Equilibria of Fully Decentralized Learning in Networked Systems

Yan Jiang, Wenqi Cui, Baosen Zhang {JIANGYAN,WENQICUI,ZHANGBAO}@UW.EDU
Department of Electrical and Computer Engineering, University of Washington, Seattle, WA 98195, USA

Jorge Cortés CORTES@UCSD.EDU
Department of Mechanical and Aerospace Engineering, University of California, San Diego, CA 92093, USA

Editors: N. Matni, M. Morari, G. J. Pappas

Abstract

Existing settings of decentralized learning either require players to have full information or the system to have certain special structure that may be hard to check and hinder their applicability to practical systems. To overcome this, we identify a structure that is simple to check for linear dynamical system, where each player learns in a fully decentralized fashion to minimize its cost. We first establish the existence of pure strategy Nash equilibria in the resulting noncooperative game. We then conjecture that the Nash equilibrium is unique provided that the system satisfies an additional requirement on its structure. We also introduce a decentralized mechanism based on projected gradient descent to have agents learn the Nash equilibrium. Simulations on a 5-player game validate our results.

Keywords: Decentralized control, multi-agent learning, Nash equilibrium, noncooperative game.

1. Introduction

Many real world systems are too large and complex for decisions to be made in a centralized fashion. Instead, there are a multitude of decision makers (or players), interacting over a system, each possessing limited knowledge and observations. Thus, the study of decentralized decision making has been a topic of interest for several decades (see, e.g., Tsitsiklis (1984); Olfati-Saber et al. (2007); and the references within). Recently, the advancement of machine learning in multi-agent setting has attracted significant attention from the control and learning communities, which gives rise to successful application of multi-agent decision learning across multiple fields, including power system operations (Yang et al., 2018; Cui et al., 2022), traffic control (Bazzan, 2009), communication networks (Han et al., 2012), and others (Li et al., 2022).

A foundational question in decentralized decision making and learning is whether the players would reach some type of equilibria. One setting that has been extensively studied is that of linear quadratic (LQ) games (Zhang et al., 2019; Mazumdar et al., 2020), which generalize the well-known linear quadratic regulator (LQR) problems. Unlike LQR problems, which can be considered to have a single agent (Fazel et al., 2018), LQ games have multiple players that interact over a linear system and all try to minimize their individual regulation and control costs. This complexity makes LQ games not enjoy guarantees of convergence to Nash equilibria, a property enjoyed by the LQR problem with policy-gradient methods. Moreover, although a player is limited in its input channels (Başar and Olsder, 1998; Engwerda, 2005; Li et al., 2022), it typically has full information (access to the full state). This assumption of the availability of full information makes it difficult to apply related results to many practical systems.

For systems where the players are limited both in input and information, several properties have been discovered to guarantee that good decentralized controllers can be found in a tractable manner. These include quadratic invariance (Rotkowitz and Lall, 2002; Lessard and Lall, 2014), spatial invariance (Bamieh et al., 2002), partially nestedness (Shah and Parrilo, 2013), and positiveness (Rantzer, 2015). These conditions, however, can be challenging to check in practice. In addition, they often require the players to have nested information, which may not hold in practice.

It is important to note that some structure of the system is necessary for decentralized learning to be analytically tractable. Even for linear systems, finding the optimal controllers is NP-hard in general (Blondel and Tsitsiklis, 2000). Even if restricted to linear controllers, the feasible set of the stabilizing controllers can have an exponential number of connected pieces (Feng and Lavaei, 2019), and understanding the equilibria or convergence of the learned controllers becomes very difficult.

In this paper, we identify a new structure for a class of systems where the behavior of decentralized learning can be explicitly characterized. In particular, we study a game over a linear and symmetric dynamical system with each player being modeled as a part of the state, where the action of the players is to choose a linear feedback control gain that minimizes quadratic costs on its own state regulation and control effort. Notably, compared to existing LQ games in the literature (Fazel et al., 2018; Zhang et al., 2019; Mazumdar et al., 2020), we adopt a fully decentralized setting, in the sense that each player only knows its own information and takes an action directly affecting its own state, which can be naturally characterized as a noncooperative game. This captures the fact that, in many settings, the edge devices are becoming increasingly intelligent and capable of making sophisticated decisions, but they still do not have access to regular real-time communication with other devices. We show that there exists at least one pure strategy Nash Equilibrium in such a noncooperative game by establishing that, in contrast to existing results, the stabilizing controllers lie in a convex region and the cost functions are convex in the control gains. This, in turn, provides a simple way for players to estimate the gradients of the cost functions in a completely decentralized way and converge to the Nash equilibrium through gradient play. We conjecture that the pure strategy Nash equilibria is in fact unique, where partial and numerical results for this conjecture are provided.

The key property used in our analysis is the symmetry of the dynamical system. That is, the state matrix is symmetric. This condition is distinct from existing ones and has the benefit that it is simple to check. The symmetry of the system is satisfied by many practical systems. For example, in power distribution systems with angle droop control (Zhang and Xie, 2016; Huang et al., 2020), the system is always symmetric since the matrix comes from the Laplacian of the underlying network.

The rest of this paper is organized as follows. Section 2 formulates the decentralized learning problem in a symmetric linear dynamical system as a n-player noncooperative game. Section 3 shows the existence of pure strategy Nash equilibria in such a game and further conjectures its uniqueness under additional conditions on the system structure. Section 4 introduces a decentralized learning mechanism based on projected gradient descent and discusses how to implement it. Section 5 presents our conclusion and ideas for future work.

2. Problem Setup

We consider a networked system with n players (or agents), whose dynamics is given by

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{u}(t)\,, \tag{1}$$

where $\boldsymbol{x}(t) := (x_i(t), i \in [n]) \in \mathbb{R}^n$ is the state vector, $\boldsymbol{u}(t) := (u_i(t), i \in [n]) \in \mathbb{R}^n$ is the control input, and $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is the state matrix.[1] We make the following assumption:

Assumption 1 (Structure assumption) *The state matrix $\boldsymbol{A}$ is symmetric and negative definite.*

Remark 1 (Structure interpretations and extensions) The symmetry of the state matrix is key to our developments since most of results in this paper critically rely on this assumption. The negative definiteness can be relaxed as discussed later in Remark 2. Assumption 1 reflects the graph structure of the system found in many applications. For example, in microgrid control (Huang et al., 2020; Cui and Zhang, 2022), $\boldsymbol{A}$ is related to a Laplacian matrix that captures the active power flow. □

To model a fully decentralized setting, we assume that each controller u_i can only depend on the state of the ith player. Namely, each player chooses an action $k_i \in \left[0, \overline{k}_i\right]$, with $\overline{k}_i > 0$ being some upper bound, such that the ith component of the control input is determined by

$$u_i(t) = -k_i x_i(t)\,. \tag{2}$$

Let $\boldsymbol{K} := \operatorname{diag}(k_i, i \in [n]) \in \mathbb{R}^{n\times n}$. Then the closed-loop system of (1) under (2) becomes

$$\dot{\boldsymbol{x}}(t) = (\boldsymbol{A} - \boldsymbol{K})\boldsymbol{x}(t)\,. \tag{3}$$

Remark 2 (Hurwitz closed-loop system matrix) Note that, with Assumption 1 on $\boldsymbol{A}$ and the restrictions that $k_i \geq 0$, the closed-loop system (3) is always stable since $\boldsymbol{A} - \boldsymbol{K} \prec 0$. However, if $\boldsymbol{A}$ is symmetric but not negative definite, the set of controller gains that make the closed-loop system (3) stable is a convex set determined by $\boldsymbol{A} - \boldsymbol{K} \prec 0$. In this case, it is easy to find a lower bound $\underline{k}_i$ such that $\boldsymbol{A} - \boldsymbol{K} \prec 0$ if $k_i > \underline{k}_i$, $\forall i \in [n]$. □

The goal of the ith player is to minimize its own expected cost $J_i(k_i, k_{-i})$ on state deviations and control effort along the trajectories of the system (3), given the actions of other players $k_{-i} := \{k_1, \dots, k_{i-1}, k_{i+1}, \dots, k_n\}$. Formally, we define the cost of the ith player as

$$J_i(k_i, k_{-i}) := \mathbb{E}\left[\int_0^\infty \left(x_i^2(t) + \rho_i u_i^2(t)\right) \mathrm{d}t\right], \quad \forall i \in [n]\,, \tag{4}$$

where $\rho_i \geq 0$ is the coefficient for tradeoff between the two components (state deviation and control effort). Note that the expectation $\mathbb{E}\left[\cdot\right]$ is taken with respect to random initial conditions $\boldsymbol{x}(0)$, where we make the common assumption (e.g., see Mendel and Gieseking (1971)) that $\mathbb{E}\left[\boldsymbol{x}(0)\boldsymbol{x}(0)^T\right] = \boldsymbol{I}_n$, that is, the components of $\boldsymbol{x}(0)$ are independent and identically distributed (i.i.d.).

Remark 3 (Infinite time-horizon) We take an infinite time-horizon in (4) for several reasons. First, it makes the analysis cleaner and thus is adopted in many settings (e.g., see Fazel et al. (2018); Dean et al. (2020); Bu et al. (2019); and references within). Second, for stable systems, finite trajectory costs are well approximated by an infinite trajectory cost if the number of time steps in a trajectory is not too small. Third, this cost is equivalent to the expected average cost in systems with persistent white noise (Mendel and Gieseking, 1971; Kwon and Han, 2006; Weitenberg et al., 2019). □

[1] Throughout the paper, vectors are denoted in bold lower case and matrices are denoted in bold upper case, while scalars are unbolded.

The setting above defines a noncooperative game, where each player has action space $k_i \in [0, \overline{k}_i]$ and cost $J_i(k_i, k_{-i})$. A pure strategy Nash equilibrium of the game is defined as an action profile of the players where no single player i can obtain a lower cost by choosing a different action, given that the actions of other players are fixed. That is, $(k_1^*, \dots, k_n^*)$ is a pure strategy Nash equilibrium if, $\forall i \in [n]$, $J_i(k_i^*, k_{-i}^*) \leq J_i(k_i', k_{-i}^*)$, $\forall k_i' \in [0, \overline{k}_i]$. It is well known that not all games have a pure strategy Nash equilibrium (Başar and Olsder, 1998), especially when the actions of the players are not explicitly reflected in the cost functions (Marden and Shamma, 2015). We study in Section 3 the existence of pure strategy Nash equilibria for this game and describe in Section 4 how the players update their actions to find them.

3. Pure Strategy Nash Equilibria

In this section, we study the existence and uniqueness of pure strategy Nash equilibria for the n-player noncooperative game introduced in Section 2.

3.1. Existence of Nash Equilibrium

Notice that each player has an action space $\left[0, \overline{k}_i\right]$ that is a closed, bounded, and convex subset of $\mathbb{R}$. Therefore, based on the well-known result (Başar and Olsder, 1998, Theorem 4.3), in order to show the existence of pure strategy Nash equilibria, it suffices to show that the cost function $J_i(k_i, k_{-i})$ is jointly continuous in all its arguments and strictly convex in k_i, for every k_{-i}. In this subsection, we proceed by presenting a sequence of results that eventually enable us to prove the following main result.

Theorem 1 (Existence of pure strategy Nash equilibria) *The n-player noncooperative game admits a pure strategy Nash equilibrium.*

We start by investigating an explicit expression for the cost function $J_i(k_i, k_{-i})$. Clearly, k_i does not explicitly show up in the definition of $J_i(k_i, k_{-i})$ given in (4), which hinders our analysis. The next result addresses this by providing an explicit expression of $J_i(k_i, k_{-i})$ in terms of k_i.

Lemma 1 (Individual cost functions) *The cost function of the ith player, $\forall i \in [n]$, is given by*

$$J_i(k_i, k_{-i}) = \frac{\left(1 + \rho_i k_i^2\right)}{2} f_i(k_i, k_{-i}) \,, \quad \textit{with } f_i(k_i, k_{-i}) := \boldsymbol{e}_i^T \left(\boldsymbol{K} - \boldsymbol{A}\right)^{-1} \boldsymbol{e}_i \,, \tag{5}$$

where $\boldsymbol{e}_i \in \mathbb{R}^n$ is the ith standard basis vector.

Proof First, substituting (2) to (4) yields

$$J_i(k_i, k_{-i}) = \left(1 + \rho_i k_i^2\right) \mathbb{E}\left[\int_0^\infty x_i^2(t) \, \mathrm{d}t\right] = \left(1 + \rho_i k_i^2\right) \mathbb{E}\left[\int_0^\infty \boldsymbol{x}(t)^T \boldsymbol{e}_i \boldsymbol{e}_i^T \boldsymbol{x}(t) \, \mathrm{d}t\right] \,, \tag{6}$$

where the second equality uses $x_i(t) = \boldsymbol{e}_i^T \boldsymbol{x}(t)$. Note that $\boldsymbol{x}(t)$ in (6) still implicitly depends on k_i. Therefore, we need to get an explicit expression of $\boldsymbol{x}(t)$ in terms of k_i to perform further analysis.

Since the solution to the closed-loop system (3) is $\boldsymbol{x}(t) = e^{(\boldsymbol{A}-\boldsymbol{K})t} \boldsymbol{x}(0)$, (6) becomes:

$$J_i(k_i, k_{-i}) \overset{\textcircled{1}}{=} \left(1 + \rho_i k_i^2\right) \mathbb{E}\left[\int_0^\infty \boldsymbol{x}(0)^T e^{(\boldsymbol{A}-\boldsymbol{K})t} \boldsymbol{e}_i \boldsymbol{e}_i^T e^{(\boldsymbol{A}-\boldsymbol{K})t} \boldsymbol{x}(0) \, \mathrm{d}t\right]$$

$$
\begin{aligned}
&= \left(1+\rho_i k_i^2\right) \mathbb{E}\left[\boldsymbol{x}(0)^T \int_0^\infty e^{(\boldsymbol{A}-\boldsymbol{K})t} \boldsymbol{e}_i \boldsymbol{e}_i^T e^{(\boldsymbol{A}-\boldsymbol{K})t} \,\mathrm{d}t \,\boldsymbol{x}(0)\right] \\
&= \left(1+\rho_i k_i^2\right) \mathbb{E}\left[\operatorname{tr}\left(\boldsymbol{x}(0)^T \int_0^\infty e^{(\boldsymbol{A}-\boldsymbol{K})t} \boldsymbol{e}_i \boldsymbol{e}_i^T e^{(\boldsymbol{A}-\boldsymbol{K})t} \,\mathrm{d}t \,\boldsymbol{x}(0)\right)\right] \\
&\overset{②}{=} \left(1+\rho_i k_i^2\right) \mathbb{E}\left[\operatorname{tr}\left(\int_0^\infty e^{(\boldsymbol{A}-\boldsymbol{K})t} \boldsymbol{e}_i \boldsymbol{e}_i^T e^{(\boldsymbol{A}-\boldsymbol{K})t} \,\mathrm{d}t \,\boldsymbol{x}(0)\boldsymbol{x}(0)^T\right)\right] \\
&= \left(1+\rho_i k_i^2\right) \operatorname{tr}\left(\int_0^\infty e^{(\boldsymbol{A}-\boldsymbol{K})t} \boldsymbol{e}_i \boldsymbol{e}_i^T e^{(\boldsymbol{A}-\boldsymbol{K})t} \,\mathrm{d}t \,\mathbb{E}\left[\boldsymbol{x}(0)\boldsymbol{x}(0)^T\right]\right) \\
&\overset{③}{=} \left(1+\rho_i k_i^2\right) \operatorname{tr}\left(\int_0^\infty e^{(\boldsymbol{A}-\boldsymbol{K})t} \boldsymbol{e}_i \boldsymbol{e}_i^T e^{(\boldsymbol{A}-\boldsymbol{K})t} \,\mathrm{d}t\right) \\
&\overset{④}{=} \left(1+\rho_i k_i^2\right) \operatorname{tr}\left(\int_0^\infty \boldsymbol{e}_i^T e^{2(\boldsymbol{A}-\boldsymbol{K})t} \boldsymbol{e}_i \,\mathrm{d}t\right) \\
&= \left(1+\rho_i k_i^2\right) \operatorname{tr}\left(\boldsymbol{e}_i^T \int_0^\infty e^{2(\boldsymbol{A}-\boldsymbol{K})t} \,\mathrm{d}t\boldsymbol{e}_i\right) = \left(1+\rho_i k_i^2\right) \boldsymbol{e}_i^T \int_0^\infty e^{2(\boldsymbol{A}-\boldsymbol{K})t} \,\mathrm{d}t \,\boldsymbol{e}_i\,, \qquad (7)
\end{aligned}
$$

where ① uses the closed-loop solution, ② and ④ use the cyclic property of the trace, and ③ uses $\mathbb{E}\left[\boldsymbol{x}(0)\boldsymbol{x}(0)^T\right] = \boldsymbol{I}_n$. Note that the integral $\int_0^\infty e^{2(\boldsymbol{A}-\boldsymbol{K})t}\,\mathrm{d}t$ is a common integral (see, e.g., Hespanha (2018)) given by

$$
\int_0^\infty e^{2(\boldsymbol{A}-\boldsymbol{K})t}\,\mathrm{d}t = \frac{(\boldsymbol{K}-\boldsymbol{A})^{-1}}{2}\,. \tag{8}
$$

Therefore, substituting (8) into (7) yields (5). ■

Lemma 1 provides an explicit expression of the cost function $J_i(k_i, k_{-i})$ in terms of the action k_i. This allows us next to characterize the convex properties of the cost function.

Lemma 2 (Strict convexity of cost functions) *The cost function $J_i(k_i, k_{-i})$, $\forall i \in [n]$, is strictly convex in k_i for each $k_{-i} \in \prod_{j\in[n]\setminus\{i\}} \left[0, \overline{k}_j\right]$.*

Proof Note that the action space $\left[0, \overline{k}_i\right]$ of each player is convex. By the second-order condition for convexity (Boyd and Vandenberghe, 2004, Chapter 3.1.4), in order to establish the strict convexity of $J_i(k_i, k_{-i})$ in k_i, it suffices to show that

$$
\frac{\partial^2 J_i(k_i, k_{-i})}{\partial k_i^2} > 0\,, \qquad \forall i \in [n]\,. \tag{9}
$$

We start by taking the partial derivative of $J_i(k_i, k_{-i})$ with respect to k_i. Direct calculations on (5) show that

$$
\frac{\partial J_i(k_i, k_{-i})}{\partial k_i} = \rho_i k_i f_i(k_i, k_{-i}) + \frac{\left(1+\rho_i k_i^2\right)}{2} \frac{\partial f_i(k_i, k_{-i})}{\partial k_i}\,. \tag{10}
$$

Next, we take the partial derivative of (10) with respect to k_i to obtain

$$
\frac{\partial^2 J_i(k_i, k_{-i})}{\partial k_i^2} = \rho_i f_i(k_i, k_{-i}) + 2\rho_i k_i \frac{\partial f_i(k_i, k_{-i})}{\partial k_i} + \frac{\left(1+\rho_i k_i^2\right)}{2} \frac{\partial^2 f_i(k_i, k_{-i})}{\partial k_i^2}\,. \tag{11}
$$

To obtain the partial derivative of $f_i(k_i, k_{-i})$, we use (5) to get

$$\begin{aligned}\frac{\partial f_i(k_i, k_{-i})}{\partial k_i} &= \boldsymbol{e}_i^T \frac{\partial \left[(\boldsymbol{K}-\boldsymbol{A})^{-1}\right]}{\partial k_i} \boldsymbol{e}_i \overset{\textcircled{1}}{=} -\boldsymbol{e}_i^T \left(\boldsymbol{K}-\boldsymbol{A}\right)^{-1} \frac{\partial \left(\boldsymbol{K}-\boldsymbol{A}\right)}{\partial k_i} \left(\boldsymbol{K}-\boldsymbol{A}\right)^{-1} \boldsymbol{e}_i \\ &= - \boldsymbol{e}_i^T \left(\boldsymbol{K}-\boldsymbol{A}\right)^{-1} \boldsymbol{e}_i \boldsymbol{e}_i^T \left(\boldsymbol{K}-\boldsymbol{A}\right)^{-1} \boldsymbol{e}_i \overset{\textcircled{2}}{=} -f_i^2(k_i, k_{-i})\,, \end{aligned} \tag{12}$$

where ① uses the formula for the derivative of an inverse matrix (Petersen and Pedersen, 2012, Chapter 2.2) and ② uses the definition of $f_i(k_i, k_{-i})$ in (5) twice. Further taking the partial derivative of (12) with respect to k_i yields

$$\frac{\partial^2 f_i(k_i, k_{-i})}{\partial k_i^2} = -2 f_i(k_i, k_{-i}) \frac{\partial f_i(k_i, k_{-i})}{\partial k_i} = 2 f_i^3(k_i, k_{-i})\,, \tag{13}$$

where the second equality uses (12). Now, substituting (12) and (13) into (11) yields

$$\begin{aligned}\frac{\partial^2 J_i(k_i, k_{-i})}{\partial k_i^2} &= \rho_i f_i(k_i, k_{-i}) - 2\rho_i k_i f_i^2(k_i, k_{-i}) + \left(1 + \rho_i k_i^2\right) f_i^3(k_i, k_{-i}) \\ &= f_i(k_i, k_{-i}) \left[\rho_i \left(1 - k_i f_i(k_i, k_{-i})\right)^2 + f_i^2(k_i, k_{-i})\right]\,. \end{aligned} \tag{14}$$

Clearly, the sign of (14) only depends on the sign of $f_i(k_i, k_{-i})$, since $\rho_i \geq 0$ and the terms inside the square brackets are squared. It follows from Remark 2 that $(\boldsymbol{K}-\boldsymbol{A})^{-1} \succ 0$, which further implies that $f_i(k_i, k_{-i}) > 0$ by its definition in (5). Hence, it follows directly from (14) that (9) holds, concluding the proof of strict convexity. ■

We now have the core element required to establish the existence of pure strategy Nash equilibria stated in Theorem 1.

Proof of Theorem 1. Recall that, by (Başar and Olsder, 1998, Theorem 4.3), since the action space $\left[0, \overline{k}_i\right]$ of each player is a closed, bounded, and convex subset of $\mathbb{R}$, the n-player noncooperative game in question admits a pure strategy Nash equilibrium if the cost function $J_i(k_i, k_{-i})$ is jointly continuous in all its arguments and strictly convex in k_i for every k_{-i}. By Lemma 2, $J_i(k_i, k_{-i})$ is strictly convex in k_i for every k_{-i}. Thus, it remains to show that $J_i(k_i, k_{-i})$ is jointly continuous in all its arguments. This can be seen clearly when one notices that $J_i(k_i, k_{-i})$ in (5) is eventually a quotient of two multivariable polynomial functions in k_j, $\forall j \in [n]$. First, since the multivariable polynomial functions can be considered as a sum of products of polynomial functions, their joint continuity directly follows from the fact that polynomial functions are continuous everywhere and a product or sum of continuous functions is continuous as well (Rudin, 2013, Theorem 4.9). Then, a quotient of such two continuous functions is also continuous everywhere except perhaps at the points which make the denominator zero (Rudin, 2013, Theorem 4.9). Yet, no such points exist in the action space since $\boldsymbol{K} - \boldsymbol{A} \succ 0$ by Remark 2. Hence, $J_i(k_i, k_{-i})$ is jointly continuous in all its arguments, concluding the proof. ■

3.2. Uniqueness of Nash Equilibrium

Having shown the existence of pure strategy Nash Equilibria in the n-player noncooperative game, here we analyze its uniqueness. A well-established condition to guarantee uniqueness is proposed by Rosen (Rosen, 1965, Theorem 6) and states that the game in question admits a unique Nash equilibrium if, $\forall \boldsymbol{k} := (k_i, i \in [n]) \in \prod_{i\in[n]} [0, \overline{k}_i]$,

$$\boldsymbol{G}(\boldsymbol{k}) + \boldsymbol{G}^T(\boldsymbol{k}) \succ 0 \,, \tag{15}$$

where $\boldsymbol{G}(\boldsymbol{k}) \in \mathbb{R}^{n\times n}$ is the Jacobian of the so-called pseudogradient defined as the stacked vector of partial derivatives of the cost function $J_i(k_i, k_{-i})$ with respect to the action k_i, i.e.,

$$\boldsymbol{g}(\boldsymbol{k}) := \left(\frac{\partial J_i(k_i, k_{-i})}{\partial k_i}, i \in [n]\right) \in \mathbb{R}^n \qquad \text{and} \qquad G_{ij}(\boldsymbol{k}) := \frac{\partial g_i(\boldsymbol{k})}{\partial k_j} \,. \tag{16}$$

In general, the condition (15) does not hold for the game in question. Yet, after performing extensive numerical tests, we have found that, if the state matrix $\boldsymbol{A}$ is symmetric strictly diagonally dominant with negative diagonal entries, (15) always holds. Thus, we formulate the conjecture below.

Conjecture 1 (Unique Nash equilibrium) *If $\boldsymbol{A}$ is symmetric strictly diagonally dominant with negative diagonal entries, then the n-player noncooperative game has a unique Nash equilibrium.*

We prove the conjecture for the 2-player case. For simplicity, we assume $\rho_1 = \rho_2 = 0$.
Proof 2-players In this case, $\boldsymbol{K} - \boldsymbol{A}$ can be parameterized as

$$\boldsymbol{K} - \boldsymbol{A} = \begin{bmatrix} k_1 - a_{11} & -a_{12} \\ -a_{12} & k_2 - a_{22} \end{bmatrix} \quad \text{and} \quad (\boldsymbol{K} - \boldsymbol{A})^{-1} = \frac{1}{\nu}\begin{bmatrix} k_2 - a_{22} & a_{12} \\ a_{12} & k_1 - a_{11} \end{bmatrix}, \tag{17}$$

with $a_{11} < -|a_{12}|$, $a_{22} < -|a_{12}|$, $k_1 \in [0, \overline{k}_1]$, $k_2 \in [0, \overline{k}_2]$, and $\nu := (k_1 - a_{11})(k_2 - a_{22}) - a_{12}^2$. Note that this parameterization ensures that

$$k_1 - a_{11} > |a_{12}| \geq 0 \qquad \text{and} \qquad k_2 - a_{22} > |a_{12}| \geq 0 \,. \tag{18}$$

We next write the pseudogradient $\boldsymbol{g}(\boldsymbol{k})$ defined in (16). Observe from (10) and (12) in the proof of Lemma 2 that, when $\rho_i = 0$, we have $\partial J_i(k_i, k_{-i})/\partial k_i = -f_i^2(k_i, k_{-i})/2$, $\forall i \in [n]$, which together with the definition of $f_i(k_i, k_{-i})$ in (5) and the expression of $(\boldsymbol{K} - \boldsymbol{A})^{-1}$ in (17) yields

$$\boldsymbol{g}(\boldsymbol{k}) = -\frac{1}{2\nu^2}\begin{bmatrix} (k_2 - a_{22})^2 \\ (k_1 - a_{11})^2 \end{bmatrix} . \tag{19}$$

Through standard calculus, we can get

$$\boldsymbol{G}(\boldsymbol{k}) + \boldsymbol{G}^T(\boldsymbol{k}) = \frac{1}{\nu^3}\begin{bmatrix} 2(k_2 - a_{22})^3 & a_{12}^2(k_1 - a_{11} + k_2 - a_{22}) \\ a_{12}^2(k_1 - a_{11} + k_2 - a_{22}) & 2(k_1 - a_{11})^3 \end{bmatrix} .$$

By Sylvester's criterion, to show $\boldsymbol{G}(\boldsymbol{k}) + \boldsymbol{G}^T(\boldsymbol{k}) \succ 0$, it suffices to show

$$\mu := 4(k_1 - a_{11})^3(k_2 - a_{22})^3 - a_{12}^4(k_1 - a_{11} + k_2 - a_{22})^2 > 0 \,, \tag{20}$$

since $\nu > 0$ and $k_2 - a_{22} > 0$ due to (18). To see why (20) holds, we can expand μ as

$$\begin{aligned} \mu &= 4(k_1 - a_{11})^3(k_2 - a_{22})^3 - a_{12}^4(k_1 - a_{11})^2 - a_{12}^4(k_2 - a_{22})^2 - 2a_{12}^4(k_1 - a_{11})(k_2 - a_{22}) \\ &= (k_1 - a_{11})^2\left[(k_1 - a_{11})(k_2 - a_{22})^3 - a_{12}^4\right] + (k_2 - a_{22})^2\left[(k_1 - a_{11})^3(k_2 - a_{22}) - a_{12}^4\right] \\ &\quad + 2(k_1 - a_{11})(k_2 - a_{22})\left[(k_1 - a_{11})^2(k_2 - a_{22})^2 - a_{12}^4\right] > 0 \,, \end{aligned}$$

where the inequality is due to (18). ■

The proof of Conjecture 1 for the general n-player game is a direction of future research.

4. Decentralized Learning of Nash Equilibrium via Projected Gradient Descent

In this section, we present a mechanism for the players to reach the Nash equilibrium of the game. Since the goal of each player is to selfishly minimize its own cost $J_i(k_i, k_{-i})$ in (4), an intuitive choice for each player as the game proceeds is to update its action k_i by modifying it in the direction where the cost $J_i(k_i, k_{-i})$ descends the fastest. More specifically, after a random initialization of the action k_i, denoted as $k_i^{(0)}$, such that $k_i^{(0)} \in \left[0, \overline{k}_i\right]$, each player updates its action $k_i^{(l)}$ at the lth stage of the game along a projected direction of cost descent, i.e.,

$$k_i^{(l)} = \left[k_i^{(l-1)} - \left. \frac{\partial J_i(k_i, k_{-i})}{\partial k_i} \right|_{(k_i^{(l-1)}, k_{-i}^{(l-1)})} \right]_0^{\overline{k}_i}, \qquad \forall l = 1, 2, \dots , \tag{21}$$

where the projection $[\cdot]_a^b := \min(\max(\cdot, a), b) \in [a, b]$ ensures that $k_i^{(l)} \in \left[0, \overline{k}_i\right]$, $\forall l = 1, 2, \dots$. If the Nash equilibrium is unique, then the gradient update converges to it (Rosen, 1965). Using the results in (Ratliff et al., 2013), it is not hard to show that each of Nash equilibria are locally stable. Hence if $\boldsymbol{k}$ is initialized close to an equilibrium, it would converge to it (we skip the details here because of length constraints).

4.1. Implementation of the Action Updating Rule

To implement the action updating rule (21), each player has to compute the partial derivative of its cost function $J_i(k_i, k_{-i})$ with respect to its action k_i at the current stage. In general, this marginal cost is not explicitly available. Here, we describe our approach to tackle this. The following result provides an explicit expression of the partial derivative in terms of the cost itself.

Proposition 1 (Estimation of marginal cost) *The marginal cost function of the ith player is*

$$\frac{\partial J_i(k_i, k_{-i})}{\partial k_i} = \frac{2J_i(k_i, k_{-i})}{1 + \rho_i k_i^2} \left(\rho_i k_i - J_i(k_i, k_{-i}) \right) , \quad \forall i \in [n] \; . \tag{22}$$

Proof We use the expressions of the partial derivatives of $J_i(k_i, k_{-i})$ and $f_i(k_i, k_{-i})$ with respect to k_i in (10) and (12), respectively, to express the marginal cost in terms of k_i and $f_i(k_i, k_{-i})$. Substituting (12) into (10) yields

$$\frac{\partial J_i(k_i, k_{-i})}{\partial k_i} = \rho_i k_i f_i(k_i, k_{-i}) - \frac{\left(1 + \rho_i k_i^2\right)}{2} f_i^2(k_i, k_{-i}) \, . \tag{23}$$

Now, by Lemma 1, we have

$$f_i(k_i, k_{-i}) = \frac{2J_i(k_i, k_{-i})}{1 + \rho_i k_i^2} \, . \tag{24}$$

Substituting (24) into (23), we get

$$\frac{\partial J_i(k_i, k_{-i})}{\partial k_i} = \rho_i k_i \frac{2J_i(k_i, k_{-i})}{1 + \rho_i k_i^2} - \frac{1 + \rho_i k_i^2}{2} \left(\frac{2J_i(k_i, k_{-i})}{1 + \rho_i k_i^2} \right)^2 = \frac{2J_i(k_i, k_{-i})}{1 + \rho_i k_i^2} \left(\rho_i k_i - J_i(k_i, k_{-i}) \right),$$

concluding the proof. ■

Proposition 1 indicates that the marginal cost for a given action k_i is computable through (22) as long as each player knows the values of its own cost $J_i(k_i, k_{-i})$ for that action at the current stage. A way to do this becomes clear when recalling the original definition of $J_i(k_i, k_{-i})$ in (4) as a selfish expected cost-to-go on state deviations and control efforts over an infinite time-horizon. It readily follows that each player can estimate $J_i(k_i, k_{-i})$ by averaging its own cost-to-go along a batch of sampled trajectories over a finite time-horizon. Specifically, $\forall l = 1, 2, \dots$, at the lth stage of the game, each player can estimate $J_i(k_i^{(l-1)}, k_{-i}^{(l-1)})$ via

$$J_i(k_i^{(l-1)}, k_{-i}^{(l-1)}) \approx \frac{1}{|\mathcal{B}|} \sum_{b=1}^{|\mathcal{B}|} \left[\int_0^{T_\mathrm{s}} \left(x_i^2(t) + \rho_i u_i^2(t) \right) \mathrm{d}t \right]^{\langle b \rangle} \Bigg|_{(k_i^{(l-1)}, k_{-i}^{(l-1)})}, \quad \forall i \in [n], \tag{25}$$

where $|\mathcal{B}|$ is the batch size, T_s is the sampling time-horizon, and, with an abuse of notation, we simply introduce a superscript $\langle b \rangle$ to denote the cost along the bth trajectory in the batch rather than accurately distinguishing $x_i(t)$ and $u_i(t)$ along different trajectories to avoid complicating the notation. As $|\mathcal{B}| \to \infty$, the error in using (25) goes to zero (law of large numbers). The rate can be bounded if we assume more information on the distribution of $\boldsymbol{x}(0)$. For example, if it has bounded moments, then the error in gradient estimate goes to zero exponentially fast (Van der Vaart, 2000).

Note that all trajectories in (25) are generated by the system (1) given that control input from each player is $u_i(t) = -k_i^{(l-1)} x_i(t)$, with the initial condition $\boldsymbol{x}(0)$ being randomly drawn from n uniform i.i.d. on $(-\sqrt{12}/2, \sqrt{12}/2)$ that have 0 as mean and 1 as variance such that the assumption $\mathbb{E}\left[\boldsymbol{x}(0)\boldsymbol{x}(0)^T\right] = \boldsymbol{I}_n$ holds. Once $J_i(k_i^{(l-1)}, k_{-i}^{(l-1)})$ has been estimated by (25), the marginal cost can be calculated through (22) as

$$\frac{\partial J_i(k_i, k_{-i})}{\partial k_i} \Bigg|_{(k_i^{(l-1)}, k_{-i}^{(l-1)})} = \frac{2 J_i(k_i^{(l-1)}, k_{-i}^{(l-1)})}{1 + \rho_i \left(k_i^{(l-1)} \right)^2} \left(\rho_i k_i^{(l-1)} - J_i(k_i^{(l-1)}, k_{-i}^{(l-1)}) \right). \tag{26}$$

Equipped with this, each player can update its action $k_i^{(l)}$ at the lth stage via (21).

Remark 4 (Decentralized action update) The implementation of the action updating rule (21) through (25) and (26) is a repeated procedure that includes two phases in each stage, where in the first phase each player collects its own trajectories to evaluate (25) and (26) for the given actions, and in the second phase all players execute (21). This procedure is decentralized in the sense that each player only needs to estimate its own marginal cost by observing its own sampled trajectories, without knowing the actions of its opponents, although the evolution of the trajectories depends on the actions taken by all players. □

4.2. Experiments of the Action Updating Rule in Noncooperative Game

With the implementation of the action updating rule (21) being made explicit, we test its performance in the noncooperative game. For a test involving 5 players, we randomly generate a symmetric strictly diagonally dominant matrix $\boldsymbol{A} \in \mathbb{R}^{5\times5}$ with negative diagonal entries as

$$\boldsymbol{A} = \begin{bmatrix} -0.0342 & -0.0111 & 0.0095 & -0.0012 & 0.0118 \\ -0.0111 & -0.0627 & 0.0098 & 0.0155 & 0.0254 \\ 0.0095 & 0.0098 & -0.0341 & -0.0065 & -0.0081 \\ -0.0012 & 0.0155 & -0.0065 & -0.0323 & -0.0081 \\ 0.0118 & 0.0254 & -0.0081 & -0.0081 & -0.1086 \end{bmatrix}.$$

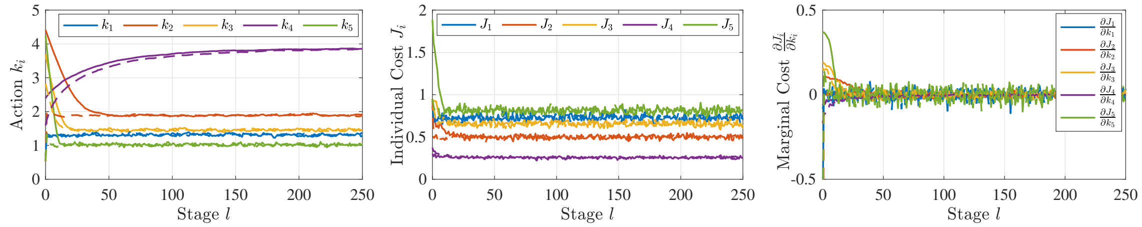

Figure 1: Evolution of individual actions (left), costs (center), and gradients (right) under the updating rule (21) with different initializations. The trajectories for different initializations are distinguished by solid and dashed lines.

Table 1: Comparison Between Two Rounds of the Game

Action		Player				
		1	2	3	4	5
$k_i^{(0)}$	Round 1	0.69	4.41	3.69	2.39	4.24
	Round 2	1.15	0.53	2.82	1.59	0.54
$k_i^{(250)}$	Round 1	1.31	1.89	1.46	3.85	1.03
	Round 2	1.29	1.88	1.49	3.85	1.03

The tradeoff coefficient ρ_i of each player is also generated randomly from $(0, 1)$ as: $\rho_1 = 0.5542$, $\rho_2 = 0.2642$, $\rho_3 = 0.4526$, $\rho_4 = 0.0664$, and $\rho_5 = 0.7990$. As the game kicks off, each player randomly initializes a positive action $k_i^{(0)}$ and updates its action $k_i^{(l)}$, $\forall l = 1, 2, \dots,$ based on (21) with batch size $|\mathcal{B}| = 500$ and sampling time-horizon $T_\mathrm{s} = 200\,\mathrm{s}$. The upper bound $\overline{k}_i$ on action is set to be sufficiently large such that it is never activated. Fig. 1 plots the evolution of individual actions, costs, and gradients in two different rounds of the aforementioned game, which shows that, although each player initializes its action differently, the same equilibrium is always reached where individual gradients all converge to zero. Table 1 confirms this observation by showing numerically that the final actions $k_i^{(250)}$ of each player are practically the same while their initial values $k_i^{(0)}$ are different in each round.

5. Conclusions and Outlook

We have formulated a fully decentralized learning problem for a symmetric linear dynamical system as a noncooperative game. We have shown the existence of pure strategy Nash equilibrium and conjectured its uniqueness under additional conditions on the state matrix. We have used projected gradient descent to have agents learn, in a fully decentralized way, the Nash equilibrium. Simulations in a 5-player game confirm our conjecture on the uniqueness of Nash equilibrium. Future work will explore the uniqueness of Nash equilibria in the general case, the analysis of the noncooperative game when the distribution of the initial states is not white, the formal characterization of the robustness of the proposed action updating rule, and the extension of the results to time-varying topologies.

Acknowledgments

This material is based in part upon work supported by the State of Washington through the University of Washington Clean Energy Institute. Yan Jiang, Wenqi Cui, and Baosen Zhang were partially supported by NSF Award ECCS-2153937. Jorge Cortés was partially supported by NSF Award IIS-2007141.

References

Bassam Bamieh, Fernando Paganini, and Munther A. Dahleh. Distributed control of spatially invariant systems. *IEEE Transactions on Automatic Control*, 47(7):1091–1107, July 2002.

Tamer Başar and Geert Jan Olsder. *Dynamic Noncooperative Game Theory*. SIAM, 1998.

Ana L.C. Bazzan. Opportunities for multiagent systems and multiagent reinforcement learning in traffic control. *Autonomous Agents and Multi-Agent Systems*, 18(3):342–375, 2009.

Vincent D. Blondel and John N. Tsitsiklis. A survey of computational complexity results in systems and control. *Automatica*, 36(9):1249–1274, Sept. 2000.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

Jingjing Bu, Afshin Mesbahi, Maryam Fazel, and Mehran Mesbahi. LQR through the lens of first order methods: Discrete-time case. *arXiv preprint:1907.08921*, 2019.

Wenqi Cui and Baosen Zhang. Equilibrium-independent stability analysis for distribution systems with lossy transmission lines. *IEEE Control Systems Letters*, 6:3349–3354, 2022.

Wenqi Cui, Jiayi Li, and Baosen Zhang. Decentralized safe reinforcement learning for inverter-based voltage control. *Electric Power Systems Research*, 211:108609, Oct. 2022.

Sarah Dean, Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu. On the sample complexity of the linear quadratic regulator. *Foundations of Computational Mathematics*, 20(4): 633–679, 2020.

Jacob Engwerda. *LQ Dynamic Optimization and Differential Games*. John Wiley & Sons, 2005.

Maryam Fazel, Rong Ge, Sham Kakade, and Mehran Mesbahi. Global convergence of policy gradient methods for the linear quadratic regulator. In *Proc. of International Conference on Machine Learning*, pages 1467–1476, 2018.

Han Feng and Javad Lavaei. On the exponential number of connected components for the feasible set of optimal decentralized control problems. In *Proc. of American Control Conference*, pages 1430–1437, July 2019.

Zhu Han, Dusit Niyato, Walid Saad, Tamer Başar, and Are Hjørungnes. *Game Theory in Wireless and Communication Networks: Theory, Models, and Applications*. Cambridge university press, 2012.

Joao P. Hespanha. *Linear Systems Theory*. Princeton university press, 2018.

Tong Huang, Hongbo Sun, Kyeong Jin Kim, Daniel Nikovski, and Le Xie. A holistic framework for parameter coordination of interconnected microgrids against disasters. In *Proc. of IEEE Power & Energy Society General Meeting*, pages 1–5, Aug. 2020.

Wook Hyun Kwon and Soo Hee Han. *Receding Horizon Control: Model Predictive Control for State Models*. Springer Science & Business Media, 2006.

Laurent Lessard and Sanjay Lall. An algebraic approach to the control of decentralized systems. *IEEE Transactions on Control of Network Systems*, 1(4):308–317, Dec. 2014.

Tao Li, Guanze Peng, Quanyan Zhu, and Tamer Başar. The confluence of networks, games, and learning a game-theoretic framework for multiagent decision making over networks. *IEEE Control Systems Magazine*, 42(4):35–67, Aug. 2022.

Jason R. Marden and Jeff S. Shamma. Game theory and distributed control. In *Handbook of game theory with economic applications*, volume 4, pages 861–899. 2015.

Eric Mazumdar, Lillian J. Ratliff, Michael I. Jordan, and S. Shankar Sastry. Policy-gradient algorithms have no guarantees of convergence in linear quadratic games. In *Proc. of the International Conference on Autonomous Agents and MultiAgent Systems*, pages 860–868, May 2020.

J. Mendel and D. Gieseking. Bibliography on the linear-quadratic-gaussian problem. *IEEE Transactions on Automatic Control*, 16(6):847–869, Dec. 1971.

Reza Olfati-Saber, J. Alex Fax, and Richard M. Murray. Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, 95(1):215–233, Jan. 2007.

Kaare Brandt Petersen and Michael Syskind Pedersen. *The Matrix Cookbook*. Technical University of Denmark, 2012.

Anders Rantzer. Scalable control of positive systems. *European Journal of Control*, 24:72–80, July 2015.

Lillian J. Ratliff, Samuel A. Burden, and S. Shankar Sastry. Characterization and computation of local nash equilibria in continuous games. In *Proc. of Annual Allerton Conference on Communication, Control, and Computing*, pages 917–924, Oct 2013.

J. B. Rosen. Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica*, 33(3):520–534, July 1965.

Michael Rotkowitz and Sanjay Lall. Decentralized control information structures preserved under feedback. In *Proc. of IEEE Conference on Decision and Control*, pages 569–575, Dec. 2002.

Walter Rudin. *Principles of Mathematical Analysis*. McGraw Hill, 3rd edition, 2013.

Parikshit Shah and Pablo A. Parrilo. H_2-optimal decentralized control over posets: A state-space solution for state-feedback. *IEEE Transactions on Automatic Control*, 58(12):3084–3096, Dec. 2013.

John Nikolas Tsitsiklis. Problems in decentralized decision making and computation. Technical report, Massachusetts Inst of Tech Cambridge Lab for Information and Decision Systems, 1984.

Aad W Van der Vaart. *Asymptotic Statistics*, volume 3. Cambridge university press, 2000.

Erik Weitenberg, Yan Jiang, Changhong Zhao, Enrique Mallada, Claudio De Persis, and Florian Dörfler. Robust decentralized secondary frequency control in power systems: Merits and tradeoffs. *IEEE Transactions on Automatic Control*, 64(10):3967–3982, Oct. 2019.

Yaodong Yang, Jianye Hao, Mingyang Sun, Zan Wang, Changjie Fan, and Goran Strbac. Recurrent deep multiagent q-learning for autonomous brokers in smart grid. In *Proc. of the International Joint Conference on Artificial Intelligence*, pages 569–575, 2018.

Kaiqing Zhang, Zhuoran Yang, and Tamer Basar. Policy optimization provably converges to nash equilibria in zero-sum linear quadratic games. In *Proc. of the International Conference on Neural Information Processing Systems*, pages 11602–11614, Dec. 2019.

Yun Zhang and Le Xie. A transient stability assessment framework in power electronic-interfaced distribution systems. *IEEE Transactions on Power Systems*, 31(6):5106–5114, Nov. 2016.