = x - J_{\frac{1}{\lambda} F} r \), \( \tilde{F} \) replaced with \( F \). Therefore, following the same argument as Lemma 5 we can construct a Lyapunov function that shows that \( x^\ast \) is asymptotically stable.

Proof of Lemma 7

Let \( v_i = (x_i, y_i) \) be the \( i \)th eigenvector of \( M \) then,
\[
\begin{pmatrix}
-A - \beta I & \beta I \\
\alpha I & -\alpha I
\end{pmatrix}
\begin{pmatrix}
x_i \\
y_i
\end{pmatrix}
= \lambda_i
\begin{pmatrix}
x_i \\
y_i
\end{pmatrix}
\]

The second row implies that \( x_i = \frac{\alpha + \lambda_i}{\alpha} y_i \). Substituting this into the first row, yields
\[
A y_i = -\frac{\lambda_i (\alpha + \beta + \lambda_i)}{\alpha + \lambda_i} y_i
\]

This equation can only hold true if \( y_i \) is an eigenvector for \( A \). With \( \rho_i \) is the corresponding eigenvalue. Therefore,
\[
\rho_i = -\frac{\lambda_i (\alpha + \beta + \lambda_i)}{\alpha + \lambda_i}
\]

and solving for the roots of the quadratic in \( \lambda_i \) gives (24).

Proof of Lemma 8

(i) From Lemma 7 we see that the characteristic polynomial is \( C_i \equiv \lambda_i^4 + 2(\alpha + \beta + \lambda_i) \lambda_i^2 
\]
\[
+ ((\alpha + \beta + \lambda_i) \lambda_i^3 
\]
\[
+ 2 (r_i + \beta + \lambda_i) \lambda_i + \alpha^2 (r_i^2 + k_i^2)
\]

are in the left half plane. The Routh array for \( C_i \) is,
\[
\begin{array}{c|c|c|c|c|c|c|c}
1 & 2 (\alpha + \beta + \lambda_i) & \alpha^2 (r_i^2 + k_i^2) \\
2 (\alpha + \beta + r_i) & 2 (\alpha r_i (\alpha + \beta + r_i) + \alpha k_i^2) & 0 \\
T_1 & T_2 & 0 & 0 \\
\alpha^2 (r_i^2 + k_i^2) & 0 & 0 & 0
\end{array}
\]
where
\[
T_1 = \begin{bmatrix}
(\alpha + \beta + r_i)^2 + \alpha r_i + \frac{\beta + r_i}{\alpha + \beta + r_i} k_i^2 \\
> 0 \\
\end{bmatrix} > 0
\]
\[
T_2 = \frac{2\alpha}{T_1} \begin{bmatrix}
r_i(\alpha + \beta + r_i)^3 + \frac{\beta + r_i}{\alpha + \beta + r_i} k_i^4 \\
> 0 \\
\end{bmatrix} + \begin{bmatrix}
(\alpha + \beta + r_i)(\beta + 2r_i) k_i^2 \\
> 0 \\
\end{bmatrix} > 0
\]

If we show that all elements in the left column in the Routh array are all positive then the roots of \(C_i\) are less than 0. The term \(2(\alpha + \beta + r_i), \alpha^2 (r_i^2 + k_i^2)\), and \(T_1\) are positive. Either \(r_i \neq 0\) or \(k_i \neq 0\), therefore one of the terms in \(T_2\) will be strictly positive making \(T_2 > 0\). Therefore, \(\lambda_i\) has real part less than 0.

(iii) The term \(\alpha^2 (r_i^2 + k_i^2)\) is always positive. By assumption, \(\alpha > 0\) and \(\beta + r_i > 0\), therefore the term \(2(\alpha + \beta + r_i)\) is positive. For the \(T_1\) term, let \(\epsilon = \beta + r_i > 0\) then,
\[
T_1 = \left[ (\alpha + \epsilon)^2 + \alpha r_i + \frac{\epsilon}{\epsilon + \alpha} k_i^2 \right]
\]

Multiplying \(T_1\) by \(\epsilon + \alpha > 0\) gives the condition,
\[
0 < \frac{[(\alpha + \epsilon)^3 + \alpha(\epsilon + \alpha)r_i + \epsilon k_i^2]}{2(\alpha + \epsilon)(\epsilon + \alpha)}
0 < (\alpha + \epsilon)^3 + \frac{\alpha(\epsilon + \epsilon)r_i + \epsilon k_i^2}{(\alpha + \epsilon)(\epsilon + \alpha)}
0 < \frac{\alpha^2(\epsilon + \alpha)^2 + \epsilon k_i^2}{(\epsilon + \alpha)T_1} \left( r_i(\epsilon + \alpha)^4 + \epsilon k_i^2 + (\alpha + \epsilon)^2 r_i \right)
\]

Since \(\frac{2}{(\epsilon + \alpha)T_1} > 0\) the condition for \(T_2 > 0\) is,
\[
r_i x^2 + (\epsilon + r_i) k_i^2 x + \epsilon k_i^4 > 0
\]
where \(x = (\epsilon + \alpha)^2\). The roots of this equation are,
\[
x = \frac{1}{2r_i} [-\epsilon + r_i k_i^2 + \sqrt{(\epsilon + r_i) k_i^4 - 4 r_i \epsilon k_i^2}]
\]
\[
= \frac{1}{2r_i} [-\epsilon + r_i k_i^2 + \epsilon (r_i - k_i^2)] = -\frac{\epsilon k_i^2}{r_i} \quad \text{or} \quad k_i^2
\]

Therefore, \(x \in (-k_i^2, \frac{-\epsilon k_i^2}{r_i})\) for \(T_2 > 0\), but \(x = (\epsilon + \alpha)^2\) so \(x = (\epsilon + \alpha)^2 \in (0, \frac{-\epsilon k_i^2}{r_i})\) which implies,
\[
0 < x < -\frac{\beta + r_i}{\epsilon} + \sqrt{-\frac{(\beta + r_i)k_i^2}{r_i}}
\]

**Proof of Theorem 5**

After performing a change of coordinates as in \(\text{[23]}\) and a decomposition as in the nonlinear case, the dynamics \(\text{[HAy]}\) can be written as,
\[
\dot{x}^\parallel = \Pi_x^\parallel (R^T A x - \beta (x - r) - cL x) 
= \Pi_x^\parallel R^T A (x^\parallel + x^\perp) - \beta (x^\parallel - r^\parallel) 
\]
\[
\dot{x}^\perp = (I - \Pi_x^\parallel) (R^T A x - \beta (x - r) - cL x) 
= (I - \Pi_x^\parallel) R^T A (x^\parallel + x^\perp) - \beta (x^\perp - r^\perp) - cL x^\perp
\]
\[
\dot{r}^\parallel = \alpha (x^\parallel - r^\parallel) 
\]
\[
\dot{r}^\perp = \alpha (x^\perp - r^\perp)
\]

Let \(w = w^\parallel + w^\perp\), \(w^\parallel = (x^\parallel, r^\parallel)\), \(w^\perp = (x^\perp, r^\perp)\), and \((w^\parallel)^\parallel = ((x^\parallel)^\parallel, (r^\parallel)^\parallel)\). The matrix \(M\) has the same structure as \(M\) (some terms scaled). From Lemma \(\text{[8]}\) we know that \(M\), for the \(\alpha\) and \(\beta\) satisfying the assumptions in the theorem, has all its eigenvalues with real part less than 0 and therefore there exists a \(P\) satisfying the Lyapunov equation. Consider the following Lyapunov function,
\[
V(w) = \frac{1}{2} \|x^\parallel\|^2 + \frac{\beta}{2\alpha} \|r^\parallel\|^2 + \sum_{i \in \mathcal{N}} \| (w^i)^\parallel - w^* \|^2_p
\]

For the first two terms in \((35)\),
\[
\frac{d}{dt} \left( \frac{1}{2} \|x^\parallel\|^2 + \frac{\beta}{2\alpha} \|r^\parallel\|^2 \right)
= (x^\perp)^T (I - \Pi_x^\parallel) R^T A (x^\parallel + x^\perp) - cL x^\perp 
- (w^\perp)^T \left[ \begin{matrix} \beta I & -\beta I \\ -\beta I & \beta I \end{matrix} \right] w^\perp
\]

Since the last term is equal to \(-\beta (x^\perp - r^\perp)^2 \leq 0\), therefore
\[
\frac{d}{dt} \left( \frac{1}{2} \|x^\parallel\|^2 + \frac{\beta}{2\alpha} \|r^\parallel\|^2 \right)
\leq (x^\perp)^T (I - \Pi_x^\parallel) R^T A (x^\parallel + x^\perp) - c\lambda_2(L) \|x^\parallel\|^2
\]

Using \(\| (I - \Pi_x^\parallel) R^T A \| = \sqrt{\frac{N-1}{N}} \|A\|\), \(\|A\| \leq L_A\) and \(R^T A x^* = 0\) yields,
\[
\frac{d}{dt} \left( \frac{1}{2} \|x^\parallel\|^2 + \frac{\beta}{2\alpha} \|r^\parallel\|^2 \right)
\leq \sqrt{\frac{N-1}{N}} L_A \left( \|x^\parallel\|^2 + \|x^\perp\|^2 \|x^\parallel - x^*\|^2 \right)
- c\lambda_2(L) \|x^\parallel\|^2
\]
\[
\leq -c\lambda_2(L) \|x^\parallel\|^2
\]

Note that \(w^\parallel\) can be written as \(w^\parallel = 1 \otimes w^\parallel\) and \((w^i)^\parallel = (w^i)^\parallel = w^\parallel\). For the third term in \((35)\), along the solution
of (33) and (34),
\[
\frac{d}{dt} \sum_{i \in N} \| (w^i)\| - w^* \|^2_P
= \sum_{i \in N} ((w^i)\| - w^*)^T (\hat{\Theta}_i + \hat{\Theta}^T) ((w^i)\| - w^*)
+ ((w^i)\| - w^*)^T (P + P^T) Q w^\perp
\]
where \( Q = \begin{bmatrix} -\frac{1}{N} A & 0 \\ 0 & 0 \end{bmatrix} \).

\[
= - \| w\| - w^* \|^2 + \sum_{i \in N} ((w^i)\| - w^*)^T (P + P^T) Q w^\perp
= - \| w\| - w^* \|^2 + (w\| - w^*)^T [I \otimes (P + P^T)] [1 \otimes Q] w^\perp.
\]
Using \( \| P \| = p \) and \( \| \frac{1}{N} 1 \otimes A \| \leq \frac{1}{\sqrt{N}} L_A \),
\[
\frac{d}{dt} \sum_{i \in N} \| (w^i)\| - w^* \|^2_V
\leq - \| w\| - w^* \|^2 + 2 p L_A \sqrt{N} \| w\| - w^* \| \| x^\perp \|
\]
Therefore, from (36) and (37) the Lyapunov function \( V \), (35), satisfies
\[
\frac{d}{dt} V(w) \leq \omega^T \begin{bmatrix} -1 & \frac{L_A}{2\sqrt{N}} (2 p + \sqrt{N - 1}) \\ * & -c \lambda_2(L) + \sqrt{\frac{N - 1}{N}} L_A \end{bmatrix} \omega
\]
where \( \omega = \text{col}(\| w\| - w^* \|, \| x^\perp \|) \). Under (27) the matrix is negative definite and using LaSalle’s Invariance Principle [32] concludes the proof.