A Distributed Algorithm For Almost-Nash Equilibria of Average Aggregative Games With Coupling Constraints

Journal Article

Author(s):
Parise, Francesca; Gentile, Basilio; Lygeros, John

Publication date:
2020-06

Permanent link:
https://doi.org/10.3929/ethz-b-000424125

Originally published in:
IEEE Transactions on Control of Network Systems 7(2), https://doi.org/10.1109/TCNS.2019.2944300

Funding acknowledgement:
787845 - Optimal control at large (EC)

A distributed algorithm for almost-Nash equilibria of average aggregative games with coupling constraints

Francesca Parise, Basilio Gentile and John Lygeros

***Abstract*—We consider the framework of average aggregative games, where the cost function of each agent depends on his own strategy and on the average population strategy. We focus on the case in which the agents are coupled not only via their cost functions, but also via a shared constraint coupling their strategies. We propose a distributed algorithm that achieves an ε-Nash equilibrium by requiring only local communications of the agents, as specified by a sparse communication network. The proof of convergence of the algorithm relies on the auxiliary class of network aggregative games. We apply our theoretical findings to a multi-market Cournot game with transportation costs and maximum market capacity.**

I. Introduction

AVERAGE aggregative games are used to describe populations of non-cooperative agents where each agent is not subject to one-to-one interactions, but is rather influenced by the average strategy of the population. These games can be used to model a vast number of technological applications ranging from traffic [1] and wireless systems [2] to electricity [3] and commodity markets [4]. Applying game theoretical concepts to such systems is challenging because the agents have private costs and constraints and may be able or willing to exchange information only with a (small) subset of the (large) population. Moreover, often the agents' strategies must collectively satisfy some physical coupling constraints, as in electricity markets [5], where the energy demand should not exceed the grid capacity, or in communication networks [6], where the package traffic should not exceed the congestion level.

Main contributions: We present a *distributed algorithm* that guarantees convergence to an ε-Nash Equilibrium (NE) of an average aggregative game *with affine coupling constraints* by using local communications over a *sparse network*. The tolerance ε can be made arbitrarily small by increasing the number of communications among the agents. Our method works for populations of heterogeneous agents with local convex constraints, a shared coupling constraint and smooth strongly convex cost functions (e.g., we do not assume that the cost functions are quadratic, as in [7]) under the assumption of strong monotonicity of the game operator. To prove algorithmic convergence we rely on two key steps: i) we show that the NE of the average aggregative game of interest can be approximated to any desired precision by the NE of an auxiliary network aggregative game (as defined in [8]); ii) we show that previously derived algorithms can be implemented in a distributed fashion to compute the NE of such auxiliary game (which is an almost-Nash equilibrium for the original game). As a side contribution, to tackle i) we derive a convergence rate result for parametric variational inequalities (VIs) with affine constraints that is applicable beyond the context of games.

To illustrate our theoretical findings, we study a Cournot game with transportation costs, as introduced in [9, Section 1.4.3]. Such setup extends the multi-market Cournot game [10, Section 7.1] by introducing transportation costs.

Comparison with the literature: We review the rapidly growing literature on NE coordination by distinguishing whether the proposed algorithms 1) can only be applied when the agent strategy sets are decoupled or allow for constraints coupling the agents' strategies and 2) employ a central coordinator (decentralized algorithms) or local communications (distributed algorithms); which type of communication structure (i.e. decentralized vs distributed) is preferable depends on the application.

A vast literature focused on the case of *decoupled* strategy sets, where the feasible strategy set of each agent is not affected by the strategies of the other agents. Distributed algorithms relying on local communications among the agents are suggested in [11], [12], [13], [14]. All these algorithms cannot be applied to the case of *shared coupling constraints*, because they build on the core assumption that the strategy sets are decoupled, as we highlight in detail in Section IV-C. To the best of our knowledge, the only distributed algorithm available in the literature for average aggregative games with coupling constraints is [15]. However such algorithm is only applicable if the coupling constraints can be expressed as a system of linear equations [15, eq. (5)], thus preventing its applicability to the cases discussed above.

Finally, we note that our work has some affinity with the distributed algorithms suggested in [10], [16], [17], [18], [19], [20], [21] to compute a NE of generic games (i.e., games that do not have the aggregative structure considered here) with coupling constraints. In [10], [16] each agent is required to communicate with all the agents affecting its cost function. Hence, these schemes require communications among all the agents in the setup of average aggregative games. The recent work [17] presents a two-level algorithm, whereas here we propose a one-level algorithm. The algorithm in [18] requires an assumption on the coupling constraint that is not met in our setup (e.g. a constraint on the sum of the strategies would violate such an assumption). Finally, the algorithms proposed in [19], [20] and [21], which appeared in parallel

The first two authors contributed equally. F. Parise is with the Laboratory for Information and Decision Systems, MIT, Cambridge, MA, USA. Email: parisef@mit.edu. B. Gentile is with Circuit Mind Limited, London, United Kingdom. Email: basilio@circuitmind.io. J. Lygeros is with the Automatic Control Laboratory, ETH Zurich, Switzerland. Email: lygeros@control.ee.ethz.ch. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme grant agreement OCAL, No. 787845 and by the Swiss National Science Foundation (grant P2EZP2_168812 and P300P2_177805).

to our work, can be applied to the setup considered in this paper. The algorithms in [19] and [20] require an additional assumption in terms of an extended pseudo gradient mapping, hence we focus our comparison with the algorithm in [21], which is the closest to our work. While in our setting multiple communications are used in between two strategy updates, in [21] strategies and estimates are updated simultaneously. The approach of [21] has the advantage of converging to an exact Nash equilibrium rather than to an ε-Nash equilibrium. On the other hand, it requires each agent to maintain and communicate a local estimate of the strategies of all the rest of the agents, which can be expensive for large populations, whereas our algorithm requires each agent to only store and communicate an estimate of the average strategy. Moreover, given that each strategy update is performed with a more accurate estimate of the average, our setting might be beneficial for cases when multiple communications are cheap and it is instead important to minimize the number of strategy updates. By deriving rigorous bounds on the dependence of ε on the number of communications we allow a tradeoff between number of communications and performance. Finally, we note that convergence in [21] is proven for undirected networks, while we here consider (doubly stochastic) directed networks.

Organization: In Section II we formulate the game setup and we introduce the proposed algorithm. In Section III we present some preliminary results. In Section IV we prove our main result. Section V focuses on the application. Section VI presents generalizations and future research directions. Appendix A is a standalone section on convergence of parametric VIs. Appendix B contains the proofs omitted in the main text.

Notation: Given $A \in \mathbb{R}^{n\times n}$ $\|A\|_\mu$ is the induced μ-norm of A, for simplicity we write $\|A\|_2 = \|A\|$; $r_{\text{step}}(A) = \|A - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n^\top\|_2$; diag$(A)$ is the diagonal matrix which has the same diagonal of A. For $a,b \in \mathbb{Z}$, $a \le b$, $\mathbb{Z}[a,b] := [a,b]\cap\mathbb{Z}$. Given N vectors of $\mathbb{R}^n$, $[x^i]_{i=1}^N := [{x^1}^\top,\dots,{x^N}^\top]^\top \in \mathbb{R}^{Nn}$ and $x^{-i} := [x_1;\dots;x_{i-1};x_{i+1};\dots;x_N] \in \mathbb{R}^{(N-1)n}$. Given sets $\mathcal{X}^1,\dots,\mathcal{X}^N \subseteq \mathbb{R}^n$, $\frac{1}{N}\sum_{i=1}^N \mathcal{X}^i := \{z \in \mathbb{R}^n | z = \frac{1}{N}\sum_{i=1}^N x^i, \text{for some } x^i \in \mathcal{X}^i\}$ and $\mathcal{X}^{-i} := \mathcal{X}^1 \times \dots \mathcal{X}^{i-1} \times \mathcal{X}^{i+1} \times \dots \mathcal{X}^N$. $\Pi_{\mathcal{X}}[x]$ is the projection of x onto $\mathcal{X}$.

II. Problem formulation and main result

A. Average aggregative games

Consider a population of $N \in \mathbb{N}$ agents, where agent $i \in \mathbb{Z}[1,N]$ chooses his decision variable x^i in his individual constraint set $\mathcal{X}^i \subseteq \mathbb{R}^n$, and interacts with the other agents via the average of their strategies. The aim of agent i is to minimize his cost function $J^i(x^i, \sigma_\infty(x)) = J^i(z_1,z_2)\,|_{z_1=x^i, z_2=\sigma_\infty(x)}$, where $J^i(z_1,z_2) : \mathcal{X}^i \times \text{conv}(\mathcal{X}^1,\dots,\mathcal{X}^N) \to \mathbb{R}$ and

$$\sigma_\infty(x) := \frac{1}{N}\sum_{j=1}^N x^j. \tag{1}$$

Note that we use as aggregator the average $\sigma_\infty(x)$ instead of the sum $\sum_{j=1}^N x^j$ of the strategies. This is without loss of generality. Moreover, the subscript ∞ does not refer to an infinite population, but to the fact that the agents interact through the *exact* average $\frac{1}{N}\sum_{j=1}^N x^j$. We set $x := [x^1;\dots;x^N] \in \mathcal{X} := \mathcal{X}^1 \times\dots\times \mathcal{X}^N \subset \mathbb{R}^{Nn}$. Besides the individual constraints, we assume that the agents have to satisfy a linear coupling constraint on the average

$$x \in \mathcal{C}_\infty := \{x \in \mathbb{R}^{Nn} \,|\, \hat{A}\sigma_\infty(x) \le \hat{b}\}, \tag{2}$$

with $\hat{A} \in \mathbb{R}^{m\times n}$, $\hat{b} \in \mathbb{R}^m$, for some $m > 0$, where $\sigma_\infty(x)$ is as in (1). The coupling constraints in (2) can model the fact that the usage level for a certain commodity cannot exceed a fixed capacity, as in [5]. The strong modeling flexibility of linear coupling constraints is further discussed in [10, Remark 3.1].

The cost and constraints give rise to the average aggregative game (AAG)

$$\mathcal{G}_\infty := \begin{cases} \text{agents}: & (1,\dots,N) \\ \text{cost of agent } i: & J^i(x^i,\sigma_\infty(x)) \\ \text{individual constraint}: & \mathcal{X}^i \\ \text{coupling constraint}: & \mathcal{C}_\infty. \end{cases} \tag{3}$$

Let us denote $\mathcal{Q}_\infty := \mathcal{X} \cap \mathcal{C}_\infty$, and $\mathcal{Q}^i_\infty(x^{-i}) := \{x^i \in \mathcal{X}^i \,|\, x \in \mathcal{C}_\infty\} = \{x^i \in \mathcal{X}^i | \hat{A}\sigma_\infty(x) \le \hat{b}\}$.

Definition 1 (Nash Equilibrium (NE))**.** *A set of strategies $\bar{x} = [\bar{x}^1;\dots;\bar{x}^N] \in \mathcal{Q}_\infty$ is an ε-Nash equilibrium of $\mathcal{G}_\infty$ if for all $i \in \mathbb{Z}[1,N]$ and all $x^i \in \mathcal{Q}^i_\infty(\bar{x}^{-i})$*

$$J^i(\bar{x}^i,\sigma_\infty(\bar{x})) \le J^i\left(x^i, \tfrac{1}{N}x^i + \textstyle\sum_{j\ne i}\tfrac{1}{N}\bar{x}^j\right) + \varepsilon\,. \tag{4}$$

If (4) holds with $\varepsilon = 0$ then $\bar{x}$ is a Nash equilibrium. □

A NE for a game with coupling constraints is usually called a generalized NE [22]; here we omit the word generalized for brevity. The following conditions on cost functions and constraints of $\mathcal{G}_\infty$ are assumed to hold throughout.

Standing assumption. *For each i, $\mathcal{X}^i \subset \mathbb{R}^n$ is convex, compact and has non-empty interior. For all $i \in \mathbb{Z}[1,N]$, the cost function $J^i(x^i,\sigma_\infty(x))$ is convex in x^i for all $x^{-i} \in \mathcal{X}^{-i}$ and $J^i(z_1,z_2)$ is continuously differentiable in z_1, z_2 and Lipschitz with constant L, that is, $|J^i(z_1,z_2) - J^i(z_1',z_2')| \le L(\|z_1 - z_1'\| + \|z_2 - z_2'\|)$ for all z_1, z_2, z_1', z_2'*[1]. □

B. Communication limitations

Our main objective is to coordinate the agents' strategies to a NE by using a distributed algorithm that utilizes communications over a (typically sparse) communication network. We model such network by its adjacency matrix $T \in [0,1]^{N\times N}$, where the element $T_{ij} \in [0,1]$ is the weight that agent i assigns to communications received from agent j, with $T_{ij} = 0$ representing no communication. For brevity, we refer to T as communication network, even though it is its adjacency matrix. Agent j is an in-neighbor of i if $T_{ij} > 0$ and an out-neighbor if $T_{ji} > 0$. We denote the sets of in- and out-neighbors of agent i as $\mathcal{N}^i_{\text{in}}$ and $\mathcal{N}^i_{\text{out}}$, respectively.

Assumption 1 (Communication network)**.** *The communication matrix T is primitive (i.e. there exists $\nu > 0$ such that T^ν is element-wise positive) and doubly stochastic (i.e. $T\mathbb{1}_N = T^\top\mathbb{1}_N = \mathbb{1}_N$).*[2] *Moreover, $r_{\text{step}}(T) = \|T - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n^\top\|_2 < 1$.* □

[1]Note that the same Lipschitz constant L is used for each agent. This is without loss of generality since N is finite, hence L can be taken as the maximum of the Lipschitz constants.

[2]A relaxation of doubly stochasticity is discussed in Section VI-1.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TCNS.2019.2944300

Graph theoretical conditions guaranteeing Assumption 1 can be found in [23]. Loosely speaking, the fact that T is primitive and doubly stochastic ensures that if the agents communicate a sufficiently large number of times over T, they are able to approximate the average to any desired precision. The requirement on $r_{\text{step}}(T)$ is automatically met if either T is symmetric [23, Theorem 10.2 and Lemma 10.3] or if all its diagonal elements are strictly positive [23, Exercise 10.1].

C. The distributed algorithm

We propose an algorithm (Algorithm 1 below), where at iteration k each agent i updates four variables:

- his strategy $x^i_{(k)}$,
- a dual variable $\lambda^i_{(k)}$ relative to the coupling constraint $\mathcal{C}_\infty$,
- a local average of his in-neighbors' strategies $\sigma^i_{\nu,(k)}$,
- a local average of his out-neighbors' dual variables $\mu^i_{\nu,(k)}$.

To overcome the fact that the communication network is sparse we assume that to compute $\sigma^i_{\nu,(k)}$ and $\mu^i_{\nu,(k)}$ the agents communicate not once but multiple times over the network T. The number of rounds of communication per update is denoted by $\nu \in \mathbb{N}$ and is a tuning parameter of the algorithm. The personal strategy (or primal variable) and the dual variable, in turn, are updated by a gradient-like step that depends on a second tuning parameter $\tau > 0$. In particular, the strategy update step is similar to that of the standard projection algorithm [9, Algorithm 12.1.1]. We finally note that both tuning parameters ν and τ are decided a priori and do not change during the algorithm execution.

Algorithm 1: Distributed algorithm for ε_ν-NE of $\mathcal{G}_\infty$

Initialize: Agent i with state $x^i_{(0)} \in \mathcal{X}^i$ and dual variable $\lambda^i_{(0)} \in \mathbb{R}^m_{\geq 0}$. A communication network T. Set $\tau > 0$, $\nu \in \mathbb{N}$, $k = 0$, $\sigma^i_{\nu,(0)} = x^i_{(0)}$.

Iterate until convergence:

Communication: Dual

$\mu^i_{\nu,(k)} \leftarrow \lambda^i_{(k)}, \forall i$

repeat ν times

$\mu^i_{\nu,(k)} \leftarrow \sum_{j\in\mathcal{N}^i_{\text{out}}} T_{ji}\, \mu^j_{\nu,(k)}, \forall i$

Update: Primal

$F^i_{\nu,(k)} \leftarrow \nabla_{x^i} J^i(x^i_{(k)}, \sigma^i_{\nu,(k)}), \forall i$

$x^i_{(k+1)} \leftarrow \Pi_{\mathcal{X}^i}[x^i_{(k)} - \tau(F^i_{\nu,(k)} + \hat{A}^\top \mu^i_{\nu,(k)})], \forall i$

Communication: Primal

$\sigma^i_{\nu,(k+1)} \leftarrow x^i_{(k+1)}, \forall i$

repeat ν times

$\sigma^i_{\nu,(k+1)} \leftarrow \sum_{j\in\mathcal{N}^i_{\text{in}}} T_{ij} \sigma^j_{\nu,(k+1)}, \forall i$

Update: Dual

$\lambda^i_{(k+1)} \leftarrow \Pi_{\mathbb{R}^m_{\geq 0}}[\lambda^i_{(k)} - \tau(\hat{b} - 2\hat{A}\sigma^i_{\nu,(k+1)} + \hat{A}\sigma^i_{\nu,(k)})], \forall i$

$k \leftarrow k + 1$

The objective of the paper is to prove convergence of Algorithm 1 to an ε_ν-Nash of $\mathcal{G}_\infty$, where $\varepsilon_\nu \to 0$ as $\nu \to \infty$. To this end, we introduce the following additional assumptions.

Definition 2 (Strong monotonicity)**.** *An operator $F : \mathcal{K} \subseteq \mathbb{R}^n \to \mathbb{R}^n$ is strongly monotone if there exists $\alpha > 0$ such that $(F(x) - F(y))^\top (x - y) \geq \alpha \|x - y\|^2$ for all $x, y \in \mathcal{K}$.* □

Assumption 2 (Regularity of cost functions)**.** *The operator $F_\infty : \mathcal{X} \to \mathbb{R}^{Nn}$, defined as*

$$x \mapsto F_\infty(x) := [\nabla_{x^i} J^i(x^i, \sigma_\infty(x))]_{i=1}^N \tag{6}$$

is strongly monotone with constant α_∞. The cost function $J^i(z_1, z_2)$ is twice continuously differentiable for all $i \in \mathbb{Z}[1, N]$. Let $L' > 0$ and $M' > 0$ be such that $\|\nabla_{z_a} J^i(z_1, z_2) - \nabla_{z_a} J^i(z_1, z_2')\|_2 \leq L'\|z_2 - z_2'\|_2$ and $\|\nabla_{z_a} J^i(z_1, z_2)\|_2 \leq M'$ for all $a \in \{1, 2\}$, all z_1, z_2, z_2' and all $i \in \mathbb{Z}[1, N]$. □

We note that sufficient conditions for Assumption 2 to hold have been discussed in [24, Lemmas 3 and 4, Corollaries 1 and 2] for specific instances of average aggregative games.

Our key idea to prove convergence of Algorithm 1 to an ε_ν-NE of $\mathcal{G}_\infty$ is to construct an auxiliary game $\mathcal{G}_\nu$, parametric in the number of communications ν, and note that:

Claim 1: Algorithm 1 converges to a specific NE of $\mathcal{G}_\nu$, called a *variational NE*;

Claim 2: the variational NE of $\mathcal{G}_\nu$ is an ε_ν-NE of $\mathcal{G}_\infty$, with $\varepsilon_\nu \to 0$ as $\nu \to \infty$.

It turns out that, thanks to the structure imposed on the game $\mathcal{G}_\nu$, Claim 1 follows easily from previous literature results (in fact also other algorithms proposed in the literature could be similarly adapted to find the variational NE of $\mathcal{G}_\nu$ by using only communications over T). Claim 2, on the other hand, is the main contribution of the paper. A convergence rate for ε_ν is provided under the following additional assumption.

Assumption 3. *There exists $G_i \in \mathbb{R}^{m_i \times n}$ and $g_i \in \mathbb{R}^{m_i}$ such that $\mathcal{X}^i := \{x^i \in \mathbb{R}^n \mid G_i x^i \leq g_i\}$. Let $G := [G_i]_{i=1}^N, g := [g_i]_{i=1}^N$. The set $\mathcal{Q}_\infty$ has non-empty interior. There exists $L'' > 0$ and $M'' > 0$ such that $\|\nabla_{z_a z_b} J^i(z_1, z_2) - \nabla_{z_a z_b} J^i(z_1, z_2')\|_2 \leq L''\|z_2 - z_2'\|_2$ and $\|\nabla_{z_a z_b} J^i(z_1, z_2)\|_2 \leq M''$ for all $a, b \in \{1, 2\}$, all z_1, z_2, z_2' and $i \in \mathbb{Z}[1, N]$.* □

The second half of Assumption 3 requires that, for each agent i, every second derivative of J^i is bounded and is Lipschitz in the second argument (uniformly in the first). Claims 1 and 2 are proven in Section IV, we start with auxiliary results.

III. Auxiliary results

We here define the game $\mathcal{G}_\nu$, introduce some basic results on variational inequalities (VI) and study the relation between the operators and the sets of the VIs associated with $\mathcal{G}_\infty$ and $\mathcal{G}_\nu$.

A. Multi-communication network aggregative games

In each iteration of Algorithm 1 the agents need to communicate ν times over T; mathematically this is equivalent to communicating once over a fictitious network with adjacency matrix T^ν. Based on T^ν, we introduce the local averages

$$\sigma^i_\nu(x) := \textstyle\sum_{j=1}^N [T^\nu]_{ij} x^j.$$

We define $\mathcal{G}_\nu$ as a game with same constraints and cost functions as in $\mathcal{G}_\infty$ except for the fact that each agent reacts to the local average $\sigma_\nu^i(x)$ instead of the global average $\sigma_\infty(x)$. Specifically, upon defining

$$\mathcal{C}_\nu := \{x \in \mathbb{R}^{Nn} \mid \hat{A}\sigma_\nu^j(x) \le \hat{b},\ \forall j \in \mathbb{Z}[1,N]\}$$

we formally introduce the multi-communication network aggregative game as

$$\mathcal{G}_\nu := \begin{cases} \text{agents :} & (1,\dots,N) \\ \text{cost of agent } i : & J^i(x^i,\sigma_\nu^i(x)) \\ \text{individual constraint :} & \mathcal{X}^i \\ \text{coupling constraint :} & \mathcal{C}_\nu . \end{cases}$$

The definition of NE for $\mathcal{G}_\nu$ is the analogous of Definition 1.

B. Variational inequalities and variational Nash equilibria

A fundamental fact used throughout the rest of the paper is that a specific class of Nash equilibria of any convex game, called variational Nash equilibria, can be obtained by solving a variational inequality constructed from the game primitives.

Definition 3 (Variational inequality). *Given a set $\mathcal{K} \subseteq \mathbb{R}^n$ and an operator $H : \mathcal{K} \to \mathbb{R}^n$, the point $\bar{x} \in \mathbb{R}^n$ is a solution of* VI$(\mathcal{K}, H)$ *if it satisfies*

$$H(\bar{x})^\top (x - \bar{x}) \ge 0, \quad \forall\, x \in \mathcal{K}.$$

A discussion on how variational inequalities generalize convex optimization programs can be found in [9, Section 1.3.1]. In the following we report a sufficient condition for existence and uniqueness of the solution of a variational inequality.

Proposition 1 ([9, Theorem 2.3.3.b]). *Consider a closed and convex set $\mathcal{K}$ and a strongly monotone operator $H : \mathcal{K} \to \mathbb{R}^n$. Then VI($\mathcal{K}$,$H$) admits a unique solution.* □

The following lemma gives a more intuitive characterization of the strong monotonicity property.

Proposition 2 ([9, Proposition 2.3.2]). *A continuously differentiable operator $H : \mathcal{K} \subseteq \mathbb{R}^n \to \mathbb{R}^n$ is strongly monotone with monotonicity constant α if and only if $\nabla_x H(x) \succeq \alpha I_n$ for all $x \in \mathcal{K}$.* □

To draw the connection between VIs and Nash equilibria, let us introduce the following quantities relative to $\mathcal{G}_\nu$

$$A_\nu := T^\nu \otimes \hat{A}, \quad b := \mathbb{1}_N \otimes \hat{b} \tag{7a}$$

$$F_\nu(x) := [\nabla_{x^i} J^i(x^i, \sigma_\nu^i(x))]_{i=1}^N, \tag{7b}$$

$$Q_\nu := \{x \in \mathcal{X} | A_\nu x \le b\}, \tag{7c}$$

$$Q_\nu^i(x^{-i}) := \{x^i \in \mathcal{X}^i | A_\nu x \le b\}, \tag{7d}$$

and recall the corresponding quantities relative to $\mathcal{G}_\infty$

$$A_\infty := \tfrac{1}{N}\mathbb{1}_N \mathbb{1}_N{}^\top \otimes \hat{A}. \tag{8a}$$

$$F_\infty(x) := [\nabla_{x^i} J^i(x^i, \sigma_\infty(x))]_{i=1}^N, \tag{8b}$$

$$Q_\infty := \{x \in \mathcal{X} | A_\infty x \le b\}, \tag{8c}$$

$$Q_\infty^i(x^{-i}) := \{x^i \in \mathcal{X}^i | A_\infty x \le b\}, \tag{8d}$$

The operator in (8b) is the same as (6) in Assumption 2 and in (8) the coupling constraint $\mathcal{C}_\infty$ is expressed in the redundant form $A_\infty x \le b$ (consisting of N repetitions of the constraint $\hat{A}\sigma_\infty(x) \le \hat{b}$) to match the structure of $A_\nu x \le b$ in (7).

In the following we specialize a well-known result of [22, Theorem 2.1] to the two games $\mathcal{G}_\infty$ and $\mathcal{G}_\nu$.

Proposition 3 (Variational NE [22, Theorem 2.1]). *Every solution $\bar{x}_\infty$ of VI(Q_∞,F_∞) is a NE of $\mathcal{G}_\infty$, called a variational NE of $\mathcal{G}_\infty$. Moreover, if $J^i(x^i, \sigma_\nu^i(x))$ is convex in x^i for all $x^{-i} \in \mathcal{X}^{-i}$, every solution $\bar{x}_\nu$ of VI(Q_ν,F_ν) is a NE of $\mathcal{G}_\nu$, called a variational NE of $\mathcal{G}_\nu$.* □

Due to the presence of the coupling constraints the converse of Proposition 3 does not hold.[3]

C. Convergence of sets and operators

Given Proposition 3, to show Claim 2 it suffices to prove that the solution to VI$(\mathcal{Q}_\nu, F_\nu)$ is a good approximation of the solution to VI$(\mathcal{Q}_\infty, F_\infty)$, for ν sufficiently large. To this end, we start by introducing three lemmas that clarify the relation between the operators F_ν, F_∞ and the sets $\mathcal{Q}_\nu, \mathcal{Q}_\infty$. Their proofs are in Appendix B. To simplify the exposition we define

$$M_2 := \max_{x\in\mathcal{X}} \|x\|_2, \qquad M_\infty := \max_{x\in\mathcal{X}} \|x\|_\infty$$

and, for any norm μ, we define $d_\mu(\nu) := \|\frac{1}{N}\mathbb{1}_N \mathbb{1}_N{}^\top - T^\nu\|_\mu$.

Lemma 1 (Operator convergence). *Under Assumption 1*

1) *$\lim_{\nu\to\infty} T^\nu = \frac{1}{N}\mathbb{1}_N \mathbb{1}_N{}^\top$, $\lim_{\nu\to\infty} A_\nu = A_\infty$ and $\lim_{\nu\to\infty} F_\nu(x) = F_\infty(x)$ uniformly in x.*
2) *For all $x \in \mathcal{X}$ and all ν, $\|F_\nu(x)\| \le F_{\max}$ and $\|F_\infty(x)\| \le F_{\max}$, with $F_{\max} := 2\sqrt{N}M'$.*
3) *$d_2(\nu) \le r_{\text{step}}(T)^\nu$ and $d_\infty(\nu) \le \sqrt{N} r_{\text{step}}(T)^\nu$.*

Under the additional Assumption 2

4) *$\|F_\infty(x) - F_\nu(x)\|_2 \le K_F r_{\text{step}}(T)^\nu$ with $K_F := 2L'M_2 + M'\sqrt{N}$.* □

The next lemma provides a sufficient condition for F_ν to be strongly monotone, thus guaranteeing uniqueness of the solution to VI$(\mathcal{Q}_\nu, F_\nu)$.

Lemma 2 (Strong monotonicity of F_ν). *If Assumptions 1 and 2 hold, there exists ν_{SMON} such that F_ν is strongly monotone for all $\nu > \nu_{\text{SMON}}$. Under the additional Assumption 3 we can use*

$$\nu_{\text{SMON}} = \frac{1}{\log(r_{\text{step}}(T))} \log\left(\frac{\alpha_\infty}{4(L''M_2 + M'')}\right).$$ □

Finally, we look at convergence of the sets $\mathcal{Q}_\nu$ to $\mathcal{Q}_\infty$ in terms of the Hausdorff distance.

Definition 4 (Hausdorff distance). *The Hausdorff distance between two subsets R and S of $\mathbb{R}^n$ is*

$$d_H(R,S) := \max\{\sup_{r\in R} \inf_{s\in S} \|r-s\|_2, \sup_{s\in S} \inf_{r\in R} \|r-s\|_2\}.$$ □

[3] In other words, there can be Nash equilibria of $\mathcal{G}_\infty$ (resp. $\mathcal{G}_\nu$) that cannot be obtained as solutions of VI(Q_∞,F_∞) (resp. VI(Q_ν,F_ν)). Within the class of Nash equilibria, variational Nash equilibria enjoy special stability and sensitivity properties and the burden of meeting the coupling constraints is divided equally among all the agents [22, Theorem 3.1]. Variational equilibria are also a subset of the *normalized equilibria* defined in [25], which are in most cases the only Nash equilibria that can be computed.

To study convergence rates we use the following quantity.

Definition 5. *[26] For any matrix $H \in \mathbb{R}^{m\times n}$, let $\beta(H)$ be the smallest number such that for each nonsingular submatrix B of H all entries of B^{-1} are at most $\beta(H)$ in absolute value. Then we define $\Delta(H) = n\sqrt{n}\beta(H)$.* □

Lemma 3 (Hausdorff convergence of sets)**.** *Under Assumption 1, the following facts hold.*

1) $d_H(\mathcal{Q}_\nu, \mathcal{Q}_\infty) \to 0$;
2) $d_H(\mathcal{Q}^i_\nu(\bar{x}^{-i}_\nu), \mathcal{Q}^i_\infty(\bar{x}^{-i}_\nu)) \to 0, \forall \bar{x}_\nu \in \mathcal{Q}_\nu, i \in \mathbb{N}[1,N]$.

If additionally Assumption 3 holds then

3) $d_H(\mathcal{Q}_\nu, \mathcal{Q}_\infty) \le K_H r_{\text{step}}(T)^\nu, \forall \nu \ge \nu^{\text{glob}}_H$;
4) $d_H(\mathcal{Q}^i_\nu(\bar{x}^{-i}_\nu), \mathcal{Q}^i_\infty(\bar{x}^{-i}_\nu)) \le 2K^i_H r_{\text{step}}(T)^\nu, \forall \nu \ge \nu^{\text{loc}}_H$, $\bar{x}_\nu \in \mathcal{Q}_\nu$, $i \in \mathbb{N}[1,N]$;

where $\nu^{\text{glob}}_H := \frac{1}{\log(r_{\text{step}}(T))}\log\left(\frac{M_{\hat{A}}}{\|\hat{A}\|_\infty M_\infty \sqrt{N}}\right)$, $K_H := \Delta(\left[\begin{smallmatrix} A_\infty \\ G\end{smallmatrix}\right])\sqrt{N}\|\hat{A}\|_\infty M_\infty$, $M_{\hat{A}} := \max_{x\in\mathcal{X}} \min_{h\in\{1,\dots,m\}}[\hat{b} - \hat{A}\sigma_\infty(x)]_h > 0$, $\nu^{\text{loc}}_H := \frac{1}{\log(r_{\text{step}}(T))}\log\left(\frac{1}{2N}\right)$, $K^i_H := \Delta(\left[\begin{smallmatrix} A^i_\infty \\ G^i\end{smallmatrix}\right])\sqrt{N}\|\hat{A}\|_\infty M_\infty \le K_H$, $A^i_\infty := \frac{1}{N}[\mathbb{1}_N \otimes \hat{A}]$. □

Note that $M_{\hat{A}}$ as defined in Lemma 3 is positive since, by Assumption 3, there exists a point $\tilde{x} \in \mathcal{X}$ such that $\hat{A}\sigma_\infty(\tilde{x}) < \hat{b}$ and thus $\min_{h\in\{1,\dots,m\}}[\hat{b} - \hat{A}\sigma_\infty(\tilde{x})]_h > 0$.

IV. Main results

A. Convergence of variational NE

We here prove Claim 2. Specifically, Theorem 1 shows that the Euclidean distance between the vectors corresponding to the variational NE of $\mathcal{G}_\nu$ and $\mathcal{G}_\infty$ converges to zero. Theorem 2 then shows that the variational NE of $\mathcal{G}_\nu$ is an ε_ν-NE of $\mathcal{G}_\nu$. We note that this latter fact is not a trivial consequence of Theorem 1 and Lipschitz continuity because the set of feasible deviations for each agent i (i.e. $\mathcal{Q}^i_\nu(\bar{x}^{-i}_\nu)$ vs $\mathcal{Q}^i_\infty(\bar{x}^{-i}_\nu)$) is different in the two games, thus requiring Lemma 3.

Theorem 1 (Convergence in strategies)**.** *Suppose that Assumptions 1 and 2 hold. Then*

1) *The game $\mathcal{G}_\infty$ has a unique variational Nash equilibrium $\bar{x}_\infty$ and for any $\nu > \nu_{\text{SMON}}$, $\mathcal{G}_\nu$ has a unique variational Nash equilibrium $\bar{x}_\nu$. Moreover,*
$$\lim_{\nu\to\infty} \bar{x}_\nu = \bar{x}_\infty. \tag{9}$$
2) *If additionally Assumption 3 holds then for $\nu > \max\{\nu_{\text{SMON}}, \nu^{\text{glob}}_H\}$ it holds*
$$\|\bar{x}_\nu - \bar{x}_\infty\|_2 \le K_X\sqrt{r_{\text{step}}(T)^\nu}$$
where $K_X := \frac{K_F + \sqrt{K_F^2 + 8\alpha_\infty F_{\max} K_H}}{2\alpha_\infty}$, with $F_{\max}, K_F$ and K_H as defined in Lemmas 1 and 3.

Proof: 1) Existence and uniqueness of $\bar{x}_\infty$ and $\bar{x}_\nu$ solutions to VI(Q_∞,F_∞) and VI(Q_ν,F_ν) respectively is guaranteed by Proposition 1, because the operator F_∞ is strongly monotone by Assumption 2 and the operator F_ν (for $\nu > \nu_{\text{SMON}}$) is strongly monotone by Lemma 2. Strong monotonicity of F_ν implies convexity of $J^i(x^i, \sigma^i_\nu(x))$ in x^i. Consequently, Proposition 3 guarantees that $\bar{x}_\infty$ and $\bar{x}_\nu$ are the unique variational NE of $\mathcal{G}_\infty$ and $\mathcal{G}_\nu$, respectively. Asymptotic convergence of $\bar{x}_\nu$ to $\bar{x}_\infty$ follows from a continuity result for the solution of parametric VIs given in [27, Theorem A(b)], whose assumptions are verified in Lemma 1 and Lemma 3.1).
2) The second statement follows from Theorem A.1, given in Appendix A, combined with Lemma 1 and Lemma 3.3) to bound $d_F(\nu)$ and $d_H(\mathcal{Q}_\nu, \mathcal{Q}_\infty)$, respectively. ■

Theorem 2 (Convergence to ε-Nash)**.** *Suppose that Assumptions 1 and 2 hold. Then*

1) *For every $\varepsilon > 0$, there exists $\nu_\varepsilon > 0$ such that, for all $\nu \ge \nu_\varepsilon$, any variational Nash equilibrium $\bar{x}_\nu$ of $\mathcal{G}_\nu$ is an ε-Nash equilibrium of $\mathcal{G}_\infty$.*
2) *Under the additional Assumption 3, we can use $\nu_\varepsilon = \max\{\nu^{\text{loc}}_H, \nu_J\}$ where*[4]
$$\nu_J := \frac{1}{\log(r_{\text{step}}(T))}\log\left(\frac{\varepsilon}{2L(2\max_i(K^i_H) + M_\infty\sqrt{Nn})}\right),$$
with ν^{loc}_H and K^i_H as defined in Lemma 3.

Proof: 1) We divide the proof of this statement into two parts: i) we prove that $\bar{x}_\nu \in Q_\infty$ for any $\nu > 0$, and ii) we prove that condition (4) is satisfied.
i) Since $\bar{x}_\nu$ is a NE for $\mathcal{G}_\nu$, $\bar{x}_\nu \in \mathcal{X}$ and $\hat{A}\sigma^i_\nu(\bar{x}_\nu) \le \hat{b}$ for all i. By summing over all i and dividing by N, we obtain
$$\hat{A}\left(\tfrac{1}{N}\textstyle\sum_{i=1}^N \sigma^i_\nu(\bar{x}_\nu)\right) \le \hat{b}. \tag{10}$$
However,
$$\begin{aligned}&\textstyle\sum_{i=1}^N \sigma^i_\nu(\bar{x}_\nu) = \sum_{i=1}^N\sum_{j=1}^N [T^\nu]_{ij}\bar{x}^j_\nu \\ &= \textstyle\sum_{j=1}^N\left(\sum_{i=1}^N [T^\nu]_{ij}\right)\bar{x}^j_\nu = \sum_{j=1}^N \bar{x}^j_\nu = N\sigma_\infty(\bar{x}_\nu),\end{aligned} \tag{11}$$
where the second to last equality holds because, by Assumption 1, T is doubly stochastic, hence so is T^ν. By substituting (11) into (10) we obtain $\hat{A}\sigma_\infty(\bar{x}_\nu) \le \hat{b}$, thus $\bar{x}_\nu \in \mathcal{Q}_\infty \forall\nu$.
ii) Since $\bar{x}_\nu$ is a NE for $\mathcal{G}_\nu$, for all $i \in \mathbb{Z}[1,N]$ and for all $x^i_\nu \in \mathcal{Q}^i_\nu(\bar{x}^{-i}_\nu)$ it holds
$$J^i(\bar{x}^i_\nu, \sum_j [T^\nu]_{ij}\bar{x}^j_\nu) \le J^i(x^i_\nu, [T^\nu]_{ii}x^i_\nu + \sum_{j\ne i}[T^\nu]_{ij}\bar{x}^j_\nu). \tag{12}$$

Recall that L is the Lipschitz constant of $J^i(z_1, z_2)$ as by standing assumption. Then
$$\begin{aligned}&J^i(\bar{x}^i_\nu, \sigma_\infty(\bar{x}_\nu)) = J^i(\bar{x}^i_\nu, \textstyle\sum_j \frac{1}{N}\bar{x}^j_\nu) \\ &\le J^i(\bar{x}^i_\nu, \textstyle\sum_j [T^\nu]_{ij}\bar{x}^j_\nu) + L\|\sum_j(1/N - [T^\nu]_{ij})\bar{x}^j_\nu\| \\ &\le J^i(\bar{x}^i_\nu, \textstyle\sum_j [T^\nu]_{ij}\bar{x}^j_\nu) + L\sum_j |1/N - [T^\nu]_{ij}|\|\bar{x}^j_\nu\| \\ &\le J^i(\bar{x}^i_\nu, \textstyle\sum_j [T^\nu]_{ij}\bar{x}^j_\nu) + L\sqrt{n}M_\infty d_\infty(\nu) \\ &\overset{(12)}{\le} J^i(x^i_\nu, [T^\nu]_{ii}x^i_\nu + \textstyle\sum_{j\ne i}[T^\nu]_{ij}\bar{x}^j_\nu) + L\sqrt{n}M_\infty d_\infty(\nu) \\ &\le J^i(x^i_\nu, \tfrac{1}{N}x^i_\nu + \textstyle\sum_{j\ne i}\frac{1}{N}\bar{x}^j_\nu) + 2L\sqrt{n}M_\infty d_\infty(\nu),\end{aligned} \tag{13}$$
for all $x^i_\nu \in \mathcal{Q}^i_\nu(\bar{x}^{-i}_\nu)$, where the last inequality follows through with a chain of inequalities similar to the previous

[4] If the quantities depending on cost functions, network, constraints do not scale with N, Theorem 2 reads $\nu_\varepsilon = \max\{c_1\log(\sqrt{N}/\varepsilon), c_2\log(\sqrt{N})\}$, where c_1 and c_2 are constants independent from N, ε.

ones. Condition (13) holds for all $x^i_\nu \in \mathcal{Q}^i_\nu(\bar{x}^{-i}_\nu)$, as we used the fact that $\bar{x}_\nu$ is a NE for $\mathcal{G}_\nu$. To show that (4) holds for all $x^i \in Q^i_\infty(\bar{x}^{-i}_\nu)$ note that, by definition of Hausdorff distance, for any any $x^i \in Q^i_\infty(\bar{x}^{-i}_\nu)$ there exists $\tilde{x}^i_\nu \in \mathcal{Q}^i_\nu(\bar{x}^{-i}_\nu)$ such that $\|x^i - \tilde{x}^i_\nu\| \leq d_H(\mathcal{Q}^i_\nu(\bar{x}^{-i}_\nu), \mathcal{Q}^i_\infty(\bar{x}^{-i}_\nu)) =: H^i(\nu)$. From (13) we know that since $\tilde{x}^i_\nu \in \mathcal{Q}^i_\nu(\bar{x}^{-i}_\nu)$ then

$$\begin{aligned} &J^i(\bar{x}^i_\nu, \sigma_\infty(\bar{x}_\nu)) \\ &\leq J^i(\tilde{x}^i_\nu, \tfrac{1}{N}\tilde{x}^i_\nu + \textstyle\sum_{j\neq i}\tfrac{1}{N}\bar{x}^j_\nu) + 2L\sqrt{n}M_\infty d_\infty(\nu) \\ &\leq J^i(x^i, \tfrac{1}{N}x^i + \textstyle\sum_{j\neq i}\tfrac{1}{N}\bar{x}^j_\nu) + \varepsilon^i(\nu). \end{aligned} \tag{14}$$

for $\varepsilon^i(\nu) := (L + \frac{L}{N})H^i(\nu) + 2L\sqrt{n}M_\infty d_\infty(\nu)$. Note that $H^i(\nu) \to 0$ by Lemma 3.2) and $d_\infty(\nu) \to 0$ by Lemma 1. Consequently $\varepsilon^i(\nu) \to 0$ and there exists ν_ε such that $\varepsilon^i(\nu) \leq \varepsilon$ for all $i \in \mathbb{N}[1, N]$ and all $\nu \geq \nu_\varepsilon$. Since (14) holds for all i and all $x^i \in Q^i_\infty(\bar{x}^{-i}_\nu)$ and given i), $\bar{x}_\nu$ is an ε-NE of $\mathcal{G}_\infty$, for all $\nu \geq \nu_\varepsilon$.

2) Under the additional Assumption 3 we get that for all $\nu \geq \nu^{\text{loc}}_H$, by Lemma 1 and Lemma 3.4),

$$\begin{aligned} \varepsilon^i(\nu) &:= (L + \tfrac{L}{N})H^i(\nu) + 2L\sqrt{n}M_\infty d_\infty(\nu) \\ &\leq 2L(2K^i_H + M_\infty\sqrt{Nn})r_{\text{step}}(T)^\nu \\ &\leq 2L(2\max_i(K^i_H) + M_\infty\sqrt{Nn})r_{\text{step}}(T)^\nu. \end{aligned}$$

Setting the right hand side equal to ε leads to ν_J. ∎

B. Convergence of Algorithm 1

Corollary 1. *Suppose that for the value of ν used in Algorithm 1 the operator F_ν in* (7b) *is strongly monotone with constant $\alpha_\nu > 0$ and Lipschitz with constant $L_\nu > 0$. Set*

$$\tau \leq \frac{-L_\nu^2 + \sqrt{L_\nu^4 + 4\alpha_\nu^2\|A_\nu\|^2}}{2\alpha_\nu\|A_\nu\|^2}. \tag{15}$$

Then for every initial condition $(x_{(0)}, \lambda_{(0)}) \in \mathcal{X} \times \mathbb{R}^{Nm}_{\geq 0}$ the sequence $(x_{(k)})^\infty_{k=1}$ produced by Algorithm 1 converges to the unique variational Nash equilibrium of $\mathcal{G}_\nu$ which, under Assumptions 1 and 2, is an ε-NE of $\mathcal{G}_\infty$ for any $\nu \geq \nu_\varepsilon$.

Proof: Let us define $x_{(k)} := [x^i_{(k)}]^N_{i=1}, \lambda_{(k)} := [\lambda^i_{(k)}]^N_{i=1}, \sigma_{\nu,(k)} := [\sigma^i_{\nu,(k)}]^N_{i=1}, \mu_{\nu,(k)} := [\mu^i_{\nu,(k)}]^N_{i=1}$. Then the communication steps are equivalent to

$$\sigma_{\nu,(k)} \leftarrow (T^\nu \otimes I_n)\, x_{(k)}, \quad \mu_{\nu,(k)} \leftarrow (T^\nu \otimes I_m)^\top \lambda_{(k)}.$$

Consequently, the update steps can be rewritten as

$$\begin{aligned} &x^i_{(k+1)} \leftarrow \Pi_{\mathcal{X}^i}[x^i_{(k)} - \tau(F^i_{\nu,(k)} + \hat{A}^\top \textstyle\sum_{j=1}^N [T^\nu]_{ji}\lambda^j_{(k)})], \\ &\lambda^i_{(k+1)} \leftarrow \Pi_{\mathbb{R}^m_{\geq 0}}[\lambda^i_{(k)} - \tau(\hat{b} - \hat{A}\textstyle\sum_{j=1}^N [T^\nu]_{ij}(2x^j_{(k+1)} - x^j_{(k)}))] \end{aligned}$$

for all $i \in \mathbb{Z}[1, N]$ or, in compact form,

$$\begin{aligned} x_{(k+1)} &\leftarrow \Pi_{\mathcal{X}}[x_{(k)} - \tau\left(F_\nu(x_{(k)}) + A_\nu{}^\top\lambda_{(k)}\right)], \\ \lambda_{(k+1)} &\leftarrow \Pi_{\mathbb{R}^{Nm}_{\geq 0}}[\lambda_{(k)} - \tau(b - 2A_\nu x_{(k+1)} + A_\nu x_{(k)})]. \end{aligned} \tag{16}$$

The update (16) coincides with one iteration of the asymmetric projection algorithm given e.g. in [24, Algorithm 2] or in [10, Algorithm 4.1] (for $\bar{L} = 0$) applied to VI($\mathcal{Q}_\nu, F_\nu$). Then [24, Theorem 3] shows that, by choosing τ as in (15), which implies $\tau < 1/\|A_\nu\|$, Algorithm 1 is guaranteed to converge to the unique solution of VI($\mathcal{Q}_\nu, F_\nu$). The conclusion then follows by Theorem 2. ∎

Remark 1. *To implement Algorithm 1 the agents need to agree on the values of ν and τ to use. Under Assumption 3, we can provide explicit ways of computing these two quantities. Regarding ν, the quantities ν_{SMON}, $\nu^{\text{loc}}_H, \nu_J$ are given in Lemmas 2 and 3, Theorem 2. Some recent works, such as [28], tackle the problem of computing $r_{\text{step}}(T)$ in a distributed way for symmetric communication networks. For asymmetric communication networks instead, one can use the bound $r_{\text{step}}(T) = \|T - (1/N)\mathbb{1}_N\mathbb{1}_N^\top\|_2 \leq \|T - (1/N)\mathbb{1}_N\mathbb{1}_N^\top\|_1\|T - (1/N)\mathbb{1}_N\mathbb{1}_N^\top\|_\infty$ and the fact that one and infinity norms can be distributedly computed as maxima over all the nodes of local quantities. N can be computed with classical distributed algorithms, while the constants M_2, M_∞ can be computed in a distributed way if each agent is willing to disclose $\max_{x^i\in\mathcal{X}^i}\|x^i\|_2$ and $\max_{x^i\in\mathcal{X}^i}\|x^i\|_\infty$. Bounds on α_∞ relying only on local information have been studied in the literature in the case of specific cost functions J^i, as in [24, Lemma 3]. Finally, the remaining constants L, L', L'', M'' are (or can be upper bounded by) the maximum of local quantities, see e.g. footnote 1, and can thus be computed in a distributed fashion. Regarding τ, the bound in* (15) *of Corollary 1 can be used. To compute the quantities in there, note that $\|A_\nu\| = \|T^\nu\|\|\hat{A}\| = \|\hat{A}\|$ (as T is doubly stochastic); an upper bound to α_ν is given in* (25) *by using $L'', M_2, M'', r_{\text{step}}(T)$; and L_ν can be upper bounded by using L' and N.*

C. Relations with the literature on distributed convergence in AAG without coupling constraints

Distributed algorithms for AAGs without coupling constraints have been derived in the literature e.g. in [11], [12], [13], [14]. We highlight here how the steps of Algorithm 1 and the proofs greatly simplify in the absence of coupling constraints.

Regarding Theorem 1, in the absence of coupling constraints the VIs of $\mathcal{G}_\nu$ and $\mathcal{G}_\infty$ feature the same non-parametric set $\mathcal{X}$. Convergence of $\bar{x}_\nu$ to $\bar{x}_\infty$ can thus be proven by using standard sensitivity analysis results for VIs (see Appendix A).

Regarding Theorem 2, in the absence of coupling constraints the fact that $\bar{x}_\nu$ is an ε-NE of $\mathcal{G}_\infty$ is a trivial consequence of (9) and of the fact that the cost functions are Lipschitz. The difficulty when introducing the coupling constraints are that i) the feasibility of $\bar{x}_\nu$ in $\mathcal{G}_\nu$ does not imply feasibility of $\bar{x}_\nu$ in $\mathcal{G}_\infty$ and ii) in the definition of NE, the set of feasible deviations $\mathcal{Q}^i_\nu(\bar{x}^{-i}_\nu)$ in $\mathcal{G}_\nu$ is different from the set of feasible deviations $\mathcal{Q}^i_\infty(\bar{x}^{-i}_\nu)$ in $\mathcal{G}_\infty$ (without coupling constraints both these sets would be $\mathcal{X}^i$). Hence to prove Theorem 2 one needs to show Hausdorff convergence of $\mathcal{Q}^i_\nu(\bar{x}^{-i}_\nu)$ to $\mathcal{Q}^i_\infty(\bar{x}^{-i}_\nu)$ as $\nu \to \infty$, as done in Lemma 3.

Regarding Algorithm 1 and Corollary 1, in the absence of coupling constraints since the constraint set $\mathcal{X}$ can be decoupled among the agents, the standard projection algorithm [9, Algorithm 12.1.1] is distributed and it is guaranteed to converge to a solution to VI($\mathcal{X}, F_\nu$), because F_ν is strongly monotone. Hence, one can run Algorithm 1 with only the primal steps and simplified strategy update $x^i_{(k+1)} \leftarrow \Pi_{\mathcal{X}^i}[x^i_{(k)} - \tau F^i_{\nu,(k)}]$.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TCNS.2019.2944300

V. Application: Cournot game with transportation costs

A. Game definition

Consider a single-commodity Cournot game with N firms and V markets, which correspond to V physical locations. Firm $i \in \mathbb{Z}[1,N]$ chooses to sell $y^i_v \in \mathbb{R}_{\geq 0}$ amount of commodity at each market $v \in \mathbb{Z}[1,V]$. Each firm i produces its commodity at a given location $\ell_i \in \mathbb{Z}[1,V]$ and then ships its commodity to the different markets over a transportation network, where the V nodes represent market locations and a directed edge connecting two nodes represents a road connecting two markets. We characterize the network by its incidence matrix $B \in \{0,1,-1\}^{V\times E}$, where E is the number of edges and $B_{v,e} = -1$ if edge e leaves node v, $B_{v,e} = 1$ if edge e enters node v and $B_{v,e} = 0$ otherwise. Denote by $r^i \in \mathbb{R}_{\geq 0}$ the total amount of commodity produced and sold by firm i (i.e., $r^i = \sum_{v\in V} y^i_v$) and by $t^i_e \in \mathbb{R}_{\geq 0}$ the amount of commodity transported by firm i over edge e, with $t^i = [t^i_e]_{e=1}^E$. Define the strategy vector of firm i as $x^i := [t^i; r^i] \in \mathbb{R}^{E+1}_{\geq 0}$, which uniquely determines $y^i := [y^i_v]_{v=1}^V$, due to the balance equation

$$y^i = Bt^i + e_{\ell_i} r^i = H^i x^i,$$

with $H^i := [B, e_{\ell_i}] \in \mathbb{R}^{V\times(E+1)}$ and e_j the j^{th} canonical vector.

1) Cost function: We assume that at each market the commodity is sold at a price that depends on the total commodity sold by the N firms. We allow for inter-market effects and use the inverse demand function $p : \mathbb{R}^V_{\geq 0} \to \mathbb{R}^V_{\geq 0}$ that maps the normalized vector $\sigma_\infty(x) = \frac{1}{N}\sum_{j=1}^N y^j = \frac{1}{N}\sum_{j=1}^N H^j x^j$ to the vector of prices of each market $p(\sigma_\infty(x)) := [p_v(\sigma_\infty(x))]_{v=1}^V$. Then the revenue of firm i is $p(\sigma_\infty(x))^\top y^i$. Moreover, for firm i transporting t^i_e commodity over an edge comes with cost

$$c^i_e(t^i_e) := \beta^i_e t^i_e - \gamma^i_e(t^i_e), \tag{17}$$

where, for all i, γ^i_e is a strongly concave, increasing function with maximum derivative smaller than β^i_e. The transportation cost in (17) can be thought of as the sum of two terms: the first is a cost proportional to the amount shipped, the second term is a discount that increases with the amount shipped.

The production cost function of firm i has a similar form

$$a^i(r^i) := \beta^i_a r^i - \gamma^i_a(r^i), \tag{18}$$

where γ^i_a is a strongly concave, increasing function with maximum derivative smaller than β^i_a. The functions (17) and (18) are strongly convex, as in [9, Section 1.4.3]. To sum up

$$J^i(x^i,\sigma_\infty(x)) := \underbrace{a^i(r^i)}_{\text{production cost}} + \underbrace{\textstyle\sum_{e=1}^E c^i_e(t^i_e)}_{\text{transportation cost}} - \underbrace{p(\sigma_\infty(x))^\top y^i}_{\text{revenue}}.$$

2) Constraints: The strategy of firm i must satisfy the individual constraints

$$\mathcal{X}^i := \{x^i \in \mathbb{R}^{E+1}_{\geq 0} | x^i \leq \bar r^i \cdot \mathbb{1}_{E+1}, y^i = H^i x^i \geq 0\}, \tag{19}$$

where $\bar r^i$ is the production capacity of firm i. Note that (19) implies $t^i_e \leq \bar r^i \ \forall e$, guaranteeing boundedness of $\mathcal{X}^i$. This constraint is without loss of generality, as the transportation costs c^i_e are increasing. We also assume that each market v is composed by retailers whose storage capacity imposes an upper bound K_v on the total commodity that can be sold at market v, leading to the coupling constraints $\sigma_\infty(x) \leq K := [K_v]_{v=1}^V$.

3) Communication network: We assume that the firms can communicate with each other according to a sparse communication network T satisfying Assumption 1. This network can model spatial proximity of firms, or the fact that they may want to share their strategies only with firms they trust.

B. Theoretical guarantees

The cost, constraints and network introduced above give rise to a game as in (3), with the only difference that the aggregate $\sigma_\infty(x)$ depends on $y^i = H^i x^i$ instead of x^i directly. Defining a new game with strategies $\tilde x^i = H^i x^i$, so that the aggregate depends only on the $\tilde x^i$, is not possible as the cost $J^i(x^i,\sigma_\infty(x))$ cannot be expressed in terms of the $\tilde x^i$s only, unless H^i is full column rank for all i. Our theoretical results can nonetheless be easily extended to cover such case.

1) Extension: Set $H_{\text{blkd}} := \text{blkdiag}(H^1,\dots,H^N) \in \mathbb{R}^{NV\times N(E+1)}$, the quantities in (7) (8), relative to $\mathcal{G}_\infty$, $\mathcal{G}_\nu$ are

$$\begin{aligned}
F_\infty(x) &:= [\nabla_{x^i} J^i(x^i,\sigma_\infty(x))]_{i=1}^N,\\
&= [\nabla_{z_1} J^i(x^i,\sigma_\infty(x)) + \tfrac{1}{N} H_i^\top \nabla_{z_2} J^i(x^i,\sigma_\infty(x))]_{i=1}^N,\\
\mathcal{Q}_\infty &:= \{x \in \mathcal{X}^1\times\cdots\times\mathcal{X}^N | A_\infty x \leq b\},\\
A_\infty &:= \left(\frac{1}{N}\mathbb{1}_N\mathbb{1}_N^\top \otimes \hat A\right) H_{\text{blkd}}, \quad b := \mathbb{1}_N \otimes \hat b,\\
F_\nu(x) &:= [\nabla_{x^i} J^i(x^i,\sigma^i_\nu(x))]_{i=1}^N,\\
&= [\nabla_{z_1} J^i(x^i,\sigma^i_\nu(x)) + [T^\nu]_{ii} H_i^\top \nabla_{z_2} J^i(x^i,\sigma^i_\nu(x))]_{i=1}^N,\\
\mathcal{Q}_\nu &:= \{x \in \mathcal{X}^1\times\cdots\times\mathcal{X}^N | A_\nu x \leq b\},\\
A_\nu &:= (T^\nu \otimes \hat A) H_{\text{blkd}}.
\end{aligned}$$

2) Verify the assumptions: Assumption 1 holds by problem statement. To guarantee that Assumption 2 holds we make the following assumption, whose sufficiency is proven in Lemma 4.

Assumption 4 (Cournot-game regularity conditions)**.** *The cost $J^i(z_1,z_2)$ is twice continuously differentiable for all i, and the inverse demand function p satisfies one of the following conditions.*
1) p is affine, i.e., $p(\sigma_\infty(x)) = -D\sigma_\infty(x) + d$, for some $D \in \mathbb{R}^{V\times V}$, $d \in \mathbb{R}^V$ and $D \succeq 0$.
2) p_v depends only on the commodity sold at v, i.e., $p(\sigma_\infty(x)) =: [p_v([\sigma_\infty(x)]_v)]_{v=1}^V$. For each v, p_v is twice continuously differentiable, strictly decreasing and satisfies

$$\min_{\substack{v\in\{1,\dots,V\}\\ z\in[0,\tilde r]}} \left(-p'_v(z) + \tfrac{\tilde r}{8} p''_v(z)\right) > 0, \quad \tilde r := \max_{i\in\mathbb{Z}[1,N]} \bar r^i. \tag{20}$$

□

Lemma 4. *If Assumption 4 holds then the Cournot game in Section V-A satisfies Assumption 2.* □

We report the proof in Appendix B.

Remark 2. *If the function p is as in Assumption 4.1) and $D = D^\top$, then $\mathcal{G}_\infty$ is a potential game. In other words, there exists a function $f : \mathcal{Q} \to \mathbb{R}$ such that $\nabla_x f(x) = F_\infty(x)$ and VI($\mathcal{Q}_\infty$,F_∞) is equivalent to* $\text{argmin}_{x\in\mathcal{Q}_\infty} f(x)$*, as described* *in [9, Section 1.3.1]. Then a NE can be found by solving the optimization program* $\mathrm{argmin}_{x\in\mathcal{Q}_\infty} f(x)$*. Assumption 4.1) is verified for instance in [7] or in the cost (5)-(6) in [12], with* $b_h = 1$*, for any* a_h, c_h*. Assumption 4.2) is satisfied if* p_v *is convex and strictly decreasing* $\forall v$*.*

Regarding the Standing Assumption, Assumption 2 implies that J^i is continuously differentiable in its arguments and that $\nabla_x[\nabla_{x^i} J^i(x^i, \sigma_\infty(x))]_{i=1}^N \succ \alpha I_{N(E+1)}$ by Proposition 2, which in turn implies convexity of J^i in x^i for all fixed x^{-i}. The sets $\mathcal{X}^i$ are trivially convex, compact and non-empty.

Finally, the next lemma (proven in [29]) shows that for the network, number of rounds of communication ν and price functions used in the next subsection, F_ν is strongly monotone.

Lemma 5. *Under Assumption 4.1), if* T *is the adjacency matrix of an undirected network, so that* $T = T^\top$*, then the operator* F_ν *is strongly monotone for any* ν *even.* □

C. Numerical analysis

We consider the transportation network illustrated in Figure 2 which consists of $V = 43$ markets and $E = 51$ (bidirectional) edges connecting them. The network is taken from [30], which provides the Cartesian coordinates of the vertexes. We consider 5 firms that differ only for their locations $\ell^i \in \{37, 20, 11, 6, 35\}$, as indicated in Figure 2. Each firm has a production capacity of $\bar{r}^i = 10$, while we consider a capacity of 1.5 for each market (i.e. $K = 1.5/5$). The production and transportation costs for firm i are $a^i(r^i) = a(r^i) = 2\left[r^i - \left(1 - \frac{1}{1+r^i}\right)\right]$, $c_e^i(t_e^i) = c_e(t_e^i) = \rho_e\left(t_e^i - \left(1 - \frac{1}{1+t_e^i}\right)\right)$, for all e where $\rho_e \in]0,1]$ is the normalized[5] length of road e. The inverse demand function p is affine, i.e. $p(\sigma) = 10 \cdot \mathbb{1}_{43} - D\sigma$ and it encodes intra-market competition via the matrix D whose component in position (h,k) is $[D]_{h,k} = 1$ if $h = k$, $[D]_{h,k} = 0.3(1-\rho_e)$, if there is a road $e = (h,k)$ between markets h and k, while $[D]_{h,k} = 0$ otherwise. In words, the price p_v at market v not only decreases when more commodity is sold at v, but also when more commodity is sold at the neighboring markets, with physically close markets being more influential. We verified numerically that $D \succeq 0$. We use the communication matrix T that corresponds to a symmetric ring, i.e., $T_{ij} = 0.5$ if $|i-j| = 1$ or $|i-j| = 4$, $T_{ij} = 0$ otherwise. Note that T satisfies Assumption 1. We run Algorithm 1 with $\tau = 0.05$, initial conditions all equal to zero and different values of ν.[6] We use $\max\{\|x_{(k)} - x_{(k-1)}\|_\infty, \|\lambda_{(k)} - \lambda_{(k-1)}\|_\infty\} < 10^{-4}$ as stopping criterion. We consider even values of ν between 2 and 20. For each ν we run Algorithm 1 and find the variational NE of $\mathcal{G}_\nu$, which is an ε_ν-NE for $\mathcal{G}_\infty$, as by Theorem 2. After convergence, ε_ν can be computed according to Definition 1. A more descriptive quantity is the relative maximum improvement $\hat{\varepsilon}_\nu$, defined as[7]

$$\hat{\varepsilon}_\nu := \max_{i, x^i \in \mathcal{Q}_\infty^i(\bar{x}_\nu^{-i})} \frac{J^i(\bar{x}_\nu^i, \sigma_\infty(\bar{x}_\nu)) - J^i(x^i, \frac{1}{N}x^i + \sum_{j\neq i}\frac{1}{N}\bar{x}_\nu^j)}{J^i(\bar{x}_\nu^i, \sigma_\infty(\bar{x}_\nu))}. \tag{21}$$

Figure 1 reports the value of $\hat{\varepsilon}_\nu$ as function of ν, thus numerically verifying Theorem 2. It also contains the value of $\|\bar{x}_\nu - \bar{x}_\infty\|_2$ as function of ν, thus numerically verifying Theorem 1. In Figure 2 we illustrate the variational NE of $\mathcal{G}_\infty$ obtained by setting $\nu = 1$ and $T = \frac{1}{N}\mathbb{1}_N\mathbb{1}_N^\top$. Each firm is the only seller at its production location, and more in general firms tend to sell close to their production location, as expected.

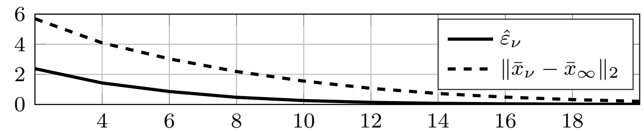

Fig. 1: Relative maximum improvement and two-norm distance against ν.

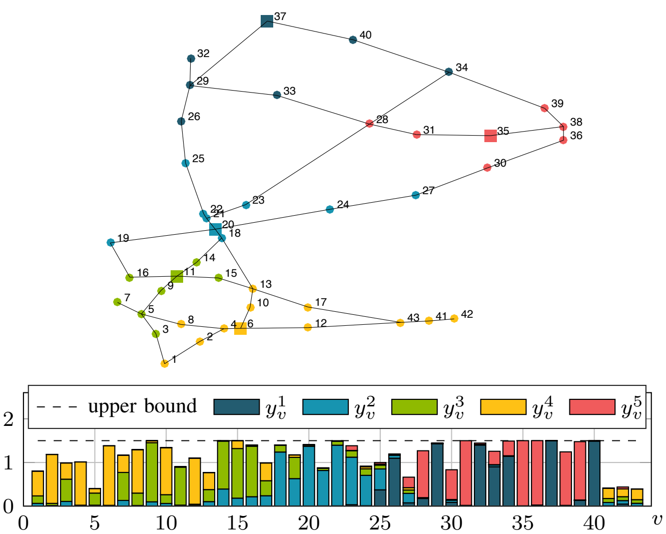

Fig. 2: Variational Nash of $\mathcal{G}_\infty$, computed by Algorithm 1 for $\nu = 1$ and $T = \frac{1}{N}\mathbb{1}_N\mathbb{1}_N^\top$. Each of the 5 firms is represented by a different color. The bottom plot reports y_v^i for each firm i and market v (from 1 to 43). Note that the coupling constraint $\sum_i y_v^i \leq 1.5$ is met in all markets v. A geographic illustration of the market is represented in the graph on top. Therein each market takes the color of the firm that sells the most commodity there. The five production locations are denoted by squares. Note that each firm has majority in the markets closer to its location (as transporting there is cheaper).

VI. Immediate generalizations and future work

1) Weighted average: The above results can be immediately generalized to aggregative games that depend on a weighted average $\sigma_\infty(x) = \sum_{i=1}^N w_i x^i$, for some $w_i > 0$ instead of the average $\sigma_\infty(x) = \frac{1}{N}\sum_{i=1}^N x^i$ used above. We can impose $\sum_i w_i = 1$ without loss of generality. Then Assumption 1 should require T to be primitive with $w > 0$ as left eigenvector relative to the eigenvalue 1 (normalized such that $w^\top \mathbb{1}_N = 1$).

[5] Defined as the absolute length divided by the maximum road length.

[6] The values in (15) can be shown to be $\alpha_\nu = (2\rho_{\min})/(1+\tilde{r})^3 = 7 \cdot 10^{-5}, \forall \nu$; $L_\nu = \lambda_{\max}(H_{\text{blkd}}^\top[(I_N \otimes D)(T^\nu \otimes I_E) + \mathrm{diag}(T^\nu) \otimes D^\top]H_{\text{blkd}}) \leq 12.89$; and $\|A_\nu\| = 1, \forall \nu$; then (15) reads $\tau \leq 4 \cdot 10^{-7}$. This is a conservative bound, we verified by simulations that the algorithm converges also for $\tau = 0.05$. Lemma 2 gives theoretical guarantees of convergence for $\nu \geq \nu_{\text{SMON}} = 67$. Note that an extension of Lemma 2 is needed due to the presence of H^i; this results in $\nu_{\text{SMON}} = \log(\alpha_\infty/((1+h)^2(L''M_2) + 2(h+h^2)M''))/\log(r_{\text{step}}(T))$, with $h := \|H_{\text{blkd}}\|_2$. We verified by simulations that F_ν is strongly monotone for all $\nu \geq 11$.

[7] Note that for any fixed $\bar{x}_\nu$, $\hat{\varepsilon}_\nu$ in (21) can be computed by solving the N optimization problems $\{\min_{x^i \in \mathcal{Q}_\infty^i(\bar{x}_\nu^{-i})} J^i(x^i, \frac{1}{N}x^i + \sum_{j\neq i}\frac{1}{N}\bar{x}_\nu^j)\}_{i=1}^N$.

2) Local strategy sets of different dimensions: Here $\mathcal{X}^i \subset \mathbb{R}^n$ for all i, but the results hold for strategy sets of different dimension, i.e, $\mathcal{X}^i \subset \mathbb{R}^{n_i}$, as in [10], with $\sigma^i(x) = \frac{1}{N}\sum_{j=1}^N [H^j x^j + h^j] \in \mathbb{R}^n$, for some $H^j \in \mathbb{R}^{n\times n_j}, h^j \in \mathbb{R}^n$.

3) Wardrop instead of Nash equilibrium: The results stated for Nash equilibrium extend to Wardrop equilibrium [24, Definition 2], called traffic equilibrium in transportation or competitive equilibrium in economics. If in the primal update of Algorithm 1 we use $F^i_{\nu,(k)} \leftarrow \nabla_{z_1} J^i(x^i_{(k)}, \sigma^i_{\nu,(k)})$, then Algorithm 1 converges to a Wardrop equilibrium.

Future directions: An extension to time-varying networks, as well as asynchronous updates would be valuable for applications where synchronization might be problematic. It would also be interesting to perform a comparison in terms of strategy and communication updates with the algorithms presented in the recent works (for generic games) [19], [21]. Finally, it would be valuable to understand if convergence to the equilibrium $\bar{x}_\infty$ can be achieved via an algorithm where the number of communications changes with the iterations.

Appendix

A. Convergence result for parametric variational inequalities

The notation in Appendix A is disjoint from the main paper.

1) Literature: We survey some literature results on convergence of the solution $\bar{x}_\theta$ of VI$(\mathcal{Q}_\theta, F_\theta)$ to the solution $\bar{x}_{\hat{\theta}}$ of VI$(\mathcal{Q}_{\hat{\theta}}, F_{\hat{\theta}})$ under the assumption that $F_{\hat{\theta}}$ is strongly monotone and F_θ converges uniformly to $F_{\hat{\theta}}$ as $\theta \to \hat{\theta}$. Such results can be divided in three classes, based on assumptions on the sets.

The first class of results focuses on fixed sets and studies convergence of the solution $\bar{x}_\theta$ of VI$(\mathcal{Q}, F_\theta)$ to the solution $\bar{x}_{\hat{\theta}}$ of VI$(\mathcal{Q}, F_{\hat{\theta}})$. If $\mathcal{Q}$ is closed and convex, F_θ is Lipschitz in θ uniformly in x and $F_{\hat{\theta}}$ is strongly monotone, then the solution is Lipschitz continuous [31, Theorem 1.14], [9, Section 5.3]. Strong monotonicity of $F_{\hat{\theta}}$ can be relaxed if $\mathcal{Q}$ is a polytope [32]. The second class of results [33], [34] focuses on parametric sets that can be described as $\mathcal{Q}_\theta := \{x \in \mathbb{R}^n \,|\, g(x,\theta) \le 0\}$ for a suitable parametric function $g(x,\theta)$. If $g(x,\theta)$ converges uniformly in x to $g(x,\hat{\theta})$ as $\theta \to \hat{\theta}$ and at $\bar{x}_{\hat{\theta}}$ the Linear Independence Constraint Qualification holds, the parametric solution $\bar{x}_\theta$ is locally Lipschitz continuous around $\hat{\theta}$. Such results have been applied to games, e.g. in [35]. The third class of results [34], [27] is the most general and only assumes that $\mathcal{Q}_\theta$ converges to $\mathcal{Q}_{\hat{\theta}}$ according to the Kuratowski set convergence definition. In this case one can prove continuity of $\bar{x}_\theta$ around $\hat{\theta}$. We are not aware of results proving Lipschitz continuity in this case.

In the following we show convergence (and derive convergence rates) without assuming any constraint qualification for VIs whose parametric sets have the special structure

$$\begin{aligned}\mathcal{Q}_\nu &:= \{x \in \mathcal{X} \subset \mathbb{R}^n | A_\nu x \le b_\nu\}, \\ \mathcal{Q}_\infty &:= \{x \in \mathcal{X} \subset \mathbb{R}^n | A_\infty x \le b_\infty\},\end{aligned} \tag{22}$$

where $A_\nu, A_\infty \in \mathbb{R}^{m\times n}$, $b_\nu, b_\infty \in \mathbb{R}^m$, $A_\nu \to A_\infty, b_\nu \to b_\infty$. Note that we focus on a discrete parameter $\nu \in \mathbb{N}$ that tends to infinity, however similar arguments can be used for a continuous parameter θ tending to $\hat{\theta}$. Our results depend on $\mathcal{X}$: 1) If $\mathcal{X}$ is convex and compact we show that $\mathcal{Q}_\nu$ converges to $\mathcal{Q}_\infty$ in Hausdorff (and thus Kuratowski) norm. Convergence of the VI solution then follows from [34], [27]; 2) If $\mathcal{X}$ is a polytope we additionally bound the distance between the solution to VI$(\mathcal{Q}_\nu, F_\nu)$ and VI$(\mathcal{Q}_\infty, F_\infty)$.

Our result in 2) can be seen as an extension of [32], [36] on parametric variational inequalities over polyhedral sets, since [32] studies VIs where the polyhedron is fixed and [36] provides convergence rates when only the right-hand side b_ν of the affine constraint defining the polyhedron is parametric. Instead, we allow the parameter to appear also in the matrix A_ν (and quantify its effect on the convergence rate).

2) Preliminaries on Hausdorff distance: We report here some preliminary facts, which follow immediately from previous literature results.

Lemma A.1. *For any matrix $H \in \mathbb{R}^{m\times n}$ there exists a constant $\Delta(H) > 0$, such that for any $x_1 \in \mathbb{R}^n$ and $h_1, h_2 \in \mathbb{R}^m$ such that $\{x \in \mathbb{R}^n \mid Hx \le h_2\} \ne \emptyset$ and $Hx_1 \le h_1$ there exists $x_2 \in \mathbb{R}^n$ such that $Hx_2 \le h_2$ and*

$$\|x_1 - x_2\|_2 \le \Delta(H)\|h_1 - h_2\|_\infty \qquad \square$$

This lemma follows immediately from the fact that the Hausdorff distance between polytopes of the form $\{x \mid Hx \le h\}$ is Lipschitz continuous in h, see e.g. [37]. Different refinements for the constant $\Delta(H)$ have been provided in different works, see e.g. [38]. Here we use the formula derived in [26, Theorem 0.1] and reported in Definition 5. The previous lemma leads to the following convergence result for the sets $\mathcal{Q}_\nu, \mathcal{Q}_\infty$ when $\mathcal{X}$ is a polytope.

Lemma A.2. *Suppose that the set $\mathcal{X}$ in (22) is a compact polytope $\mathcal{X} := \{x \in \mathbb{R}^n | Gx \le g\}$. Let $c(\nu) := \max_{x\in\mathcal{X}}\{\|(A_\infty - A_\nu)x + b_\nu - b_\infty\|_\infty\}$. The following holds.*

1) If $\mathcal{Q}_\nu, \mathcal{Q}_\infty$ are non-empty, then

$$d_H(\mathcal{Q}_\nu, \mathcal{Q}_\infty) \le \max\left\{\Delta(\left[\begin{smallmatrix}A_\infty\\ G\end{smallmatrix}\right]), \Delta(\left[\begin{smallmatrix}A_\nu\\ G\end{smallmatrix}\right])\right\} \cdot c(\nu);$$

2) If additionally, $\{x \in \mathcal{X} \mid A_\infty x \le b_\infty - c(\nu)\mathbb{1}_m\}$ is non-empty then

$$d_H(\mathcal{Q}_\nu, \mathcal{Q}_\infty) \le \Delta(\left[\begin{smallmatrix}A_\infty\\ G\end{smallmatrix}\right]) \cdot c(\nu).$$

Proof: Take any point $x_1 \in \mathcal{Q}_\nu$. By definition of $\mathcal{Q}_\nu$ it holds $\left[\begin{smallmatrix}A_\nu\\ G\end{smallmatrix}\right] x_1 \le \left[\begin{smallmatrix}b_\nu\\ g\end{smallmatrix}\right]$ equivalently $H_1 x_1 := \left[\begin{smallmatrix}A_\infty\\ G\end{smallmatrix}\right] x_1 \le \left[\begin{smallmatrix}b_\nu + (A_\infty - A_\nu)x_1\\ g\end{smallmatrix}\right] =: h_1$. If we set $h_2 = \left[\begin{smallmatrix}b_\infty\\ g\end{smallmatrix}\right]$, by Lemma A.1 we know that there exists $x_2 \in \mathcal{Q}_\infty$ (non-empty by assumption) such that $\|x_1 - x_2\|_\infty \le \Delta(\left[\begin{smallmatrix}A_\infty\\ G\end{smallmatrix}\right])\|(A_\infty - A_\nu)x_1 + b_\nu - b_\infty\|_\infty \le \Delta(\left[\begin{smallmatrix}A_\infty\\ G\end{smallmatrix}\right])c(\nu)$. Similarly, for any point $x_1 \in \mathcal{Q}_\infty$ by Lemma A.1 there exists $x_2 \in \mathcal{Q}_\nu$ such that $\|x_1 - x_2\|_\infty \le \Delta(\left[\begin{smallmatrix}A_\nu\\ G\end{smallmatrix}\right])c(\nu)$, thus concluding the first part of the proof. If the set $\{x \in \mathcal{X} \mid A_\infty x \le b_\infty - c(\nu)\mathbb{1}_m\}$ is non-empty then an alternative way to associate a point $x_2 \in \mathcal{Q}_\nu$ to any $x_1 \in \mathcal{Q}_\infty$ is to set $h_2 = \left[\begin{smallmatrix}b_\infty\\ g\end{smallmatrix}\right] - \left[\begin{smallmatrix}c(\nu)1\\ 0\end{smallmatrix}\right]$. By definition of $\mathcal{Q}_\infty$ it holds $Hx_1 := \left[\begin{smallmatrix}A_\infty\\ G\end{smallmatrix}\right] x_1 \le \left[\begin{smallmatrix}b_\infty\\ g\end{smallmatrix}\right] =: h_1$. By Lemma A.1 there exists x_2 such that

$$\left[\begin{smallmatrix}A_\infty\\ G\end{smallmatrix}\right] x_2 \le \left[\begin{smallmatrix}b_\infty\\ g\end{smallmatrix}\right] - \left[\begin{smallmatrix}c(\nu)\mathbb{1}_m\\ 0\end{smallmatrix}\right] \tag{23}$$

and $\|x_1 - x_2\|_\infty \le \Delta(\left[\begin{smallmatrix}A_\infty\\ G\end{smallmatrix}\right])c(\nu)$. Let us now show that $x_2 \in \mathcal{Q}_\nu$. Inequality (23) implies $\left[\begin{smallmatrix}A_\nu\\ G\end{smallmatrix}\right] x_2 \le \left[\begin{smallmatrix}b_\nu\\ g\end{smallmatrix}\right] + \left[\begin{smallmatrix}(A_\nu - A_\infty)x_2 + (b_\infty - b_\nu)\\ 0\end{smallmatrix}\right] - \left[\begin{smallmatrix}c(\nu)\mathbb{1}_m\\ 0\end{smallmatrix}\right]$. From $Gx_2 \le g$ it follows

$x_2 \in \mathcal{X}$, hence $\|(A_\nu - A_\infty)x_2 + (b_\infty - b_\nu)\|_\infty \le c(\nu)$ and the previous inequality implies $\left[\begin{smallmatrix} A_\nu \\ G \end{smallmatrix}\right] x_2 \le \left[\begin{smallmatrix} b_\nu \\ g \end{smallmatrix}\right]$. Hence $x_2 \in \mathcal{Q}_\nu$. This leads to the second statement. ∎

If $\mathcal{X}$ is not a polytope, we prove asymptotic convergence.

Lemma A.3. *Suppose that the set $\mathcal{X} \subset \mathbb{R}^n$ is convex, compact and has non-empty interior. Moreover, $\lim_{\nu\to\infty} A_\nu = A_\infty$, $\lim_{\nu\to\infty} b_\nu = b_\infty$. Then $d_H(\mathcal{Q}_\nu, \mathcal{Q}_\infty) \to 0$ as $\nu \to \infty$.*

Proof: Since $\mathcal{X}$ is compact there exists a polytope $\mathcal{X}^c := \{x \in \mathbb{R}^n \mid Gx \le g\}$ such that $\mathcal{X} \subseteq \mathcal{X}^c$. We define $\mathcal{S}_\nu := \{x \in \mathcal{X}^c | A_\nu x \le b_\nu\}$, $\mathcal{S}_\infty := \{x \in \mathcal{X}^c | A_\infty x \le b_\infty\}$. Since $\Delta(\left[\begin{smallmatrix} A_\nu \\ G \end{smallmatrix}\right]) \to \Delta(\left[\begin{smallmatrix} A_\infty \\ G \end{smallmatrix}\right])$, implying that $\Delta(\left[\begin{smallmatrix} A_\nu \\ G \end{smallmatrix}\right])$ is bounded, and $c(\nu) \to 0$, Lemma A.2 implies that $d_H(\mathcal{S}_\nu, \mathcal{S}_\infty) \to 0$. The conclusion follows since $\mathcal{Q}_\nu = \mathcal{S}_\nu \cap \mathcal{X}$ and $\mathcal{Q}_\infty = \mathcal{S}_\infty \cap \mathcal{X}$ and $\mathcal{X}$ is convex, compact and has non-empty interior, see [27, Lemma 1.4] and [39, Theorem 3]. ∎

3) Convergence result: Convergence of the solution of VI$(\mathcal{Q}_\nu, F_\nu)$ to the solution of VI$(\mathcal{Q}_\infty, F_\infty)$ can be proven immediately by using Lemma A.3 and [27, Theorem A(b)] since Hausdorff implies Kuratowski convergence in compact spaces. In the next theorem we provide a refinement of [27, Theorem A(b)] that gives an explicit convergence rate.

Theorem A.1. *Suppose that F_∞ is strongly monotone with constant $\alpha_\infty > 0$ and that the sets $\mathcal{Q}_\nu$, $\mathcal{Q}_\infty$ are convex and compact. Moreover, assume that*
i) $\|F_\nu(x) - F_\infty(x)\| \le d_F(\nu)$ for all x and ν,
ii) there exists $F_{\max}$ s.t. $\|F_\infty(x)\| \le F_{\max}$ and $\|F_\nu(x)\| \le F_{\max}$ for all $x \in \mathcal{X}$ and all ν.
Let $\bar{x}_\infty$ be the unique solution of VI$(\mathcal{Q}_\infty, F_\infty)$ and $\bar{x}_\nu$ any solution of VI$(\mathcal{Q}_\nu, F_\nu)$, then

$$\|\bar{x}_\nu - \bar{x}_\infty\| \le \frac{d_F(\nu) + \sqrt{d_F(\nu)^2 + 8\alpha_\infty F_{\max} d_H(\mathcal{Q}_\nu, \mathcal{Q}_\infty)}}{2\alpha_\infty}.$$

In particular, if $d_H(\mathcal{Q}_\nu, \mathcal{Q}_\infty) \to 0$ and $d_F(\nu) \to 0$ as $\nu \to \infty$ then $\bar{x}_\nu \to \bar{x}_\infty$.

Proof: By definition of solution to a VI it holds

$$\begin{aligned} F_\infty(\bar{x}_\infty)^\top (x_\infty - \bar{x}_\infty) \ge 0,\ \forall x_\infty \in \mathcal{Q}_\infty \\ F_\nu(\bar{x}_\nu)^\top (x_\nu - \bar{x}_\nu) \ge 0,\ \forall x_\nu \in \mathcal{Q}_\nu \end{aligned} \tag{24}$$

By definition of Hausdorff distance there exists $\tilde{x}_\infty \in \mathcal{Q}_\infty$ and $\tilde{x}_\nu \in \mathcal{Q}_\nu$ such that $\|\tilde{x}_\infty - \bar{x}_\nu\| \le d_H(\mathcal{Q}_\nu, \mathcal{Q}_\infty) =: d_H$ and $\|\tilde{x}_\nu - \bar{x}_\infty\| \le d_H$. Plugging these values instead of the generic x_∞ and x_ν in (24) and adding and subtracting $\bar{x}_\nu, \bar{x}_\infty$ we get

$$\begin{aligned} F_\infty(\bar{x}_\infty)^\top (\tilde{x}_\infty - \bar{x}_\nu) \ge F_\infty(\bar{x}_\infty)^\top (\bar{x}_\infty - \bar{x}_\nu) \\ F_\nu(\bar{x}_\nu)^\top (\tilde{x}_\nu - \bar{x}_\infty) \ge -F_\nu(\bar{x}_\nu)^\top (\bar{x}_\infty - \bar{x}_\nu) \end{aligned}$$

adding the two inequalities leads to $T_1 \ge T_2$ with $T_1 := F_\infty(\bar{x}_\infty)^\top (\tilde{x}_\infty - \bar{x}_\nu) + F_\nu(\bar{x}_\nu)^\top (\tilde{x}_\nu - \bar{x}_\infty), T_2 := [F_\infty(\bar{x}_\infty) - F_\nu(\bar{x}_\nu)]^\top (\bar{x}_\infty - \bar{x}_\nu)$. $T_1 \le (\|F_\infty(\bar{x}_\infty)\| + \|F_\nu(\bar{x}_\nu)\|) d_H \le 2F_{\max} d_H T_2 := [F_\infty(\bar{x}_\infty) - F_\infty(\bar{x}_\nu) + F_\infty(\bar{x}_\nu) - F_\nu(\bar{x}_\nu)]^\top (\bar{x}_\infty - \bar{x}_\nu) \ge \alpha_\infty \|\bar{x}_\infty - \bar{x}_\nu\|^2 - d_F(\nu)\|\bar{x}_\infty - \bar{x}_\nu\|$. Combining we get $\alpha_\infty \|\bar{x}_\infty - \bar{x}_\nu\|^2 - d_F(\nu)\|\bar{x}_\infty - \bar{x}_\nu\| - 2F_{\max} d_H \le 0$, as desired. ∎

Remark 3. *This proof is based on [31, Theorem 1.14], where however only the VI operator is parametric. If $\mathcal{X}$ is a polytope, $d_H(\mathcal{Q}_\nu, \mathcal{Q}_\infty)$ can be bound as in Lemma A.2. Hence Theorem A.1 can be used to bound the VI solutions distance.*

B. Omitted proofs

Proof of Lemma 1

1) The fact that $\lim_{\nu\to\infty} T^\nu = \frac{1}{N}\mathbb{1}_N\mathbb{1}_N^\top$ is proven in [23, Theorem 2.13]. Convergence of A_ν to A_∞ follows immediately from the definitions (7a), (8a) and the properties of the Kronecker product. Note that

$$\begin{aligned} F_\infty(x) &= [\nabla_{z_1} J^i(x^i, \sigma_\infty(x)) + \tfrac{1}{N} \nabla_{z_2} J^i(x^i, \sigma_\infty(x))]_{i=1}^N, \\ F_\nu(x) &= [\nabla_{z_1} J^i(x^i, \sigma_\nu^i(x)) + [T^\nu]_{ii} \nabla_{z_2} J^i(x^i, \sigma_\nu^i(x))]_{i=1}^N. \end{aligned}$$

Uniform convergence of F_ν to F_∞ follows by continuity of $\nabla_{z_1} J^i(z_1, z_2)$ and $\nabla_{z_2} J^i(z_1, z_2)$ in z_1, z_2 for all i, by $[T^\nu]_{ii} \to \frac{1}{N}$, and by $\sigma_\nu^i(x) \to \sigma_\infty(x)$ uniformly in x.
2) Since $\|\nabla_{z_1} J^i(z_1, z_2)\|$ and $\|\nabla_{z_2} J^i(z_1, z_2)\|$ are continuous functions over the compact set $\mathcal{X}^i \times \operatorname{conv}\{\mathcal{X}^1, \dots, \mathcal{X}^N\}$, there exists $M' > 0$ such that $\|\nabla_{z_1} J^i(z_1, z_2)\| \le M'$ and $\|\nabla_{z_2} J^i(z_1, z_2)\| \le M'$, for all $(z_1, z_2) \in \mathcal{X}^i \times \operatorname{conv}\{\mathcal{X}^1, \dots, \mathcal{X}^N\}$. Note that $[T^\nu]_{ii} \le 1$ for all $i \in \mathbb{Z}[1, N]$ and for all $\nu > 0$, since T and thus T^ν are non-negative and doubly stochastic. Then $\|F_\nu(x)\|^2 = \sum_{i=1}^N \|\nabla_{z_1} J^i(x^i, \sigma_\nu^i(x)) + [T^\nu]_{ii} \nabla_{z_2} J^i(x^i, \sigma_\nu^i(x))\|^2 \le \sum_{i=1}^N (M'^2 + M'^2 + 2M'^2) = 4NM'^2 = F_{\max}^2$, for all $x \in \mathcal{X}$. The bound $\|F_\infty(x)\| \le F_{\max}$ can be proven in the same way.
3) The result on $d_2(\nu)$ follows from [23, Lemma 10.3] and it implies the one on $d_\infty(\nu)$, because $\|A\|_\infty \le \sqrt{N}\|A\|_2$ for any matrix $A \in \mathbb{R}^{N\times N}$.
4) Under Assumption 2, $\nabla_{z_a} J^i(z_1, z_2)$ is Lipschitz continuous for $a \in \{1, 2\}$ with constant L'. Note that $F_\infty(x) - F_\nu(x) = v_1 + v_2 + v_3$ with $v_1 := [\nabla_{z_1} J^i(x^i, \sigma_\infty(x)) - \nabla_{z_1} J^i(x^i, \sigma_\nu^i(x))]_{i=1}^N$, $v_2 := \frac{1}{N}[\nabla_{z_2} J^i(x^i, \sigma_\infty(x)) - \nabla_{z_2} J^i(x^i, \sigma_\nu^i(x))]_{i=1}^N$ and $v_3 := [\frac{1}{N}\nabla_{z_2} J^i(x^i, \sigma_\nu^i(x)) - [T^\nu]_{ii} \nabla_{z_2} J^i(x^i, \sigma_\nu^i(x))]_{i=1}^N$. Moreover, $\|v_1\|_2^2 = \sum_{i=1}^N \|\nabla_{z_1} J^i(x^i, \sigma_\infty(x)) - \nabla_{z_1} J^i(x^i, \sigma_\nu^i(x))\|^2 \le \sum_{i=1}^N (L')^2 \|\sigma_\infty(x) - \sigma_\nu^i(x)\|^2 = (L')^2 \|[(\frac{1}{N}\mathbb{1}_N\mathbb{1}_N^\top - T^\nu) \otimes I_n]x\|^2 \le (L')^2 d_2(\nu)^2 M_2^2$. Similarly, $\|v_2\|_2 \le \frac{1}{N} L' d_2(\nu) M_2$. Finally, $\|v_3\|_2^2 = \sum_{i=1}^N \|(\frac{1}{N} - [T^\nu]_{ii}) \nabla_{z_2} J^i(x^i, \sigma_\nu^i(x))\|^2 \le (M')^2 \sum_{i=1}^N (\frac{1}{N} - [T^\nu]_{ii})^2 \le (M')^2 \|\frac{1}{N}\mathbb{1}_N\mathbb{1}_N^\top - T^\nu\|_F^2 \le (M')^2 N d_2(\nu)^2$, where for any matrix $A \in \mathbb{R}^{N\times N}$, $\|A\|_F$ denotes its Frobenius norm and we used $\|A\|_F \le \sqrt{N}\|A\|_2$. Overall, using the triangular inequality, we get $\|F_\infty(x) - F_\nu(x)\| \le (2L'M_2 + M'\sqrt{N}) d_2(\nu)$.

Proof of Lemma 2

By the mean value theorem[8], there exists $\bar{x} \in [x, y]$ such that

$$\begin{aligned} &(F_\nu(x) - F_\nu(y))^\top (x - y) = (F_\infty(x) - F_\infty(y))^\top (x - y) + \\ &(F_\nu(x) - F_\infty(x) - (F_\nu(y) - F_\infty(y)))^\top (x - y) \ge \\ &(\alpha_\infty - \|\nabla_x F_\infty(\bar{x}) - \nabla_x F_\nu(\bar{x})\|)\|x - y\|^2 =: \alpha_\nu \|x - y\|^2. \end{aligned}$$

[8] Applied to the function $\phi(t) = (x - y)^\top \tilde{F}(tx + (1 - t)y)$ with $\tilde{F}(\cdot) = F_\nu(\cdot) - F_\infty(\cdot)$, so that $(x - y)^\top [\tilde{F}(x) - \tilde{F}(y)] = \phi(1) - \phi(0) = (x - y)^\top \nabla \tilde{F}(\bar{t}x + (1 - \bar{t})y)(x - y)$.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TCNS.2019.2944300

As $J^i(z_1,z_2)$ is twice-continuously differentiable, $T^\nu \to \frac{1}{N}\mathbb{1}_N\mathbb{1}_N^\top$ and $\sigma_\nu^i(x)\to\sigma_\infty(x)$, then $\nabla_x F_\infty(x) - \nabla_x F_\nu(x) \to 0$, uniformly in x because $\mathcal{X}$ is compact (Heine-Cantor theorem). Hence for ν large enough $\alpha_\nu > 0$ and F_ν is strongly monotone. To derive an exact bound, under the additional Assumption 3, note that

$$\nabla_x F_\nu(x) = K_\nu^{(11)} + K_\nu^{(12)}[W_\nu^{(1)}\otimes I_n] + K_\nu^{(22)}[W_\nu^{(2)}\otimes I_n]$$
$$\nabla_x F_\infty(x) = K_\infty^{(11)} + K_\infty^{(12)}[W_\infty^{(1)}\otimes I_n] + K_\infty^{(22)}[W_\infty^{(2)}\otimes I_n]$$

where $W_\nu^{(1)} := (\text{diag}(T^\nu)+T^\nu)$, $W_\nu^{(2)} := (\text{diag}(T^\nu)T^\nu)$, $W_\infty^{(1)} := \frac{1}{N}(I_N + \mathbb{1}_N\mathbb{1}_N^\top)$, $W_\infty^{(2)} := \frac{1}{N^2}\mathbb{1}_N\mathbb{1}_N^\top$, $K_\nu^{(ab)} := \text{blkdiag}[\nabla_{z_a z_b} J^i(x^i,\sigma_\nu^i(x))]_{i=1}^N$ for $a,b\in\{1,2\}$ and similar for $K_\infty^{(ab)}$. Since $\|K_\nu^{(ab)} - K_\infty^{(ab)}\|_2 \le L'' M_2 d_2(\nu)$, $\|W_\infty^{(a)} - W_\nu^{(a)}\|_2 \le 2d_2(\nu)$, $\|K_{\nu/\infty}^{(ab)}\|_2 \le M''$, $\|W_{\nu/\infty}^{(1)}\|_2 \le 2$ and $\|W_{\nu/\infty}^{(2)}\|_2 \le 1$ it holds

$$\|\nabla_x F_\infty(\bar{x}) - \nabla_x F_\nu(\bar{x})\|_2 \le 4(L''M_2 + M'')d_2(\nu). \quad (25)$$

Hence for strong monotonicity it suffices to impose $d_2(\nu) < \frac{\alpha_\infty}{4(L''M_2+M'')}$. The conclusion follows from $d_2(\nu)\le r_{\text{step}}(T)^\nu$, as shown in Lemma 1.

Proof of Lemma 3

Statements 1) and 2) follow from Lemma A.3. To prove 3), note that $\nu > \nu_H^{\text{glob}}$ implies

$$d_2(\nu) \le \frac{M_{\hat{A}}}{\|\hat{A}\|_\infty M_\infty \sqrt{N}}. \quad (26)$$

Moreover, according to the definition of $\mathcal{Q}_\nu, \mathcal{Q}_\infty$ in (7c) and (8c), we have that $b_\nu = b_\infty$. Consequently, for $c(\nu)$ as defined in Lemma A.2 it holds

$$c(\nu) = \max_{x\in\mathcal{X}} \|[(\tfrac{\mathbb{1}_N\mathbb{1}_N^\top}{N} - T^\nu)\otimes\hat{A}]x\|_\infty \le d_\infty(\nu)\|\hat{A}\|_\infty M_\infty. \quad (27)$$

Note that a sufficient condition for $\{x\in\mathcal{X} \mid A_\infty x \le b_\infty - c(\nu)\mathbb{1}_{Nm}\}$ to be non-empty is that $c(\nu)\le M_{\hat{A}}$. Combining with (27) we get that a sufficient condition is $d_\infty(\nu)\|\hat{A}\|_\infty M_\infty \le M_{\hat{A}}$, which is met for $\nu \ge \nu_H^{\text{glob}}$ as shown in (26). The conclusion then follows from Lemma A.2.2).

For part 4) we cannot apply the same reasoning because the non-empty condition might not be met. However, Lemma A.2.1) implies that

$$d_H(\mathcal{Q}_\nu(\bar{x}_\nu^{-i}),\mathcal{Q}_\infty(\bar{x}_\nu^{-i})) \le \max\left\{\Delta(\left[\begin{smallmatrix}A_\infty^i\\G^i\end{smallmatrix}\right]),\Delta(\left[\begin{smallmatrix}A_\nu^i\\G^i\end{smallmatrix}\right])\right\}\cdot c^i(\nu),$$

where $c^i(\nu) = \max_{x^i\in\mathcal{X}^i}\{\|(A_\infty^i - A_\nu^i)x^i + \sum_{j\neq i}(A_\infty^j - A_\nu^j)\bar{x}_\nu^j\|_\infty\} \le c(\nu)$ and $A_\nu^i = [[T^\nu]_{1i}\hat{A};\ldots;[T^\nu]_{Ni}\hat{A}]$, $A_\infty^i = [\frac{1}{N}\hat{A};\ldots;\frac{1}{N}\hat{A}]$. Because of this structure for any submatrix B_ν of $\left[\begin{smallmatrix}A_\nu^i\\G^i\end{smallmatrix}\right]$ and corresponding submatrix B_∞ of $\left[\begin{smallmatrix}A_\infty^i\\G^i\end{smallmatrix}\right]$ it holds $B_\nu^{-1} = B_\infty^{-1}D$, where D is a diagonal matrix with elements that are either 1 (corresponding to rows selected from G^i) or $\frac{1}{T_{ji}^\nu N}$ (corresponding to rows selected from A^i). Consequently, for each element e_ν of B_ν^{-1} and corresponding element e_∞ of B_∞^{-1} it holds that either $e_\nu = e_\infty$ or $e_\nu = e_\infty\frac{1}{T_{ji}^\nu N}$. Note that in the latter case it holds $|e_\nu| = |e_\infty|\frac{1}{T_{ji}^\nu N}$. For $\nu\ge\nu_H^{\text{loc}}$ it holds that $|T_{ji}^\nu - \frac{1}{N}| \le d_2(\nu) \le \frac{1}{2N}$ or equivalently $|NT_{ji}^\nu - 1| \le \frac{1}{2}$. Hence $\frac{1}{T_{ji}^\nu N} \le 2$ and in all cases $|e_\nu| \le 2|e_\infty|$. This implies $\max\left\{\Delta(\left[\begin{smallmatrix}A_\infty^i\\G^i\end{smallmatrix}\right]),\Delta(\left[\begin{smallmatrix}A_\nu^i\\G^i\end{smallmatrix}\right])\right\} \le 2\Delta(\left[\begin{smallmatrix}A_\infty^i\\G^i\end{smallmatrix}\right])$. Finally, $K_H^i \le K_H$ since $\Delta(\left[\begin{smallmatrix}A_\infty^i\\G^i\end{smallmatrix}\right]) \le \Delta(\left[\begin{smallmatrix}A_\infty\\G\end{smallmatrix}\right])$ given that $\left[\begin{smallmatrix}A_\infty^i\\G^i\end{smallmatrix}\right]$ is a submatrix of $\left[\begin{smallmatrix}A_\infty\\G\end{smallmatrix}\right]$.

Proof of Lemma 4

We start by computing the operator $F_\infty(x)$.

$$F_\infty(x) = [\nabla_{x^i}\left(a^i(r^i) + \sum_{e=1}^E c_e^i(t_e^i)\right)]_{i=1}^N + P(x),$$

where $P(x) := -[\nabla_{x^i}(p(\sigma_\infty(x))^\top y^i)]_{i=1}^N$. Since for each i the functions a^i and c_e^i are strongly convex and continuously differentiable, by Proposition 2 and [40, equation (12)] there exists $\alpha>0$ such that

$$\nabla_x([\nabla_{x^i}\left(a^i(r^i) + \sum_{e=1}^E c_e^i(t_e^i)\right)]_{i=1}^N) \succ \alpha I_{N(E+1)},\ \forall x\in\mathcal{X}.$$

We now prove that $\nabla_x P(x) \succeq 0$ under either of the two conditions stated.

1) We have

$$N\cdot P(x) := \left[\sum_j H_i^\top D H_j x^j + H_i^\top D^\top H_i x^i - N H_i^\top d\right]_{i=1}^N$$
$$= [H^\top D H + H_{\text{blkd}}^\top D^\top H_{\text{blkd}}]x - N[H_i^\top d]_{i=1}^N,$$

with $H := ([H_i^\top]_{i=1}^N)^\top$ and $H_{\text{blkd}} = \text{blkdiag}(H_1,\ldots,H_N)$. Moreover, since $D\succeq 0$, then $\nabla_x P(x) = \frac{1}{N}(H^\top D H + H_{\text{blkd}}^\top D^\top H_{\text{blkd}}) \succeq 0$.

2) Let $\tilde{P}(y) := -[\nabla_{y^i}(p(\frac{1}{N}\sum_{j=1}^N y^j)^\top y^i)]_{i=1}^N$. It was shown in [24, Corollary 1], that if (20) holds, then there exists $\alpha' > 0$ such that $\nabla_y\tilde{P}(y) \succ \alpha' I_{NE}$. Moreover, from $p(\sigma_\infty(x))^\top y^i = p(\frac{1}{N}\sum_{j=1}^N H^j x^j)^\top H^i x^i$ one immediately gets that $P(x) = H_{\text{blkd}}^\top\tilde{P}(H_{\text{blkd}}x)$. It follows that for any x and corresponding $y = H_{\text{blkd}}x$, $\nabla_x P(x) = (H_{\text{blkd}}^\top\nabla_y\tilde{P}(y)H_{\text{blkd}})_{|y=H_{\text{blkd}}x} \succeq 0$. We have proven that $\nabla_x F_\infty(x) \succ \alpha I_{N(E+1)}$ for all x. Consequently, F_∞ is strongly monotone by Proposition 2 and Assumption 2 holds.

References

[1] J. R. Correa, A. S. Schulz, and N. E. Stier-Moses, "Selfish routing in capacitated networks," *Mathematics of Operations Research*, vol. 29, no. 4, pp. 961–976, 2004.
[2] T. Alpcan, T. Başar, R. Srikant, and E. Altman, "CDMA uplink power control as a noncooperative game," *Wireless Networks*, vol. 8, no. 6, pp. 659–670, 2002.
[3] Z. Ma, D. S. Callaway, and I. A. Hiskens, "Decentralized charging control of large populations of plug-in electric vehicles," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 1, pp. 67–78, 2013.
[4] R. Johari and J. N. Tsitsiklis, "Efficiency loss in a network resource allocation game," *Mathematics of Operations Research*, vol. 29, no. 3, pp. 407–435, 2004.
[5] S. Kar and G. Hug, "Distributed robust economic dispatch in power systems: A consensus + innovations approach," in *IEEE Power and Energy Society General Meeting*, 2012.
[6] Y. Pan and L. Pavel, "Games with coupled propagated constraints in optical networks with multi-link topologies," *Automatica*, vol. 45, no. 4, pp. 871–880, 2009.
[7] S. Grammatico, F. Parise, M. Colombino, and J. Lygeros, "Decentralized convergence to Nash equilibria in constrained deterministic mean field control," *IEEE Transactions on Automatic Control*, vol. 61, no. 11, pp. 3315–3329, 2016.

[8] F. Parise, B. Gentile, S. Grammatico, and J. Lygeros, "Network aggregative games: Distributed convergence to Nash equilibria," in *Proceedings of the IEEE Conference on Decision and Control*, 2015, pp. 2295–2300.
[9] F. Facchinei and J. Pang, *Finite-dimensional variational inequalities and complementarity problems*. Springer Science & Business Media, 2007.
[10] P. Yi and L. Pavel, "An operator splitting approach for distributed generalized Nash equilibria computation," *Automatica*, vol. 102, pp. 111–121, 2019.
[11] J. Koshal, A. Nedić, and U. V. Shanbhag, "A gossip algorithm for aggregative games on graphs," in *Proceedings of the IEEE Conference on Decision and Control*, 2012, pp. 4840–4845.
[12] H. Chen, Y. Li, R. H. Louie, and B. Vucetic, "Autonomous demand side management based on energy consumption scheduling and instantaneous load billing: An aggregative game approach," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 1744–1754, 2014.
[13] F. Parise, S. Grammatico, B. Gentile, and J. Lygeros, "Network aggregative games and distributed mean field control via consensus theory," *arXiv preprint arXiv:1506.07719*, 2015.
[14] J. Koshal, A. Nedić, and U. V. Shanbhag, "Distributed algorithms for aggregative games on graphs," *Operations Research*, vol. 64, no. 3, pp. 680–704, 2016.
[15] S. Liang, P. Yi, and Y. Hong, "Distributed Nash equilibrium seeking for aggregative games with coupled constraints," *arXiv preprint arXiv:1609.02253*, 2016.
[16] H. Yin, U. V. Shanbhag, and P. G. Mehta, "Nash equilibrium problems with scaled congestion costs and shared constraints," *IEEE Transactions on Automatic Control*, vol. 56, no. 7, pp. 1702–1708, 2011.
[17] P. Yi and L. Pavel, "Distributed generalized nash equilibria computation of monotone games via double-layer preconditioned proximal-point algorithms," *IEEE Transactions on Control of Network Systems*, 2018.
[18] F. Salehisadaghiani and L. Pavel, "Generalized nash equilibrium problem by the alternating direction method of multipliers," *arXiv preprint arXiv:1703.08509*, 2017.
[19] P. Yi and L. Pavel, "Asynchronous distributed algorithms for seeking generalized nash equilibria under full and partial decision information," *arXiv preprint arXiv:1801.02967*, 2018.
[20] L. Pavel, "A doubly-augmented operator splitting approach for distributed GNE seeking over networks," in *IEEE Annual Conference on Decision and Control (CDC)*, 2018.
[21] L. Pavel, "Distributed GNE seeking under partial-decision information over networks via a doubly-augmented operator splitting approach," *arXiv preprint arXiv:1808.04465*, 2018.
[22] F. Facchinei, A. Fischer, and V. Piccialli, "On generalized Nash games and variational inequalities," *Operations Research Letters*, vol. 35, no. 2, pp. 159–164, 2007.
[23] F. Bullo, *Lectures on Network Systems*. Kindle Direct Publishing, 2019, with contributions by J. Cortes, F. Dorfler, and S. Martinez. [Online]. Available: http://motion.me.ucsb.edu/book-lns
[24] D. Paccagnan, B. Gentile, F. Parise, M. Kamgarpour, and J. Lygeros, "Nash and wardrop equilibria in aggregative games with coupling constraints," *IEEE Transactions on Automatic Control*, 2018.
[25] J. B. Rosen, "Existence and uniqueness of equilibrium points for concave N-person games," *Econometrica: Journal of the Econometric Society*, pp. 520–534, 1965.
[26] C. Bergthaller and I. Singer, "The distance to a polyhedron," *Linear algebra and its applications*, vol. 169, pp. 111–129, 1992.
[27] U. Mosco, "Convergence of convex sets and of solutions of variational inequalities," *Advances in Mathematics*, vol. 3, no. 4, pp. 510–585, 1969.
[28] C. Li and Z. Qu, "Distributed estimation of algebraic connectivity of directed networks," *Systems & Control Letters*, vol. 62, no. 6, pp. 517–524, 2013.
[29] F. Parise, B. Gentile, and J. Lygeros, "A distributed algorithm for average aggregative games with coupling constraints," *arXiv preprint arXiv:1706.04634*, 2017.
[30] T. Brinkhoff, "A framework for generating network-based moving objects," *GeoInformatica*, vol. 6, no. 2, pp. 153–180, 2002.
[31] A. Nagurney, *Network economics: A variational inequality approach*. Springer Science & Business Media, 2013, vol. 10.
[32] Y. Qiu and T. L. Magnanti, "Sensitivity analysis for variational inequalities defined on polyhedral sets," *Mathematics of Operations Research*, vol. 14, no. 3, pp. 410–432, 1989.
[33] J. Kyparisis, "Sensitivity analysis framework for variational inequalities," *Mathematical Programming*, vol. 38, no. 2, pp. 203–213, 1987.
[34] S. Dafermos, "Sensitivity analysis in variational inequalities," *Mathematics of Operations Research*, vol. 13, no. 3, pp. 421–434, 1988.
[35] M. Patriksson and R. T. Rockafellar, "Sensitivity analysis of aggregated variational inequality problems, with application to traffic equilibria," *Transportation Science*, vol. 37, no. 1, pp. 56–68, 2003.
[36] N. D. Yen, "Lipschitz continuity of solutions of variational inequalities with a parametric polyhedral constraint," *Mathematics of Operations Research*, vol. 20, no. 3, pp. 695–708, 1995.
[37] A. J. Hoffman, "On approximate solutions of systems of linear inequalities," *Journal of Research of the National Bureau of Standards*, vol. 49, no. 4, 1952.
[38] O. L. Mangasarian and T.-H. Shiau, "Lipschitz continuity of solutions of linear inequalities, programs and complementarity problems," *SIAM Journal on Control and Optimization*, vol. 25, no. 3, pp. 583–595, 1987.
[39] R. W. G. Salinetti, "On the convergence of sequences of convex sets in finite dimensions." *SIAM Review*, vol. 22, no. 4, pp. 18–33, 980.
[40] G. Scutari, D. P. Palomar, F. Facchinei, and J. Pang, "Convex optimization, game theory, and variational inequality theory," *Signal Processing Magazine, IEEE*, vol. 27, no. 3, pp. 35–49, 2010.

Francesca Parise was born in Verona, Italy, in 1988. She received the B.Sc. and M.Sc. degrees (cum Laude) in Information and Automation Engineering from the University of Padova, Italy, in 2010 and 2012, respectively. She conducted her Master's thesis research at Imperial College London, UK, in 2012. She graduated from the Galilean School of Excellence, University of Padova, Italy, in 2013. She defended her PhD at the Automatic Control Laboratory, ETH Zürich, Switzerland in 2016 and she is currently a Postdoctoral researcher at the Laboratory for Information and Decision Systems, M.I.T., USA. Her research focuses on the identification, analysis and control of multi-agent systems composed by a large number of autonomous users, with application to transportation, financial and social networks.

Basilio Gentile received his Bachelor's degree in Information Engineering and Master's degree in Automation Engineering from the University of Padova, as well as a Master's degree in Mathematical Modeling and Computation from the Technical University of Denmark. He carried out his Master's Thesis at the University of California Santa Barbara and then completed his PhD at the Automatic Control Laboratory of ETH Zurich in 2018. After that, he cofounded Circuit Mind in London, where is currently Chief Technical Officer. His research focuses on aggregative games and network games with applications to traffic networks and to smart charging of electric vehicles.

John Lygeros completed a B.Eng. degree in electrical engineering in 1990 and an M.Sc. degree in Systems Control in 1991, both at Imperial College of Science Technology and Medicine, London, U.K.. In 1996 he obtained a Ph.D. degree from the Electrical Engineering and Computer Sciences Department, University of California, Berkeley. After postdoctoral researcher appointments at M.I.T. and U.C. Berkeley in 2000 he joined the Department of Engineering, University of Cambridge, U.K., as a Lecturer and Churchill College as a Fellow. Between 2003 and 2006 he was an Assistant Professor at the Department of Electrical and Computer Engineering, University of Patras, Greece. In July 2006 he joined the Automatic Control Laboratory at ETH Zurich, first as an Associate Professor, and since January 2010 as a Full Professor. His research interests include modelling, analysis, and control of hierarchical, hybrid, and stochastic systems, with applications to biochemical networks, automated highway systems, air traffic management, power grids and camera networks. John Lygeros is a Fellow of the IEEE, and a member of the IET and the Technical Chamber of Greece; since 2013 he is serving as the Treasurer of the International Federation of Automatic Control and a member of the IFAC Council.