Tracking-Based Distributed Equilibrium Seeking for Aggregative Games

Guido Carnevale, *Graduate Student Member, IEEE*, Filippo Fabiani, Filiberto Fele, Kostas Margellos, *Member, IEEE*, Giuseppe Notarstefano, *Member, IEEE*

***Abstract*—We propose fully-distributed algorithms for Nash equilibrium seeking in aggregative games over networks. We first consider the case where only local constraints are present and we design an algorithm combining, for each agent, (i) the projected pseudo-gradient descent and (ii) a tracking mechanism to locally reconstruct the aggregative variable. To handle coupling constraints arising in generalized settings, we propose another distributed algorithm based on (i) a recently emerged augmented primal-dual scheme and (ii) two tracking mechanisms to reconstruct, for each agent, both the aggregative variable and the coupling constraint satisfaction. Leveraging tools from singular perturbations analysis, we prove linear convergence to the Nash equilibrium for both schemes. Finally, we run extensive numerical simulations to confirm the effectiveness of our methods, also showing that they outperform the current state-of-the-art distributed equilibrium seeking algorithms.**

***Index Terms*—Game theory, Optimization algorithms, Network analysis and control, Distributed algorithms.**

I. Introduction

RECENT years have seen an increasing attention to the computation of (generalized) Nash equilibria in games over networks [1]–[3]. Indeed, numerous applications falling within different domains such as smart grids management [4], [5], economic market analysis [6], cooperative control of robots [7], electric vehicles charging [8]–[10], network congestion control [11], and synchronization of coupled oscillators in power grids [12] can be modelled as networks of selfish agents – aiming at optimizing their strategy according to an associated individual cost function – that compete with each other over shared resources.

Among these examples, one can often find instances modelled as an *aggregative* game, where the strategies of all the agents in the network are coupled through the so-called aggregative variable (expressing, e.g., the mean strategy), upon which each agent's cost function depends; see, e.g., [13]–[15] for a comprehensive overview. Our work investigates such a framework proposing novel distributed algorithms for generalized Nash equilibrium (GNE) seeking under *partial information*, i.e., assuming that each agent is only aware of its own local information (e.g., its strategy set and cost function) and can communicate only with few agents in the network. This restriction naturally calls for the design of fully-distributed mechanisms for GNE seeking.

Our approach is motivated by recent developments in cooperative optimization, where agents in a network collaborate to minimize the sum of individual objective functions depending both on local decision variables and an aggregative variable [16]–[20].

A. Related work

In the context of NE problems in aggregative form, first attempts to design equilibrium seeking algorithms involve semi-decentralized approaches in which a central entity gathers and shares global quantities (such as the aggregative variable and/or a dual multiplier) with all the agents [21]–[28].

To relax the communication requirements, [29] proposes a gradient-based algorithm for non-generalized games with diminishing step-size that relies on dynamic averaging consensus (see, e.g., [30], [31]) to reconstruct the aggregative variable in each agent. Such a method has been refined in [32] to deal with privacy issues and, as a consequence, only guaranteeing approximate equilibrium computations. In [33], the distributed computation of an approximate Nash equilibrium is guaranteed through a best-response-based algorithm requiring multiple communication exchanges per iteration. In [34], instead, an asynchronous distributed algorithm based on proximal dynamics is proposed.

Looking at GNE problems where the agents' strategies are coupled also by means of constraints, in [35] the distributed computation of an approximate NE is guaranteed through an algorithm requiring, however, several communication exchanges per iteration. Exact convergence is instead guaranteed in [36], where a distributed algorithm with diminishing step-size is proposed, combining dynamic tracking mechanisms, monotone operator splitting, and the Krasnosel'skii-Mann fixed-point iteration. An exactly convergent distributed equilibrium-seeking algorithm with constant step-size is given in [37], where the authors propose a distributed method based on a forward-backward splitting of two preconditioned operators requiring a double communication exchange per iteration.

This result is part of the project "Distributed Optimization for Cooperative Machine Learning in Complex Networks" (No PGR10067) that has received funding from the Ministero degli Affari Esteri e della Cooperazione Internazionale.

G. Carnevale and G. Notarstefano are with the Department of Electrical, Electronic and Information Engineering, Alma Mater Studiorum - Universita' di Bologna, Bologna, Italy (name.lastname@unibo.it) F. Fabiani is with the IMT School for Advanced Studies Lucca, Piazza San Francesco 19, 55100 Lucca, Italy (filippo.fabiani@imtlucca.it)

B. Contributions

The main contribution of the paper is the design and the analysis of novel, fully-distributed iterative – i.e., discrete-time – algorithms for (generalized) NE seeking in aggregative games over networks. First, to address the case where only local constraints are present, we combine a projected pseudo-gradient method with a local, auxiliary variable that compensates the lack of knowledge of the aggregative variable in each agent. Successively, to deal also with coupling constraints we take inspiration from a recent augmented primal-dual scheme for centralized, continuous-time optimization [38] and resort to (i) an averaging step to enforce consensus among the agents' multipliers and (ii) two auxiliary variables to locally reconstruct both the aggregative variable and the coupling constraint status. Both iterative schemes are analyzed from a system-theoretic perspective that allows us to establish linear convergence to the NE. To the best of the authors' knowledge, the proposed algorithms are the first distributed schemes in literature guaranteeing linear convergence to the NE.

As a side technical contribution, in contrast with existing methods, our algorithms (i) do not require compactness of the local feasible sets and (ii) allow for a general form of the aggregative variable, thus not necessarily requiring the mean of the agents' strategies to operate. To better classify our work within the existing literature, Tables I and II compare it with the most relevant works. Specifically, Table I considers the framework without coupling constraints, while Table II the one with coupling constraints (GNE). From the comparison, we stress that our algorithms, i.e., Primal TRacking-based Aggregative Distributed Equilibrium Seeking (TRADES), local constraints, and Primal-Dual TRADES, generalized setting, are the only distributed schemes enjoying linear convergence to the (G)NE (note that some of the technical conditions and variables appearing in the table entries will become clear in the sequel).

The analysis of our iterative algorithms is carried out by relying on a *singular perturbations* approach that allows us to see each procedure as the interconnection between a *slow* subsystem and a *fast* one. Specifically, the slow dynamics are produced by the update of the strategies and, in the case with coupling constraints, of the mean of the multipliers over the network. The fast dynamics, instead, describe the evolution of the auxiliary variables used to compensate the lack of knowledge of the global quantities and, in the case with coupling constraints, the consensus error among the agents' multipliers. Based on this interpretation, we construct two auxiliary, simplified subsystems, known as *boundary layer* and *reduced* system, to separately study the fast and slow dynamics, respectively. Leveraging this connection, we first provide the convergence properties of these auxiliary dynamics with Lyapunov-based arguments, and then we merge the obtained results to establish linear convergence to the NE of the whole interconnection. This last step relies on a general theorem (cf. Theorem II.5) considering a class of singularly perturbed systems that includes the proposed iterative algorithms. In detail, this theorem shows that global exponential stability results for the interconnection can be achieved, while typical results in literature only provide semi-global properties (see [39, Prop. 8.1], or [40, Ch. 11] for the continuous-time case). To the best of our knowledge, similar results are not yet available in the literature: besides the construction of a novel, fully-distributed iterative mechanism with appealing features for their practical implementation, they offer a new proof line for equilibrium seeking problems.

Finally, we provide detailed numerical simulations to confirm the effectiveness of our methods and show that they clearly outperform the state-of-the-art distributed NE seeking algorithms in terms of convergence rate.

C. Paper organization

In Section II we introduce aggregative games over networks, while in Section III-A we propose and analyze a novel distributed algorithm to find NE when only local constraints are present. In Section IV we devise a novel distributed GNE seeking algorithm to address the case of linear coupling constraints. Finally, in Section V we provide detailed numerical simulations to test our methods. The proof of the result on singular perturbations – instrumental in the derivation of our main theorems – is deferred to Appendix A; Appendices B – C gather the proofs of all other technical results and lemmas.

Notation: A matrix $M \in \mathbb{R}^{n\times n}$ is Schur if all its eigenvalues lie in the open unit disc. The identity matrix in $\mathbb{R}^{m\times m}$ is I_m, while 0_m is the all-zero matrix in $\mathbb{R}^{m\times m}$. The vector of N ones is denoted by 1_N, while $\mathbf{1}_{N,d} \coloneqq 1_N \otimes I_d$ with $\otimes$ being the Kronecker product. Dimensions are omitted whenever clear from the context. Given a function of two variables $f : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \to \mathbb{R}$, we denote as $\nabla_1 f \in \mathbb{R}^{n_1}$ the gradient of f with respect to its first argument and as $\nabla_2 f \in \mathbb{R}^{n_2}$ the gradient of f with respect to the second one. The vertical concatenation of column vectors $v_1, \dots, v_N \in \mathbb{R}^n$ is $\text{COL}(v_1, \dots, v_N)$. $\mathbb{R}^n_+$ identifies the positive orthant in $\mathbb{R}^n$. We denote with $\text{diag}(v_1, \dots, v_n)$ the diagonal matrix whose i-th diagonal element is given by v_i, and with $\text{blkdiag}(M_1, \dots, M_N)$ the block diagonal matrix whose i-th block is $M_i \in \mathbb{R}^{n_i\times n_i}$. Given a vector $x \in \mathbb{R}^n$ and a set $X \subseteq \mathbb{R}^n$, $P_X[x]$ denotes the projection of x on X. Given $x \in \mathbb{R}^n$ and a symmetric, positive definite matrix $M \in \mathbb{R}^{n\times n}$, $\|x\|_M = \sqrt{x^\top M x}$. For matrix (resp., vector) $A \in \mathbb{R}^{m\times n}$ ($v \in \mathbb{R}^n$), we denote as $[A]_j$ ($[v]_j$) its j-th row (j-th component). Given two matrices $A, B \in \mathbb{R}^{m\times m}$, $A \succ B$ (resp. $A \succcurlyeq B$) is equivalent to saying that $A - B$ is positive definite (resp. semidefinite).

II. Mathematical preliminaries

A. Problem definition and main assumptions

We consider a population of $N \in \mathbb{N}$ agents – designated by the set $\mathcal{I} \coloneqq \{1, \dots, N\}$ – who, given all other agents' strategies, aim at finding a local strategy solving the optimization problem:

$$\forall i \in \mathcal{I} : \begin{cases} \min\limits_{x_i \in X_i} & J_i(x_i, \sigma(x)) \\ \text{s.t.} & A_i x_i + \sum\limits_{j\in\mathcal{I}\setminus\{i\}} A_j x_j \leq \sum\limits_{i\in\mathcal{I}} b_i, \end{cases} \tag{1}$$

	[29]	[33]	[32]	Our algorithm
Linear rate	✗	✗	✗	✓
step-size	diminishing	-	diminishing	constant
Exactness	✓	✗	✗	✓
Communications per iterate	1	v	1	1
Equilibrium assumptions	F strictly monotone	$\exists!$ equilibrium, $x_{i,\text{br}}(\cdot)$ non-expansive	F strongly monotone	F strongly monotone
Local constraint set	Compact and convex	Compact and convex	Unconstrained	Closed and convex
Gradient unboundedness	✓	✓	✗	✓
Aggregative variable	$1/N\sum_{i=1}^N x_i$	$1/N\sum_{i=1}^N x_i$	$1/N\sum_{i=1}^N x_i$	$1/N\sum_{i=1}^N \phi_i(x_i)$
Algorithmic structure	Gradient-based	Best-response-based	Gradient-based	Gradient-based
Graph	Undirected, time-varying	Directed	Undirected, time-varying	Directed

TABLE I: Setup without coupling constraints.

	[35]	[36]	[37]	Our algorithm
Linear rate	✗	✗	✗	✓
step-size	constant	diminishing	constant	constant
Exactness	✗	✓	✓	✓
Communications per iterate	v	1	2	1
Equilibrium assumptions	F strongly monotone	F cocoercive	F strongly monotone	F strongly monotone
Local constraints	Compact and convex	Compact and convex	Compact and convex	Closed and convex
Coupling constraints	Not specified	Slater's constraint qualification	Slater's constraint qualification	A full row rank
Gradient unboundedness	✗	✓	✓	✓
Aggregative variable	$1/N\sum_{i=1}^N x_i$	$1/N\sum_{i=1}^N x_i$	$1/N\sum_{i=1}^N x_i$	$1/N\sum_{i=1}^N \phi_i(x_i)$
Algorithmic structure	Gradient-based	Gradient-based	Gradient-based	Gradient-based
Graph	Directed	Undirected, time-varying	Undirected	Directed

TABLE II: Setup with coupling constraints.

where $X_i \subseteq \mathbb{R}^{n_i}$, $A_i \in \mathbb{R}^{m\times n_i}$, and $b_i \in \mathbb{R}^m$ model the feasible strategy set for agent i, while the cost function $J_i : \mathbb{R}^{n_i} \times \mathbb{R}^d \to \mathbb{R}$ depends on the i-th individual strategy $x_i \in \mathbb{R}^{n_i}$, as well as on the *aggregative variable* $\sigma(x) \in \mathbb{R}^d$, with $x := \text{COL}(x_1,\dots,x_N) \in \mathbb{R}^n$, $n := \sum_{i\in\mathcal{I}} n_i$. We consider $m \leq n$. In particular, the aggregative variable $\sigma(\cdot)$ formally reads as

$$\sigma(x) := \frac{1}{N}\sum_{i\in\mathcal{I}}\phi_i(x_i), \tag{2}$$

where each *aggregation rule* $\phi_i : \mathbb{R}^{n_i} \to \mathbb{R}^d$ models the contribution of the corresponding strategy x_i to the aggregate $\sigma(x)$. We define the constraint functions $c_i : \mathbb{R}^{n_i} \to \mathbb{R}^m$, $c_{-i} : \mathbb{R}^{n-n_i} \to \mathbb{R}^m$, and $c : \mathbb{R}^n \to \mathbb{R}^m$ as follows:

$$c_i(x_i) = A_i x_i - b_i, \tag{3a}$$

$$c_{-i}(x_{-i}) = \sum_{j\in\mathcal{I}\setminus\{i\}}(A_j x_j - b_j), \tag{3b}$$

$$c(x) = c_i(x_i) + c_{-i}(x_{-i}) = Ax - b, \tag{3c}$$

where $x_{-i} := \text{COL}(x_1,\dots,x_{i-1},x_{i+1},\dots,x_N) \in \mathbb{R}^{n-n_i}$, $A := [A_1 \dots A_N] \in \mathbb{R}^{m\times n}$, and $b := \sum_{i\in\mathcal{I}} b_i$. Then, the collective vector of strategies x belongs to the feasible set $\mathcal{C} := \{x \in X \mid c(x) \leq 0\} \subseteq \mathbb{R}^n$, where $X := \prod_{i\in\mathcal{I}} X_i \subseteq \mathbb{R}^n$.

We refer to any equilibrium solution to the collection of inter-dependent optimization problems (1) as aggregative GNE [3] (or simply GNE), and to the problem of finding such an equilibrium as GNE problem (GNEP) in aggregative form – as opposed to a NE problem (NEP) which is characterized by local constraints only. We will design distributed algorithms to find aggregative GNEs, which formally correspond to the following definition:

Definition II.1 (Generalized Nash equilibrium [3])**.** *A collective vector of strategies $x^\star \in \mathcal{C}$ is a GNE of (1) if, for all $i \in \mathcal{I}$, we have:*

$$J_i(x_i^\star, \sigma(x^\star)) \leq \min_{x_i\in\mathcal{C}_i(x_{-i}^\star)} J_i(x_i, \tfrac{1}{N}\phi_i(x_i) + \sigma_{-i}(x_{-i}^\star)),$$

with $\mathcal{C}_i(x_{-i}) := \{x_i \in X_i \mid A_i x_i \leq b_i - c_{-i}(x_{-i})\}$.

We note that the definition of NE follows directly from the above by replacing $\mathcal{C}_i(x_{-i}^\star)$ simply with X_i.

An equivalent definition of GNE requires one to find a fixed-point of the *best response* mapping $x_{i,\text{br}} : \mathbb{R}^{n-n_i} \to \mathbb{R}^{n_i}$ of each agent, which is formally defined as:

$$\begin{aligned} x_{i,\text{br}}(x_{-i}) &\in \operatorname*{arg\,min}_{x_i\in\mathcal{C}_i(x_{-i})} J_i\left(x_i, \sigma(x)\right) \\ &= \operatorname*{arg\,min}_{x_i\in\mathcal{C}_i(x_{-i})} J_i\left(x_i, \tfrac{1}{N}\phi_i(x_i) + \sigma_{-i}(x_{-i})\right), \end{aligned}$$

In fact, a collective vector of strategies $x^\star$ is a GNE if, for all $i \in \mathcal{I}$, $x_i^\star = x_{i,\text{br}}(x_{-i}^\star)$. Next, we formalize customary assumptions that establish the regularity of some local quantities in (1).

Standing Assumption II.2 (Local feasible sets and cost functions)**.** *For all $i \in \mathcal{I}$, we have that:*

(i) *The feasible set X_i is nonempty, closed, and convex;*
(ii) *The function $J_i(\cdot, \phi_i(\cdot)/N + \sigma_{-i}(x_{-i}))$ is of class $\mathcal{C}^1$, i.e., its derivative exists and is continuous, for all $x_{-i} \in \mathbb{R}^{n-n_i}$.*

A key device in this game-theoretic framework is the so-called *pseudo-gradient mapping* $F : \mathbb{R}^n \to \mathbb{R}^n$:

$$F(x) := \text{COL}(\nabla_{x_1} J_1(x_1, \sigma(x)), \dots, \nabla_{x_N} J_N(x_N, \sigma(x))). \tag{4}$$

With this regard, we also make the following assumption.

Standing Assumption II.3 (Strong monotonicity and Lipschitz continuity)**.** *F is μ-strongly monotone, i.e., there exists $\mu > 0$*

such that

$$(F(x) - F(y))^\top (x-y) \geq \mu \left\| x-y \right\|^2,$$

for any $x, y \in \mathbb{R}^n$*. Moreover, given any* $x_i, x_i' \in \mathbb{R}^{n_i}$ *and* $y, y' \in \mathbb{R}^{n-n_i}$*, for all* $i \in \mathcal{I}$*, we assume that*

$$\begin{aligned} &\|\nabla_{x_i} J_i(x_i, \phi_i(x_i)/N + y) - \nabla_{x_i'} J_i(x_i', \phi_i(x_i')/N + y')\| \\ &\qquad\qquad \leq \beta_1 \|\mathrm{COL}(x_i, y) - \mathrm{COL}(x_i', y')\|, \\ &\|\nabla_1 J_i(x_i, y) - \nabla_1 J_i(x_i', y')\| \leq \beta_1 \|\mathrm{COL}(x_i, y) - \mathrm{COL}(x_i', y')\|, \\ &\|\nabla_2 J_i(x_i, y) - \nabla_2 J_i(x_i', y')\| \leq \beta_2 \|\mathrm{COL}(x_i, y) - \mathrm{COL}(x_i', y')\|, \\ &\qquad \|\phi_i(x_i) - \phi_i(x_i')\| \leq \beta_3 \left\| x_i - x_i' \right\|. \end{aligned}$$

While assumptions on strong monotonicity and Lipschitz continuity of the game mapping are quite standard in the literature [15], [25], [26], in the second part of Standing Assumption II.3 we further specialize the Lipschitz properties of the gradients of the cost functions in both the local and aggregate variables, as well as of each single aggregation rule $\phi_i(\cdot)$.

Note that we assume partial information, i.e., each agent i is only aware of its own local information x_i, J_i, ϕ_i, X_i, A_i, and b_i. Moreover, each agent can exchange information with a subset of $\mathcal{I}$ only. Specifically, we consider a network of agents whose communication is performed according to a directed graph $\mathcal{G} = (\mathcal{I}, \mathcal{E})$, with $\mathcal{E} \subset \mathcal{I}^2$ such that i can receive information from agent j only if the edge $(j, i) \in \mathcal{E}$. The set of in-neighbors of i is represented by $\mathcal{N}_i := \{j \in \mathcal{I} \mid (j, i) \in \mathcal{E}\}$ (where also $i \in \mathcal{N}_i$), while $\mathcal{N}_i^{\text{out}} := \{j \in \mathcal{I} \mid (i, j) \in \mathcal{E}\}$ denotes the set of out-neighbors of the agent i. Graph $\mathcal{G}$ is associated with a weighted adjacency matrix $\mathcal{W} \in \mathbb{R}^{N \times N}$ whose entries satisfy $w_{ij} > 0$ whenever $(j, i) \in \mathcal{E}$ and $w_{ij} = 0$ otherwise. The following assumption characterizes the communication graphs considered:

Standing Assumption II.4 (Network)**.** *The graph* $\mathcal{G}$ *is strongly connected, i.e., for every pair of nodes* $(i, j) \in \mathcal{I}^2$ *there exists a path of directed edges that goes from* i *to* j*, and the matrix* $\mathcal{W}$ *is doubly stochastic, namely it holds that:*

$$\mathcal{W} 1_N = 1_N, \quad 1_N^\top \mathcal{W} = 1_N^\top.$$

B. A key result on singularly perturbed systems

The convergence analysis of the iterative schemes introduced in Section III and IV exploits a system-theoretic perspective based on *singular perturbation*, that strongly relies on the following crucial result proved in Appendix A.

Theorem II.5 (Global exponential stability for singularly perturbed systems)**.** *Consider the system*

$$x^{t+1} = x^t + \delta f(x^t, w^t) \tag{5a}$$
$$w^{t+1} = g(w^t, x^t, \delta), \tag{5b}$$

with $x^t \in \mathcal{D} \subseteq \mathbb{R}^n$*,* $w^t \in \mathbb{R}^m$*,* $f : \mathcal{D} \times \mathbb{R}^m \to \mathbb{R}^n$*,* $g : \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^m$*,* $\delta > 0$*. Assume that* f *and* g *are Lipschitz continuous with respect to both arguments with Lipschitz constants* $L_f > 0$ *and* $L_g > 0$*, respectively. Assume that there exists* $x^\star \in \mathbb{R}^n$ *and* $h : \mathbb{R}^n \to \mathbb{R}^m$ *such that for any* $x \in \mathbb{R}^n$

$$\begin{aligned} 0 &= \delta f(x^\star, h(x^\star)), \\ h(x) &= g(h(x), x, \delta), \end{aligned}$$

with h *being Lipschitz continuous with Lipschitz constant* $L_h > 0$*. Let*

$$x^{t+1} = x^t + \delta f(x^t, h(x^t)) \tag{6}$$

be the reduced system *and*

$$\psi^{t+1} = g(\psi^t + h(x), x, \delta) - h(x) \tag{7}$$

be the boundary layer system *with* $\psi^t \in \mathbb{R}^m$*.*

Assume that there exists a continuous function $U : \mathbb{R}^m \to \mathbb{R}$ *and* $\bar{\delta}_1 > 0$ *such that, for any* $\delta \in (0, \bar{\delta}_1)$ *(cf. (5)), there exist* $b_1, b_2, b_3, b_4 > 0$ *such that for any* $\psi, \psi_1, \psi_2 \in \mathbb{R}^m$*,* $x \in \mathbb{R}^n$*,*

$$b_1 \left\|\psi\right\|^2 \leq U(\psi) \leq b_2 \left\|\psi\right\|^2 \tag{8a}$$
$$U(g(\psi + h(x), x, \delta) - h(x)) - U(\psi) \leq -b_3 \left\|\psi\right\|^2 \tag{8b}$$
$$\begin{aligned} |U(\psi_1) - U(\psi_2)| &\leq b_4 \left\|\psi_1 - \psi_2\right\| \left\|\psi_1\right\| \\ &\quad + b_4 \left\|\psi_1 - \psi_2\right\| \left\|\psi_2\right\|. \end{aligned} \tag{8c}$$

Further, assume there exists a continuous function $W : \mathcal{D} \to \mathbb{R}$ *and* $\bar{\delta}_2 > 0$ *such that, for any* $\delta \in (0, \bar{\delta}_2)$*, there exist* $c_1, c_2, c_3, c_4 > 0$ *such that for any* $x, x_1, x_2, x_3 \in \mathcal{D}$

$$c_1 \left\|x - x^\star\right\|^2 \leq W(x) \leq c_2 \left\|x - x^\star\right\|^2 \tag{9a}$$
$$W(x + \delta f(x, h(x))) - W(x) \leq -c_3 \left\|x - x^\star\right\|^2 \tag{9b}$$
$$\begin{aligned} |W(x_1) - W(x_2)| &\leq c_4 \left\|x_1 - x_2\right\| \left\|x_1 - x^\star\right\| \\ &\quad + c_4 \left\|x_1 - x_2\right\| \left\|x_2 - x^\star\right\|. \end{aligned} \tag{9c}$$

Then, there exist $\bar{\delta} \in (0, \min\{\bar{\delta}_1, \bar{\delta}_2\})$*,* $a_1 > 0$*, and* $a_2 > 0$ *such that, for all* $\delta \in (0, \bar{\delta})$*, it holds*

$$\left\| \begin{bmatrix} x^t - x^\star \\ w^t - h(x^t) \end{bmatrix} \right\| \leq a_1 \left\| \begin{bmatrix} x^0 - x^\star \\ w^0 - h(x^0) \end{bmatrix} \right\| e^{-a_2 t},$$

for any $(x^0, w^0) \in \mathcal{D} \times \mathbb{R}^m$*.*

We will show how our algorithms, namely Primal TRADES and Primal-Dual TRADES, can be recast in the form of the interconnected system (5) while satisfying all assumptions of Theorem II.5, and hence prove their convergence in Theorems III.1 and IV.3, respectively, provided in the next sections. Compared with traditional approaches, taking a singular perturbation view offers a novel proof line for (generalized) equilibrium seeking problems.

III. AGGREGATIVE GAMES OVER NETWORKS WITHOUT COUPLING CONSTRAINTS

A. Primal TRADES

In this section we introduce and analyze Primal TRacking-based Aggregative Distributed Equilibrium Seeking (TRADES), a fully-distributed iterative NE seeking algorithm for aggregative games given by (1) without coupling constraints, i.e.,

$$\forall i \in \mathcal{I} : \min_{x_i \in X_i} J_i(x_i, \sigma(x)). \tag{10}$$

Our distributed algorithm is then able to steer the strategies of the network to a NE of the game.

The proposed scheme is iterative with t denoting the iteration index. Let $x_i^t \in \mathbb{R}^{n_i}$ be the strategy chosen by each agent i at iteration $t \geq 0$. Taking its convex combination with a projected pseudo-gradient step may be an effective way to steer each agent's strategy to the best response $x_{i,\text{br}}(\sigma_{-i}(x_{-i}^t))$. When applied to problem (10), it reads as

$$x_i^{t+1} = x_i^t + \delta\left(P_{X_i}\left[x_i^t - \gamma\left(\nabla_{x_i} J_i(x_i^t, \sigma(x^t))\right)\right] - x_i^t\right), \tag{11}$$

where $\delta \in (0,1)$ is a constant performing the combination and $\gamma > 0$ plays the role of the gradient step-size. We point out that the chain rule and the definition of $\sigma(x^t)$ (cf. (2)) lead to $\nabla_{x_i} J_i(x_i^t, \sigma(x^t)) = \nabla_1 J_i(x_i^t, \sigma(x^t)) + \frac{\nabla\phi_i(x_i^t)}{N}\nabla_2 J_i(x_i^t, \sigma(x^t))$. In our distributed setting, however, agent i cannot access the global aggregate variable $\sigma(x^t)$. To compensate this lack of information, we rely on the locally available $\phi_i(x_i^t)$ and the auxiliary variable $z_i^t \in \mathbb{R}^d$. Thus, for all $i \in \mathcal{I}$, we introduce the operator $\tilde{F}_i : \mathbb{R}^{n_i} \times \mathbb{R}^d \to \mathbb{R}^{n_i}$ defined as

$$\tilde{F}_i(x_i, s) := \nabla_1 J_i(x_i, s) + \frac{\nabla\phi_i(x_i)}{N}\nabla_2 J_i(x_i, s),$$

and, in accordance, we modify the update (11) as

$$x_i^{t+1} = x_i^t + \delta\left(P_{X_i}\left[x_i^t - \gamma\tilde{F}_i\left(x_i^t, \phi_i(x_i^t) + z_i^t\right)\right] - x_i^t\right), \tag{12}$$

which can be directly implemented without violating the distributed nature of the algorithm. In case

$$z_i^t \to -\phi_i(x_i^t) + \sigma(x^t), \tag{13}$$

then the implementable law (12) coincides with the desired one given in (11). Note that z_i^t encodes the estimate of $\sigma(x_i^t) - \phi_i(x_i^t)$, i.e., the aggregate of all other agents' strategies except for the i-th one. For this reason, we update each auxiliary variable z_i^t according to the following causal version of the perturbed average consensus scheme (see, e.g., [41], where a similar scheme has been used to locally compensate the missing knowledge of the global gradient of a distributed consensus optimization problem):

$$z_i^{t+1} = \sum_{j \in \mathcal{N}_i} w_{ij} z_j^t + \sum_{j \in \mathcal{N}_i} w_{ij}\phi_j(x_j^t) - \phi_i(x_i^t). \tag{14}$$

This is implementable in a fully-distributed fashion since it only requires communication with neighboring agents $j \in \mathcal{N}_i$. We report the whole algorithmic structure in Algorithm 1 and, from now on, we will refer to it as Primal TRADES. We note that Algorithm 1 requires the initialization $z_i^0 = 0$ for all $i \in \mathcal{I}$; we will discuss in the sequel the interpretation of this particular initialization. The local update (15) leads to the stacked vector form of Primal TRADES, namely

$$x^{t+1} = x^t + \delta\Bigg(P_X\Bigg[x^t - \gamma\tilde{F}\left(x^t, \phi(x^t) + z^t\right)\Bigg] - x^t\Bigg), \tag{16a}$$

$$z^{t+1} = \mathcal{W}_d z^t + (\mathcal{W}_d - I)\phi(x^t), \tag{16b}$$

with $\mathcal{W}_d := \mathcal{W} \otimes I_d \in \mathbb{R}^{Nd}$, $z^t := \text{COL}(z_{1,t}, \ldots, z_{N,t})$, $\phi(x^t) := \text{COL}(\phi_1(x_1^t), \ldots, \phi_N(x_N^t))$, and $\tilde{F}(x^t, \phi(x^t) + z^t) := \text{COL}(\tilde{F}_1(x_1^t, \phi_1(x_1^t) + z_1^t), \ldots, \tilde{F}_N(x_N^t, \phi_N(x_N^t) + z_N^t))$. We establish next the properties of Primal TRADES in computing the NE of problem (10).

Algorithm 1 Primal TRADES (Agent i)

Initialization: $x_i^0 \in X_i, z_i^0 = 0$.
for $t = 1, 2, \ldots$ **do**

$$x_i^{t+1} = x_i^t + \delta\left(P_{X_i}\left[x_i^t - \gamma\tilde{F}_i\left(x_i^t, \phi_i(x_i^t) + z_i^t\right)\right] - x_i^t\right) \tag{15a}$$

$$z_i^{t+1} = \sum_{j \in \mathcal{N}_i} w_{ij} z_j^t + \sum_{j \in \mathcal{N}_i} w_{ij}\phi_j(x_j^t) - \phi_i(x_i^t). \tag{15b}$$

end for

Theorem III.1. *Consider the dynamics in* (16). *There exist constants* $\bar{\delta}, \bar{\gamma}, a_1, a_2 > 0$ *such that, for any* $\delta \in (0, \bar{\delta})$, $\gamma \in (0, \bar{\gamma})$ *and* $(x^0, z^0) \in \mathbb{R}^{n+Nd}$ *such that* $\mathbf{1}_{N,d}^\top z^0 = 0$, *it holds*

$$\left\|x^t - x^\star\right\| \leq a_1 e^{-a_2 t}.$$

The proof of Theorem III.1 relies on a *singular perturbation* analysis of system (16), and will be given in the next subsection.

B. Proof of Theorem III.1

We build the framework to prove Theorem III.1 by analyzing (16) under a singular perturbations lens. We therefore establish the related proof in five steps:

1. Bringing (16) *in the form of* (5)*:* We leverage the initialization z^0 so that $\mathbf{1}_{N,d}^\top z^0 = 0$ to introduce coordinates $\bar{z} \in \mathbb{R}^d$ and $z_\perp \in \mathbb{R}^{(N-1)d}$ defined as:

$$\begin{bmatrix} \bar{z} \\ z_\perp \end{bmatrix} := \begin{bmatrix} \frac{\mathbf{1}_{N,d}^\top}{N} \\ R_d^\top \end{bmatrix} z \implies z = \mathbf{1}_{N,d}\bar{z} + R_d z_\perp, \tag{17}$$

where $R_d \in \mathbb{R}^{Nd \times (N-1)d}$ with $\|R_d\| = 1$ is such that

$$R_d R_d^\top = I - \frac{\mathbf{1}_{N,d}\mathbf{1}_{N,d}^\top}{N} \quad \text{and} \quad R_d^\top \mathbf{1}_{N,d} = 0. \tag{18}$$

Then, by using the definition of $\bar{z}$ given in (17), the associated dynamics reads as

$$\begin{aligned} \bar{z}^{t+1} &= \frac{\mathbf{1}_{N,d}^\top}{N} z^{t+1} \overset{(a)}{=} \frac{\mathbf{1}_{N,d}^\top}{N}\mathcal{W}_d z^t + \frac{\mathbf{1}_{N,d}^\top}{N}(\mathcal{W}_d - I)\phi(x^t) \\ &\overset{(b)}{=} \frac{\mathbf{1}_{N,d}^\top}{N} z^t \overset{(c)}{=} \frac{\mathbf{1}_{N,d}^\top}{N}\left(\mathbf{1}_{N,d}\bar{z}^t + R_d z_\perp^t\right) \overset{(d)}{=} \bar{z}^t, \end{aligned} \tag{19}$$

where in (a) we exploit the update (16), in (b) we use the facts that, in view of Standing Assumption II.4, (i) $\mathbf{1}_{N,d}^\top \mathcal{W}_d = \mathbf{1}_{N,d}^\top$ and (ii) $\mathbf{1}_{N,d}^\top(\mathcal{W}_d - I) = 0$, in (c) we rewrite z^t according to (17), and in (d) we use the fact that $\mathbf{1}_{N,d}^\top R_d = 0$. Thus, (19) leads to $\bar{z}^{t+1} \equiv \bar{z}^0 \equiv 0$ for all $t \geq 0$, where the last equality follows by the initialization $\mathbf{1}_{N,d}^\top z^0 = 0$ and the definition of $\bar{z}$ (cf. (17)). We are thus entitled to ignore the null dynamics of $\bar{z}^t$ and, according to (17), we equivalently rewrite (16) as

$$x^{t+1} = x^t + \delta\left(P_X\left[x^t - \gamma\tilde{F}(x^t, \phi(x^t) + R_d z_\perp^t)\right] - x^t\right), \tag{20a}$$

$$z_\perp^{t+1} = R_d^\top \mathcal{W}_d R_d z_\perp^t + R_d^\top(\mathcal{W}_d - I)\phi(x^t). \tag{20b}$$

For any $t \geq 0$, the interconnected system (20) can thus be obtained from (5) by setting

$$\begin{aligned} w^t &:= z_\perp^t, \\ f(x^t, w^t) &:= P_X\left[x^t - \gamma \tilde{F}(x^t, \phi(x^t) + R_d w^t)\right] - x^t, \\ g(w^t, x^t) &:= R_d^\top \mathcal{W}_d R_d w^t + R_d^\top (\mathcal{W}_d - I)\phi(x^t). \end{aligned} \tag{21}$$

In particular, we refer to the subsystem (20a) as the slow system, while we refer to (20b) as the fast one.

2. Equilibrium function h : For any $x^t \in \mathbb{R}^n$, under the expression for $R_d R_d^\top$ in (18) and since $\mathcal{W}$ is doubly stochastic (cf. Standing Assumption II.4) notice that for any $x^t = x \in \mathbb{R}^n$,

$$z_\perp = h(x) := -R_d^\top \phi(x) \tag{22}$$

constitutes an equilibrium of (20b). Since $R_d^\top \mathcal{W}_d R_d$ is Schur in view of Standing Assumption II.4, we interpret (20b) as a strictly stable linear system with nonlinear input $R_d^\top(\mathcal{W}_d - I)\phi(x^t)$ parametrizing the equilibrium of the subsystem. The role of γ is to slow down the variation of x^t so that the stability of $h(x^t)$ for (20b) is preserved.

3. Boundary layer system and satisfaction of (8): The so-called boundary layer system associated to (20) can be constructed by fixing $x^t = x$ for all $t \geq 0$, for some arbitrary $x \in \mathbb{R}^n$ in (20b), and rewriting it according to the error coordinates $\tilde{z}^t := z_\perp^t - h(x^t)$. Using (18), we obtain that

$$\tilde{z}^{t+1} = R_d^\top \mathcal{W}_d R_d \tilde{z}^t. \tag{23}$$

Notice that the latter is in the form of (7) with $\psi = \tilde{z}^t$, and $g(\psi + h(x), x) - h(x) = R_d^\top \mathcal{W}_d R_d \tilde{z}^t$. The next lemma provides a Lyapunov function for (23).

Lemma III.2. *Consider system* (23)*. Then, there exists a continuous function $U : \mathbb{R}^{(N-1)d} \to \mathbb{R}$ satisfying* (8) *with $\tilde{z}$ in place of ψ.*

4. Reduced system and satisfaction of (9): The so-called reduced system can be obtained by plugging into (20a) the fast state at its steady state equilibrium, i.e., we consider $z^t = h(x^t)$ for any $t \geq 0$. We thus have

$$x^{t+1} = x^t + \delta\left(P_X\left[x^t - \gamma \tilde{F}(x^t, \phi(x^t) + R_d h(x^t))\right] - x^t\right). \tag{24}$$

Due to (18) we have that $\tilde{F}(x^t, \phi(x^t) + R_d h(x^t)) = \tilde{F}(x^t, \mathbf{1}_{N,d}\sigma(x^t)) = F(x^t)$, so (24) is equivalent to

$$x^{t+1} = x^t + \delta\left(P_X\left[x^t - \gamma F(x^t)\right] - x^t\right). \tag{25}$$

The next lemma provides a Lyapunov function for (24).

Lemma III.3. *Consider system* (24)*. Let $x^\star \in \mathbb{R}^n$ be such that $f(x^\star, h(x^\star)) = 0$ with f defined as in* (21)*. Then, there exist a continuous function $W : \mathbb{R}^n \to \mathbb{R}$, $\bar{\gamma} > 0$ and $\bar{\delta}_2 > 0$ such that, for any $\gamma \in (0, \bar{\gamma})$ and any $\delta \in (0, \bar{\delta}_2)$, W satisfies* (9)*.*

5. Lipschitz continuity of f, g and h: As we will be invoking Theorem II.5, we need to ensure that the Lipschitz continuity assumptions required by the theorem are satisfied. In particular, we require f and g in (21) to be Lipschitz continuous with respect to both arguments, and h in (22) to be Lipschitz continuous with respect to x.

Lipschitz continuity of f follows by the fact that ∇J_i is Lipschitz continuous due to Standing Assumption II.3. To show Lipschitz continuity of g in (21) notice that for any $w, w' \in \mathbb{R}^{(N-1)d}$ and any $x, x' \in \mathbb{R}^n$,

$$\begin{aligned} &\left\|R_d^\top \mathcal{W}_d R_d (w - w') + R_d^\top (\mathcal{W}_d - I)(\phi(x) - \phi(x'))\right\| \\ &\quad \leq \left\|R_d^\top \mathcal{W}_d R_d\right\| \|w - w'\| + \beta_3 \left\|R_d^\top (\mathcal{W}_d - I)\right\| \|x - x'\|, \end{aligned}$$

where the inequality is due to triangle inequality and the fact that by Standing Assumption II.3, ϕ is Lipschitz continuous with Lipschitz constant β_3. To show Lipschitz continuity of h, notice that for any $x, x' \in \mathbb{R}^n$,

$$\|h(x) - h(x')\| \leq \beta_3 \|R_d\| \|x - x'\| = \beta_3 \|x - x'\|,$$

where the inequality follows from (22) and Lipschitz continuity of ϕ, while the equality from the fact that $\|R_d\| = 1$.

By combining Lemma III.2 and III.3 with the Lipschitz conditions expressed above, Theorem II.5 can therefore be applied. Thus, there exists $\bar{\delta} \in (0, \bar{\delta}_2)$ so that $(x^\star, h(x^\star))$ is an exponentially stable equilibrium for (20). ■

IV. Generalized Nash equilibrium problems in aggregative form

A. Primal-Dual TRADES

In this section we introduce the Primal-Dual TRADES algorithm, i.e., a distributed iterative methodology to find a GNE in aggregative games with local and linear coupling constraints as formalized in (1).

In addition to the assumptions made in Section II, we need some further conditions for our mathematical developments.

Assumption IV.1 (Feasibility)**.** *The set $\mathcal{C} \neq \emptyset$ and, for all $i \in \mathcal{I}$, for any $x_{-i} \in \mathbb{R}^{n-n_i}$, (i) $\mathcal{C}_i(c_{-i}(x_{-i})) \neq \emptyset$ and (ii) $J_i(x_{i,br}(x_{-i}), \phi_i(x_{i,br}(x_{-i}))/N + \sum_{j \neq i} \phi_j(x_j)/N) > -\infty$.*

Consider the following variational inequality, defined by the mapping F in (4) and the domain $\mathcal{C}$:

$$F(x^\star)^\top (x - x^\star) \geq 0, \quad \text{for all } x \in \mathcal{C}. \tag{26}$$

It is known that every point $x^\star \in \mathcal{C}$ for which (26) holds is a GNE of the game (1) and, specifically, a *variational* GNE (v-GNE) (cf. [3, Th. 2.1]). The converse, however, does not hold in general due to the presence of the coupling constraints. Since F is strongly monotone (cf. Standing Assumption II.3) and $\mathcal{C}$ nonempty, closed and convex (cf. Assumption IV.1), a v-GNE is guaranteed to exist and it is also unique by [2, Th. 3].

We will devise an iterative algorithm that will asymptotically return the (unique) v-GNE of (1). Inspired by [38], where an augmented primal-dual scheme was used for continuous-time, centralized optimization, we require the following additional condition on the matrix A of the coupling constraints (cf. (1)):

Assumption IV.2 (Full-row rank)**.** *Matrix A satisfies $\operatorname{rank}(A) = m$, and there exist $\kappa_1, \kappa_2 > 0$ such that $\kappa_1 I_m \preccurlyeq AA^\top \preccurlyeq \kappa_2 I_m$.*

Following [38], for all $i \in \mathcal{I}$ we consider the augmented Lagrangian function $L_i : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ defined as

$$L_i(x, \lambda) := J_i(x_i, \sigma(x)) + \underbrace{\sum_{\ell=1}^{m} H_\ell([Ax-b]_\ell, [\lambda]_\ell)}_{=:H(Ax-b,\lambda)}, \quad (27)$$

where

$$H_\ell([Ax-b]_\ell, [\lambda]_\ell) :=$$
$$\begin{cases} [Ax-b]_\ell[\lambda]_\ell + \frac{\rho}{2}([Ax-b]_\ell)^2 & \text{if } \rho([Ax-b]_\ell) + [\lambda]_\ell \geq 0 \\ -\frac{1}{2\rho}[\lambda]_\ell^2 & \text{if } \rho([Ax-b]_\ell) + [\lambda]_\ell < 0, \end{cases}$$

with $\lambda \in \mathbb{R}^m$ being the multiplier associated to the coupling constraints, and $\rho > 0$ a constant. We therefore address the v-GNE seeking problem by obtaining a saddle point of (27) through the discrete-time dynamics:

$$x_i^{t+1} = x_i^t + \delta\bigg(P_{X_i}\bigg[x_i^t - \gamma\nabla_{x_i}J_i(x_i^t, \sigma(x^t)) - \gamma\nabla_{x_i}H(Ax^t - b, \lambda^t)\bigg] - x_i^t\bigg), \quad (28a)$$

$$\lambda^{t+1} = \lambda^t + \delta\gamma\nabla_\lambda H(Ax^t - b, \lambda^t), \quad (28b)$$

where x_i^t, δ, and γ have the same meaning as in (11), $\lambda^t \in \mathbb{R}^m$ is the multiplier at $t \geq 0$, and the explicit form of the gradients $\nabla_{x_i}H(Ax^t - b, \lambda^t)$ and $\nabla_\lambda H(Ax^t - b, \lambda^t)$ reads as

$$\nabla_{x_i}H(Ax^t - b, \lambda^t) = \sum_{\ell=1}^{m}\nabla_{x_i}H_\ell([Ax^t-b]_\ell, [\lambda^t]_\ell)$$
$$= \sum_{\ell=1}^{m}\max\left\{\rho([Ax^t-b]_\ell) + [\lambda^t]_\ell, 0\right\}[A_i]_\ell^\top, \quad (29a)$$

$$\nabla_\lambda H(Ax^t - b, \lambda^t) = \sum_{\ell=1}^{m}\nabla_\lambda H_\ell([Ax^t-b]_\ell, [\lambda^t]_\ell)$$
$$= \sum_{\ell=1}^{m}\frac{1}{\rho}e_\ell(\max\left\{\rho([Ax^t-b]_\ell) + [\lambda^t]_\ell, 0\right\} - [\lambda^t]_\ell), \quad (29b)$$

where $e_\ell \in \mathbb{R}^m$ is the ℓ-th vector of the canonical basis of $\mathbb{R}^m$, $\ell \in \{1, \ldots, m\}$. The stacked-column form of (28) is

$$x^{t+1} = x^t + \delta\bigg(P_X\left[x^t - \gamma F(x) - \gamma\nabla_x H(Ax^t - b, \lambda^t)\right] - x^t\bigg), \quad (30a)$$

$$\lambda^{t+1} = \lambda^t + \delta\gamma\nabla_\lambda H(Ax^t - b, \lambda^t), \quad (30b)$$

where $\nabla_x H(Ax^t - b, \lambda^t) := \text{COL}(\nabla_{x_1}H(Ax^t - b, \lambda^t), \ldots, \nabla_{x_N}H(Ax^t - b, \lambda^t))$.

However, since agent i does not have access neither to $\sigma(x^t)$ nor to $Ax^t - b$, the scheme in (28) cannot be directly implemented. Moreover, dynamics (28) requires a central unit that can compute the global quantity $Ax^t - b$ and communicate the multiplier λ^t to all the agents. For this reason, in Algorithm 2 we introduce for all $i \in \mathcal{I}$ (i) two additional variables $z_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}^m$ to compensate the local unavailability of $\sigma(x^t)$ and $Ax^t - b$, respectively, (ii) a copy $\lambda_i \in \mathbb{R}^m$ of the multiplier λ, and (iii) an additional averaging step to enforce consensus among the multipliers λ_i, (cf. (32b)-(32d)). We choose causal perturbed consensus dynamics to update z_i and y_i. For all $i \in \mathcal{I}$, we then introduce operators $G_{x,i} : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^{n_i}$ and $G_{\lambda,i} : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$ as

$$G_{x,i}(s_1, s_2) := \sum_{\ell=1}^{m}\max\{\rho([s_1]_\ell) + [s_2]_\ell, 0\}[A_i]_\ell^\top,$$
$$G_{\lambda,i}(s_1, s_2) := \frac{1}{\rho}\sum_{\ell=1}^{m}(\max\{\rho([s_1]_\ell) + [s_2]_\ell, 0\} - [s_2]_\ell)\, e_\ell. \quad (31)$$

Algorithm 2 Primal-Dual TRADES (Agent i)

Initialization: $x_i^0 \in X_i, \lambda_i^t \in \mathbb{R}_+^m, z_i^0 = 0, y_i^0 = 0$.

for $t = 0, 1, \ldots$ **do**

$$x_i^{t+1} = x_i^t + \delta\bigg(P_{X_i}\bigg[x_i^t - \gamma\tilde{F}_i(x_i^t, \phi_i(x_i^t) + z_i^t) - \gamma G_{x,i}(N(A_i x_i^t - b_i) + y_i^t, \lambda_i^t)\bigg] - x_i^t\bigg) \quad (32a)$$

$$\lambda_i^{t+1} = \sum_{j\in\mathcal{N}_i} w_{ij}\lambda_j^t + \delta\gamma G_{\lambda,i}(N(A_i x_i^t - b_i) + y_i^t, \lambda_i^t) \quad (32b)$$

$$z_i^{t+1} = \sum_{j\in\mathcal{N}_i} w_{ij}z_j^t + \sum_{j\in\mathcal{N}_i} w_{ij}\phi_j(x_j^t) - \phi_i(x_i^t) \quad (32c)$$

$$y_i^{t+1} = \sum_{j\in\mathcal{N}_i} w_{ij}y_j^t + \sum_{j\in\mathcal{N}_i} w_{ij}N(A_j x_j^t - b_j) - N(A_i x_i^t - b_i), \quad (32d)$$

end for

In Algorithm 2, these operators encode the component of the gradients in (29) available to agent i at iteration t, plus the auxiliary variable y_i^t that is used to track $Ax^t - b$ (see (32a) and (32b) in Algorithm 2). The main steps of the proposed method are hence summarized in Algorithm 2 from the perspective of agent i, which is then referred as Primal-Dual TRADES. Note that all the quantities involved in the agent's calculations are purely local, thus making Algorithm 2 fully distributed.

Differently from customary primal-dual schemes, (32b) does not need the projection over the positive orthant $\mathbb{R}_+^m$ due to the chosen augmented Lagrangian functions L_i (see (27)). We only need to initialize $\lambda_i^0 \geq 0$ for all $i \in \mathcal{I}$, and choose parameters δ, γ, and ρ appropriately so that we avoid situations where $\lambda_i^t \geq 0$ implies $\lambda_i^{t+1} < 0$. To see this notice first that if $\lambda_i^t = 0$, then it is easy to check $G_{\lambda,i}(N(A_i x_i^t - b_i) + y_i^t, \lambda_i^t) \geq 0$ and, thus, $\lambda_i^{t+1} \geq 0$. The critical scenario for agent i occurs when all the multipliers of its neighbors are zero, namely $\lambda_j^t = 0$ for any $j \in \mathcal{N}_i$, and when $\max\{\rho([N(A_i x_i^t - b_i + y_i^t]_\ell) + [\lambda_i^t]_\ell, 0\} = 0$ for at least one $\ell \in \{1, \ldots, m\}$. Indeed, specializing (32b) for this case leads to the following update of that ℓ-th component of λ_i^t

$$[\lambda_i^{t+1}]_\ell = \left(w_{ii} - \frac{\delta\gamma}{\rho}\right)[\lambda_i^t]_\ell. \quad (33)$$

From (33), we conclude that $[\lambda_i^{t+1}]_\ell$ remains non-negative if $[\lambda_i^t]_\ell$ is non-negative, thus alleviating the need for a projection, as long as δ, γ, and ρ satisfy $w_{ii} > \delta\gamma/\rho$.

As in the case without coupling constraints, the purpose of the initialization step will become clear in the next subsection. The steps of Algorithm 2 in (32) can be compactly written as:

$$x^{t+1} = x^t + \delta f_X(x^t, \lambda^t, z^t, y^t), \tag{34a}$$
$$\lambda^{t+1} = \mathcal{W}_m\lambda^t + \delta\gamma G_\lambda(N(\bar{A}x^t - \bar{b}) + y^t, \lambda^t), \tag{34b}$$
$$z^{t+1} = \mathcal{W}_d z^t + (\mathcal{W}_d - I)\phi(x^t), \tag{34c}$$
$$y^{t+1} = \mathcal{W}_m y^t + (\mathcal{W}_m - I)N(\bar{A}x^t - \bar{b}). \tag{34d}$$

where $f_X : \mathbb{R}^n \times \mathbb{R}^{Nm} \times \mathbb{R}^{Nd} \times \mathbb{R}^{Nm} \to \mathbb{R}^n$ is defined as

$$f_X(x, \lambda, z, y) := P_X\left[x - \gamma\tilde{F}(x, \phi(x)+z) - \gamma G_x(N(\bar{A}x - \bar{b}) + y, \lambda)\right] - x,$$

and, similarly to (16), λ : $\mathrm{COL}(\lambda_1, \ldots, \lambda_N)$, $\mathcal{W}_d := \mathcal{W} \otimes I_d$, $\mathcal{W}_m := \mathcal{W} \otimes I_m$, $G_x(N(\bar{A}x^t - \bar{b}) + y^t, \lambda^t) := \mathrm{COL}(G_{x,1}(N(A_1x_1^t - b_1) + y_1^t, \lambda_1^t), \ldots, G_{x,N}(N(A_Nx_N^t - b_N) + y_N^t, \lambda_N^t))$, and $G_\lambda(N(\bar{A}x^t - \bar{b}) + y^t, \lambda^t) := \mathrm{COL}(G_{\lambda,1}(N(A_1x_1^t - b_1) + y_1^t, \lambda_1^t), \ldots, G_{\lambda,N}(N(A_Nx_N^t - b_N) + y_N^t, \lambda_N^t))$.

The next theorem establishes the convergence properties of Primal-Dual TRADES in computing the v-GNE of (1).

Theorem IV.3. *Consider 34 and Assumptions IV.1, IV.2. Let* $(x^0, \lambda^0, z^0, y^0) \in X \times \mathbb{R}_+^{Nm} \times \mathbb{R}^{Nd} \times \mathbb{R}^{Nm}$ *satisfy* $\mathbf{1}_{N,d}^\top z^0 = 0$ *and* $\mathbf{1}_{N,m}^\top y^0 = 0$*. Then, there exist constants* $\bar{\delta}, \bar{\gamma}, a_1, a_2 > 0$ *such that, for any* $\delta \in (0, \bar{\delta})$, $\gamma \in (0, \bar{\gamma})$, *with* $w_{ii} > \frac{\delta\gamma}{\rho}$ *for all* $i \in \{1, \ldots, N\}$, *it holds*

$$\|x^t - x^\star\| \leq a_1 e^{-a_2 t}.$$

Note that the additional condition $w_{ii} > \delta\gamma/\rho$ needs to be satisfied by δ and γ, given ρ, to ensure the dual variables remain non-negative, as discussed below (33). As in the case of NE seeking without coupling constraints, the proof of Theorem IV.3 relies on a *singular perturbations* analysis of system (34). We provide this in the next subsection.

B. Proof of Theorem IV.3

As with the proof of Theorem III.1, we show that the setting of Theorem IV.3 fits the framework of Theorem II.5, and organize its proof in five steps.

1. Bringing (34) *in the form of* (5)*:* We introduce the change of coordinates

$$\begin{bmatrix} \bar{z}^t \\ z_\perp^t \end{bmatrix} = \begin{bmatrix} \frac{\mathbf{1}_{N,d}^\top}{N} \\ R_d^\top \end{bmatrix} z^t, \quad \begin{bmatrix} \bar{y}^t \\ y_\perp^t \end{bmatrix} = \begin{bmatrix} \frac{\mathbf{1}_{N,m}^\top}{N} \\ R_m^\top \end{bmatrix} y^t,$$
$$\begin{bmatrix} \bar{\lambda}^t \\ \lambda_\perp^t \end{bmatrix} = \begin{bmatrix} \frac{\mathbf{1}_{N,m}^\top}{N} \\ R_m^\top \end{bmatrix} \lambda^t, \tag{35}$$

where $R_d \in \mathbb{R}^{Nd \times (N-1)d}$, $R_m \in \mathbb{R}^{Nm \times (N-1)m}$, $\|R_d\| = 1$, $\|R_m\| = 1$, and

$$R_d R_d^\top = I - \frac{\mathbf{1}_{N,d}\mathbf{1}_{N,d}^\top}{N}, \; R_m R_m^\top = I - \frac{\mathbf{1}_{N,m}\mathbf{1}_{N,m}^\top}{N}. \tag{36}$$

As in the proof of Theorem IV.3, we use the initialization $\mathbf{1}_{N,d}^\top z^0 = 0$ and $\mathbf{1}_{N,m}^\top y^0 = 0$ to ensure that $\bar{z}^t = 0$ and $\bar{y}^t = 0$ for all $t \geq 0$. In view of (35), we can therefore rewrite (34) by ignoring the dynamics of $\bar{z}^t$ and $\bar{y}^t$, thus obtaining the system

$$\chi^{t+1} = \chi^t + \delta f(\chi^t, w^t), \tag{37a}$$
$$w^{t+1} = Sw^t + K(\delta, \gamma)u(\chi^t). \tag{37b}$$

in which

$$\chi^t := \begin{bmatrix} x^t \\ \bar{\lambda}^t \end{bmatrix}, \quad w^t := \begin{bmatrix} \lambda_\perp^t \\ z_\perp^t \\ y_\perp^t \end{bmatrix}, \tag{38a}$$

$$f(\chi^t, w^t) := \begin{bmatrix} f_X(x^t, \mathbf{1}_{N,m}\bar{\lambda}^t + R_m\lambda_\perp^t, R_d z_\perp^t, R_m y_\perp^t) \\ \gamma\frac{\mathbf{1}_{N,m}^\top}{N} G_\lambda(N(\bar{A}x^t - \bar{b}) + R_m y_\perp^t, \mathbf{1}_{N,m}\bar{\lambda}^t + R_m\lambda_\perp^t) \end{bmatrix}, \tag{38b}$$

$$S := \begin{bmatrix} R_m^\top \mathcal{W}_m R_m & 0 & 0 \\ 0 & R_d^\top \mathcal{W}_d R_d & 0 \\ 0 & 0 & R_m^\top \mathcal{W}_m R_m \end{bmatrix}, \tag{38c}$$

$$K(\delta, \gamma) := \begin{bmatrix} \delta\gamma R_m^\top & 0 & 0 \\ 0 & R_d^\top(\mathcal{W}_d - I) & 0 \\ 0 & 0 & R_m^\top(\mathcal{W}_m - I) \end{bmatrix}, \tag{38d}$$

$$u(\chi^t) := \begin{bmatrix} G_\lambda(N(\bar{A}x^t - \bar{b}) + R_m y_\perp^t, \mathbf{1}_{N,m}\bar{\lambda}^t + R_m\lambda_\perp^t) \\ \phi(x^t) \\ N(\bar{A}x^t - \bar{b}) \end{bmatrix}. \tag{38e}$$

where We view (37) as a singularly perturbed system, namely the interconnection between the slow dynamics (37a) and the fast one (37b). Indeed, system (37) can be obtained from (5) by considering χ^t as the state of (5a) and setting

$$g(\chi^t, w^t, \delta) := Sw^t + K(\delta, \gamma)u(\chi^t). \tag{39}$$

2. Equilibrium function h*:* Under the double stochasticity condition of $\mathcal{W}$ due to Standing Assumption II.4 and using (36), for any $\chi^t = \chi$,

$$h(\chi) := \begin{bmatrix} 0 \\ -R_d^\top\phi\left(\begin{bmatrix} I_n & 0 \end{bmatrix}\chi\right) \\ -R_m^\top N\left(\bar{A}\begin{bmatrix} I_n & 0 \end{bmatrix}\chi - \bar{b}\right) \end{bmatrix} \tag{40}$$

constitutes an equilibrium of (37b) (parametrized by x).

3. Boundary layer system and satisfaction of (8)*:* The so-called boundary layer system associated to (37) can be constructed by fixing $\chi^t = \chi = \mathrm{COL}(x, \bar{\lambda})$ for some arbitrary $(x, \bar{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^m$, and rewriting it according to the error coordinates $\tilde{w} := \mathrm{COL}(\tilde{\lambda}_\perp, \tilde{z}_\perp, \tilde{y}_\perp) := w - h(\chi)$. Using (36), we then obtain that

$$\tilde{w}^{t+1} = S\tilde{w}^t + \delta\gamma\tilde{u}(\chi, \tilde{w}^t), \tag{41}$$

where

$$\tilde{u}(\chi, \tilde{w}^t) := \begin{bmatrix} R_m^\top G_\lambda\left(\mathbf{1}_{N,m}(Ax - b) + R_m\tilde{y}_\perp^t, \mathbf{1}_{N,m}\bar{\lambda} + R_m\tilde{\lambda}_\perp^t\right) \\ 0 \\ 0 \end{bmatrix}.$$

The next lemma provides a Lyapunov function for (41).

Lemma IV.4. *Consider system* (41). *Then, there exists a continuous function* $U : \mathbb{R}^{(N-1)(2m+d)} \to \mathbb{R}$ *and* $\bar{\delta}_1 > 0$ *such that for any* $\delta \in (0, \bar{\delta}_1)$ *and any* $\gamma > 0$*,* U *satisfies* (8) *with* $\tilde{w}$ *in place of* ψ*.*

4. Reduced system and satisfaction of (9)*:* The so-called reduced system can be obtained by considering the fast dynamics in (37a) at steady state, i.e., $w^t = h(\chi^t)$ for any $t \geq 0$. We thus have

$$\chi^{t+1} = \chi^t + \delta f(\chi^t, h(\chi^t)). \tag{42}$$

Let us expand (42). Using (36), we obtain

$$\begin{aligned} x^{t+1} = x^t + \delta\bigg(& P_X\left[x^t - \gamma \tilde{F}\left(x^t, \mathbf{1}_{N,d}\sigma(x^t)\right)\right. \\ & \left. - \gamma G_x\left(\mathbf{1}_{N,m}(Ax^t - b), \mathbf{1}_{N,m}\bar{\lambda}^t\right)\right] - x^t\bigg), \end{aligned} \tag{43a}$$

$$\bar{\lambda}^{t+1} = \bar{\lambda}^t + \delta\gamma \frac{\mathbf{1}_{N,m}^\top}{N} G_\lambda\left(\mathbf{1}_{N,m}(Ax^t - b), \mathbf{1}_{N,m}\bar{\lambda}^t\right). \tag{43b}$$

Notice that

$$\tilde{F}\left(x, \mathbf{1}_{N,d}\sigma(x)\right) = F(x),$$
$$G_x\left(\mathbf{1}_{N,m}(Ax^t - b), \mathbf{1}_{N,m}\bar{\lambda}^t\right) = \nabla_x H(Ax^t - b, \bar{\lambda}^t),$$

and also

$$\frac{\mathbf{1}_{N,m}^\top}{N} G_\lambda\left(\mathbf{1}_{N,m}(Ax^t - b), \mathbf{1}_{N,m}\bar{\lambda}^t\right) = \nabla_\lambda H(Ax^t - b, \bar{\lambda}^t).$$

Therefore, (42) is identical to the original update (30). Given the unique v-GNE $x^\star$ of (1) (see Assumptions IV.1, IV.2) and the associated multiplier $\lambda^\star \in \mathbb{R}^m$, the next lemma provides a Lyapunov function for (42), hence for (30).

Lemma IV.5. *Consider system* (42) *and Assumptions IV.1, IV.2. Then, there exist a continuous function* $W : \mathbb{R}^{n+m} \to \mathbb{R}$*,* $\bar{\delta} > 0$*, and* $\bar{\gamma} > 0$ *such that for any* $\delta \in (0, \bar{\delta})$ *and* $\gamma \in (0, \bar{\gamma})$*,* W *satisfies* (9) *with* χ *in place of* x*.*

5. Lipschitz continuity of f, g and h: As we will be invoking Theorem II.5, we need to ensure that the required Lipschitz continuity assumptions are satisfied. In particular, we need to show that f, g in (38b) and (39), respectively, and h in (40) are Lipschitz with respect to their arguments. This is guaranteed by the Lipschitz continuity of the aggregation rules and the gradients of the cost functions (cf. Standing Assumption II.3), the nonexpansiveness of the projection operator (since X is closed and convex, see Standing Assumption II.2), and the Lipschitz continuity of G_x and G_λ (that appear in f and g), which is ensured as shown in (63) within the proof of LemmaIV.4.

By combining Lemmas IV.4 and IV.5 with the Lipschitz continuity properties expressed above, Theorem II.5 can be applied. Then, there exists $\bar{\delta} \in (0, \min(\bar{\delta}_1, \bar{\delta}_2))$ so that, for any $\delta \in (0, \bar{\delta})$, $\text{COL}(x^\star, \lambda^\star, h(x^\star, \lambda^\star))$ is an exponentially stable equilibrium point for (37). ■

V. Numerical Examples

We demonstrate the efficacy of Primal TRADES and Primal-Dual TRADES, and compare them with the most closely related distributed equilibrium seeking algorithms from the literature. First, we consider the case with local constraints only, and then we focus also on problems with coupling constraints.

A. Example without coupling constraints

In this subsection, we consider an instance of problem (10) and perform a numerical simulations in which we compare Primal TRADES with Algorithm 2 proposed in [33].

We consider the multi-agent demand response problem considered in [33]. Consider N loads whose electricity consumption $x_i := \text{COL}(x_{i,1}, \ldots, x_{i,T}) \in \mathbb{R}^T$ with $T \in \mathbb{N}$ has to be chosen to solve

$$\forall i \in \mathcal{I} : \min_{x_i \in X_i} \; \rho_i \left\|x_i - \hat{u}_i\right\|^2 + (\lambda\sigma(x) + p_0)^\top x_i,$$

where $\hat{u}_i \in \mathbb{R}^T$ denotes some nominal energy profile, $\rho_i > 0$ is a constant weighting parameter, the term $\lambda\sigma(x) + p_0$ with $\lambda \in \mathbb{R}$, $p_0 \in \mathbb{R}^T$ models the unit price which is taken to be an affine increasing function of the aggregate (average) energy demand $\sigma(x) = (1/N)\sum_{i \in \mathcal{I}} x_i$. As for the local feasible set $X_i \subseteq \mathbb{R}^T$, for all $i \in \mathcal{I}$, we pick

$$X_i := \bigg\{ x_i \in \mathbb{R}^T \; | s_{i,\tau+1}(x_i) \in \mathcal{S}_i \text{ and } x_{i,\tau} \in \mathcal{U}_i \; \forall \tau \in \{1, \ldots, T\}, \; \sum_{\tau=1}^{T} x_{i,\tau} = \sum_{\tau=1}^{T} \hat{u}_{i,\tau} \bigg\},$$

where $\mathcal{U}_i \subseteq \mathbb{R}$, $\mathcal{S}_i \subseteq \mathbb{R}$, and $s_{i,\tau}(x_i)$ is the state of the i-th load at time τ that, given the parameters $a_i, b_i \in \mathbb{R}$, is computed according to the linear dynamics

$$s_{i,\tau} = a_i^{\tau-1} s_{i,1} + \sum_{k=1}^{\tau-1} a^{k-1} b_i x_{i,\tau-k},$$

where $s_{i,1} \in \mathcal{S}_i$ is the initial condition of the state of the i-th load. To instantiate the problem, we set $T = 24$ and randomly generate values for $\hat{u}_i$, ρ_i, λ, p_0, a_i, b_i, $s_{i,1}$ and initial strategies $x_{i,1}$ from uniform distributions. As for the sets $\mathcal{U}_i$ and $\mathcal{S}_i$, we pick the intervals $[0, 1]$ and $[0, 10]$, respectively. We consider a network with $N = 10$ players communicating according to an undirected, connected Erdős-Rényi graph with parameter 0.3.

This setting satisfies our standing assumptions. We compare our scheme, namely, Primal TRADES with Algorithm 2 in [33]. We tune the latter with $v_1 = v_2 = 50$ communication rounds per iterate and update the auxiliary variable z^t according to $z^{t+1} = (1 - \lambda)z^t + \lambda\mathcal{A}_{v_1,v_2}$ with $\lambda = 0.01$ (the quantity $\mathcal{A}_{v_1,v_2}$ is a proxy for the unavailable aggregative variable $\sigma(x)$, see [33] for more details.) As for the parameters of our scheme, we set $\delta = \gamma = 0.1$. Fig. 1 shows the evolution of the normalized distance $\|x^t - x^\star\| / \|x^\star\|$ from the NE $x^\star$ as the communication rounds (corresponding to iterations) progress. Our algorithm exhibits faster convergence and achieves higher accuracy in the calculation of the equilibrium $x^\star$. This was anticipated as the method in [33] is not guaranteed to converge to the exact NE (see Table I).

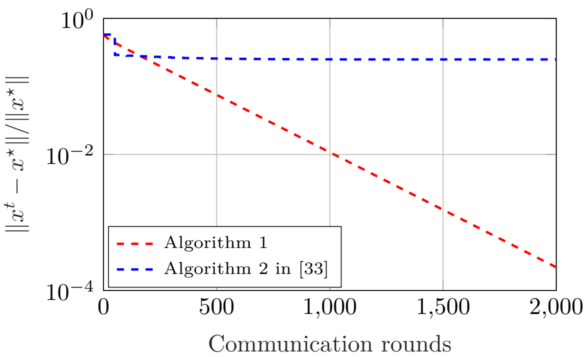


Fig. 1: Comparison in terms of the normalized distance of the iterates from the NE between Primal TRADES (Algorithm 1) and the algorithm by [33], on a case study introduced in [33].

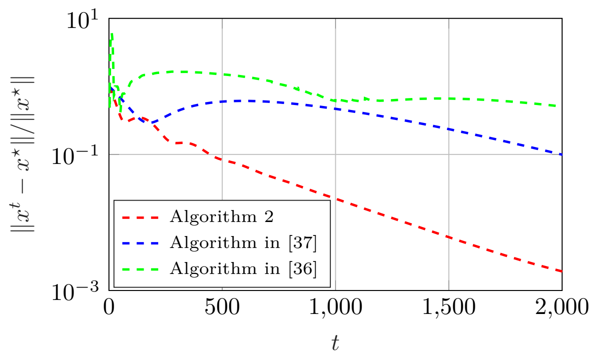


Fig. 2: Comparison in terms of the normalized distance of the iterates from the GNE between Primal-Dual TRADES (Algorithm 2), and the algorithms by [36] and [37], on a case study introduced in [36].

B. Example with coupling constraints

We address two Nash-Cournot games formulated as in (1) to compare our Primal-Dual TRADES algorithm with the distributed methods proposed in [36] and [37]. For a fair comparison we test the scheme by [36] with a constant step-size even if convergence was theoretically proven only with a diminishing one (see Table II); note that slower convergence is expected by using a diminishing step-size.

We first compare the algorithms in [36] and [37] with Algorithm 2 on a case study from [36]. Consider N firms that compete over n_m markets. In particular, for each market $k \in \mathcal{M} := \{1, \dots, n_m\}$, firm i is characterized by a production $g_{i,k} \geq 0$ and sales $s_{i,k} \geq 0$. For each $i \in \mathcal{I}$ and $k \in \mathcal{M}$, the cost of production amounts to

$$f_{i,k}(g_{i,k}) = q_{i,k} g_{i,k}^2 + c_{i,k} g_{i,k}.$$

The revenue of firm i at market k is modelled as $(d_k - \bar{s}_k)s_{i,k}$, where $d_k > 0$ is the total demand for location k, and $\bar{s}_k := \sum_{i\in\mathcal{I}} s_{i,k}$ represents the aggregate sales at location k. For all firms $i \in \mathcal{I}$ and markets $k \in \mathcal{M}$, we assume a production limitation $u_{i,k}$. Moreover, in each market k, the total production $\sum_{i\in\mathcal{I}} g_{i,k}$ must cover the demand d_k without exceeding a maximum capacity r_k. We can thus cast this setting as an instance of the GNEP in (1) with each strategy vector given by $x_i := \mathrm{COL}(g_{i,1}, \dots, g_{i,n_m}, s_{i,1}, \dots, s_{i,n_m}) \in \mathbb{R}^{2n_m}$, and cost function

$$J_i(x, \sigma(x)) = x_i^\top Q_i x_i + \ell_i^\top x_i + (\Delta\sigma(x))^\top x_i,$$

where $Q_i := \mathrm{diag}(q_{i,1}, \dots, q_{i,n_m}, 0, \dots, 0) \in \mathbb{R}^{2n_m \times 2n_m}$, $\ell_i := \mathrm{COL}(c_{i,1}, \dots, c_{i,n_m}, -d_1, \dots, -d_{n_m}) \in \mathbb{R}^{2n_m}$, $\Delta = \mathrm{blkdiag}(0_{n_m}, N I_{n_m})$, and aggregation rule $\phi_i = I$ for all $i \in \{1, \dots, N\}$. As for the constraints, for all $i \in \mathcal{I}$, we have the local constraint set $X_i := \{x_i \in \mathbb{R}^{2n_m} \mid \begin{bmatrix} -1_{2n_m}^\top & 1_{2n_m}^\top \end{bmatrix} x_i \leq 0, 0 \leq g_{i,k} \leq u_{i,k}, 0 \leq s_{i,k}, k = 1, \dots, n_m\}$, and the coupling constraints defined by $b_i := \frac{1}{N}\mathrm{COL}(r_1, \dots, r_{n_m}, -d_1, \dots, -d_{n_m})$ and

$$A_i := \begin{bmatrix} I_{n_m} & 0_{n_m} \\ -I_{n_m} & 0_{n_m} \end{bmatrix}.$$

Following [36], we choose $N = 20$, $n_m = 10$, an undirected and connected graph with doubly stochastic weighted adjacency

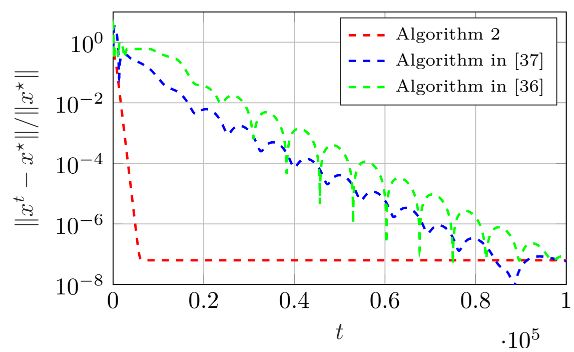


Fig. 3: Comparison in terms of the normalized distance of the iterates from the GNE between Primal-Dual TRADES (Algorithm 2), and the algorithms by [36] and [37], on a case study introduced in [37].

matrix chosen according to the Metropolis rule, and we generate values for the parameters of the problem from uniform distributions. Note that this game satisfies our standing assumptions. In particular, for all $i \in \mathcal{I}$ and $k \in \mathcal{M}$, we pick $q_{i,k} \in [2, 3]$, $c_{i,k} \in [2, 12]$, $u_{i,k} \in [50, 100]$, $d_k \in [90, 100]$, and $r_k \in [d_k, 2d_k]$. We tune the algorithm as suggested in [36], i.e., with $\delta = \min(1, 1/L)$, $\tau = 1.05/(2\delta)$, $\gamma = 1$, $\alpha_i \leq 0.95/(\|A_i + \tau\|)$, and $\beta_i \leq 1/(\|A\| + \tau)$ for all $i \in \{1, \dots, N\}$, where $L > 0$ denotes the Lipschitz constant of the pseudo-gradient of the problem. To instantiate the algorithm in [37], we choose $c = 4$, $k = 1/200$, $\tau = 1/800$, $\alpha = 1/120$, and $v = 1/120$, while we implement our scheme with $\delta = 0.25$, $\gamma = 0.01$, and $\rho = 0.1$. Fig. 2 compares the performance of these algorithms with our proposed Algorithm 2 in terms of the normalized distance $\|x^t - x^\star\| \,/\, \|x^\star\|$ from the GNE $x^\star$. We observe from Fig. 2 that Algorithm 2 outperforms the others in terms of accuracy and convergence speed.

We now focus on the case study considered in [37]. Specifically, we consider a Nash-Cournot game over a network for a single market with production constraints and globally coupling capacity constraints, which can be formulated as an instance of (1). In particular, we consider $N = 20$, $X_i = [0, 10]$

for all $i \in \mathcal{I}$, $A = [1 \ldots 1]$, $b = 20$, and the cost function

$$J_i(x_i, \sigma(x)) = (1 + 2(i-1))x_i - x_i \left(60 - \sigma(x) - \frac{1}{2}x_i\right).$$

As in [37], we consider a graph with ring topology. To achieve a fair comparison with [37], we follow the authors' tuning and choose $c = 4$, $k = 1/200$, $\tau = 1/800$, $\alpha = 1/120$, and $v = 1/120$, while we tune the scheme in [36] as above. As for the parameters of our algorithm, we empirically tune them as $\delta = \gamma = \rho = 0.1$. In Fig. 3, we compare the performance of the algorithms in [36] and [37] with Algorithm 2 in terms of the normalized distance $\|x^t - x^\star\| / \|x^\star\|$ from the GNE $x^\star$. Also in this case the proposed scheme exhibits faster convergence.

VI. Conclusions

We proposed two novel fully-distributed algorithms for (generalized) equilibrium seeking in aggregative games over networks. The first algorithm is designed to address the case where only local constraints are present. The second method, instead, can also handle coupling constraints which characterize generalized Nash games. Both schemes are studied by means of singular perturbations analysis in which slow and fast dynamics are identified and separately investigated to demonstrate the linear convergence of the whole interconnection to the (generalized) Nash equilibrium. Finally, we performed detailed numerical simulations illustrating the effectiveness of the proposed methods, also showing that they outperform state-of-the-art distributed algorithms.

Appendix

A. Proof of Theorem II.5

Define $\tilde{w}^t := w^t - h(x^t)$ and, in accordance, rewrite system (5) as

$$x^{t+1} = x^t + \delta f(x^t, \tilde{w}^t + h(x^t)) \tag{44a}$$
$$\tilde{w}^{t+1} = g(\tilde{w}^t + h(x^t), x^t, \delta) - h(x^t) + \Delta h(x^{t+1}, x^t), \tag{44b}$$

where $\Delta h(x^{t+1}, x^t) := -h(x^{t+1}) + h(x^t)$. Pick W as in (9). By evaluating $\Delta W(x^t) := W(x^{t+1}) - W(x^t)$ along the trajectories of (44a), we obtain

$$\begin{aligned}
\Delta W(x^t) &= W(x^t + \delta f(x^t, \tilde{w}^t + h(x^t))) - W(x^t) \\
&\overset{(a)}{=} W(x^t + \delta f(x^t, h(x^t))) - W(x^t) \\
&\quad + W(x^t + \delta f(x^t, \tilde{w}^t + h(x^t))) \\
&\quad - W(x^t + \delta f(x^t, h(x^t))) \\
&\overset{(b)}{\leq} -c_3 \left\|x^t - x^\star\right\|^2 + W(x^t + \delta f(x^t, \tilde{w}^t + h(x^t))) \\
&\quad - W(x^t + \delta f(x^t, h(x^t))) \\
&\overset{(c)}{\leq} -c_3 \left\|x^t - x^\star\right\|^2 + 2\delta c_4 L_f \left\|\tilde{w}^t\right\| \left\|x^t - x^\star\right\| \\
&\quad + \delta^2 c_4 L_f \left\|\tilde{w}^t\right\| \left\|f(x^t, \tilde{w}^t + h(x^t))\right\| \\
&\quad + \delta^2 c_4 L_f \left\|\tilde{w}^t\right\| \left\|f(x^t, h(x^t))\right\|,
\end{aligned} \tag{45}$$

where in (a) we add and subtract the term $W(x^t + \delta f(x^t, h(x^t)))$, in (b) we exploit (9b) to bound the difference of the first two terms, in (c) we use (9c), the Lipschitz continuity of f, and the triangle inequality. By recalling that $f(x^\star, h(x^\star)) = 0$ we can thus write

$$\begin{aligned}
\|f(x^t, \tilde{w}^t + h(x^t))\| &= \left\|f(x^t, \tilde{w}^t + h(x^t)) - f(x^\star, h(x^\star))\right\| \\
&\overset{(a)}{\leq} L_f \left\|x^t - x^\star\right\| + L_f \left\|\tilde{w}^t + h(x^t) - h(x^\star)\right\|, \\
&\overset{(b)}{\leq} L_f(1 + L_h) \left\|x^t - x^\star\right\| + L_f \left\|\tilde{w}^t\right\|,
\end{aligned} \tag{46}$$

where in (a) we use the Lipschitz continuity of f and h, and in (b) we use the Lipschitz continuity of h together with the triangle inequality. With similar arguments, we have

$$\left\|f(x^t, h(x^t))\right\| \leq L_f(1 + L_h) \left\|x^t - x^\star\right\|. \tag{47}$$

Using inequalities (46) and (47) we then bound (45) as

$$\begin{aligned}
\Delta W(x^t) &\leq -c_3 \left\|x^t - x^\star\right\|^2 + 2\delta c_4 L_f \left\|\tilde{w}^t\right\| \left\|x^t - x^\star\right\| \\
&\quad + \delta^2 c_4 L_f^2 \left\|\tilde{w}^t\right\|^2 \\
&\quad + 2\delta^2 c_4 L_f^2 (1 + L_h) \left\|\tilde{w}^t\right\| \left\|x^t - x^\star\right\| \\
&\leq -c_3 \left\|x^t - x^\star\right\|^2 + \delta^2 k_3 \left\|\tilde{w}^t\right\|^2 \\
&\quad + (\delta k_1 + \delta^2 k_2) \left\|\tilde{w}^t\right\| \left\|x^t - x^\star\right\|,
\end{aligned} \tag{48}$$

where we introduce the constants

$$k_1 := 2c_4 L_f, \quad k_2 := 2c_4 L_f^2 (1 + L_h), \quad k_3 := c_4 L_f^2.$$

We now pick U as in (8). By evaluating $\Delta U(\tilde{w}^t) := U(\tilde{w}^{t+1}) - U(\tilde{w}^t)$ along the trajectories of (44b), we obtain

$$\begin{aligned}
&\Delta U(\tilde{w}) \\
&= U(g(\tilde{w}^t + h(x^t), x^t, \delta) - h(x^t) + \Delta h(x^{t+1}, x^t)) - U(\tilde{w}^t) \\
&\overset{(a)}{\leq} U(g(\tilde{w}^t + h(x^t), x^t, \delta) - h(x^t)) - U(\tilde{w}^t) \\
&\quad - U(g(\tilde{w}^t + h(x^t), x^t, \delta) - h(x^t)) \\
&\quad + U(g(\tilde{w}^t + h(x^t), x^t, \delta) - h(x^t) + \Delta h(x^{t+1}, x^t)) \\
&\overset{(b)}{\leq} -b_3 \left\|\tilde{w}^t\right\|^2 - U(g(\tilde{w}^t + h(x^t), x^t, \delta) - h(x^t)) \\
&\quad + U(g(\tilde{w}^t + h(x^t), x^t, \delta) - h(x^t) + \Delta h(x^{t+1}, x^t)) \\
&\overset{(c)}{\leq} -b_3 \left\|\tilde{w}^t\right\|^2 \\
&\quad + b_4 \left\|\Delta h(x^{t+1}, x^t)\right\| \\
&\qquad \times \left\|g(\tilde{w}^t + h(x^t), x^t, \delta) - h(x^t) + \Delta h(x^{t+1}, x^t)\right\| \\
&\quad + b_4 \left\|\Delta h(x^{t+1}, x^t)\right\| \left\|g(\tilde{w}^t + h(x^t), x^t, \delta\right\|) - h(x^t) \\
&\overset{(d)}{\leq} -b_3 \left\|\tilde{w}^t\right\|^2 + b_4 \left\|\Delta h(x^{t+1}, x^t)\right\|^2 \\
&\quad + 2b_4 \left\|\Delta h(x^{t+1}, x^t)\right\| \left\|g(\tilde{w}^t + h(x^t), x^t, \delta) - h(x^t)\right\|,
\end{aligned} \tag{49}$$

where in (a) we add and subtract $U(g(\tilde{w}^t + h(x^t), x^t, \delta) - h(x^t))$, in (b) we exploit (8b) to bound the first two terms, in (c) we use (8c) to bound the the difference of the last two terms, and in (d) we use the triangle inequality. By exploiting the definition of $\Delta h(x^{t+1}, x^t)$ and the Lipschitz continuity of

h, we have that

$$\begin{aligned}\left\|\Delta h(x^{t+1}, x^t)\right\| &\le L_h \left\|x^{t+1} - x^t\right\| \\ &\overset{(a)}{\le} \delta L_h \left\|f(x^t, \tilde{w}^t + h(x^t))\right\| \\ &\overset{(b)}{\le} \delta L_h \left\|f(x^t, \tilde{w}^t + h(x^t)) - f(x^\star, h(x^\star))\right\| \\ &\overset{(c)}{\le} \delta L_h L_f (1+L_h)\left\|x^t - x^\star\right\| + \delta L_h L_f \left\|\tilde{w}^t\right\|,\end{aligned} \tag{50}$$

where in (a) we use the update (44a), in (b) we add the term $f(x^\star, h(x^\star))$ since this is zero, and in (c) we use the triangle inequality and the Lipschitz continuity of f and h. Moreover, since $g(h(x^t), x^t, \delta) = h(x^t)$, we obtain

$$\begin{aligned}&\left\|g(\tilde{w}^t + h(x^t), x^t, \delta) - h(x^t)\right\| \\ &= \left\|g(\tilde{w}^t + h(x^t), x^t, \delta) - g(h(x^t), x^t, \delta)\right\| \le L_g \left\|\tilde{w}^t\right\|,\end{aligned} \tag{51}$$

where the inequality is due to the Lipschitz continuity of g. Using inequalities (50) and (51), we then bound (49) as

$$\begin{aligned}&\Delta U(\tilde{w}) \\ &\le -b_3 \left\|\tilde{w}^t\right\|^2 + 2\delta b_4 L_h L_g L_f (1 + L_h) \left\|x^t - x^\star\right\| \left\|\tilde{w}^t\right\| \\ &\quad + 2\delta b_4 L_h L_g L_f \left\|\tilde{w}^t\right\|^2 + \delta^2 b_4 L_h^2 L_f^2 (1 + L_h)^2 \left\|x^t - x^\star\right\|^2 \\ &\quad + 2\delta^2 b_4 L_h^2 L_f^2 (1 + L_h) \left\|x^t - x^\star\right\| \left\|\tilde{w}^t\right\| + \delta^2 b_4 L_h^2 L_f^2 \left\|\tilde{w}^t\right\|^2 \\ &\le (-b_3 + \delta k_6 + \delta^2 k_7) \left\|\tilde{w}^t\right\|^2 + \delta^2 k_8 \left\|x^t - x^\star\right\|^2 \\ &\quad + (\delta k_4 + \delta^2 k_5) \left\|x^t - x^\star\right\| \left\|\tilde{w}^t\right\|,\end{aligned} \tag{52}$$

where we introduce the constants

$$\begin{aligned}&k_4 := 2b_4 L_h L_g L_f (1 + L_h), && k_5 := 2b_4 L_h^2 L_f^2 (1 + L_h), \\ &k_6 := 2b_4 L_h L_g L_f, && k_7 := b_4 L_h^2 L_f^2, \\ &k_8 := b_4 L_h^2 L_f^2 (1 + L_h)^2.\end{aligned}$$

We pick the following Lyapunov candidate $V : \mathcal{D} \times \mathbb{R}^m \to \mathbb{R}$:

$$V(x^t, \tilde{w}^t) = W(x^t) + U(\tilde{w}^t).$$

By evaluating $\Delta V(x^t, \tilde{w}^t) := V(x^{t+1}, \tilde{w}^{t+1}) - V(x^t, \tilde{w}^t) = \Delta W(x^t) + \Delta U(\tilde{w}^t)$ along the trajectories of (44), we can use the results (48) and (52) to write

$$\Delta V(x^t, \tilde{w}^t) \le - \begin{bmatrix}\|x^t - x^\star\| \\ \|\tilde{w}^t\|\end{bmatrix}^\top Q(\delta) \begin{bmatrix}\|x^t - x^\star\| \\ \|\tilde{w}^t\|\end{bmatrix}, \tag{53}$$

where we define the matrix $Q(\delta) = Q(\delta)^\top \in \mathbb{R}^2$ as

$$Q(\delta) := \begin{bmatrix} c_3 - \delta^2 k_8 & q_{21} \\ q_{21} & b_3 - \delta k_6 - \delta^2 (k_3 + k_7)\end{bmatrix},$$

with $q_{21} := -\frac{1}{2}(\delta(k_1 + k_4) + \delta^2(k_2 + k_5))$. By relying on the Sylvester criterion [40], we know that $Q \succ 0$ if and only if

$$c_3 b_3 > p(\delta) \tag{54}$$

where the polynomial $p(\delta)$ is defined as

$$\begin{aligned}p(\delta) := q_{21}^2 + \delta c_3 k_6 + \delta^2(c_3(k_3 + k_7) + b_3 k_8) \\ - \delta^3 k_6 k_8 - \delta^4 k_8 (k_3 + k_7).\end{aligned} \tag{55}$$

We note that p is a continuous function of δ and $\lim_{\delta \to 0} p(\delta) = 0$. Hence, there exists some $\bar{\delta} \in (0, \min\{\bar{\delta}_1, \bar{\delta}_2\})$ – recall that $\bar{\delta}_1$ and $\bar{\delta}_2$ exist as U and W are taken to satisfy (8) and (9) – so that (54) is satisfied for any $\delta \in (0, \bar{\delta})$. Under such a choice of δ, and denoting by $q > 0$ the smallest eigenvalue of $Q(\delta)$, we can bound (53) as

$$\Delta V(x^t, \tilde{w}^t) \le -q \left\| \begin{bmatrix}\|x^t - x^\star\| \\ \|\tilde{w}^t\|\end{bmatrix}\right\|^2,$$

which allows us to conclude, in view of [42, Theorem 13.2], that $(x^\star, 0)$ is an exponentially stable equilibrium point for system (44). The theorem's conclusion follows then by considering the definition of exponentially stable equilibrium point and by reverting to the original coordinates (x^t, w^t). ■

B. Proofs of technical lemmas of Section III-B

Proof of Lemma III.2: System (23) is a linear autonomous system whose state matrix $R_d^\top \mathcal{W}_d R_d \in \mathbb{R}^{(N-1)d \times (N-1)d}$ is Schur. Hence, there exists $P \in \mathbb{R}^{(N-1)d \times (N-1)d}$, $P = P^\top \succ 0$ for the candidate Lyapunov function $U(\tilde{z}^t) = (\tilde{z}^t)^\top P \tilde{z}^t$, solving the Lyapunov equation

$$(R_d^\top \mathcal{W}_d R_d)^\top P R_d^\top \mathcal{W}_d R_d - P = -Q. \tag{56}$$

for any $Q \in \mathbb{R}^{(N-1)d \times (N-1)d}$, $Q = Q^\top \succ 0$. Condition (8a) follows then from the fact that U is quadratic with $P \succ 0$ so b_1 and b_2 can be chosen to be its minimum and maximum eigenvalue, respectively. The left-hand side of (8b) becomes $(\tilde{z}^t)^\top((R_d^\top \mathcal{W}_d R_d)^\top P R_d^\top \mathcal{W}_d R_d - P)\tilde{z}^t = -(\tilde{z}^t)^\top Q \tilde{z}^t$, where the equality is due to (56). Hence, (8b) is satisfied by taking b_3 to be the smallest eigenvalue of Q. To see (8c) notice that

$$\begin{aligned}&\left\|U(\tilde{z}_1^t) - U(\tilde{z}_2^t)\right\| = \left\|(\tilde{z}_1^t)^\top P \tilde{z}_1^t - (\tilde{z}_2^t)^\top P \tilde{z}_2^t\right\| \\ &\le \left\|(\tilde{z}_1^t)^\top P \tilde{z}_1^t - (\tilde{z}_1^t)^\top P \tilde{z}_2^t\right\| + \left\|(\tilde{z}_2^t)^\top P \tilde{z}_1^t - (\tilde{z}_2^t)^\top P \tilde{z}_2^t\right\| \\ &\le \|P\| \left\|\tilde{z}_1^t - \tilde{z}_2^t\right\| \left\|\tilde{z}_1^t\right\| + \|P\| \left\|\tilde{z}_1^t - \tilde{z}_2^t\right\| \left\|\tilde{z}_2^t\right\|\end{aligned} \tag{57}$$

where the first inequality follows from adding and subtracting $(\tilde{z}_1^t)^\top P \tilde{z}_2^t$ and using the triangle inequality, while the second one from the Cauchy-Schwarz inequality. The bound in (8c) follows then from (57) by taking b_4 to be the largest eigenvalue of P (recall it is symmetric). ■

We provide here the following technical lemma which is used in the proof of Lemma III.3.

Lemma B.1 (Contraction of strongly monotone operator)**.** *Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be μ-strongly monotone and L-Lipschitz continuous. If $\gamma \in (0, 2\mu/L^2)$, then for any $x, x' \in \mathbb{R}^n$ it holds that*

$$\|x - \gamma F(x) - x' + \gamma F(x')\| \le (1 - \bar{\mu}) \|x - x'\|,$$

where $\bar{\mu} := 1 - \sqrt{1 - \gamma(2\mu - \gamma L^2)} \in (0, 1]$. □

Proof. We have that

$$\begin{aligned}&\|x - \gamma F(x) - x' + \gamma F(x')\|^2 \\ &= \|x - x'\|^2 + \gamma^2 \|F(x) - F(x')\|^2 \\ &\quad - 2\gamma(x - x')^\top (F(x) - F(x')) \\ &\overset{(a)}{\le} \|x - x'\|^2 - \gamma(2\mu - \gamma L^2) \|x - x'\|^2,\end{aligned} \tag{58}$$

where in (a) we use the strong monotonicity and the Lipschitz continuity of F. By construction, $\bar{\mu} \in (0, 1]$ is equivalent to

$\gamma(2\mu-\gamma L^2)>0$ and $\gamma(2\mu-\gamma L^2)\le 1$. The former holds since $\gamma\in(0,2\mu/L^2)$. To see the latter, notice that, by definition of μ-strong monotonicity and L-Lipschitz continuity, we have

$$\begin{aligned}\mu\left\|x-x'\right\|^2 &\le (F(x)-F(x'))^\top(x-x')\\ &\le \|F(x)-F(x')\|\,\|x-x'\|\le L\left\|x-x'\right\|^2,\end{aligned}$$

for any x,x', hence $\mu\le L$. Thus, for any γ, it holds that $1-2\mu\gamma+\gamma^2L^2\ge 1-2\gamma L+\gamma^2L^2=(1-\gamma L)^2\ge 0$. ■

Proof of Lemma III.3: Pick the function $W:\mathbb{R}^n\to\mathbb{R}$ defined as

$$W(x^t)=\frac{1}{2}\left\|x^t-x^\star\right\|^2.$$

Since W is a quadratic function, conditions (9a) and (9c) are satisfied. To show (9b) we evaluate $\Delta W(x^t):=W(x^{t+1})-W(x^t)$ along (25). We then have

$$\begin{aligned}&\Delta W(x^t)\\ &=\frac{1}{2}\left\|(1-\delta)x^t+\delta\left(P_X\left[x^t-\gamma F(x^t)\right]\right)-x^\star\right\|^2\\ &\quad-\frac{1}{2}\left\|x^t-x^\star\right\|^2\\ &\overset{(a)}{\le}\frac{(1-\delta)^2}{2}\left\|x^t-x^\star\right\|^2-\frac{1}{2}\left\|x^t-x^\star\right\|^2\\ &\quad+(\delta-\delta^2)\left\|x^t-x^\star\right\|\left\|P_X\left[x^t-\gamma F(x^t)\right]-P_X\left[x^\star-\gamma F(x^\star)\right]\right\|\\ &\quad+\frac{\delta^2}{2}\left\|P_X\left[x^t-\gamma F(x^t)\right]-P_X\left[x^\star-\gamma F(x^\star)\right]\right\|^2\\ &\overset{(b)}{\le}\frac{(1-\delta)^2}{2}\left\|x^t-x^\star\right\|^2-\frac{1}{2}\left\|x^t-x^\star\right\|^2\\ &\quad+(\delta-\delta^2)\left\|x^t-x^\star\right\|\left\|x^t-\gamma F(x^t)-x^\star+\gamma F(x^\star)\right\|\\ &\quad+\frac{\delta^2}{2}\left\|x^t-\gamma F(x^t)-x^\star+\gamma F(x^\star)\right\|^2,\end{aligned}\tag{59}$$

where in (a) we introduce the quantity $\delta(x^\star-P_X\left[x^\star-\gamma F(x^\star)\right])$ within the first norm, as this is zero due to the definition of $x^\star$, expand the square, and use the Cauchy-Schwarz inequality. Inequality (b) follows by the fact that for any a,b, we have that $\|P_X[a]-P_X[b]\|\le\|a-b\|$, since the projection operator is nonexpansive.

Since F is μ-strongly monotone and β_1 Lipschitz continuous (cf. Standing Assumption II.3), set $\bar\gamma=2\mu/(\beta_1)^2$ and choose $\gamma\in(0,\bar\gamma)$. Applying Lemma B.1 yields

$$\left\|x^t-\gamma F(x^t)-x^\star+\gamma F(x^\star)\right\|\le(1-\bar\mu)\left\|x^t-x^\star\right\|,\tag{60}$$

with $\bar\mu=1-\sqrt{1-\gamma(2\mu-\gamma(\beta_1)^2)}\in(0,1]$. Thus, by using the inequality in (60), we can bound (59) as follows

$$\begin{aligned}\Delta W(x^t)&\le\frac{(1-\delta)^2}{2}\left\|x^t-x^\star\right\|^2-\frac{1}{2}\left\|x^t-x^\star\right\|^2\\ &\quad+(\delta-\delta^2)(1-\bar\mu)\left\|x^t-x^\star\right\|^2\\ &\quad+\frac{\delta^2(1-\bar\mu)^2}{2}\left\|x^t-x^\star\right\|^2\\ &=-\delta\bar\mu(1-\frac{\delta\bar\mu}{2})\left\|x^t-x^\star\right\|^2.\end{aligned}\tag{61}$$

where the equality is obtained by rearranging the right-hand side of the inequality. Thus, for any $\delta\in(0,\bar\delta_2)$ with $\bar\delta_2:=2/\bar\mu$, $\delta\bar\mu(1-\delta\bar\mu/2)>0$ in (61), thus establishing condition (9b) and concluding the proof. ■

C. Proofs of technical lemmas of Section IV-B

Proof of Lemma IV.4: Since $R_m^\top\mathbf{1}_{N,m}=0$, we can write

$$\begin{aligned}&R_m^\top G_\lambda\left(\mathbf{1}_{N,m}(Ax-b)+R_m\tilde y_\perp^t,\mathbf{1}_{N,m}\bar\lambda+R_m\tilde\lambda_\perp^t\right)\\ &=R_m^\top\Big(G_\lambda\left(\mathbf{1}_{N,m}(Ax-b)+R_m\tilde y_\perp^t,\mathbf{1}_{N,m}\bar\lambda+R_m\tilde\lambda_\perp^t\right)\\ &\qquad-\mathbf{1}_{N,m}\nabla_\lambda H(Ax-b,\bar\lambda)\Big)\\ &=R_m^\top\Big(G_\lambda\left(\mathbf{1}_{N,m}(Ax-b)+R_m\tilde y_\perp^t,\mathbf{1}_{N,m}\bar\lambda+R_m\tilde\lambda_\perp^t\right)\\ &\qquad-G_\lambda(\mathbf{1}_{N,m}(Ax-b),\mathbf{1}_{N,m}\bar\lambda)\Big),\end{aligned}\tag{62}$$

where in the last equality we used $\mathbf{1}_{N,m}\nabla_\lambda H(Ax-b,\bar\lambda)=G_\lambda(\mathbf{1}_{N,m}(Ax-b),\mathbf{1}_{N,m}\bar\lambda)$. Following [38, Lemma 3], notice that, for any $r_1,r_2\in\mathbb{R}$, there exists $\epsilon(r_1,r_2)\in[0,1]$ so that[1]

$$\max\{r_1,0\}-\max\{r_2,0\}=\epsilon(r_1,r_2)(r_1-r_2).\tag{63}$$

Let us introduce

$$\begin{aligned}q_i^t&:=\sum_{\ell=1}^m[R_m\tilde y_\perp^t]_{\ell+(i-1)m}e_\ell\\ p_i^t&:=\sum_{\ell=1}^m[R_m\tilde\lambda_\perp^t]_{\ell+(i-1)m}e_\ell,\end{aligned}\tag{64}$$

and use them to define

$$\begin{aligned}r_{1,i}^t&:=\rho(Ax-b+q_i^t)+\bar\lambda+p_i^t\\ r_{2,i}&:=\rho(Ax-b)+\bar\lambda.\end{aligned}\tag{65}$$

By the definition of $\tilde u(\chi,\tilde w^t)$ we have that its norm $\|\tilde u(\chi,\tilde w^t)\|$ is equal to the norm of the quantity in (62). As such, for any $\chi\in\mathbb{R}^{n+m}$ and $\tilde w^t\in\mathbb{R}^{(N-1)(2m+d)}$, we use the definition of G_λ in (31), $r_{1,i}^t$ and $r_{2,i}$ in (65), and apply (63) for each component of $\tilde u(\chi,\tilde w^t)$ obtaining

$$\begin{aligned}&\left\|\tilde u(\chi,\tilde w^t)\right\|\\ &\le\left\|R_m^\top\frac{1}{\rho}\mathrm{COL}\Big(\sum_{\ell=1}^m\epsilon([r_{1,i}^t]_\ell,[r_{2,i}]_\ell)\big([r_{1,i}^t-\bar\lambda-p_i^t]_\ell\right.\\ &\qquad\left.-[r_{2,i}-\bar\lambda]_\ell\big)e_\ell\Big)_{i=1}^N\right\|\\ &\overset{(a)}{\le}\left\|R_m^\top\frac{1}{\rho}\mathrm{COL}\left(\sum_{\ell=1}^m\left([r_{1,i}^t-\bar\lambda-p_i^t]_\ell-[r_{2,i}-\bar\lambda]_\ell\right)e_\ell\right)_{i=1}^N\right\|\\ &\overset{(b)}{=}\left\|R_m^\top\frac{1}{\rho}\mathrm{COL}\left(\sum_{\ell=1}^m\rho[q_i^t]_\ell e_\ell\right)_{i=1}^N\right\|\\ &\overset{(c)}{=}\left\|R_m^\top R_m\tilde y_\perp^t\right\|\overset{(d)}{\le}\left\|\tilde w^t\right\|,\end{aligned}\tag{66}$$

where in (a) we use the fact that $\epsilon([r_{1,i}^t]_\ell,[r_{2,i}]_\ell)\in[0,1]$ for all $\ell\in\{1,\dots,m\}$ and $i\in\mathcal{I}$, (b) uses the definitions

[1] If $r_1\ne r_2$, pick $\epsilon=\frac{\max\{r_1,0\}-\max\{r_2,0\}}{r_1-r_2}$, otherwise set $\epsilon=0$.

in (65) to simplify the terms, (c) follows from (64), and (d) uses $R_m^\top R_m = I$ and $\|\tilde{y}_\perp^t\| \le \|\tilde{w}^t\|$ that holds since $\tilde{y}_\perp^t$ is a component of $\tilde{w}^t$.

Pick now $U : \mathbb{R}^{(N-1)(2m+d)} \to \mathbb{R}$ defined as

$$U(\tilde{w}) = (\tilde{w})^\top M \tilde{w},$$

where $M \in \mathbb{R}^{(N-1)(2m+d)\times(N-1)(2m+d)}$ with $M = M^\top \succ 0$, such that

$$S^\top M S - M = -I. \tag{67}$$

We remark that such a matrix M always exists because, in light of Standing Assumption II.4, both $R_d^\top \mathcal{W}_d R_d$ and $R_m^\top \mathcal{W}_m R_m$ are Schur matrices and, thus, S is Schur as well. Under this choice of U, conditions (8a) and (8c) are satisfied. To show (8b) we evaluate $\Delta U(\tilde{w}^t) := U(\tilde{w}^{t+1}) - U(\tilde{w}^t)$ along the trajectories of (41), obtaining

$$\begin{aligned}\Delta U(\tilde{w}^t) &= (S\tilde{w}^t + \delta\gamma\tilde{u}(\chi,\tilde{w}^t))^\top M (S\tilde{w}^t + \delta\gamma\tilde{u}(\chi,\tilde{w}^t)) \\ &\quad - (\tilde{w}^t)^\top M \tilde{w}^t \\ &= -\left\|\tilde{w}^t\right\|^2 + 2\delta\gamma(\tilde{w}^t)^\top S^\top M \tilde{u}(\chi,\tilde{w}^t) \\ &\quad + \delta^2\gamma^2 \tilde{u}(\chi,\tilde{w}^t)^\top M \tilde{u}(\chi,\tilde{w}^t) \\ &\le -(1 - \delta\gamma\mu_1 - \delta^2\gamma^2\mu_2)\left\|\tilde{w}^t\right\|^2,\end{aligned} \tag{68}$$

where the second equality is due to (67), and the inequality is due to (66) and the Cauchy-Schwarz inequality , with the constants $\mu_1 := 2\|S\|\|M\|$ and $\mu_2 := \|M\|$. Thus, there always exists $\bar{\delta}_1 > 0$ small enough so that $(1 - \delta\gamma\mu_1 - \delta^2\gamma^2\mu_2) > 0$ for any $\delta \in (0, \bar{\delta}_1)$ and $\gamma > 0$, concluding the proof. ∎

Proof of Lemma IV.5: The proof is inspired by [38, Theorem 2, Lemma 3, Lemma 4], adapted to our framework. Let $\mathcal{F} : \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$ and $\mathcal{H} : \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$ be defined as

$$\mathcal{F}(\chi^t) := \begin{bmatrix} F\left(\begin{bmatrix} I & 0\end{bmatrix}\chi^t\right) \\ 0 \end{bmatrix}, \tag{69a}$$

$$\mathcal{H}(\chi^t) := \begin{bmatrix} \nabla_x H\left(A\begin{bmatrix} I & 0\end{bmatrix}\chi^t - b, \begin{bmatrix} 0 & I\end{bmatrix}\chi^t\right) \\ -\nabla_\lambda H\left(A\begin{bmatrix} I & 0\end{bmatrix}\chi^t - b, \begin{bmatrix} 0 & I\end{bmatrix}\chi^t\right)\end{bmatrix}. \tag{69b}$$

Applying (63) to each of the components of $\mathcal{H}(\chi^t) - \mathcal{H}(\chi^\star)$, for any $\chi^t \in \mathbb{R}^{n+m}$ we obtain

$$\begin{aligned}&\mathcal{H}(\chi^t) - \mathcal{H}(\chi^\star) \\ &= \begin{bmatrix} \rho A^\top E(\chi^t,\chi^\star)A & A^\top E(\chi^t,\chi^\star) \\ -E(\chi^t,\chi^\star)A & -\frac{1}{\rho}(E(\chi^t,\chi^\star) - I)\end{bmatrix}(\chi^t - \chi^\star),\end{aligned} \tag{70}$$

where $E(\chi^t,\chi^\star) := \operatorname{diag}(\epsilon_1(\chi^t,\chi^\star),\dots,\epsilon_m(\chi^t,\chi^\star))$ and $\epsilon_\ell(\chi^t,\chi^\star) \in [0,1]$ so that

$$\begin{aligned}&\max\{\rho([Ax^t - b]_\ell) + [\bar{\lambda}^t]_\ell, 0\} - \max\{\rho([Ax^\star - b]_\ell) + [\lambda^\star]_\ell, 0\} \\ &\quad = \epsilon_\ell(\chi^t,\chi^\star)(\rho([Ax^t - b]_\ell - [Ax^\star - b]_\ell) + [\bar{\lambda}^t]_\ell - [\lambda^\star]_\ell),\end{aligned}$$

for all $\ell \in \{1,\dots,m\}$ and $\chi^t := \mathrm{COL}(x^t, \bar{\lambda}^t) \in \mathbb{R}^{n+m}$. Moreover, for any $x^t \in \mathbb{R}^n$, we have

$$\begin{aligned}F(x^t) - F(x^\star) &= \int_0^1 \nabla F((1-\nu)x^\star + \nu x^t)(x^t - x^\star)d\nu \\ &\overset{(a)}{=} \left[\int_0^1 \nabla F((1-\nu)x^\star + \nu x^t)d\nu\right](x^t - x^\star) \\ &\overset{(b)}{=} B(x^t,x^\star)(x^t - x^\star).\end{aligned} \tag{71}$$

where in (a) we have extracted the term $(x^t - x^\star)$ from the integral and in (b) we have introduced $B(x^t,x^\star) := \int_0^1 \nabla F((1-\nu)x^\star + \nu x^t)d\nu$. Since F is μ-strongly monotone and β_1-Lipschitz continuous (cf. Standing Assumption II.3), we can uniformly bound the integrand term of (71) as

$$\mu I \preccurlyeq \nabla F((1-\nu)x^\star + \nu x^t) \preccurlyeq \beta_1 I,$$

which leads to

$$\mu I \preccurlyeq \int_0^1 \mu I d\nu \preccurlyeq B(x^t,x^\star) \preccurlyeq \int_0^1 \beta_1 I d\nu \preccurlyeq \beta_1 I. \tag{72}$$

Combining the definitions (69) with (70) and (71), we can write

$$\begin{aligned}&-\mathcal{F}(\chi^t) + \mathcal{F}(\chi^\star) - (\mathcal{H}(\chi^t) - \mathcal{H}(\chi^\star)) \\ &\qquad = D(\chi^t,\chi^\star)(\chi^t - \chi^\star),\end{aligned} \tag{73}$$

where $D(\chi^t,\chi^\star) \in \mathbb{R}^{(n+m)\times(n+m)}$ is given by

$$\begin{aligned}&D(\chi^t,\chi^\star) \\ &:= \begin{bmatrix} -B(\chi^t,\chi^\star) - \rho A^\top E(\chi^t,\chi^\star)A & -A^\top E(\chi^t,\chi^\star) \\ E(\chi^t,\chi^\star)A & \frac{1}{\rho}(E(\chi^t,\chi^\star) - I)\end{bmatrix}.\end{aligned}$$

Consider now $M \in \mathbb{R}^{(n+m)\times(n+m)}$ defined as

$$M := \begin{bmatrix} cI & A^\top \\ A & cI \end{bmatrix}, \tag{74}$$

and notice that choosing c such that $c^2 > \kappa_2$ (cf. Assumption IV.2) ensures that $M \succ 0$ (see also [38, Theorem 1]). Now, let

$$\mathcal{P}_\mathcal{X}[\chi] := P_{X\times\mathbb{R}^m}[\chi]. \tag{75}$$

We can employ matrix M to show that $\|\mathcal{P}_\mathcal{X}[\chi^t - \gamma\mathcal{F}(\chi^t) - \gamma\mathcal{H}(\chi^t)] - \chi^\star\|_M$ enjoys certain contraction properties under the M-weighted norm. Note that $\mathcal{P}_\mathcal{X}[\chi^t - \gamma\mathcal{F}(\chi^t) - \gamma\mathcal{H}(\chi^t)]$ combines both the projected descent step for x^{t+1} and the ascent step for λ^{t+1} in (34); this also justifies the opposite sign in the two block rows of $\mathcal{H}(\chi^t)$ (cf. (69b)) and hence also of $D(\chi^t,\chi^\star)$.

We then have that

$$\begin{aligned}&\left\|\mathcal{P}_\mathcal{X}[\chi^t - \gamma\mathcal{F}(\chi^t) - \gamma\mathcal{H}(\chi^t)] - \chi^\star\right\|_M^2 \\ &\overset{(a)}{\le} \left\|\chi^t - \chi^\star - \gamma(\mathcal{F}(\chi^t) - \mathcal{F}(\chi^\star)) - \gamma(\mathcal{H}(\chi^t) - \mathcal{H}(\chi^\star))\right\|_M^2 \\ &\overset{(b)}{\le} \left\|\chi^t - \chi^\star + \gamma D(\chi^t,\chi^\star)(\chi^t - \chi^\star)\right\|_M^2 \\ &\overset{(c)}{=} \left\|\chi^t - \chi^\star\right\|_M^2 + \gamma^2\left\|D(\chi^t,\chi^\star)(\chi^t - \chi^\star)\right\|_M^2 \\ &\quad + \gamma(\chi^t - \chi^\star)^\top(D(\chi^t,\chi^\star)^\top M + MD(\chi^t,\chi^\star))(\chi^t - \chi^\star),\end{aligned} \tag{76}$$

where in (a) we use the relation $\chi^\star = \mathcal{P}_\mathcal{X}[\chi^\star - \gamma\mathcal{F}(\chi^\star) - \gamma\mathcal{H}(\chi^\star)]$, and the non-expansiveness property of the projection since X is closed and convex (cf. Standing Assumption II.2), in (b) we use (73), and in (c) we expand $\|\cdot\|_M^2$. In light of (72), selecting $c := 20\beta_1\left(\max\left\{\frac{\rho\kappa_2}{\mu}, \frac{\beta_1}{\mu}\right\}\right)^2\left(\max\left\{\frac{1}{\beta_1\rho}, \frac{\beta_1}{\mu}\right\}\right)^2\frac{\kappa_2}{\kappa_1}$ and $\tau := \frac{\kappa_1}{2c}$, we can apply [38, Lemma 4] to $D(\chi^t,\chi^\star)$, obtaining

$$D(\chi^t,\chi^\star)^\top M + MD(\chi^t,\chi^\star) \le -\tau M. \tag{77}$$

We then have that for any $\chi^t \in \mathbb{R}^{n+m}$,

$$\left\| D(\chi^t, \chi^\star)(\chi^t - \chi^\star) \right\|_M^2 \leq \mu_1 \left\| \chi^t - \chi^\star \right\|_M^2, \tag{78}$$

where $\mu_1 := \left(\max\left\{\beta_1 + \rho \left\|A\right\|^2, \frac{1}{\rho}\right\}\right)^2$ and the inequality follows by inspection of $D(\chi^t, \chi^\star)(\chi^t - \chi^\star)$ and using $\|E(\chi^t, \chi^\star)\| \leq 1$. Thus, we bound the right-hand side of (76) as

$$\begin{aligned}&\left\|\mathcal{P}_{\mathcal{X}}\left[\chi^t - \gamma\mathcal{F}(\chi^t) - \gamma\mathcal{H}(\chi^t)\right] - \chi^\star\right\|_M^2 \\ &\leq (1 - \gamma\tau + \gamma^2\mu_1)\left\|\chi^t - \chi^\star\right\|_M^2.\end{aligned} \tag{79}$$

Setting $\bar{\gamma} = \frac{\tau}{\mu_1}$, for any $\gamma \in (0, \bar{\gamma})$, we have that $0 < 1 - \gamma\tau + \gamma^2\mu_1 < 1$. Therefore,

$$\begin{aligned}&\left\|\mathcal{P}_{\mathcal{X}}\left[\chi^t - \gamma\mathcal{F}(\chi^t) - \gamma\mathcal{H}(\chi^t)\right] - \chi^\star\right\|_M \\ &\leq (1 - \bar{\mu})\left\|\chi^t - \chi^\star\right\|_M,\end{aligned} \tag{80}$$

where $\bar{\mu} := 1 - \sqrt{1 - \gamma\tau + \gamma^2\mu_1} \in (0, 1)$.

Consider now $W : \mathbb{R}^{n+m} \to \mathbb{R}$ defined as

$$W(\chi) = (\chi - \chi^\star)^\top M(\chi - \chi^\star), \tag{81}$$

where M is as in (74). Since $M \succ 0$, W satisfies conditions (9a) and (9c). To show (9b) we evaluate $\Delta W(\chi^t) := W(\chi^{t+1}) - W(\chi^t)$ along the trajectories of (42), obtaining

$$\begin{aligned}
&\Delta W(\chi^t) \\
&= \left\|\chi^t + \delta f(\chi^t, h(\chi^t)) - \chi^\star\right\|_M^2 - \left\|\chi^t - \chi^\star\right\|_M^2 \\
&\overset{(a)}{\leq} \left\|\chi^t + \delta\left(\mathcal{P}_{\mathcal{X}}\left[\chi^t - \gamma\mathcal{F}(\chi^t) - \gamma\mathcal{H}(\chi^t)\right] - \chi^t - \chi^\star\right)\right\|_M^2 \\
&\quad - \left\|\chi^t - \chi^\star\right\|_M^2 \\
&\overset{(b)}{\leq} (1-\delta)^2\left\|\chi^t - \chi^\star\right\|_M^2 - \left\|\chi^t - \chi^\star\right\|_M^2 \\
&\quad + 2(\delta - \delta^2)\left\|\chi^t - \chi^\star\right\|_M \left\|\mathcal{P}_{\mathcal{X}}[\chi^t - \gamma\mathcal{F}(\chi^t) - \gamma\mathcal{H}(\chi^t)] - \chi^\star\right\|_M \\
&\quad + \delta^2\left\|\mathcal{P}_{\mathcal{X}}\left[\chi^t - \gamma\mathcal{F}(\chi^t) - \gamma\mathcal{H}(\chi^t)\right] - \chi^\star\right\|_M^2 \\
&\overset{(c)}{\leq} (1-\delta)^2\left\|\chi^t - \chi^\star\right\|_M^2 - \left\|\chi^t - \chi^\star\right\|_M^2 \\
&\quad + 2(\delta - \delta^2)(1 - \bar{\mu})\left\|\chi^t - \chi^\star\right\|_M^2 \\
&\quad + \delta^2(1 - \bar{\mu})^2\left\|\chi^t - \chi^\star\right\|_M^2 \\
&\overset{(d)}{\leq} -\delta\bar{\mu}(2 - \delta\bar{\mu})\left\|\chi^t - \chi^\star\right\|_M^2,
\end{aligned} \tag{82}$$

where (a) uses the definitions of f and $\mathcal{P}_{\mathcal{X}}$ (cf. (38b), (75)) to explicitly write the update, in (b) we expand the squared norm, (c) follows by (80), while in (d) we rearrange the terms. Setting $\bar{\delta} := 2/\bar{\mu}$, (82) ensures that for any $\delta \in (0, \bar{\delta})$, W satisfies (9b), and the proof follows. ■

References

[1] I. Menache and A. Ozdaglar, "Network games: Theory, models, and dynamics," *Synthesis Lectures on Communication Networks*, vol. 4, no. 1, pp. 1–159, 2011.

[2] G. Scutari, F. Facchinei, J.-S. Pang, and D. P. Palomar, "Real and complex monotone communication games," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 4197–4231, 2014.

[3] F. Facchinei and C. Kanzow, "Generalized nash equilibrium problems," *Annals of Operations Research*, vol. 175, no. 1, pp. 177–211, 2010.

[4] A.-H. Mohsenian-Rad, V. W. Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia, "Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid," *IEEE transactions on Smart Grid*, vol. 1, no. 3, pp. 320–331, 2010.

[5] G. Belgioioso, W. Ananduta, S. Grammatico, and C. Ocampo-Martinez, "Energy management and peer-to-peer trading in future smart grids: A distributed game-theoretic approach," in *2020 European Control Conference (ECC)*. IEEE, 2020, pp. 1324–1329.

[6] K. Okuguchi and F. Szidarovszky, *The theory of oligopoly with multi-product firms*. Springer Science & Business Media, 2012.

[7] F. Fabiani, D. Fenucci, and A. Caiti, "A distributed passivity approach to auv teams control in cooperating potential games," *Ocean Engineering*, vol. 157, pp. 152–163, 2018.

[8] L. Deori, K. Margellos, and M. Prandini, "Price of anarchy in electric vehicle charging control games: When nash equilibria achieve social welfare," *Automatica*, vol. 96, pp. 150–158, 2018.

[9] F. Fele and K. Margellos, "Probably approximately correct nash equilibrium learning," *IEEE Transactions on Automatic Control*, vol. 66, no. 9, pp. 4238–4245, 2020.

[10] C. Cenedese, F. Fabiani, M. Cucuzzella, J. M. Scherpen, M. Cao, and S. Grammatico, "Charging plug-in electric vehicles as a mixed-integer aggregative game," in *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 4904–4909.

[11] J. Barrera and A. Garcia, "Dynamic incentives for congestion control," *IEEE Transactions on Automatic Control*, vol. 60, no. 2, pp. 299–310, 2014.

[12] H. Yin, P. G. Mehta, S. P. Meyn, and U. V. Shanbhag, "Synchronization of coupled oscillators is a game," *IEEE Transactions on Automatic Control*, vol. 57, no. 4, pp. 920–935, 2011.

[13] M. K. Jensen, "Aggregative games and best-reply potentials," *Economic theory*, vol. 43, no. 1, pp. 45–66, 2010.

[14] F. Parise and A. Ozdaglar, "Analysis and interventions in large network games," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 4, pp. 455–486, 2021.

[15] G. Belgioioso, P. Yi, S. Grammatico, and L. Pavel, "Distributed generalized nash equilibrium seeking: An operator-theoretic perspective," *IEEE Control Systems Magazine*, vol. 42, no. 4, pp. 87–102, 2022.

[16] X. Li, L. Xie, and Y. Hong, "Distributed aggregative optimization over multi-agent networks," *IEEE Transactions on Automatic Control*, 2021.

[17] X. Li, X. Yi, and L. Xie, "Distributed online convex optimization with an aggregative variable," *IEEE Transactions on Control of Network Systems*, 2021.

[18] G. Carnevale, A. Camisa, and G. Notarstefano, "Distributed online aggregative optimization for dynamic multi-robot coordination," *IEEE Transactions on Automatic Control*, 2022.

[19] G. Carnevale, N. Mimmo, and G. Notarstefano, "Aggregative feedback optimization for distributed cooperative robotics," *IFAC-PapersOnLine*, vol. 55, no. 13, pp. 7–12, 2022.

[20] T. Wang and P. Yi, "Distributed projection-free algorithm for constrained aggregative optimization," *arXiv preprint arXiv:2207.11885*, 2022.

[21] 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, 2015.

[22] G. Belgioioso and S. Grammatico, "Semi-decentralized nash equilibrium seeking in aggregative games with separable coupling constraints and non-differentiable cost functions," *IEEE control systems letters*, vol. 1, no. 2, pp. 400–405, 2017.

[23] 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, 2019.

[24] S. Grammatico, "Dynamic control of agents playing aggregative games with coupling constraints," *IEEE Transactions on Automatic Control*, vol. 62, no. 9, pp. 4537–4548, 2017.

[25] 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*, vol. 64, no. 4, pp. 1373–1388, 2018.

[26] P. Yi and L. Pavel, "An operator splitting approach for distributed generalized nash equilibria computation," *Automatica*, vol. 102, pp. 111–121, 2019.

[27] H. Kebriaei, S. J. Sadati-Savadkoohi, M. Shokri, and S. Grammatico, "Multipopulation aggregative games: Equilibrium seeking via mean-field control and consensus," *IEEE Transactions on Automatic Control*, vol. 66, no. 12, pp. 6011–6016, 2021.

[28] G. Belgioioso and S. Grammatico, "Semi-decentralized generalized nash equilibrium seeking in monotone aggregative games," *IEEE Transactions on Automatic Control*, 2021.

[29] 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.

[30] M. Zhu and S. Martínez, "Discrete-time dynamic average consensus," *Automatica*, vol. 46, no. 2, pp. 322–329, 2010.

[31] S. S. Kia, B. Van Scoy, J. Cortes, R. A. Freeman, K. M. Lynch, and S. Martinez, "Tutorial on dynamic average consensus: The problem, its applications, and the algorithms," *IEEE Control Systems Magazine*, vol. 39, no. 3, pp. 40–72, 2019.

[32] M. Ye, G. Hu, L. Xie, and S. Xu, "Differentially private distributed nash equilibrium seeking for aggregative games," *IEEE Transactions on Automatic Control*, 2021.

[33] F. Parise, S. Grammatico, B. Gentile, and J. Lygeros, "Distributed convergence to nash equilibria in network and average aggregative games," *Automatica*, vol. 117, p. 108959, 2020.

[34] C. Cenedese, G. Belgioioso, Y. Kawano, S. Grammatico, and M. Cao, "Asynchronous and time-varying proximal type dynamics in multiagent network games," *IEEE Transactions on Automatic Control*, vol. 66, no. 6, pp. 2861–2867, 2020.

[35] F. Parise, B. Gentile, and J. Lygeros, "A distributed algorithm for almost-nash equilibria of average aggregative games with coupling constraints," *IEEE Transactions on Control of Network Systems*, vol. 7, no. 2, pp. 770–782, 2019.

[36] G. Belgioioso, A. Nedić, and S. Grammatico, "Distributed generalized nash equilibrium seeking in aggregative games on time-varying networks," *IEEE Transactions on Automatic Control*, vol. 66, no. 5, pp. 2061–2075, 2020.

[37] D. Gadjov and L. Pavel, "Single-timescale distributed gne seeking for aggregative games over networks via forward–backward operator splitting," *IEEE Transactions on Automatic Control*, vol. 66, no. 7, pp. 3259–3266, 2020.

[38] G. Qu and N. Li, "On the exponential stability of primal-dual gradient dynamics," *IEEE Control Systems Letters*, vol. 3, no. 1, pp. 43–48, 2018.

[39] N. Bof, R. Carli, and L. Schenato, "Lyapunov theory for discrete time systems," *arXiv preprint arXiv:1809.05289*, 2018.

[40] H. K. Khalil, "Nonlinear systems," *Upper Saddle River*, 2002.

[41] G. Carnevale and G. Notarstefano, "Nonconvex distributed optimization via lasalle and singular perturbations," *IEEE Control Systems Letters*, vol. 7, pp. 301–306, 2022.

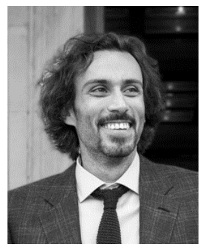
[42] V. Chellaboina and W. M. Haddad, *Nonlinear dynamical systems and control: A Lyapunov-based approach*. Princeton University Press, 2008.

Guido Carnevale (S'21) received the M.Sc. degree "summa cum laude" in Automation Engineering in 2019 from the Alma Mater Studiorum Università di Bologna, Italy. He is a Ph.D. student within the Ph.D Programme "Biomedical, Electrical, and Systems Engineering" at the Department of Electrical, Electronic and Information Engineering, Alma Mater Studiorum Università di Bologna, Italy. He was a visiting scholar at the University of Oxford in 2022. His research interests include distributed optimization.

Filippo Fabiani is an Assistant Professor at the IMT School for Advanced Studies Lucca, IT. He received the B.Sc. degree in Bio-Engineering, the M.Sc. degree in Automatic Control Engineering, and the Ph.D. degree in Automatic Control, all from the University of Pisa, IT, in 2012, 2015, and 2019 respectively. In 2018-2019 he was post-doctoral Research Fellow in the Delft Center for Systems and Control at TU Delft, NL, while in 2019-2022 he was a post-doctoral Research Assistant in the Control Group at the Department of Engineering Science, University of Oxford, UK.

His research interests include game theory, optimization and control of complex uncertain systems, with applications in generation and load side control for power networks and automated driving.

Filiberto Fele Filiberto Fele received the MSc in Mechatronics Engineering from Politecnico di Torino, Italy, in 2008, and the Ph.D. in Control Engineering from the University of Seville, Spain, in 2017. After holding a first postdoctoral appointment at the Electrical and Electronic Engineering Dept., Imperial College London, he joined the Department of Engineering Science, University of Oxford in 2018, where he is currently a Research Associate with the Energy and Power Group and the Control Group. Since 2019, he has been serving as Lecturer at Christ Church and St Edmund Hall Colleges. His interests include data-driven robust optimisation and control of multi-agent systems, with a focus on the smart grid and the transition to net zero.



Kostas Margellos received the Diploma in electrical engineering from the University of Patras, Greece, in 2008, and the Ph.D. in control engineering from ETH Zurich, Switzerland, in 2012. He spent 2013, 2014 and 2015 as a postdoctoral researcher at ETH Zurich, UC Berkeley and Politecnico di Milano, respectively. In 2016 he joined the Control Group, Department of Engineering Science, University of Oxford, where he is currently an Associate Professor. He is also a Fellow in AI & Machine Learning at Reuben College and a Lecturer at Worcester College. He is currently serving as Associate Editor in Automatica and in the *IEEE Control Systems Letters*, and is part of the Conference Editorial Board of the IEEE Control Systems Society and EUCA. His research interests include optimization and control of complex uncertain systems, with applications to energy and transportation networks.

Giuseppe Notarstefano (M'11) received the Laurea degree summa cum laude in electronics engineering from the Università di Pisa, Italy, in 2003 and the Ph.D. degree in automation and operation research from the Università di Padova, Italy, in 2007. He is a Professor with the Department of Electrical, Electronic, and Information Engineering G. Marconi, Alma Mater Studiorum Università di Bologna, Italy. He was Associate Professor (from 2016 to 2018) and previously Assistant Professor, Ricercatore (from 2007), with the Università del Salento, Italy. He has been Visiting Scholar at the University of Stuttgart, University of California Santa Barbara, CA, USA and University of Colorado Boulder, CO, USA. His research interests include distributed optimization, cooperative control in complex networks, applied nonlinear optimal control, and trajectory optimization and maneuvering of aerial and car vehicles. Dr. Notarstefano serves as an Associate Editor for *IEEE Transactions on Automatic Control*, *IEEE Transactions on Control Systems Technology*, and *IEEE Control Systems Letters*. He has been also part of the Conference Editorial Board of IEEE Control Systems Society and EUCA. He received the *IEEE Transactions on Control of Network Systems* outstanding paper award 2021 and he was a recipient of an ERC Starting Grant.