Nash equilibrium seeking under partial-decision information over directed communication networks

Bianchi, Mattia; Grammatico, Sergio

Citation (APA)
Bianchi, M., & Grammatico, S. (2020). Nash equilibrium seeking under partial-decision information over directed communication networks. In *Proceedings of the 59th IEEE Conference on Decision and Control, CDC 2020* (pp. 3555-3560). IEEE. https://doi.org/10.1109/CDC42340.2020.9304267

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Nash equilibrium seeking under partial-decision information over directed communication networks

Mattia Bianchi and Sergio Grammatico

Abstract— We consider the Nash equilibrium problem in a partial-decision information scenario. Specifically, each agent can only receive information from some neighbors via a communication network, while its cost function depends on the strategies of possibly all agents. In particular, while the existing methods assume undirected or balanced communication, in this paper we allow for non-balanced, directed graphs. We propose a fully-distributed pseudo-gradient scheme, which is guaranteed to converge with linear rate to a Nash equilibrium, under strong monotonicity and Lipschitz continuity of the game mapping. Our algorithm requires global knowledge of the communication structure, namely of the Perron-Frobenius eigenvector of the adjacency matrix and of a certain constant related to the graph connectivity. Therefore, we adapt the procedure to setups where the network is not known in advance, by computing the eigenvector online and by means of vanishing step sizes.

I. INTRODUCTION

Game theory is a powerful tool to model and control the decision-making process of selfish agents, that aim at optimizing their individual, but inter-dependent, objective functions. This scenario arises in several relevant engineering applications, such as congestion control in traffic networks [1], smart-grid management [2], demand response in competitive markets [3] and analysis of social dynamics [4]. Often, the goal (either of the agents or of a coordinator that pursues network regulation by imposing incentives or behavioral rules) is the attainment of a Nash equilibrium (NE), a joint strategy from which it is not convenient for any agent to unilaterally deviate.

In fact, a recent part of the literature focuses on designing distributed NE seeking algorithms, where the computational effort is partitioned among the agents [5]–[7]. Nonetheless, typically these methods still assume the presence of a central coordinator that can broadcast some data – for instance, the average of all the agents' strategies, in the case of aggregative games [7]. Unfortunately, this requirement is impractical in some domains [8]. To overcome this limitation, we consider *fully-distributed* schemes, where the agents only rely on the information locally exchanged over a network, via peer-to-peer communication. In particular, the main challenge is that the cost function of each agent may depend on the strategies of some other non-neighboring agents. One example is the Cournot competition model described in [9], where the profit of each of a group of firms depends not only on its own production, but also on the total supply,

a quantity not directly accessible by any of the firms. To remedy the lack of knowledge, each agent can estimate and eventually reconstruct the strategies of all the competitors (or an aggregation value), based on the data received from its neighbors.

Such a *partial-decision information* setup has only been introduced very recently. Most of the available results resort to (projected) pseudo-gradient and consensus dynamics [9]–[14]. Alternatively, schemes based on a proximal-point iteration were studied in [15]; a fully-distributed fictitious play algorithm was proposed in [8]. These approaches assume undirected communication, which might be unrealistic, e.g., in wireless systems, if the agents send signals at different power levels, implying unilateral transmission capability. Fewer works deal with asymmetric networks. Under the assumption of balanced weights, continuous-time dynamics were proposed in [16] for aggregative games; most recently, we also addressed generally-coupled-cost games via a fixed-step forward-backward method [17]. To the best of our knowledge, the only discrete-time NE seeking algorithm that takes into account non-balanced digraphs is the asynchronous gossip-based scheme in [18].

Even in the context of distributed optimization, most algorithms are designed with doubly stochastic adjacency matrices, which enjoy several convenient properties, not least that the average of the agents' estimates is preserved over time. However, doubly stochastic weights cannot be easily assigned over directed networks. An alternative is to rely on column stochastic graphs, which maintain the average invariance and only require the agents to know their out-degree. Yet, this is impractical in setups where the agents broadcast some information, but ignoring which of the other nodes can receive it; or if some of the communication links can fail. In contrast, distributed design of row stochastic matrices is straightforward, as it suffices for each agent to locally assign appropriate weights to the incoming information. However, the use of row stochastic graphs comes with technical challenges, since many properties of doubly stochastic matrices are lost. Of major interest for this work is the approach in [19]: to correct the imbalance caused by employing row stochastic weights, the algorithm exploits the information contained in the Perron-Frobenius (PF) eigenvector of the adjacency matrix, which is computed online.

Contribution: Motivated by the above, we design the first synchronous, fully-distributed algorithm to compute a NE over directed non-balanced communication networks. Our contributions are summarized as follows:

- We prove that any row stochastic primitive matrix with

positive diagonal enjoys a contractivity property, in a Hilbert space weighted by its PF eigenvector. We later exploit this general result to prove convergence of our equilibrium seeking dynamics (§II);

- We design a fully-distributed, fixed-step gradient algorithm to seek a NE over strongly connected directed graphs, which is guaranteed to converge with liner rate under strong monotonicity of the game mapping. In our method, the pseudo-gradient component is divided by the entries of the PF eigenvector of the network. Although this technique has already been adopted in distributed optimization [19], we give a new, powerful, monotone-operator-theoretic interpretation, which greatly simplifies our analysis (§III-A);

- We show that convergence is retained even if the graph is not known in advance and the PF eigenvector is computed online, provided that a small-enough step size is chosen. Since computing the upper bound distributedly can be troublesome, we also provided convergence guarantees for vanishing steps (§III-B).

Basic notation: $\mathbb{R}_{>0}$ denotes the set of positive real numbers. $\mathbf{0}_n$ $(\mathbf{1}_n) \in \mathbb{R}^n$ denotes the vector with all elements equal to 0 (1); $I_n \in \mathbb{R}^{n \times n}$ denotes an identity matrix; we may omit the subscripts if there is no ambiguity. $e_i^n \in \mathbb{R}^n$ denotes a vector with all elements equal to 0 except the i-th element, which is 1. For a function $g : \mathbb{R}^n \to \mathbb{R}$, $\nabla_x g(x)$ denotes its gradient. For a matrix $A \in \mathbb{R}^{m \times n}$, $[A]_{i,j}$ represents the element on row i and column j; $\sigma_{\min}(A) = \sigma_1(A) \leq \cdots \leq \sigma_n(A) =: \sigma_{\max}(A)$ are its singular values. If $A \in \mathbb{R}^{n \times n}$ is symmetric, $\lambda_{\min}(A) = \lambda_1(A) \leq \cdots \leq \lambda_n(A) =: \lambda_{\max}(A)$ denote its eigenvalues; $A \succ 0$ stands for positive definite matrix. $\otimes$ denotes the Kronecker product. $\mathrm{diag}(A_1, \ldots, A_N)$ denotes the block diagonal matrix with $A_1, \ldots, A_N$ on its diagonal. Given $P \succ 0$, $\langle x, y \rangle_P = x^\top P y$ and $\|x\|_P = \sqrt{x^\top P x}$ denote the P-weighted Euclidean inner product and norm, respectively; $\|A\|_P := \sup_{x \neq \mathbf{0}} \frac{\|Ax\|_P}{\|x\|_P}$ is the P-induced norm of $A \in \mathbb{R}^{n \times n}$; we omit the subscripts if $P = I$. $\mathcal{H}_P := (\mathbb{R}^n, \langle \cdot, \cdot \rangle_P)$ is the Hilbert space obtained by endowing $\mathbb{R}^n$ with the P-weighted inner product.

Operator-theoretic notation: An operator $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^n$ is (μ-strongly) monotone in $\mathcal{H}_P$ if $\langle \mathcal{A}(x) - \mathcal{A}(y), x - y \rangle_P \geq 0$ ($\geq \mu\|x - y\|_P^2$), for all $x, y \in \mathbb{R}^n$. $\mathcal{A}$ is ℓ-Lipschitz continuous in $\mathcal{H}_P$ if $\|\mathcal{A}(x) - \mathcal{A}(y)\|_P \leq \ell\|x - y\|_P$, for all $x, y \in \mathbb{R}^n$; if $\ell \leq 1$ ($\ell < 1$), $\mathcal{A}$ is nonexpansive (contractive) in $\mathcal{H}_P$. We omit the indication "in $\mathcal{H}_P$" if $P = I$. $\mathrm{proj}_S^P : \mathbb{R}^n \to S$ is the Euclidean P-weighted projection onto a closed convex set $S \subseteq \mathbb{R}^n$, i.e. $\mathrm{proj}_S^P(x) := \mathrm{argmin}_{y \in S} \|y - x\|_P$; we omit the superscript if $P = I$.

II. MATHEMATICAL SETUP

A. The game

We consider a set of agents, $\mathcal{I} := \{1, \ldots, N\}$, where each agent $i \in \mathcal{I}$ shall choose its decision variable (i.e., strategy) x_i from its local decision set $\Omega_i \subseteq \mathbb{R}^{n_i}$. Let $x := \mathrm{col}((x_i)_{i \in \mathcal{I}}) \in \Omega$ denote the stacked vector of all the agents' decisions, with $\Omega := \Omega_1 \times \cdots \times \Omega_N \subseteq \mathbb{R}^n$ the overall action space and $n := \sum_{i \in \mathcal{I}} n_i$. The goal of agent $i \in \mathcal{I}$ is to minimize its objective function $J_i(x_i, x_{-i})$, which depends both on the local variable x_i and on the decision variables of the other agents $x_{-i} := \mathrm{col}((x_j)_{j \in \mathcal{I} \setminus \{i\}})$. The game is then represented by the inter-dependent optimization problems

$$\forall i \in \mathcal{I}: \quad \underset{y_i \in \Omega_i}{\mathrm{argmin}}\, J_i(y_i, x_{-i}). \tag{1}$$

The technical problem we consider here is the distributed computation of a NE, as formalized next.

Definition 1: A collective strategy $x^* = \mathrm{col}((x_i^*)_{i \in \mathcal{I}})$ is a Nash equilibrium if, for all $i \in \mathcal{I}$,

$$J_i\left(x_i^*, x_{-i}^*\right) \leq \inf\{J_i\left(y_i, x_{-i}^*\right) \mid (y_i, x_{-i}^*) \in \Omega\}. \quad \square$$

Next, we postulate common regularity assumptions for the constraint sets and cost functions [14, Ass. 1], [10, Ass. 1].

Standing Assumption 1: For each $i \in \mathcal{I}$, the set Ω_i is non-empty, closed and convex; J_i is continuous and $J_i(\cdot, x_{-i})$ is convex and continuously differentiable for every x_{-i}. $\square$

Under Standing Assumption 1, a collective strategy x^* is a NE of the game in (1) if and only if it is a solution of the variational inequality $\mathrm{VI}(F, \Omega)^1$ [20, Prop. 1.4.2], where F is the *pseudo-gradient* mapping of the game:

$$F(x) := \mathrm{col}\left((\nabla_{x_i} J_i(x_i, x_{-i}))_{i \in \mathcal{I}}\right). \tag{2}$$

Equivalently, x^* is a NE if and only if

$$\forall i \in \mathcal{I}: \quad x_i^* = \mathrm{proj}_{\Omega_i}(x_i^* - \beta_i \nabla_{x_i} J_i(x_i^*, x_{-i}^*)), \tag{3}$$

for arbitrary positive scalars β_i's [20, Prop. 1.5.8]. A sufficient condition for the existence and uniqueness of a NE is the strong monotonicity of the pseudo-gradient [20, Th. 2.3.3], as postulated next. This assumption has always been used for NE seeking under partial-decision information with fixed step sizes, e.g., [14, Ass. 2], [11, Ass. 4], [10, Ass. 2].

Standing Assumption 2: The pseudo-gradient mapping F in (2) is μ-strongly monotone and ℓ_0-Lipschitz continuous, for some $\mu, \ell_0 > 0$. $\square$

B. Network communication

The agents can exchange information with some neighbors over a directed communication network $\mathcal{G}(\mathcal{I}, \mathcal{E})$. The ordered pair (i, j) belongs to the set of edges, $\mathcal{E}$, if and only if agent i can receive information from agent j. We denote $W \in \mathbb{R}^{N \times N}$ the weighted adjacency matrix of $\mathcal{G}$ and $w_{i,j} := [W]_{i,j}$, with $w_{i,j} > 0$ if $(i, j) \in \mathcal{E}$, $w_{i,j} = 0$ otherwise; $d_i = \sum_{j=1}^N w_{i,j}$ and $\mathcal{N}_i = \{j \mid (i, j) \in \mathcal{E}_k\}$ the in-degree and the set of in-neighbors of agent i, respectively.

Standing Assumption 3: The communication graph $\mathcal{G}$ is strongly connected. $\square$

Standing Assumption 4: The adjacency matrix W satisfies the following conditions:

(i) *Self-loops:* $w_{i,i} > 0$ for all $i \in \mathcal{I}$;
(ii) *Row stochasticity:* $W_k \mathbf{1}_N = \mathbf{1}_N$. $\square$

Remark 1: Standing Assumption 4 can be fulfilled on any digraph, if the agents can access their own in-degree, by *locally* assigning weights to the received information. $\square$

---

[1] Given a set $S \subseteq \mathbb{R}^m$ and a mapping $\psi : S \to \mathbb{R}^m$, the $\mathrm{VI}(\psi, S)$ is the problem of finding $\omega^* \in S$ such that $\langle \psi(\omega^*), \omega - \omega^* \rangle \geq 0$, for all $\omega \in S$.

Under Standing Assumptions 3-4, by the PF theorem, W has a simple eigenvalue in 1; all the other (complex) eigenvalues of W have absolute value strictly smaller than 1. Besides, there exist a vector $q = \mathrm{col}((q_i)_{i \in \mathcal{I}})$ such that

$$q \in \mathbb{R}^N_{>0}, \quad q^\top W = q^\top, \quad \mathbf{1}_N^\top q = 1. \tag{4}$$

We call q the (left) *Perron-Frobenius eigenvector* of W. Let

$$Q := \mathrm{diag}((q_i)_{i \in \mathcal{I}}). \tag{5}$$

Clearly, $Q \succ 0$. Unless W is doubly stochastic, W is not nonexpansive in $\mathcal{H}_I$, i.e., $\sigma_{\max}(W) > 1$. This is one of the main technical challenges to face when studying fixed-point iterations over directed graphs [18]. To deal with this complication, it was shown in [21, Lemma 1] that W is nonexpansive (averaged, indeed) in $\mathcal{H}_Q$. Next, we provide an additional contractivity result, which we exploit later on.

Lemma 1: For any $y \in \mathbb{R}^N$, $\|W(y - \mathbf{1}_N q^\top y)\|_Q \leq \bar{\sigma}\|y - \mathbf{1}_N q^\top y\|_Q$, where $\bar{\sigma} := \sigma_{N-1}(Q^{\frac{1}{2}} W Q^{-\frac{1}{2}}) < 1$. $\square$

If W is also column stochastic, Lemma 1 holds with $q = \frac{1}{N}\mathbf{1}_N$ and $Q = I_N$, and we recover a well-known property of doubly stochastic matrices [17, Eq. 4].

Remark 2: [19, Lemma 1] states that there exist a norm and $\bar{\sigma} > 0$ such that the property in Lemma 1 holds; instead, we explicitly characterized both the norm and $\bar{\sigma}$, which proves very advantageous in our analysis, see §III. $\square$

C. Partial-decision information scenario

In our setup, agent $i \in \mathcal{I}$ can only access its own feasible set Ω_i and an analytic expression of its own cost function J_i. However, the agents cannot evaluate the actual value of the cost $J_i(x_i, x_{-i})$ (or the partial derivative $\nabla_{x_i} J_i(x_i, x_{-i})$), since they cannot access the strategies of all the competitors x_{-i}. Instead, the agents only rely on the information exchanged locally with their neighbors over the communication graph $\mathcal{G}$. To cope with the lack of knowledge, the general assumption for this partial-decision information scenario is that each agent keeps an estimate of all other agents' actions [14], [9], [11]. Then, the agents aim at reconstructing the actual values, based on the data received from their neighbors. We denote $\boldsymbol{x}_i = \mathrm{col}((\boldsymbol{x}_{i,j})_{j \in \mathcal{I}}) \in \mathbb{R}^n$, where $\boldsymbol{x}_{i,i} := x_i$ and $\boldsymbol{x}_{i,j}$ is agent i's estimate of agent j's action, for all $j \neq i$; $\boldsymbol{x}_{j,-i} = \mathrm{col}((\boldsymbol{x}_{j,l})_{l \in \mathcal{I} \setminus \{i\}})$; $\boldsymbol{x} = \mathrm{col}((\boldsymbol{x}_i)_{i \in \mathcal{I}})$. As in [14, Eq.13-14], we define

$$\mathcal{R}_i := \left[\begin{array}{ccc} 0_{n_i \times n_{<i}} & I_{n_i} & \mathbf{0}_{n_i \times n_{>i}} \end{array} \right], \tag{6}$$

where $n_{<i} := \sum_{j<i, j \in \mathcal{I}} n_j$, $n_{>i} := \sum_{j>i, j \in \mathcal{I}} n_j$. In simple terms, $\mathcal{R}_i$ selects the i-th n_i-dimensional component from an n-dimensional vector, i.e., $\mathcal{R}_i \boldsymbol{x}_i = \boldsymbol{x}_{i,i} = x_i$. We denote by $\mathcal{R} := \mathrm{diag}((\mathcal{R}_i)_{i \in \mathcal{I}})$; thus, we have $x = \mathcal{R}\boldsymbol{x}$. Moreover, we define the *extended pseudo-gradient* mapping $\boldsymbol{F}$ as

$$\boldsymbol{F}(\boldsymbol{x}) := \mathrm{col}\left((\nabla_{x_i} J_i(x_i, \boldsymbol{x}_{i,-i}))_{i \in \mathcal{I}}\right). \tag{7}$$

Lemma 2 ([22, Lemma 3]): The mapping $\boldsymbol{F}$ in (7) is ℓ-Lipschitz continuous, for some $\ell \in [\mu, \ell_0]$: for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{Nn}$, $\|\boldsymbol{F}(\boldsymbol{x}) - \boldsymbol{F}(\boldsymbol{y})\| \leq \ell\|\boldsymbol{x} - \boldsymbol{y}\|$. $\square$

We remark that in (7), each agent i evaluates its partial gradients $\nabla_{x_i} J_i(x_i, \boldsymbol{x}_{i,-i})$ on the local estimate $\boldsymbol{x}_{i,-i}$, not

on the actual strategies x_{-i}. Only when the estimates of all the agents coincide with the actual value, i.e., $\boldsymbol{x} = \mathbf{1}_N \otimes x$, we have that $\boldsymbol{F}(\boldsymbol{x}) = F(x)$. As a consequence, the mapping $\mathcal{R}^\top \boldsymbol{F}$ is not monotone, not even under strong monotonicity of the game mapping F in Standing Assumption 2. Indeed, the loss of monotonicity is the main technical difficulty arising in the partial-decision information scenario [14], [10].

III. FULLY-DISTRIBUTED NASH EQUILIBRIUM SEEKING

In this section, we present a pseudo-gradient method (along with some variants) to seek a NE in a fully-distributed way. Before going into details, we need some definitions. Let

$$\bar{Q} := \mathrm{diag}((q_i I_{n_i})_{i \in \mathcal{I}}), \quad \boldsymbol{Q} := Q \otimes I_n. \tag{8}$$

We define the consensus subspace as $\boldsymbol{E} = \{\boldsymbol{y} \in \mathbb{R}^{Nn} | \boldsymbol{y} = \mathbf{1}_N \otimes y, y \in \mathbb{R}^n\}$ and its orthogonal complement in $\mathcal{H}_{\boldsymbol{Q}}$ as $\boldsymbol{E}_\perp^{\boldsymbol{Q}} = \{\boldsymbol{y} \in \mathbb{R}^{Nn} | (q \otimes I_n)^\top \boldsymbol{y} = \mathbf{0}_n\}$. Thus, any vector of estimates $\boldsymbol{x} \in \mathbb{R}^{Nn}$ can be written as $\boldsymbol{x} = \boldsymbol{x}_\| + \boldsymbol{x}_\perp$, where $\boldsymbol{x}_\| = \mathrm{proj}_{\boldsymbol{E}}^{\boldsymbol{Q}}(\boldsymbol{x}) = (\mathbf{1}_N q^\top \otimes I_n)\boldsymbol{x}$, $\boldsymbol{x}_\perp = \mathrm{proj}_{\boldsymbol{E}_\perp^{\boldsymbol{Q}}}^{\boldsymbol{Q}}(\boldsymbol{x})$, and it holds that $\langle \boldsymbol{x}_\|, \boldsymbol{x}_\perp \rangle_{\boldsymbol{Q}} = 0$. Clearly, if the estimates of the agents $\boldsymbol{x} \in \boldsymbol{E}$, then $\boldsymbol{x}_i = x$ for all $i \in \mathcal{I}$, namely the estimate of each agent coincides with the actual collective strategy x.

A. Case 1: Known q and $\bar{\sigma}$

Our basic fully-distributed NE seeking algorithm is summarized in Algorithm 1, where α is a fixed step size. Each agent update its estimates according to consensus dynamics, then its strategy via a projected pseudo-gradient step. We remark that each agent computes the partial gradient of its cost in its local estimate, not on the actual joint strategy x.

Compared to similar pseudo-gradient dynamics proposed in the literature [10], [17], the novelty of Algorithm 1 is that the cost related components $\nabla_{x_i} J_i$ are weighted by the reciprocal of the elements q_i of the PF eigenvector. This operation enables convergence on row stochastic graphs, and in fact it is not necessary for doubly stochastic graphs, for which $q = \mathbf{1}$. The idea behind this key modification is that $(W - \mathbf{1}q^\top)$ is contractive in $\mathcal{H}_Q$, while the game-mapping F is strongly monotone in $\mathcal{H}_I$; instead, we would like both properties to hold in the same space. Division by the PF eigenvector achieves this goal, as we show next. Let

$$\bar{F}(x) =: \bar{Q}^{-1} F(x), \quad \bar{\boldsymbol{F}}(\boldsymbol{x}) := \bar{Q}^{-1} \boldsymbol{F}(\boldsymbol{x}). \tag{9}$$

Lemma 3: $\bar{F}$ is $\bar{\mu}$-strongly monotone and $\bar{\ell}_0$-Lipschitz continuous in $\mathcal{H}_{\bar{Q}}$, for some $\bar{\mu}, \bar{\ell}_0 > 0$; $\bar{\boldsymbol{F}}$ is $\bar{\ell}$-Lipschitz continuous from $\mathcal{H}_{\boldsymbol{Q}}$ to $\mathcal{H}_{\bar{Q}}$, for some $\bar{\ell} > 0$, i.e., for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{Nn}$, $\|\bar{\boldsymbol{F}}(\boldsymbol{x}) - \bar{\boldsymbol{F}}(\boldsymbol{y})\|_{\bar{Q}} \leq \bar{\ell}\|\boldsymbol{x} - \boldsymbol{y}\|_{\boldsymbol{Q}}$. $\square$

Remark 3: Lemmas 1 and 3 provide a general, operator-theoretic interpretation of the approach in [19], where a similar technique is used in the context of distributed optimization. $\square$

In compact form, Algorithm 1 reads as

$$\boldsymbol{x}^{k+1} = \mathrm{proj}_{\boldsymbol{\Omega}}(\mathcal{F}(\boldsymbol{x}^k)), \tag{10}$$

where $\boldsymbol{\Omega} := \{\boldsymbol{x} \in \mathbb{R}^{Nn} \mid \mathcal{R}\boldsymbol{x} \in \Omega\}$, $\boldsymbol{W} := W \otimes I_n$ and

$$\mathcal{F}(\boldsymbol{x}) := \boldsymbol{W}\boldsymbol{x} - \alpha \mathcal{R}^\top \bar{\boldsymbol{F}}(\boldsymbol{W}\boldsymbol{x}). \tag{11}$$

3557

Algorithm 1

Initialization: $\forall i \in \mathcal{I}$, set $x_i^0 \in \Omega_i$, $\boldsymbol{x}_{i,-i}^0 \in \mathbb{R}^{N-n_i}$.

Iterate until convergence: each agent $i \in \mathcal{I}$ does:

$$\hat{\boldsymbol{x}}_i^k = \sum_{j \in \mathcal{N}_i} w_{i,j}^k \boldsymbol{x}_j^k$$

$$x_i^{k+1} = \text{proj}_{\Omega_i}(\hat{\boldsymbol{x}}_{i,i}^k - \frac{\alpha}{q_i} \nabla_{x_i} J_i(\hat{\boldsymbol{x}}_i^k))$$

$$\boldsymbol{x}_{i,-i}^{k+1} = \hat{\boldsymbol{x}}_{i,-i}^k.$$

The following Lemma shows a contractivity property of the operator $\mathcal{F}$ and represents the cornerstone we use to prove convergence of our NE seeking schemes. The result is based on the strong monotonicity of $\bar{F}$ in $\mathcal{H}_{\bar{Q}}$ and on Lemma 1.

Lemma 4: Let

$$M_\alpha := \begin{bmatrix} 1-2\alpha\bar{\mu}\lambda_{\min}(Q)+\alpha^2\bar{\ell}^2 & (2\alpha\bar{\ell})\bar{\sigma} \\ (2\alpha\bar{\ell})\bar{\sigma} & (1+2\alpha\bar{\ell}+\alpha^2\bar{\ell}^2)\bar{\sigma}^2 \end{bmatrix} \quad (12)$$

If the step size $\alpha > 0$ is chosen such that

$$\rho_\alpha := \lambda_{\max}(M_\alpha) = \|M_\alpha\| < 1, \quad (13)$$

then the operator $\mathcal{F}$ in (11) is $\sqrt{\rho_\alpha}$-restricted contractive in $\mathcal{H}_{\boldsymbol{Q}}$ with respect to the consensus subspace $\boldsymbol{E}$, i.e., for any $\boldsymbol{x} \in \mathbb{R}^{Nn}$, $\boldsymbol{y} \in \boldsymbol{E}$, it holds that $\|\mathcal{F}(\boldsymbol{x}) - \mathcal{F}(\boldsymbol{y})\|_{\boldsymbol{Q}} \leq \sqrt{\rho_\alpha}\|\boldsymbol{x} - \boldsymbol{y}\|_{\boldsymbol{Q}}$. $\qquad\square$

Remark 4: The condition in (13) can always be satisfied by choosing α small enough; an explicit upper bound can be obtained as in [17, Lemma 2]. $\qquad\square$

Theorem 1: Let $\alpha > 0$ satisfy the condition in (13). Then, for any initial condition, the sequence $(\boldsymbol{x}^k)_{k \in \mathbb{N}}$ generated by Algorithm 1 converges to $\boldsymbol{x}^* = \boldsymbol{1}_N \otimes x^*$, where x^* is the NE of the game in (1), with linear rate: for all $k \in \mathbb{N}$,

$$\|\boldsymbol{x}^k - \boldsymbol{x}^*\|_{\boldsymbol{Q}} \leq (\sqrt{\rho_\alpha})^k \|\boldsymbol{x}^0 - \boldsymbol{x}^*\|_{\boldsymbol{Q}}. \qquad\square$$

Proof: By (3), we infer that x^* is the NE if and only if $x^* = \text{proj}_\Omega(x^* - \alpha Q^{-1}F(x^*))$. Together with $\boldsymbol{W}\boldsymbol{x}^* = \boldsymbol{x}^*$ and $\boldsymbol{F}(\boldsymbol{x}^*) = F(x^*)$, this implies that $\boldsymbol{x}^*$ is a fixed point for the iteration in (10). Therefore we can write

$$\begin{aligned}
\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\|_{\boldsymbol{Q}} &= \|\text{proj}_{\boldsymbol{\Omega}}(\mathcal{F}(\boldsymbol{x}^k)) - \text{proj}_{\boldsymbol{\Omega}}(\mathcal{F}(\boldsymbol{x}^*))\|_{\boldsymbol{Q}} \\
&= \|\text{proj}_{\boldsymbol{\Omega}}^{\boldsymbol{Q}}(\mathcal{F}(\boldsymbol{x}^k)) - \text{proj}_{\boldsymbol{\Omega}}^{\boldsymbol{Q}}(\mathcal{F}(\boldsymbol{x}^*))\|_{\boldsymbol{Q}} \\
&\leq \|\mathcal{F}(\boldsymbol{x}^k) - \mathcal{F}(\boldsymbol{x}^*)\|_{\boldsymbol{Q}} \leq \sqrt{\rho_\alpha}\|\boldsymbol{x}^k - \boldsymbol{x}^*\|_{\boldsymbol{Q}},
\end{aligned}$$

where the second equality follows by $\boldsymbol{Q} = Q \otimes I_n$ and the definition of $\boldsymbol{\Omega}$ (note that $\text{proj}_{\Omega_i}^{q_i I_{n_i}} = \text{proj}_{\Omega_i}$), the first inequality follows by nonexpansiveness of the projection [23, Prop. 4.16], and the second inequality by Lemma 4. ∎

We note that Algorithm 1 requires a priori knowledge of the communication graph $\mathcal{G}$, both to compute the PF eigenvector q and to tune the step size α. In the next subsection, we relax this hypothesis.

B. Case 2: Online computation of q

When the PF eigenvalue q is not known in advance, it can be computed online in a distributed fashion. The procedure is illustrated in Algorithm 2. Each agent $i \in \mathcal{I}$ keeps an extra variable $\hat{q}_i = \text{col}((\hat{q}_{i,j})_{j \in \mathcal{I}})$, which is an estimate of q, initialized as the i-th vector of the canonical basis $e_i^N \in \mathbb{R}^N$.

Algorithm 2

Initialization: $\forall i \in \mathcal{I}$, set $x_i^0 \in \Omega_i$, $\boldsymbol{x}_{i,-i}^0 \in \mathbb{R}^{N-n_i}$, $\hat{q}_i^0 = e_i^N$.

Iterate until convergence: each agent $i \in \mathcal{I}$ does:

$$\hat{q}_i^{k+1} = \sum_{j \in \mathcal{N}_i} w_{i,j}^k \hat{q}_j^k$$

$$x_i^{k+1} = \text{proj}_{\Omega_i}(\hat{\boldsymbol{x}}_{i,i}^k - \alpha^k (\hat{q}_{i,i}^k)^{-1} \nabla_{x_i} J_i(\hat{\boldsymbol{x}}_i^k))$$

$$\boldsymbol{x}_{i,-i}^{k+1} = \hat{\boldsymbol{x}}_{i,-i}^k \qquad \hat{\boldsymbol{x}}_i^k = \sum_{j \in \mathcal{N}_i} w_{i,j}^k \boldsymbol{x}_j^k.$$

Notably, each estimate $\hat{q}_i$ converges to the real value q. In fact, the updates in Algorithm 2 can be written compactly as

$$\hat{\boldsymbol{q}}^{k+1} = (W \otimes I_N)\hat{\boldsymbol{q}}^k, \quad (14)$$

where $\hat{\boldsymbol{q}} := \text{col}((\hat{q}_i)_{i \in \mathcal{I}})$. Therefore, by the PF theorem (and by Standing Assumptions 3-4), $\hat{\boldsymbol{q}}^k$ converges linearly to $(\boldsymbol{1}_N q^\top \otimes I_N)\hat{\boldsymbol{q}}^0 = \boldsymbol{1}_N \otimes q$. In particular, $\hat{q}_{i,i}^k \to q_i$. Also, $\hat{q}_{i,i}^k > 0$ for all $k \geq 0$, since $\hat{q}_{i,i}^0 > 0$ and W is nonnegative with positive diagonal. As such, Algorithm 2 is always well defined. We first show its convergence for a fixed step size.

Theorem 2: Let $\alpha > 0$ satisfy the condition in (13), and $\alpha^k = \alpha \; \forall k \in \mathbb{N}$. Then, for any initial condition, the sequence $(\boldsymbol{x}^k)_{k \in \mathbb{N}}$ generated by Algorithm 2 converges to $\boldsymbol{x}^* = \boldsymbol{1}_N \otimes x^*$, where x^* is the NE of the game in (1), with linear rate: for any $\epsilon > 0$, there exists $K > 0$ such that, for all $k \in \mathbb{N}$,

$$\|\boldsymbol{x}^k - \boldsymbol{x}^*\|_{\boldsymbol{Q}} \leq K \left(\sqrt{\rho_\alpha} + \epsilon\right)^k \|\boldsymbol{x}^0 - \boldsymbol{x}^*\|_{\boldsymbol{Q}}. \qquad\square$$

While in Algorithm 2 the PF eigenvector is estimated online, the upper bound on α in Theorem 2 is still a function of the network parameter $\bar{\sigma}$, which can be difficult to compute distributedly. Upper/lower bounds might be available for some classes of networks, e.g., unweighted graphs. This is analogous to [19, Th. 2], where q is computed online, but the step size depends on global, not easily accessible, information. In fact, this notion of fixed but small-enough step sizes is not uncommon in distributed algorithms literature.

When estimating a step α that satisfies (13) is impossible, convergence to a NE can still be guaranteed by allowing for diminishing step sizes. In this case, also the information on the game (i.e., Lipschitz and monotonicity constants of the pseudo-gradient) is not needed for the tuning.

Theorem 3: Let $(\alpha^k)_{k \in \mathbb{N}}$ be a positive nonincreasing sequence such that $\sum_{k \in \mathbb{N}} \alpha^k = \infty$ and $\lim_{k \to \infty} \alpha^k = 0$. Then, for any initial condition, the sequence $(\boldsymbol{x}^k)_{k \in \mathbb{N}}$ generated by Algorithm 2 converges to $\boldsymbol{x}^* = \boldsymbol{1}_N \otimes x^*$, where x^* is the NE of the game in (1). $\qquad\square$

IV. NUMERICAL EXAMPLE: A NASH-COURNOT GAME

We consider the Cournot competition model in [14, §6]. N firms produce an uniform commodity that is sold to m markets. Each firm $i \in \mathcal{I} = \{1, \ldots, N\}$ is allowed to participate in $n_i \leq m$ of the markets; its decision variable is the vector $x_i \in \mathbb{R}^{n_i}$ of quantities of product to be delivered to each of the n_i markets, bounded by the local constraints $\boldsymbol{0}_{n_i} \leq x_i \leq X_i$. Let $A_i \in \mathbb{R}^{m \times n_i}$ such that $[A_i]_{k,j} = 1$ if $[x_i]_j$ is the amount of commodity sent to the k-th market by agent i, $[A_i]_{k,j} = 0$ otherwise, for all $j = 1, \ldots, n_i$, $k = 1, \ldots, m$. Hence, $Ax = \sum_{i=1}^N A_i x_i \in \mathbb{R}^m$, where $A :=$
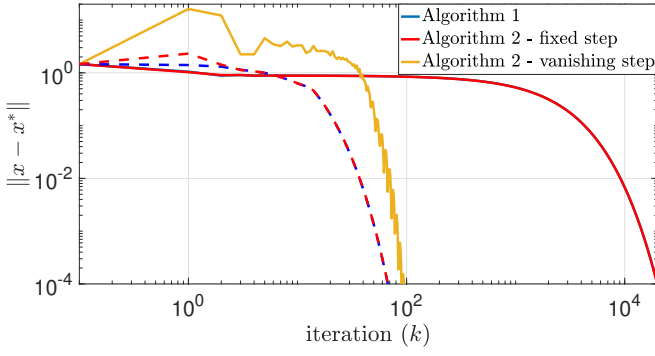
3558

Fig. 1. Distance from the Nash equilibrium, with the step sizes that ensure convergence (solid lines) and with a fixed step size chosen 400 times bigger than the theoretical upper bound (dashed lines).

$[A_1 \ldots A_N]$, are the quantities of product delivered to each market. Firm i aims at maximizing its profit, i.e., minimizing the cost $J_i(x_i, x_{-i}) = c_i(x_i) - p(Ax)^\top A_i x_i$. Here, $c_i(x_i) = x_i^\top Q_i x_i + q_i^\top x_i$ is firm i's production cost, with $Q_i \succ 0$; $p : \mathbb{R}^m \to \mathbb{R}^m$ associates to each market a price that depends on the amount of product delivered to that market. Specifically, for $k = 1, \ldots, m$, $[p(x)]_k = \bar{P}_k - \chi_k[Ax]_k$, where $\bar{P}_k, \chi_k > 0$. We set $N = 20$, $m = 7$. The market structure (i.e., which firms participate in each market) is defined as in [14, Fig. 1]. Therefore, $x = \text{col}((x_i))_{i \in \mathcal{I}} \in \mathbb{R}^n$ and $n = 32$. We select randomly with uniform distribution r_k in $[1, 2]$, Q_i diagonal with diagonal entries in $[14, 16]$, q_i with elements in $[1, 2]$, $\bar{P}_k$ in $[10, 20]$, χ_k in $[1, 3]$, X_i in $[5, 10]$, for all $i \in \mathcal{I}$, $k = 1, \ldots, m$. This setup satisfies Standing Assumptions 1-2 [14, §6]. The firms communicate over a randomly generated strongly connected row stochastic directed network, but cannot access the production of all the competitors. We set $\alpha \approx 3 \times 10^{-5}$ to satisfy the condition in (13). We compare the performance of Algorithms 1 and Algorithm 2, the latter both with a fixed ($\alpha^k = \alpha$) and vanishing step size ($\alpha^k = \frac{1}{k+1}$), in figure 1. Due to the small α, the schemes with fixed step are almost indistinguishable, and diminishing step sizes result in faster convergence. The good performance obtained with vanishing step suggests that the choice of α is quite conservative. Indeed, Algorithms 1-2 still converge, and much faster, with a fixed step size 400 times larger than its theoretical upper bound (dashed lines).

V. CONCLUSION

Certain properties of doubly stochastic matrices carry on to row stochastic matrices, but in a different Hilbert space, weighted by their left Perron-Frobenius eigenvector. We exploited one such contractivity property to solve, in a fully-distributed way, Nash equilibrium problems over directed networks. Any requirement for global knowledge of the graph and of the game mapping can be avoided in the case of vanishing step sizes.

The extension of our results to generalized games, where the agents share some common constraints, is left as future research. It would be also valuable to relax our connectivity and monotonicity assumptions, namely allowing for jointly connected networks and (strictly) monotone game mappings.

APPENDIX

1) Proof of Lemma 1: By $W\mathbf{1}_N = \mathbf{1}_N$, it suffices to show that $\|W - \mathbf{1}_N q^\top\|_Q = \bar{\sigma} < 1$. Let $p := \text{col}((\sqrt{q_i})_{i \in \mathcal{I}})$. Then,

$$\|W - \mathbf{1}_N q^\top\|_Q^2 = \|Q^{\frac{1}{2}}(W - \mathbf{1}_N q^\top)Q^{-\frac{1}{2}}\|^2$$
$$= \lambda_{\max}((Q^{\frac{1}{2}}WQ^{-\frac{1}{2}} - pp^\top)^\top(Q^{\frac{1}{2}}WQ^{-\frac{1}{2}} - pp^\top))$$
$$\overset{\text{(a)}}{=} \lambda_{\max}(Q^{-\frac{1}{2}}W^\top QWQ^{-\frac{1}{2}} - pp^\top) := \lambda_{\max}(M - pp^\top),$$

where in (a) we used $p^\top p = 1$, and $M = Q^{-\frac{1}{2}}W^\top QWQ^{-\frac{1}{2}}$. Since M is symmetric and $Mp = p$, M has a basis of eigenvectors, say $\{v_1, \ldots, v_{N-1}, p\}$, with associate eigenvalues $\{s_1, \ldots, s_{N-1}, 1\}$. By orthogonality and $p^\top p = 1$, it follows that the eigenvalues of $M - pp^\top$ are $\{s_1, \ldots, s_{N-1}, 0\}$, with associate eigenvectors $\{v_1, \ldots, v_{N-1}, p\}$. Since $M \succeq 0$, it suffices to show that $s_i < 1$, for $i = 1, \ldots, N-1$. Seeking a contradiction, let $j \in \{1, \ldots, N-1\}$ such that $s_j \geq 1$, and $\bar{v} := Q^{-\frac{1}{2}}v_j$. Thus, we have $\|W\bar{v}\|_Q^2 = v_j^\top Q^{-\frac{1}{2}}W^\top QWQ^{-\frac{1}{2}}v_j = v_j^\top M v_j \geq v_j^\top v_j = \bar{v}Q\bar{v} = \|\bar{v}\|_Q^2$. By [21, Lemma 1], it also holds, for some $\gamma > 0$, for any $y \in \mathbb{R}^N$, that $\|Wy\|_Q \leq \|y\|_Q - \gamma\|(I_N - W)y\|_Q$. Hence, by Standing Assumption 3, it must hold that $\bar{v} = \beta\mathbf{1}_N$, for some $\beta \neq 0$. Equivalently, $v_j = \beta p$. This is a contradiction, since p and v_j must be orthogonal. The conclusion follows with $\bar{\sigma} = \sqrt{\lambda_{N-1}(M)}$. ∎

2) Proof of Lemma 3: For any $x, y \in \mathbb{R}^n$ it holds that $\langle \bar{Q}^{-1}(F(x) - F(y)), x - y \rangle_{\bar{Q}} = \langle F(x) - F(y), x - y \rangle \geq \mu\|x - y\|^2 \geq \frac{\mu}{\lambda_{\max}(\bar{Q})}\|x - y\|_{\bar{Q}}^2$, and that $\|\bar{Q}^{-1}(F(x) - F(y))\|_{\bar{Q}}^2 = \|F(x) - F(y)\|_{\bar{Q}^{-1}}^2 \leq \frac{\lambda_{\max}(\bar{Q}^{-1})}{\lambda_{\min}(\bar{Q})}\ell_0^2\|x - y\|_{\bar{Q}}^2$. Analogously, by Lemma 2, it holds that, for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{Nn}$, $\|\bar{Q}^{-1}(\boldsymbol{F}(\boldsymbol{x}) - \boldsymbol{F}(\boldsymbol{y}))\|_{\bar{Q}}^2 \leq \frac{\lambda_{\max}(\bar{Q}^{-1})}{\lambda_{\min}(\boldsymbol{Q})}\ell^2\|\boldsymbol{x} - \boldsymbol{y}\|_{\boldsymbol{Q}}^2$. ∎

3) Proof of Lemma 4: We use the shorthand notation $\bar{\boldsymbol{F}}\boldsymbol{x}$ and $\bar{F}x$ in place of $\bar{\boldsymbol{F}}(\boldsymbol{x})$ and $\bar{F}(x)$. Let $\boldsymbol{x} \in \mathbb{R}^{Nn}$, $\boldsymbol{y} = \mathbf{1} \otimes y \in \boldsymbol{E}$, and $\hat{\boldsymbol{x}} := \boldsymbol{W}\boldsymbol{x} = \hat{\boldsymbol{x}}_\parallel + \hat{\boldsymbol{x}}_\perp = \mathbf{1}_N \otimes \hat{x}_\parallel + \hat{\boldsymbol{x}}_\perp \in \mathbb{R}^{Nn}$, with $\hat{\boldsymbol{x}}_\perp \in \boldsymbol{E}_\perp^Q$. Thus, we have

$$\|\mathcal{F}(\boldsymbol{x}) - \mathcal{F}(\boldsymbol{y})\|_{\boldsymbol{Q}}^2$$
$$= \|(\hat{\boldsymbol{x}} - \alpha\mathcal{R}^\top\bar{\boldsymbol{F}}\hat{\boldsymbol{x}}) - (\boldsymbol{y} - \alpha\mathcal{R}^\top\bar{\boldsymbol{F}}\boldsymbol{y})\|_{\boldsymbol{Q}}^2$$
$$= \|\hat{\boldsymbol{x}}_\parallel - \boldsymbol{y}\|_{\boldsymbol{Q}}^2 + \|\hat{\boldsymbol{x}}_\perp\|_{\boldsymbol{Q}}^2 + \alpha^2\|\mathcal{R}^\top(\bar{\boldsymbol{F}}\hat{\boldsymbol{x}} - \bar{\boldsymbol{F}}\boldsymbol{y})\|_{\boldsymbol{Q}}^2$$
$$\quad - 2\alpha\langle\hat{\boldsymbol{x}}_\perp, \mathcal{R}^\top(\bar{\boldsymbol{F}}\hat{\boldsymbol{x}} - \bar{\boldsymbol{F}}\boldsymbol{y})\rangle_{\boldsymbol{Q}}$$
$$\quad - 2\alpha\langle\hat{\boldsymbol{x}}_\parallel - \boldsymbol{y}, \mathcal{R}^\top(\bar{\boldsymbol{F}}\hat{\boldsymbol{x}} - \bar{\boldsymbol{F}}\hat{\boldsymbol{x}}_\parallel)\rangle_{\boldsymbol{Q}} \qquad (15)$$
$$\quad - 2\alpha\langle\hat{\boldsymbol{x}}_\parallel - \boldsymbol{y}, \mathcal{R}^\top(\bar{\boldsymbol{F}}\hat{\boldsymbol{x}}_\parallel - \bar{\boldsymbol{F}}\boldsymbol{y})\rangle_{\boldsymbol{Q}}$$
$$\leq \|\hat{\boldsymbol{x}}_\parallel - \boldsymbol{y}\|_{\boldsymbol{Q}}^2 + \|\hat{\boldsymbol{x}}_\perp\|_{\boldsymbol{Q}}^2 + \alpha^2\bar{\ell}^2(\|\hat{\boldsymbol{x}}_\perp\|_{\boldsymbol{Q}}^2 + \|\hat{\boldsymbol{x}}_\parallel - \boldsymbol{y}\|_{\boldsymbol{Q}}^2)$$
$$\quad + 2\alpha\bar{\ell}\|\hat{\boldsymbol{x}}_\perp\|(\|\hat{\boldsymbol{x}}_\perp\|_{\boldsymbol{Q}} + \|\hat{\boldsymbol{x}}_\parallel - \boldsymbol{y}\|_{\boldsymbol{Q}})$$
$$\quad + 2\alpha\bar{\ell}\|\hat{\boldsymbol{x}}_\parallel - \boldsymbol{y}\|_{\boldsymbol{Q}}\|\hat{\boldsymbol{x}}_\perp\|_{\boldsymbol{Q}} - 2\alpha\bar{\mu}\lambda_{\min}(Q)\|\hat{\boldsymbol{x}}_\parallel - \boldsymbol{y}\|_{\boldsymbol{Q}}^2,$$

and to bound the addends in (15) we used:

- 3rd, 4th, 5th terms: Lipschitz continuity of $\bar{\boldsymbol{F}}$, the Cauchy-Schwartz inequality, $\|\mathcal{R}^\top v\|_{\boldsymbol{Q}} = \|v\|_{\bar{Q}}$ for any $v \in \mathbb{R}^n$, $\|\hat{\boldsymbol{x}} - \boldsymbol{y}\|_{\boldsymbol{Q}}^2 = \|\hat{\boldsymbol{x}}_\parallel - \boldsymbol{y}\|_{\boldsymbol{Q}}^2 + \|\hat{\boldsymbol{x}}_\perp\|_{\boldsymbol{Q}}^2$ (by orthogonality);
- 7th term: $\langle\hat{\boldsymbol{x}}_\parallel - \boldsymbol{y}, \mathcal{R}^\top(\bar{\boldsymbol{F}}\hat{\boldsymbol{x}}_\parallel - \bar{\boldsymbol{F}}\boldsymbol{y})\rangle_{\boldsymbol{Q}} = \langle\hat{x}_\parallel - y, \bar{F}\hat{x}_\parallel - \bar{F}y\rangle_{\bar{Q}} \geq \bar{\mu}\|\hat{x}_\parallel - y\|_{\bar{Q}}^2 \geq \bar{\mu}\lambda_{\min}(\bar{Q})\|\hat{x}_\parallel - y\|^2 = \bar{\mu}\lambda_{\min}(\bar{Q})\|\hat{\boldsymbol{x}}_\parallel - \boldsymbol{y}\|_{\boldsymbol{Q}}^2$, and the last equality follows since $\hat{\boldsymbol{x}}_\parallel, \boldsymbol{y} \in \boldsymbol{E}$, $\boldsymbol{Q} = Q \otimes I_n$ and $\mathbf{1}_N^\top q = 1$.

3559

Besides, for every $\boldsymbol{x} = \boldsymbol{x}_{\|} + \boldsymbol{x}_{\perp} \in \mathbb{R}^{Nn}$, with $\boldsymbol{x}_{\|} \in \boldsymbol{E}$ and $\boldsymbol{x}_{\perp} \in \boldsymbol{E}_{\perp}^{Q}$, it holds that $\hat{\boldsymbol{x}} = \boldsymbol{W}\boldsymbol{x} = \boldsymbol{x}_{\|} + \boldsymbol{W}\boldsymbol{x}_{\perp}$, where $\boldsymbol{W}\boldsymbol{x}_{\perp} \in \boldsymbol{E}_{\perp}^{Q}$ (since $(q \otimes I_n)^{\top}\boldsymbol{W}\boldsymbol{x}_{\perp} = (q \otimes I_n)^{\top}\boldsymbol{x}_{\perp} = \boldsymbol{0}_n$, by definition of $\boldsymbol{W}$ and q). Consequently, by Lemma 1 and by $\boldsymbol{x}_{\perp} = (I_{Nn} - \boldsymbol{1}_N q^{\top} \otimes I_n)\boldsymbol{x}$, we have $\|\hat{\boldsymbol{x}}_{\perp}\|_{\boldsymbol{Q}} = \|\boldsymbol{W}\boldsymbol{x}_{\perp}\|_{\boldsymbol{Q}} \le \bar{\sigma}\|\boldsymbol{x}_{\perp}\|_{\boldsymbol{Q}}$. Therefore, we can finally write

$$\|\mathcal{F}(\boldsymbol{x}) - \mathcal{F}(\boldsymbol{y}))\|_{\boldsymbol{Q}}^2 \le \begin{bmatrix} \|\boldsymbol{x}_{\|} - \boldsymbol{y}\|_{\boldsymbol{Q}} \\ \|\boldsymbol{x}_{\perp}\|_{\boldsymbol{Q}} \end{bmatrix}^{\top} M_{\alpha} \begin{bmatrix} \|\boldsymbol{x}_{\|} - \boldsymbol{y}\|_{\boldsymbol{Q}} \\ \|\boldsymbol{x}_{\perp}\|_{\boldsymbol{Q}} \end{bmatrix}$$
$$\le \lambda_{\max}(M_{\alpha})(\|\boldsymbol{x}_{\|} - \boldsymbol{y}\|_{\boldsymbol{Q}}^2 + \|\boldsymbol{x}_{\perp}\|_{\boldsymbol{Q}}^2)$$
$$= \lambda_{\max}(M_{\alpha})\|\boldsymbol{x} - \boldsymbol{y}\|_{\boldsymbol{Q}}^2. \qquad \blacksquare$$

4) Proof of Theorem 2: We recast Algorithm 2 as

$$\boldsymbol{x}^{k+1} = \text{proj}_{\boldsymbol{\Omega}}(\hat{\mathcal{F}}^k(\boldsymbol{x}^k)),$$

where $\hat{\mathcal{F}}^k(\boldsymbol{x}^k) := \boldsymbol{W}\boldsymbol{x}^k - \alpha\mathcal{R}^{\top}(\bar{Q} + \tilde{Q}^k)^{-1}\boldsymbol{F}(\boldsymbol{W}\boldsymbol{x}^k)$, and $\tilde{Q}^k = \text{diag}(((\hat{q}_{i,i}^k - q_i)I_{n_i})_{i \in \mathcal{I}})$. We noted in §III-B that $(\bar{Q} + \tilde{Q}^k) = \text{diag}((\hat{q}_{i,i}^k I_{n_i})_{i \in \mathcal{I}}) \succ 0$, for all k; also, $\hat{q}_{i,i}^k - q_i \to 0$, for all $i \in \mathcal{I}$. Intuitively, Theorem 2 is based on the fact that $\hat{\mathcal{F}}^k$ approaches $\mathcal{F}$ in (11) asymptotically (i.e., when $\tilde{Q}^k \approx \boldsymbol{0}$), hence a contractivity property similar to Lemma 4 can be ensured for any big-enough k. Specifically, we note that $(\bar{Q} + \tilde{Q}^k)^{-1} = \bar{Q}^{-1} - (\bar{Q}(\bar{Q} + \tilde{Q}^k))^{-1}\tilde{Q}^k =: \bar{Q}^{-1} - P^k$, since the matrices involved are diagonal. Therefore $\hat{\mathcal{F}}^k(\boldsymbol{x}) = \mathcal{F}(\boldsymbol{x}) + \alpha\tilde{\mathcal{F}}^k(\boldsymbol{W}\boldsymbol{x})$, with $\mathcal{F}$ as in (11) and $\tilde{\mathcal{F}}^k(\boldsymbol{x}) := \mathcal{R}^{\top}P^k\boldsymbol{F}(\boldsymbol{x})$. Analogously to Lemma 3, it can be shown that $\tilde{\mathcal{F}}^k$ is $\tilde{\ell}^k$-Lipschitz in $\mathcal{H}_{\boldsymbol{Q}}$, with $\tilde{\ell}^k := \lambda_{\max}(P^k)\ell\sqrt{\lambda_{\max}(Q)/\lambda_{\min}(Q)}$. Then, by Lemma 4, $\hat{\mathcal{F}}$ is $(\sqrt{\rho_{\alpha}} + \alpha\tilde{\ell}^k)$-restricted Lipschitz in $\mathcal{H}_{\boldsymbol{Q}}$ with respect to $\boldsymbol{E}$ (cf. Lemma 4). Then, analogously to Theorem 1, it holds, for all $k \in \mathbb{N}$, that

$$\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\|_{\boldsymbol{Q}} \le (\sqrt{\rho_{\alpha}} + \alpha\tilde{\ell}^k)\|\boldsymbol{x}^k - \boldsymbol{x}^*\|_{\boldsymbol{Q}}.$$

We remark that $\tilde{\ell}^k \to 0$, since $\tilde{Q}^k \to \boldsymbol{0}$. Hence, for any $\epsilon > 0$, the conclusion follows with $K = (\prod_{k=1}^{\bar{k}} \max\{\sqrt{\rho_{\alpha}} + \alpha\tilde{\ell}^k, 1\})(\sqrt{\rho_{\alpha}} + \epsilon)^{-\bar{k}}$, where $\bar{k} := \max\{k \mid \alpha\tilde{\ell}^k > \epsilon\}$. $\qquad \blacksquare$

5) Proof of Theorem 3: Analogously to the proof of Theorem 2, for all $k \in \mathbb{N}$, it holds that $\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\|_{\boldsymbol{Q}} \le \delta^k\|\boldsymbol{x}^k - \boldsymbol{x}^*\|_{\boldsymbol{Q}}$, $\delta^k := (\sqrt{\rho_{\alpha^k}} + \alpha^k\tilde{\ell}^k)$, with ρ_{α^k} as in (13) and $(\tilde{\ell}^k)_{k \in \mathbb{N}}$ a vanishing nonnegative sequence. The conclusion follows because $\prod_{k=0}^{\infty} \delta^k = 0$, as we show next. By explicit computation of the quantity in (13) and Taylor expansion at $\alpha = 0$, it holds, in a neighborhood V_0 of $\alpha = 0$, that $\sqrt{\rho_{\alpha}} = 1 - \bar{\mu}\lambda_{\min}(Q)\alpha + o(\alpha)$, where $o(\alpha)$ is a series of monomial terms at least quadratic in α. Take $\bar{k}$ such that, for all $k \ge \bar{k}$, $\tilde{\ell}^k \le \ell^* < \bar{\mu}\lambda_{\min}(Q)$ for some ℓ^*, $\alpha^k \in \boldsymbol{V}_0$ and $\delta^k < 1$ (which is always possible, because $\tilde{\ell}^k \to 0$, $\alpha^k \to 0$ and $\delta^k = 1 - (\bar{\mu}\lambda_{\min}(Q) - \tilde{\ell}^k)\alpha^k + o(\alpha^k)$ if $\alpha^k \in \boldsymbol{V}_0$). Then, $\prod_{k=\bar{k}}^{\infty} \delta^k = 0$ if and only if $\sum_{k=\bar{k}}^{\infty} -\log(\delta^k) = \infty$. In turn, by the asymptotic comparison theorem and by the Taylor expansion at $\alpha^k = 0$, the latter series diverges if the series $\sum_{k=\bar{k}}^{\infty} \alpha^k(\bar{\mu}\lambda_{\min}(Q) - \tilde{\ell}^*)$ diverges, which holds by the assumption on $(\alpha^k)_{k \in \mathbb{N}}$. $\qquad \blacksquare$

REFERENCES

[1] J. Barrera and A. Garcia, "Dynamic incentives for congestion control," *IEEE Transactions on Automatic Control*, vol. 60, no. 2, pp. 299–310, 2015.

[2] W. Saad, Z. Han, H. V. Poor, and T. Basar, "Game-theoretic methods for the smart grid: An overview of microgrid systems, demand-side management, and smart grid communications," *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 86–105, 2012.

[3] N. Li, L. Chen, and M. A. Dahleh, "Demand response using linear supply function bidding," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1827–1838, 2015.

[4] J. Ghaderi and R. Srikant, "Opinion dynamics in social networks with stubborn agents: Equilibrium and convergence rate," *Automatica*, vol. 50, no. 12, pp. 3209 – 3215, 2014.

[5] J. S. Shamma and G. Arslan, "Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria," *IEEE Transactions on Automatic Control*, vol. 50, no. 3, pp. 312–327, 2005.

[6] C. De Persis and S. Grammatico, "Continuous-time integral dynamics for a class of aggregative games with coupling constraints," *IEEE Transactions on Automatic Control*, vol. 65, no. 5, pp. 2171–2176, 2020.

[7] G. Belgioioso and S. Grammatico, "Projected-gradient algorithms for generalized equilibrium seeking in aggregative games are pre-conditioned forward-backward methods," in *2018 European Control Conference*, 2018, pp. 2188–2193.

[8] B. Swenson, S. Kar, and J. Xavier, "Empirical centroid fictitious play: An approach for distributed learning in multi-agent games," *IEEE Transactions on Signal Processing*, vol. 63, no. 15, pp. 3888–3901, 2015.

[9] J. Koshal, A. Nedić, and U. V. Shanbhag, "Distributed algorithms for aggregative games on graphs," *Operations Research*, vol. 64, pp. 680–704, 2016.

[10] T. Tatarenko, W. Shi, and A. Nedić, "Accelerated gradient play algorithm for distributed Nash equilibrium seeking," in *2018 IEEE Conference on Decision and Control (CDC)*, 2018, pp. 3561–3566.

[11] C. De Persis and S. Grammatico, "Distributed averaging integral Nash equilibrium seeking on networks," *Automatica*, vol. 110, p. 108548, 2019.

[12] M. Ye and G. Hu, "Distributed Nash equilibrium seeking by a consensus based approach," *IEEE Transactions on Automatic Control*, vol. 62, no. 9, pp. 4811–4818, 2017.

[13] G. Belgioioso, A. Nedić, and S. Grammatico, "Distributed generalized Nash equilibrium seeking in aggregative games on time-varying networks," *IEEE Transactions on Automatic Control, DOI:10.1109/TAC.2020.3005922*, 2020.

[14] L. Pavel, "Distributed GNE seeking under partial-decision information over networks via a doubly-augmented operator splitting approach," *IEEE Transactions on Automatic Control*, vol. 65, no. 4, pp. 1584–1597, 2020.

[15] M. Bianchi, G. Belgioioso, and S. Grammatico, "Fast generalized Nash equilibrium seeking under partial-decision information," *arXiv preprint 2003.09335*, 2020.

[16] Z. Deng and X. Nian, "Distributed generalized Nash equilibrium seeking algorithm design for aggregative games over weight-balanced digraphs," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 3, pp. 695–706, 2019.

[17] M. Bianchi and S. Grammatico, "Fully distributed Nash equilibrium seeking over time-varying communication networks with linear convergence rate," *IEEE Control Systems Letters*, vol. 5, no. 2, pp. 499–504, 2021.

[18] F. Salehisadaghiani and L. Pavel, "Nash equilibrium seeking with non-doubly stochastic communication weight matrix," *EAI Endorsed Trans. on Collaborative Computing*, vol. 4, no. 13, pp. 1– 11, 2019.

[19] C. Xi, V. S. Mai, R. Xin, E. H. Abed, and U. A. Khan, "Linear convergence in optimization over directed graphs with row-stochastic matrices," *IEEE Transactions on Automatic Control*, vol. 63, no. 10, pp. 3558–3565, 2018.

[20] F. Facchinei and J. Pang, *Finite-dimensional variational inequalities and complementarity problems*. Springer, 2007.

[21] C. Cenedese, G. Belgioioso, Y. Kawano, S. Grammatico, and M. Cao, "Asynchronous and time-varying proximal type dynamics in multi-agent network games," *IEEE Transactions on Automatic Control, DOI:10.1109/TAC.2020.3011916*, 2020.

[22] M. Bianchi and S. Grammatico, "A continuous-time distributed generalized Nash equilibrium seeking algorithm over networks for double-integrator agents," in *2020 European Control Conference*, 2020, pp. 1474–1479.

[23] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*. Springer, 2017, vol. 2011.